technische universität
dortmund

Bachelorarbeit

Gesangserkennung in polyphonen Musikaufnahmen

Florian Scholz

November 2016

Gutachter:
Prof. Dr. Günter Rudolph
Dr. Igor Vatolkin

Technische Universität Dortmund
Fakultät für Informatik
Lehrstuhl 11
https://ls11-www.cs.uni-dortmund.de

Inhaltsverzeichnis

1 Einleitung und verwandte Arbeiten **1**

1.1 Gesangsdefinition . . . 1

1.2 Nutzen und Möglichkeiten . . . 3

1.3 Verwandte Arbeiten . . . 4

1.4 Aufbau der Arbeit . . . 6

2 Grundlagen und Methoden **7**

2.1 Audiomerkmale . . . 7

2.2 Klassifikatoren . . . 9

2.2.1 Random Forest . . . 11

2.2.2 Neuronale Netze . . . 13

2.3 Ansatz 1: J. Schlüter und T. Grill . . . 16

2.4 Ansatz 2: B. Lehner, G. Widmer und R. Sonnleiter . . . 16

2.5 Ansatz 3: AMUSE-Merkmale . . . 17

3 Ziele der Arbeit **19**

3.1 Testdaten . . . 19

3.2 Parameterevaluation . . . 21

4 Experimente **23**

4.1 Annotation der Testdaten . . . 23

4.2 Testverfahren . . . 25

4.3 Ansatz 1: J. Schlüter und T. Grill . . . 26

4.4 Ansatz 2: B. Lehner, G. Widmer und R. Sonnleiter . . . 28

4.4.1 Kreuzvalidierung und Parameterevaluation . . . 29

4.4.2 „Leave one out“ . . . 30

4.4.3 Genre Konfusion . . . 32

4.5 Ansatz 3: AMUSE-Merkmale . . . 37

4.5.1 Kreuzvalidierung und Parameterevaluation . . . 39

4.5.2 „Leave one out“ . . . 41

4.5.3 Genre Konfusion . . . 43

4.6 Vergleich der Ansätze 52

5 Zusammenfassung **55**

5.1 Fazit 55

5.2 Ausblick 56

A Anhang **59**

A.1 Verwendeter Testdatensatz 59

A.2 Auswertungstabellen Ansatz 2 67

A.3 Auswertungstabellen Ansatz 3 72

A.4 AMUSE-Merkmale 94

Abbildungsverzeichnis **97**

Tabellenverzeichnis **99**

Literaturverzeichnis **103**

Danksagungen **105**

Erklärung **107**

Kapitel 1

Einleitung und verwandte Arbeiten

Sprachassistenten, wie Google Now, Cortana oder Siri, sind in der heutigen Zeit fester Bestandteil in modernen Smartphones und Desktop-Betriebssystemen. Was früher noch „Science-Fiction“ war, ist heute möglich: verbale Kommunikation mit einem Computer. Der Nutzer gibt ein Kommando, die Maschine „hört zu“ und führt den Befehl aus.

Doch nicht nur bei Sprachassistenten ist die Erkennung von Sprache von besonderer Bedeutung. In der heutigen Zeit, geprägt von gigantischen Musikdatenbanken und Streaming Diensten, wird auch die Erkennung von Gesang in Musikstücken, als Spezialfall der Spracherkennung, immer wichtiger.

1.1 Gesangsdefinition

Bevor sich eine Maschine um das Problem kümmern kann, was eine Person konkret gesagt oder gesungen hat, muss ein viel grundlegenderes Problem gelöst werden: Wann enthält das Eingabesignal Sprache oder Gesang und wann sind es nur Hintergrundgeräusche oder Musik?

Im ersten Moment scheint dieses Problem, zumindest für einen Menschen, recht einfach zu lösen. Doch sollte zuerst geklärt werden, wie genau Gesang definiert wird, denn hier gibt es durchaus Spielraum für verschiedene Interpretationen. Es gibt Definitionen aus der Literatur sowie aus der Zoologie, die sich auf literarische Texte bzw. tierischen Gesang beziehen. Die treffendste Definition findet sich jedoch im Brockhaus:

Gesang als Element der Musik ist die *„Bezeichnung sowohl für die Tätigkeit des Singens wie auch für eine abgeschlossene und wiederholbare musikalische Einheit. Gesang kann sowohl von mehreren Singenden gemeinsam (Chor-Gesang) als auch von einem Einzelnen (Solo-Gesang) ausgeübt werden. In der Regel ist Gesang an Worte oder Texte mit deutlich geprägtem Sinnzusammenhang gebunden; es ist aber auch möglich, auf für sich genommen sinnleere Laute oder Silben zu singen (Jodeln, Vokalise, Skat). Geschichtsepochen, Völker*

und Kulturen sowie soziale Schichten haben verschiedene Vorstellungen und Praktiken des Singens [...]“ [1, S.594].

Dieser Definition ist zu entnehmen, dass der Begriff „Gesang“ eine sehr breites Spektrum abdeckt. Demnach kann Gesang von einer einzelnen Person stammen oder aber von mehreren Personen gemeinsam als Chor. Abbildung 1.1 zeigt das Spektogramm des Gesangs einer einzelnen Person. Das für Gesang typische Wellenmuster ist gut zu erkennen. In Abbildung 1.2 ist ein Ausschnitt eines Chorgesangs zu sehen. Hier überlagern sich die Wellenmuster, sodass sie schwieriger zu erkennen sind. Laut Definition sollte jedoch beides als Gesang erkannt werden.

Um ein möglichst breites Spektrum abzudecken, wird in dieser Arbeit der obigen Definition entsprochen. Somit zählen Gesang einer einzelnen Person sowie Chorgesang zu den gesuchten Passagen innerhalb eines Musikstücks. Auch „sinnleere Laute oder Silben“ werden als Gesang gewertet. Die getesteten Ansätze machen in ihrer Beschreibung keinerlei Einschränkungen, wo eine Grenze bei der Definition von Gesang zu ziehen ist. Daher sollten alle Ansätze auch mit der am weitesten ausgelegten Definition von Gesang funktionieren.

Nach der Festlegung einer treffenden Definition für Gesang, bleibt noch das Problem, diese Definition für den Computer greifbar zu machen, sie in für ihn verwertbare Zahlen und Formeln zu verpacken. Hierzu werden spezielle Audiomerkmale verwendet, die besondere

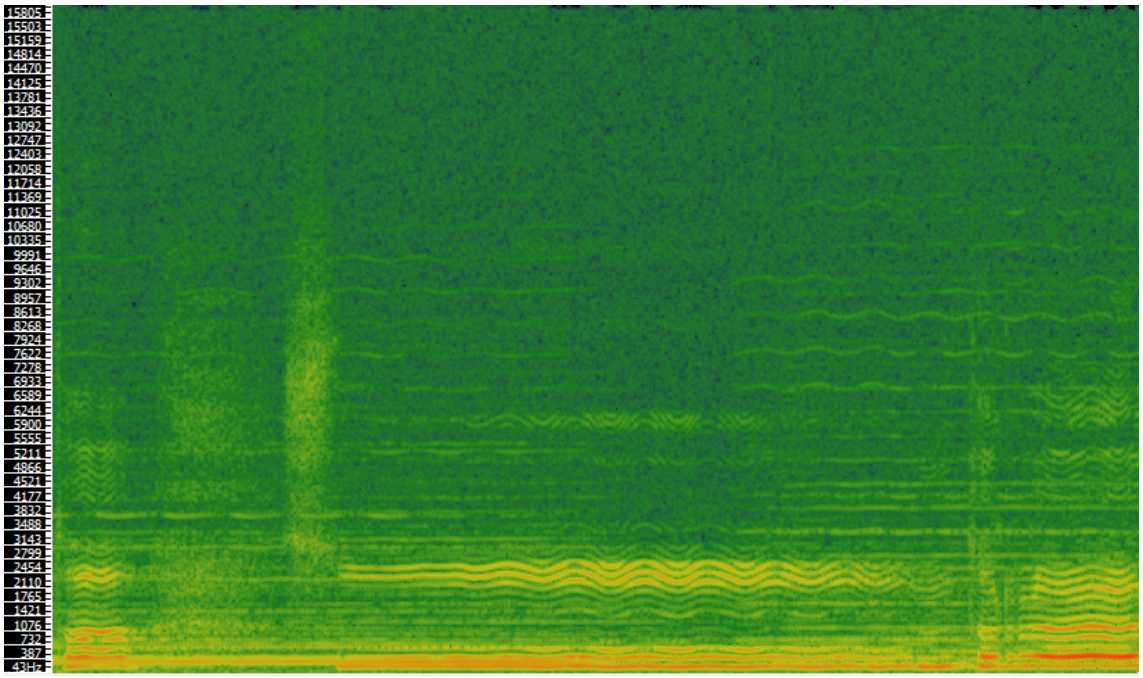

Abbildung 1.1: Spektogramm aus „Carl Orff - O Fortuna“: Ein Sänger, Zeitabschnitt von ca. 2 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse. Farbverlauf von Schwarz über grün und gelb bis hin zu rot. Schwarz entspricht fehlender Frequenz, rot zeigt eine stark vorhandene Frequenz an.

Eigenschaften der Musik auf Kennzahlen abbilden, die der Computer dann verarbeiten kann. Diese werden in Kapitel 2 noch genauer beschrieben.

1.2 Nutzen und Möglichkeiten

Die Erkennung von Gesang innerhalb eines Musikstücks ist der erste und grundlegende Schritt bei der Analyse und Datengewinnung aus Gesang. Wie zuvor bereits angedeutet, muss für eine Gesangsanalyse das zu verarbeitende Material erst einmal gefunden werden. Ist dieser Schritt geschafft, eröffnet dies dem Nutzer ein breites Feld an Möglichkeiten.

Neben der Erkennung des gesungenen Textes könnte zum Beispiel untersucht werden, in welcher Tonlage gesungen wird, welcher Künstler singt und ob es sich bei dem Künstler um einen Mann oder eine Frau handelt. Für den Fall, dass ein völlig unbekanntes Lied im Radio läuft, könnten mittels der Gesangserkennung Interpret sowie Titel des Liedes bestimmt werden. Des Weiteren lässt sich mit dem Wissen, an welchen Stellen gesungen wird, leicht eine Karaoke-Version eines Musikstücks erstellen, indem an den betreffenden Passagen der Gesang entfernt wird. Außerdem könnten gewonnene Informationen wie Text, Gesangsstil und Interpret dazu beitragen, ein Musikstück einem bestimmten Genre zuzuordnen.

Besonders für Musikdatenbanken und Online-Streaming-Dienste eröffnen sich so neue Möglichkeiten. Angenommen, ein jüngerer Nutzer ist auf die Musik von Phil Collins ge-

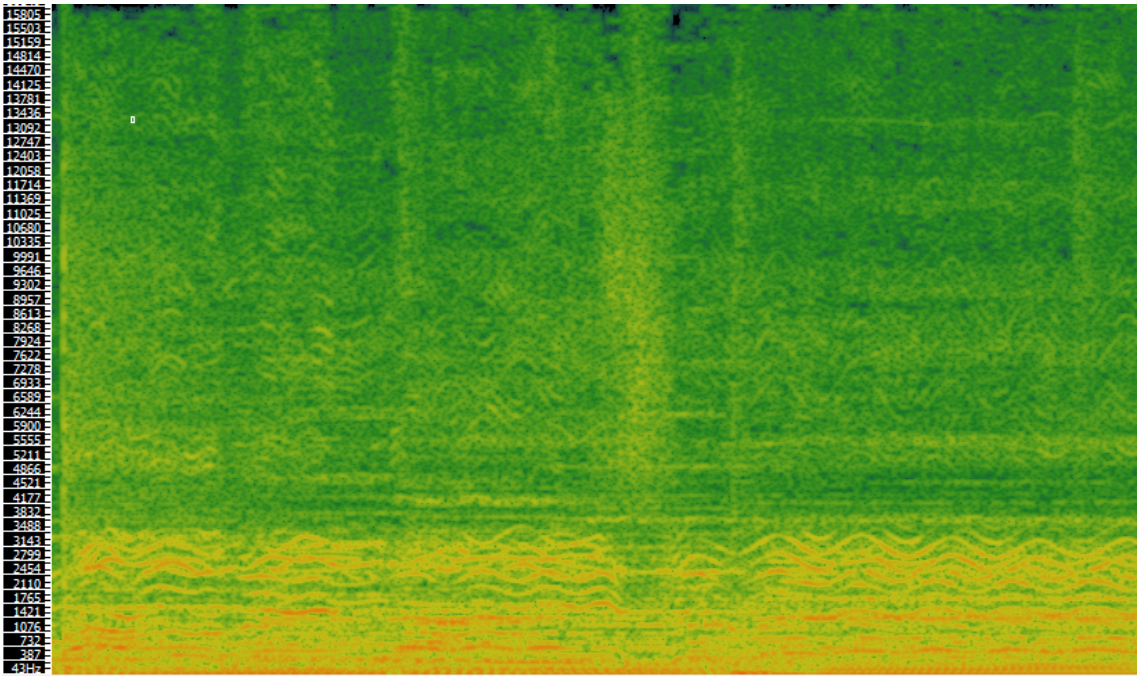

Abbildung 1.2: Spektogramm aus „Il Divo - Everytime I Look At You“: Chorgesang, Zeitabschnitt von ca. 2 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse. Farbverlauf von Schwarz über grün und gelb bis hin zu rot. Schwarz entspricht fehlender Frequenz, rot zeigt eine stark vorhandene Frequenz an.

stoßen. Aufgrund seines noch recht jungen Alters ist dem Nutzer der Künstler und sein Werdegang nicht bekannt. Mit der Möglichkeit, Interpreten an ihrer Stimme zu erkennen, könnte ihm der Streaming-Dienst Titel der Band Genesis empfehlen, der Phil Collins angehörte, da die Software dessen Stimme in den Genesis-Aufnahmen wiedererkannt hat.

1.3 Verwandte Arbeiten

Schon seit mehreren Jahren wird mit verschiedenen Methoden Gesangserkennung in Musik betrieben. Hier werden nun einige interessante Ansätze kurz beschrieben, um einen Überblick über den aktuellen Stand der Forschung zu geben.

Eine der ersten Arbeiten zum Thema Gesangserkennung wurde 2001 von A. L. Berenzweig und D. P. W. Ellis unter dem Titel „Locating singing voice segments within music signals“ [2] veröffentlicht. Als Klassifikator kam ein *„Hidden Markov Modell (HMM)“* zu Einsatz. Des Weiteren wurden unter anderem *„posterior probability features“* und *„PLP cepstral coefficients“* als Merkmale verwendet. Der Testdatensatz bestand aus 246 zufällig gewählten Radiomitschnitten mit einer Länge von 15 Sekunden, die 1996 von einem FM-Radio aufgenommen wurden. Mit ihrem Verfahren konnten sie den Klassifikationsfehler auf 18,8% reduzieren.

2008 verfassten M. Ramona et al. eine Arbeit mit dem Titel „Vocal detection in music with support vector machines“ [12], in der sie Gesang mittels einer *„Support Vector Machine (SVM)“* klassifizierten. Als Merkmale wurden unter anderem *„Mel Frequency Cepstral Coefficients (MFCC)“* und *„Linear Predictive Coding Coefficients (LPCC)“* eingesetzt. Der Testdatensatz bestand aus 93 zufällig gewählten Liedern aus der freien Musikdatenbank Jamendo[1]. Dieser Datensatz wird noch heute bei Neuveröffentlichungen für Vergleichszwecke genutzt. Mit ihrem Ansatz konnten sie eine *„accuracy“* von 74,8% und ein *„f-measure“* von 72,6% erreichen.

2009 veröffentlichten L. Regnier und G. Peeters ihre Arbeit „Singing voice detection in music tracks using direct voice vibrato detection“ [13], in der sie Gesangspassagen innerhalb eines Liedes nur anhand von Tremolo und Vibrato erkannten. Als Testdatensatz wurde der in [12] eingeführte Jamendo-Datensatz verwendet. Zur Klassifikation wurde ein simples Schwellwertverfahren genutzt. Nach der Klassifikation wurden alle gesangslosen Segmente unter einer Sekunde herausgefiltert, wodurch sich der *„recall“* von 61,69% auf 83,57% erhöhte. Die Werte für das *„f-measure“* betrugen 66,54% bzw. 76,83% nach der Filterung. Die *„precision“* erhöhte sich leicht von 70,56% auf 71,09%.

Im Jahr 2014 untersuchten J. Urbano et al. in ihrer Arbeit „What is the effect of audio quality on the robustness of MFCCs and chroma features?“ [16], wie sich verlustbehaftete Audiokompression auf die Klassifikation von Gesang auswirkt. Sie extrahierten aus 400 Musikstücken, verteilt auf zehn Genres, jeweils einen Auszug von 30 Sekunden. Diese Aus-

[1] https://www.jamendo.com/, Abrufdatum: 28.10.2016

züge bildeten die Testmenge. Dabei konnten sie bestätigen, dass sowohl MFCC Merkmale als auch Chroma-Merkmale robust gegenüber Audiokompression sind. Eine deutliche Verschlechterung der Mekrmalsqualität stellte sich erst bei einer Bitrate von unter 160Kbps ein.

Ein Jahr später publizierten C. Dittmar et al. ihre Arbeit mit dem Titel „Cross-version singing voice detection in classical opera recordings“ [4], in der sie eine neue Methode zur Gesangserkennung speziell für klassische Opernstücke vorstellten. Als Merkmale kamen MFCCs, *„vocal variance“*, *„fluctogram variance“*, *„spectral contraction variance“*, *„spectral flatness mean“* und *„polynomial shape spectral contrast“* in Verbindung mit einem „Random Forest“ Klassifikator zum Einsatz. Ihr Ansatz beruht darauf, dass das zu testende Musikstück in mehreren Varianten, von verschiedenen Künstlern interpretiert, vorliegt. Als Teststück diente die Oper „Der Freischütz“ in mehrfacher Interpretation.

Durch die Kombination der Testergebnisse konnte die Erkennungsgenauigkeit erkennbar gesteigert werden. Da es jedoch fast ausschließlich im klassischen Bereich Musikstücke gibt, die von verschiedenen Künstlern interpretiert wurden, eignet sich dieser Ansatz nicht für die Gesangserkennung in Liedern abseits der Klassik.

Ebenfalls 2015 untersuchten S. D. You et al. in ihrer Studie „Comparative study of siniging voice detection methods“ [19], welche Methode zur Gesangserkennung sich am besten eignet. Dabei testeten sie unter anderem, ob eine Optimierung an der Formel für die Mel-Frequenzen zu besseren Klassifikationsergebnissen führt. Hierzu veränderten sie die Gleichung so, dass sich mehr Bänder auf die niedrigeren Frequenzbereiche verteilten. Neben MFCCs kamen außerdem *„Linear Predictive Coefficients (LPC)“* sowie *„Linear Predictive Cepstral Coefficients“* zum Einsatz. Als Klassifikator wurde ein HMM genutzt. Zur Durchführung der Tests wurde eine Sammlung aus 300 Musikstücken, über mehrere Genres verteilt, zusammengestellt.

Die Tests ergaben, dass eine Kombination aus mehreren Audiomerkmalen die Erkennungsgenauigkeit erhöhen kann. Die Kombination aus MFCCs und LPCCs erreichte eine *„accuracy“* von 95,8%. Wird ein Merkmal alleine genutzt, so eignen sich MFCCs am besten mit einer *„accuracy“* von 94,3%. Hier ist jedoch zu beachten, dass die von You et al. vorgenommenen Optimierungen nicht zu einem besseren Ergebnis führten.

S. Leglaive et al. klassifizierten 2015 Gesang mittels eines rekurrenten neuronalen Netzes in ihrer Arbeit „Singing voice detection with deep recurrent neural networks“[8]. Das neuronale Netz bestand aus drei verdeckten Schichten und 80 Eingabeparametern. Sie verwendeten außerdem „low-level“-Merkmale, die sie mittels einer Filter-Bank auf die Mel-Skala abbildeten. Damit erreichten sie auf dem Jamendo-Datensatz einen *„recall“* von 92,6% und eine *„precision“* von 89,5%. Das *„f-measure“* betrug 91,0%.

1.4 Aufbau der Arbeit

Diese Arbeit umfasst insgesamt fünf Kapitel. Im folgenden zweiten Kapitel wird eine Einführung in die Grundlagen und Methoden gegeben, die bei der Gesangserkennung angewendet werden. Es wird zwischen Audiomerkmalen und Klassifikatoren für die Auswertung unterschieden. Des Weiteren werden die drei Testansätze beschrieben.

Im dritten Kapitel folgt die Beschreibung der Ziele dieser Arbeit sowie eine Erklärung zum verwendeten Testdatensatz und dessen Zusammensetzung.

Kapitel vier widmet sich den durchgeführten Experimenten. Es werden das Vorgehen bei der Annotation des Testdatensatzes sowie die Testverfahren beschrieben. Im Anschluss folgt je Testansatz ein Abschnitt zur Auswertung und Analyse der Ergebnisse. Am Ende des Kapitels werden die Ergebnisse der drei Testansätze direkt miteinander verglichen.

Im letzten Kapitel wird ein Fazit gezogen, indem die wichtigsten Ergebnisse zusammengefasst werden. Den Abschluss bildet ein Ausblick auf mögliche weiterführende Arbeiten und Problemstellungen.

Kapitel 2

Grundlagen und Methoden

Zur Klassifikation von Musikausschnitten in die Klassen „Gesang“ und „kein Gesang“ sind grundsätzlich immer zwei Schritte notwendig. In einem ersten Schritt müssen geeignete Audiomerkmale gefunden und berechnet werden, welche den Musikausschnitt repräsentieren. In einem zweiten Schritt können diese Repräsentationen dann an einen Klassifikator übergeben werden, der den Ausschnitt nun einer der beiden Klassen zuordnet. Dieses Verfahren wird im Folgenden nun genauer beschrieben.

2.1 Audiomerkmale

Um die Eingabedaten, also in diesem Falle die Musik, in eine geeignete Form zu bringen, die ein Klassifikator verarbeiten kann, ist es notwendig, dass Audiomerkmale berechnet werden. Dazu ist es sinnvoll, sich vor Augen zu führen, welche Form ein digitalisiertes Musikstück hat.

Digitalisierte Musik besteht im wesentlichen aus einer Reihe von Zahlen, den Abtastwerten. Jeder Abtastwert hat eine Länge von 16 Bit, somit bewegt sich dessen Größe zwischen 0 und $2^{16} - 1$. Die Größe eines Abtastwertes repräsentiert so die Amplitude des Musiksignals. Auf einer handelsüblichen Audio-CD wird mit einer Abtastrate von 44100Hz gearbeitet. Das bedeutet, dass das Audiosignal innerhalb einer Sekunde 44100 mal abgetastet wird und somit für eine Sekunde 44100 Abtastwerte vorhanden sind, die den Verlauf der Signalkurve darstellen. Die Anzahl der Abtastwerte ist hierbei keineswegs zufällig. Sie basiert auf dem Abtasttheorem [11, S.65], welches besagt, dass die Abtastfrequenz mehr als doppelt so hoch sein muss, wie die höchste im Signal enthaltene Frequenz. Somit können Frequenzen bis knapp über 22000Hz digitalisiert werden, womit der Wahrnehmungsbereich des menschlichen Ohres abgedeckt ist [14, S.80].

Jeder dieser Abtastwerte kann auch als Funktionswert verstanden werden. Die zugrunde liegende Funktion ist das Audiosignal. Diese Abtastwerte genügen einem Computer, um

daraus später wieder das Ausgangssignal zu approximieren - die Grundvoraussetzung, um die Musik abzuspielen.

Der bloße Verlauf der Signalkurve reicht jedoch bei Weitem nicht aus, um damit einen Klassifikator zu trainieren. Hier kommen nun die Audiomerkmale ins Spiel. Alle Merkmale basieren auf den Abtastwerten des Signals oder anderen Merkmalen, die zuvor aus den Abtastwerten gewonnen wurden. Die spezifische Fenstergröße eines Merkmals gibt an, wie viele der zugrunde liegenden Abtastwerte zur Berechnung dieses Merkmals verwendet wurden. Ein Merkmal mit einer Fenstergröße von 512 Frames nutzt zur Berechnung folglich 512 Abtastwerte. Durch geschickte Berechnungen können nun weitere Informationen generiert werden, die aus den Abtastwerten nicht ersichtlich sind. Mittels einer Fourier-Transformation [11, S.63] lässt sich beispielsweise das Spektogramm des Audiosignals berechnen, sodass die im Signal enthaltenen Frequenzen zu entnehmen sind.

Die für die Spracherkennung sowie auch für die Gesangserkennung wichtigsten Merkmale sind die *„Mel Frequency Cepstral Coefficients (MFCC)“*, zu deutsch *Mel-Frequenz Cepstrum-Koeffizienten* [11, S.213]. Sie orientieren sich am Modell des menschlichen Gehörs sowie der von Menschen empfundenen Tonhöhe. Bei der empfundenen Tonhöhe handelt es sich um die subjektive Bewertung mehrerer Testpersonen. Zwischen der tatsächlichen physikalischen Tonhöhe, gemessen in Hz, und der so genannten *Tonheit*, gemessen in mel, besteht jedoch kein linearer Zusammenhang, da das menschliche Gehör die Lautstärke nicht allein für eine isolierte Frequenz, sondern immer für eine Gruppe aus mehreren Frequenzen bestimmt. Im Frequenzbereich von 20 Hz bis hin zu 16000 Hz bildet das menschliche Gehör so insgesamt 24 Gruppen. Diese Gruppen werden von den MFCCs repräsentiert [11, S.214].

Neben MFCCs gibt es viele weitere Merkmale, die zur Analyse von Musik verwendet werden können. Nach [17, S.30] lassen sich die Merkmale in drei große Gruppen aufteilen: *„Klangfarbe und Energie“*, *„Chroma und Harmonie“* sowie *„Tempo, Rhythmus und Struktur“*. Die Klangfarbe beschreibt die Eigenart eines Instruments oder eines Spielstils, einen Ton mit gleicher Höhe und Lautstärke anders klingen zu lassen. Energiemerkmale konzentrieren sich auf die Lautstärke der Musikstücke [17, S.30]. Zur Extraktion werden meist kleine Fenstergrößen mit einer Breite von 512 Frames verwendet [17, S.135f]. Chroma und Harmonie beschreiben die Art und Weise, wie einzelne Töne angeordnet sind und wie sie zueinander in Beziehung stehen. Es können so beispielsweise Akkorde erkannt werden [17, S.31]. Die Merkmale Tempo, Rhythmus und Struktur werden meist über das gesamte Musikstück berechnet. Strukturmerkmale beschreiben den Aufbau und die Gliederung des Musikstücks [17, S.31]. Das Tempo eines Musikstücks gibt an, in welcher Geschwindigkeit die Noten des Liedes gespielt werden müssen [20, S.40]. Der Rhythmus beschreibt, in welchem relativen Verhältnis die Spieldauer der einzelnen Noten zueinander stehen [7, S.445].

2.2 Klassifikatoren

Die Aufgabe eines Klassifikators ist es, eine Menge von Elementen einer bestimmten Klasse zuzuordnen. Er entscheidet also, welcher Klasse das jeweilige Element angehört. Bei den Klassifikatoren in dieser Arbeit handelt es sich ausschließlich um binäre Klassifikatoren. Sie entscheiden also zwischen zwei Klassen - „Gesang“ und „kein Gesang“. Jedes Element erhält nach der Klassifikation ein *„Label“*, welches eindeutig anzeigt, welcher Klasse das Element angehört. In den meisten Fällen wird den Elementen schlicht der Name der Klasse zugeordnet.

Es gibt eine Vielzahl an Klassifikatoren, die mittels verschiedenster Verfahren entscheiden, welcher Klasse ein einzelnes Element angehört. Der simpelste Klassifikator ist der Zufallsklassifikator, welcher ein Element zufällig einer Klasse zuordnet. Bei zwei Klassen und einem ausgewogenen Datensatz können so ca. 50% der Elemente korrekt klassifiziert werden.

Ebenfalls recht simpel, jedoch deutlich genauer als der Zufall, ist die Idee in [13] über einen zuvor definierten Schwellwert eines Merkmals eine Entscheidung zu treffen. Dieser Klassifikator bezieht die zu klassifizierenden Daten also in seine Berechnungen ein.

Um das Ergebnis weiter zu verbessern, ist jedoch ein komplexerer Klassifikator notwendig. Die in dieser Arbeit getesteten Klassifikatoren beruhen alle auf dem Ansatz des *maschinellen Lernens*. Für diese Technik sind zwei Schritte nötig. In einem ersten Schritt muss der Klassifikator trainiert werden, bevor er schließlich in einem zweiten Schritt die Klassifikation vornehmen kann.

Hierzu wird der gesamte Datensatz in eine Trainings- und eine Testmenge unterteilt. Mit den Audiomerkmalen der Trainingsmenge wird der Klassifikator nun trainiert. Er erhält also Beispiele für Passagen mit und ohne Gesang, in denen er nun Muster suchen kann, die auf eine der beiden Klassen hinweisen. Dieser Vorgang wird auch *überwachtes Lernen* genannt. Der Lernprozess des Klassifikators kann durch die bereits bekannte Zuordnung zu einer Klasse überwacht werden. Im Gegensatz dazu steht das *unüberwachte Lernen*. Hierbei bildet der Klassifikator eigenständig Klassen, in die er die Beispiele einteilt. Grundlage für diese Einteilung sind dann ausschließlich die Gemeinsamkeiten, die in den Daten erkannt werden.

Nach dem Training kann nun die unbekannte Testmenge überprüft werden. Dazu werden die Audiomerkmale der Testelemente an dem Klassifikator übergeben. Dieser entscheidet nun, zu welcher der beiden Klassen das jeweilige Testelement am besten passt. Als Ergebnis liefert der Klassifikator für jedes Testelement ein eigenes Label.

Die Ausgabe des Klassifikators lässt sich in vier Gruppen *„True Positive (TP)“*, *„False Positive (FP)“*, *„True Negative (TN)“* und *„False Negative (FN)“* teilen. „TP“ beschreibt alle Elemente, die der Klassifikator korrekt als Gesang erkannt hat. „FP“ bezeichnet alle Elemente, die vom Klassifikator mit Gesang markiert wurden, jedoch keinen enthalten.

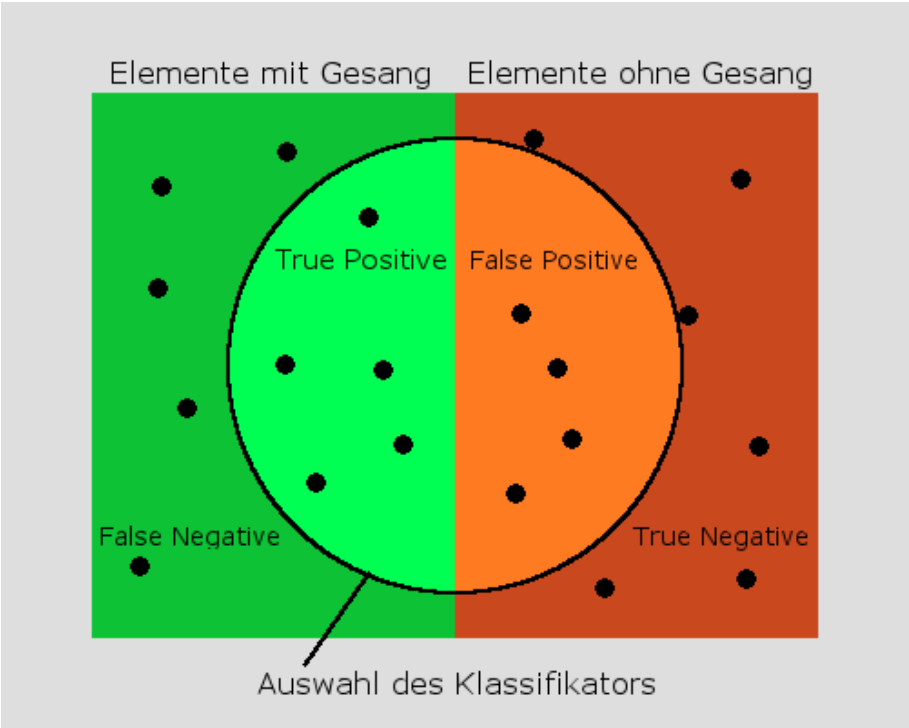

Abbildung 2.1: Unterteilung des Klassifikationsergebnises in „True Positive“, „False Positive“, „True Negative“ und „False Negative“.

„TN“ beschreibt die Elemente, die korrekt als gesangslos markiert wurden. Der letzten Gruppe „FN“ gehören alle Elemente an, die Gesang enthalten und vom Klassifikator als gesangslos erkannt wurden. Die Zusammenhänge werden in Abbildung 2.1 verdeutlicht.

Zur Bewertung des Klassifikators können nun mehrere Kennzahlen mit diesen vier Gruppen gebildet werden. Die wichtigsten sind *„recall“*, *„precision“* und *„f-measure“*. Der „recall“ beschreibt das Verhältnis von gefundenen Elementen zu den tatsächlich vorhandenen Elementen. Er ist in Gleichung 2.1 definiert.

Gleichung 2.2 zeigt die „precision“. Sie repräsentiert das Verhältnis von relevanten zu irrelevanten Elementen in der Ergebnismenge des Klassifikators.

Das „f-measure“ in Gleichung 2.3 bildet das harmonische Mittel aus „recall“ und „precision“. Somit stellt das „f-measure“ ein Maß dar, das die Qualität des Klassifikators an der Kombination aus „recall“ und „precision“ festmacht.

$$recall = \frac{TP}{TP + FN} \tag{2.1}$$

$$precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$fmeasure = 2 * \frac{precision * recall}{precision + recall} \tag{2.3}$$

Um zu verstehen, weshalb zur Bewertung drei Maße benötigt werden und nicht nur ein einziges ausreicht, sind die beiden folgenden Beispiele hilfreich. Angenommen, die Testmenge besteht zu 50% aus Gesangselementen und zu 50% aus gesangslosen Elementen. Ein Klassifikator, der die gesamte Testmenge als Gesang klassifiziert, hätte einen perfekten „recall“ mit 100%. Doch gut ist dieser Klassifikator nicht, denn die „precision“ ist, bedingt durch dieses Vorgehen, sehr schlecht. Ein zweiter Klassifikator erkennt in der gesamten Testmenge nur ein einziges Element als Gesang und er liegt damit richtig. Daraus ergibt sich eine „precision“ von 100%, da es keine falsch positiven Elemente gibt. Doch auch dieser Klassifikator arbeitet nicht gut, da sein „recall“ durch die vielen falsch negativen Elemente sehr schlecht ist. Erst das „f-measure“ zeigt in solchen Fällen, wie der Klassifikator in beiden Disziplinen abschneidet. Um den Klassifikator als ganzes zu bewerten, eignet sich somit das „f-measure“ besonders gut. Um das Verhalten des Klassifikators genauer zu untersuchen, können dann zusätzlich „recall“ und „precision“ betrachtet werden.

2.2.1 Random Forest

Der erste Klassifikator, der in dieser Arbeit verwendet wird, nennt sich Random Forest wurde 2001 von L. Breiman vorgestellt [3]. Dabei handelt es sich um eine Erweiterung eines seiner früheren Ansätze, dem *Bagging*, den er 1996 entwickelte [6, S.602]. Ziel war es, einen Klassifikator zu konstruieren, der mehrere ausgewogene Modelle kombiniert, um so die Klassifikationsgenauigkeit zu erhöhen [6, S.587].

Genau diese Idee setzt der Random Forest um. Er generiert aus den Eingabedaten mehrere Bäume. Ein *Baum* ist eine Datenstruktur, die ihren Namen der großen Ähnlichkeit zu einem Baum aus der Wirklichkeit verdankt. Am Startpunkt des Baumes, der *Wurzel*, befindet sich das erste Element. Von dort aus zweigen zwei *Äste* ab. An jedem dieser Äste befinden sich *Knoten*, in denen sich weitere Elemente befinden. Von diesen Knoten aus gehen weitere Äste mit weiteren Knoten ab. An diesen können sich ebenfalls weitere Äste befinden. Die letzten Knoten, an denen keine weiteren Äste hängen, werden *Blätter* genannt. Um zu den Blättern zu gelangen, muss der Computer einen Weg von der Wurzel aus bis zum Blatt durchlaufen. Mit den Informationen in den einzelnen Knoten entscheidet sich der Computer für einen der beiden Äste und gelangt so über den Ast zum nächsten Knoten.

Jeder Baum ist ein eigenständiges Modell, das bei einer Anfrage ein unabhängiges Ergebnis liefert. Um die Ungenauigkeit, verursacht durch Datenrauschen, zu unterdrücken, werden die Ergebnisse der einzelnen Modelle gemittelt, d.h. bei der Klassifikation wird die Klasse per Mehrheitsentscheid ausgewählt.

Der Aufbau eines einzelnen Baumes erfolgt dabei in mehreren Schritten. Im ersten Schritt wird eine zufällige Teilmenge des Trainingsdatensatzes ausgewählt, auf welcher der Baum aufgebaut werden soll; ein Baum nutzt folglich nie den gesamten Datensatz, sondern

immer nur einen kleinen Teil („*bootstrap sample*“) [6, S.588]. Die Wurzel des Baumes wird nun gebildet, indem zufällig m Merkmale der Datensätze ausgewählt werden. Typische Werte sind hier $\sqrt{p}$ oder 1, wobei p die Anzahl aller Merkmale ist [6, S.589]. Aus diesen Merkmalen wird nun das Merkmal gewählt, das, für sich genommen, eine korrekte Klassifikation am besten erreicht. Es wird nun in die Wurzel gesetzt. Zum Schluss werden zwei Äste mit je einem Knoten an die Wurzel gehangen, für die nun jeweils m neue Merkmale gezogen werden. Erneut wird das Merkmal mit dem besten Ergebnis ausgewählt und im jeweiligen Knoten abgelegt. Im Anschluss werden abermals zwei Kindknoten angehangen.

Das Verfahren wird nun so lange wiederholt, bis die minimale Anzahl an Knoten erreicht ist [6, S.588]. Zuletzt enthalten die Blätter des Baumes die Namen der Klassen. Abbildung 2.2 zeigt die Struktur eines einzelnen Baumes. Nachdem nun ein einzelner Baum aufgebaut wurde, wird die gesamte Prozedur für jeden weiteren Baum wiederholt. Der so entstehende Wald besteht nun, bedingt durch die zufällige Variablenauswahl, aus voneinander unabhängigen Bäumen. Für die Klassifikation kann nun jeder Baum ausgewertet werden, wobei das Endergebnis die Klasse ist, für die sich die meisten Bäume entschieden haben.

Im ersten Moment erscheint die Art und Weise, wie die Bäume erstellt werden, wahllos zu sein. Doch es verbirgt sich ein System hinter dem zufälligen Zusammenwürfeln der Bäume. Für eine genügend große Anzahl an Bäumen ($\geq$100) werden alle Fehlentscheidungen, die durch das Verwenden von falschen bzw. schlechten Variablen getroffen wurden, gemittelt, sodass Fehler immer weniger ins Gewicht fallen. Gleichzeitig treten aber jene Bäume, bei denen treffende Variablen gewählt wurden, hervor. Die Wahrscheinlichkeit,

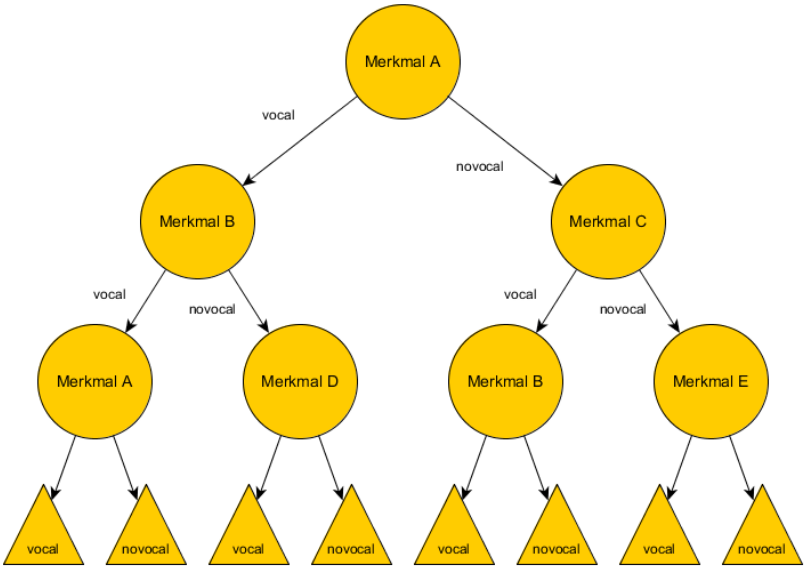

Abbildung 2.2: Schaubild eines einzelnen Entscheidungsbaumes bestehend aus sieben Knoten in drei Schichten. Zur Entscheidungsfindung werden die Merkmale A-E verwendet. Durch die zufallsbedingte Auswahl der Merkmale werden die Merkmale A und B mehrfach verwendet.

dass ein einzelner Baum einen Fehler macht, bleibt für jeden Baum gleich. Durch den Mehrheitsentscheid über viele Bäume kann die Fehlerwahrscheinlichkeit für den gesamten Wald jedoch erheblich reduziert werden.

Zuletzt bietet ein Random Forest zwei weitere wichtige Vorteile. Sowohl das Training, als auch das Testen ist parallelisierbar. Aufgrund der Tatsache, dass die Bäume auf Basis eines gemeinsamen Datensatzes erstellt werden, ansonsten aber keinerlei Zusammenhang haben, kann jeder Baum für sich erstellt und zur Klassifikation ausgewertet werden. Somit können moderne Mehrkernprozessoren optimal ausgelastet werden.

2.2.2 Neuronale Netze

Als zweiter Klassifikator kommt in dieser Arbeit ein neuronales Netz zum Einsatz. Die Funktionsweise des Netzes ist den Neuronen in einem menschlichen Gehirn nachempfunden. Es gibt sie in vielen Variationen, jedes für einen besonderen Einsatzzweck. Einen guten Überblick über neuronale Netze und deren verschiedene Typen geben die Werke [6] und [11]. Beide sind frei verfügbar. Bei dem von J. Schlüter und T. Grill in [15] verwendeten neuronalen Netz handelt es sich um ein *„Convolutional neural network (CNN)“*, zu deutsch „faltendes neuronales Netzwerk“, eine Sonderform eines neuronalen Netzes. Die Funktionsweise wird im Folgenden beschrieben.

Jedes neuronale Netz besteht aus mehreren *Knoten* oder auch *Neuronen*. Ein einzelnes Neuron setzt sich aus mehreren Eingängen zusammen, welche jeweils mit einem eigenen Kantengewicht multipliziert werden. Alle gewichteten Eingaben werden dann aufsummiert und erneut mit einem Gewicht versehen. Die Summe wird nun von einer nichtlinearen Funktion verarbeitet, deren Ergebnis der Ausgabewert des Neurons ist [11, S.383f]. Häufig wird die Sigmoid-Funktion als nichtlineare Funktion zur Realisierung eines Schwellwertes eingesetzt [11, S.385]. Abbildung 2.3 zeigt den Aufbau eines Neurons.

Mehrere Neuronen bilden gemeinsam eine *Schicht*. Mehrere Schichten ergeben dann zusammen ein *neuronales Netz*. Jedes Netz hat eine Eingabeschicht, sowie eine Ausgabeschicht. Die Eingabeschicht nimmt die zu verarbeitenden Daten entgegen und hat für jede Eingabevariable ein eigenes Neuron, die Ausgabeschicht gibt lediglich das Ergebnis aus. Sie hat für jede Ausgabevariable ein Neuron. Zwischen diesen beiden Schichten befinden sich ein oder mehrere versteckte Schichten („hidden layers“), die für die Verarbeitung der Daten verantwortlich sind. Sie haben keinen direkten Zugriff auf die Eingabe sowie die Ausgabe des Netzes. Jede Schicht innerhalb des Netzes ist nur mit der ihr folgenden Schicht verbunden. Es gibt keine Verbindungen zur vorherigen Schicht, auch Kanten zur übernächsten Schicht sind verboten. In den meisten Fällen verfügt jedes Neuron einer beliebigen Schicht über eine Kante zu jedem Neuron der Folgeschicht. Ausnahme ist hier nur die Ausgabeschicht [11, S.384]. Abbildung 2.4 zeigt den Aufbau eines kleinen neuronalen Netzes.

Bei der Verarbeitung einer Eingabe wird das Netz Schicht für Schicht ausgewertet. Für jedes Neuron werden die Eingaben gewichtet, aufsummiert und dann die nichtlineare Funktion auf sie angewendet. Das Ergebnis kann dann zusammen mit den Ergebnissen der anderen Neuronen dieser Schicht als Eingabe für die Neuronen der folgenden Schicht verwendet werden, bis die Ausgabeschicht schließlich das endgültige Ergebnis ausgibt [11, S.385f]. Hierbei können Berechnung der Neuronen aus der selben Schicht parallel durchgeführt werden. Die Neuronen der folgenden Schicht lassen sich jedoch erst berechnen, wenn die gesamte vorherige Schicht vollständig berechnet wurde.

Für eine korrekte Klassifikation muss das Netz zuvor jedoch noch trainiert werden. Mittels einiger Beispiele werden die Gewichte an den verdeckten Schichten so angepasst, dass bei Eingabe eines Trainingsbeispiels die gewünschte Klasse ausgegeben wird bzw. der Fehler so klein wie möglich wird. Somit findet auch hier, ähnlich wie beim Random Forest Klassifikator, ein überwachtes Lernen statt. Der Vorgang wird mehrfach wiederholt, bis der Ausgabefehler das gewünschte Niveau erreicht. Zur Klassifikation wird in der Regel für jede Klasse ein eigener Ausgabeknoten angelegt. Der Ausgabenknoten der korrekten Klasse sollte nach der Berechnung einen Wert möglichst nahe an Eins, die anderen Knoten eine Null

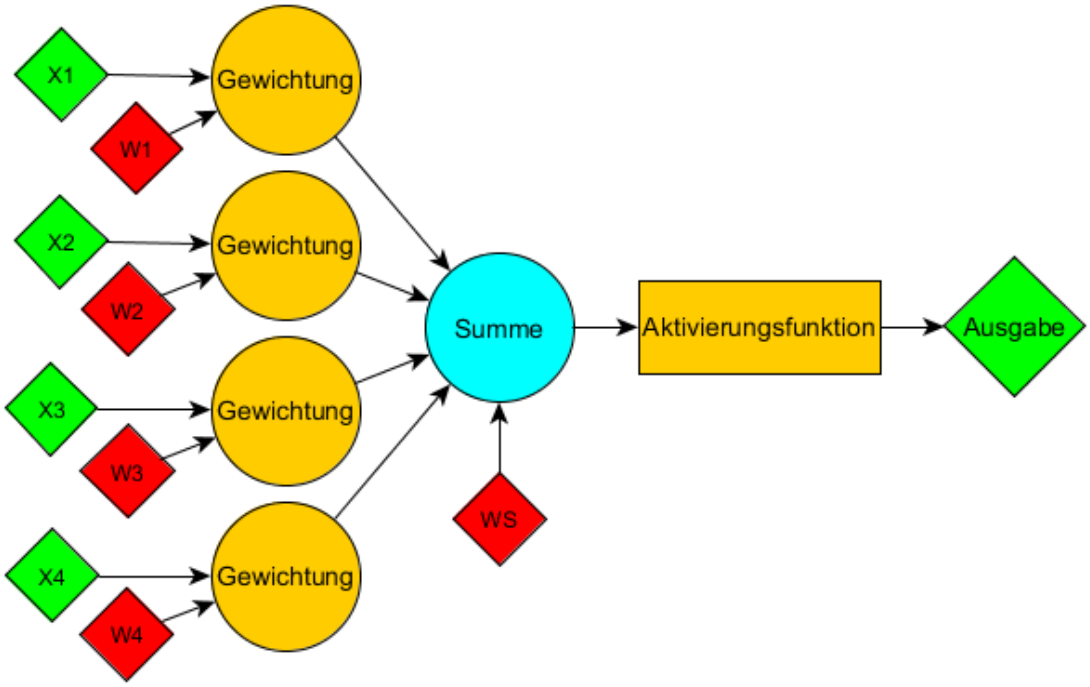

Abbildung 2.3: Aufbau eines Neurons mit vier Eingaben X1 bis X4. Sie werden mit den vier Gewichten W1 bis W4 verrechnet und aufsummiert. Anschließend wird die Summe mit WS erneut gewichtet, bevor die Aktivierungsfunktion die Ausgabe des Neurons berechnet. Die grün gefärbten Elemente zeigen Eingabe- und Ausgabefelder, die rot gefärbten Elemente sind Variablen, welche im Lernprozess angepasst werden können.

liefern. [11, S.387] Der Vorgang des Trainings wird *„error-back-propagation algorithm“*, zu deutsch *Fehlerrückführungs-Algorithmus*, genannt, da nach der Eingabe und Berechnung eines Trainingsbeispiels die Gewichte Schichtweise von der Ausgabeschicht aus bis hin zur Eingabeschicht neu berechnet werden. Der Fehler wird also in umgekehrter Richtung durch das Netz gereicht [11, S.389]. Durch das mehrfache Wiederholen dieses Vorgangs werden die Gewichte so lange angepasst, bis sich der Fehler minimiert. Dabei spielt auch die Anzahl der verwendeten Schichten eine große Rolle. Nach [11, S.389] sind zur Nachbildung einer logischen Funktion, zur Approximation einer nichtlinearen Funktion und zur Definition beliebiger Klassengrenzen bereits zwei gewichtete Schichten hinreichend. Eine höhere Anzahl an Schichten kann jedoch die Konvergenz des Trainings beschleunigen. Zur Klassifikation werden häufig drei gewichtete Schichten empfohlen [11, S.389]. In [6, S.400] wird eine simple Regel genannt: Besser zu viele verdeckte Knoten, als zu wenige.

Die Besonderheit eines faltenden neuronalen Netzes liegt nun darin, dass es zwischen der Eingabeschicht und der ersten verdeckten Schicht noch weitere Schichten enthält, die mehrere Eingaben miteinander kombinieren und diese danach an die Klassifikationsschichten übergeben werden. Durch diesen Vorgang, auch Faltung genannt, können, vor der eigentlichen Verarbeitung durch das Netz, Filter auf die Eingabedaten angewendet werden. In [15] beispielsweise, soll das CNN Spektogrammausschnitte verarbeiten, welche als

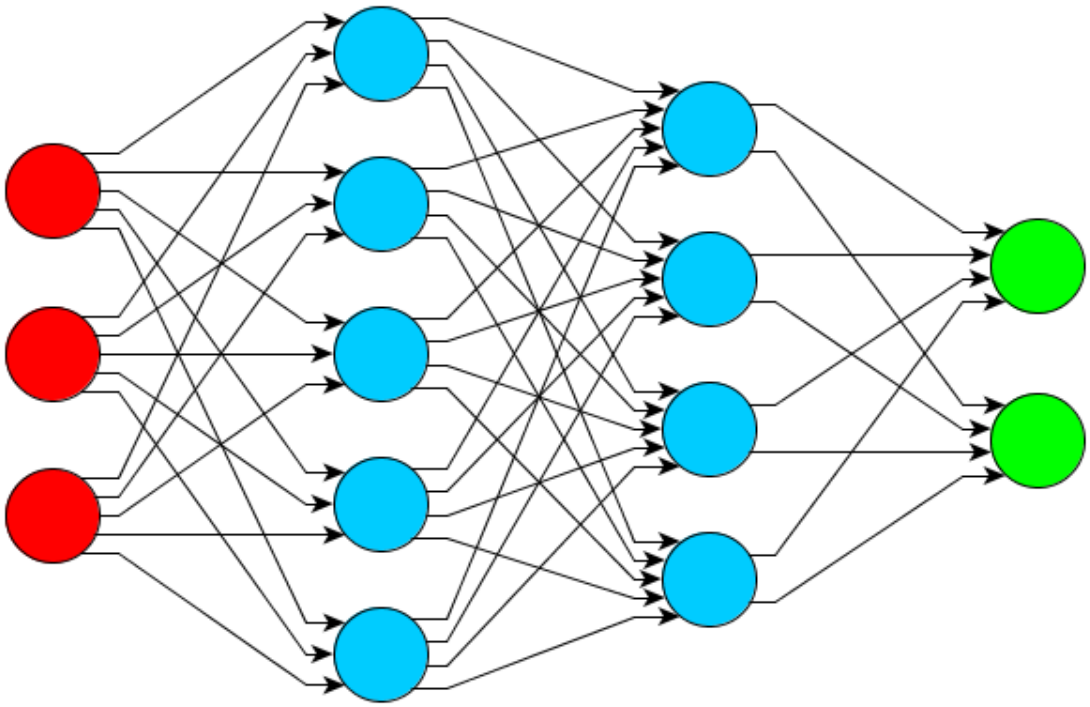

Abbildung 2.4: Aufbau eines neuronalen Netzes mit drei Eingaben und zwei Ausgaben: In rot die Eingabeschicht, in blau zwei verdeckte Schichten und in grün die Ausgabeschicht.

Pixelmatrix eingegeben werden. Zuvor werden die Daten durch eine Filterschicht jedoch noch einer Unterabtastung unterzogen, um die Datenmenge für das Netz zu reduzieren. Des Weiteren können solche Filter beispielsweise verschieden große Bilder auf eine einheitliche Größe bringen, eine Schärfe-Filter auf Bilder anwenden oder Musikdaten normalisieren.

2.3 Ansatz 1: J. Schlüter und T. Grill

Der erste Ansatz, der in dieser Arbeit verwendet wird, stammt von J. Schlüter und T. Grill und wurde 2015 in ihrer Arbeit *„Exploring data augmentation for improved singing voice detection with neural networks“* [15] beschrieben. Zur Gesangserkennung wurden Spektogrammauszüge der Testobjekte in Verbindung mit einem CNN verwendet. Insgesamt wurden vier Testdatensätze genutzt. Neben dem Jamendo-Datensatz und den ebenfalls frei verfügbaren Datensatz aus [5] wurden zwei weitere Datensätze angelegt. Der eine bestand aus 188 Hörproben mit einer Länge von jeweils 30 Sekunden, die einer Musikplattform entnommen wurden. Der andere bestand aus 149 Stücken Rockmusik.

Die Datensätze wurden außerdem zusätzlich mit mehreren Verfahren nachbearbeitet, um das Klassifikationsergebnis zu steigern. Zu den Verfahren zählten das zufällige Entfernen von Eingaben, das Einfügen von Datenrauschen, die Erhöhung der Tonhöhe, eine Streckung der Zeit, Erhöhung der *„loudness“* und das Anwenden eines Frequenzfilters.

Das neuronale Netz erreicht auf dem Jamendo-Datensatz ohne Nachbearbeitung einen „recall“ von 90,8% und eine *„specificity“* von 90,5%. Mit Nachbearbeitung der Trainings- sowie Testdaten verbesserte sich die „specificity“ auf 94,1%. Der „recall“ ging auf 90,3% zurück.

2.4 Ansatz 2: B. Lehner, G. Widmer und R. Sonnleiter

Der zweite Ansatz, der in dieser Arbeit getestet werden soll, wurde 2014 von B. Lehner, G. Widmer und R. Sonnleiter entwickelt. Sie stellten ihn in Ihrer Arbeit *„On the reduction of false positives in singing voice detection“* [9] vor. Sie stellten bei der Auswertung früherer Ergebnisse fest, dass das größte Problem bei der Gesangserkennung mittels MFCCs nicht etwa der „recall“ war, sondern die „precision“. Das System lieferte häufig falsch-positive Ergebnisse, die durch Instrumente hervorgerufen wurden, die die gleichen Charakteristika in der Tonhöhe aufweisen, wie die menschliche Stimme. Zu diesen Instrumenten zählen unter anderem Violinen und E-Gitarren.

Um genau solche Fehlklassifikationen zu vermeiden, wurden drei neue Merkmale entwickelt, die zusammen mit klassischen MFCCs ein deutlich verbessertes Ergebnis liefern sollen: Das *„Fluctogramm“* repräsentiert charakteristische Schwankungen in der Tonhöhe, die *„Spectral Contraction (SC)“* soll die Zuverlässigkeit eines bestimmten Frequenzban-

des angeben und die *„Vocal Variance (VV)“* soll die Varianz in Teilbereichen der MFCCs angeben.

Die Klassifikation erfolgt mittels eines Random Forest. Neben einem eigenen Testdatensatz bestehend aus 75 Stücken zum trainieren und 147 rein instrumentalen Stücken zum testen, wurde auch der Jamendo-Datensatz verwendet. Auf diesem konnte ein „recall“ von 86,2% erreicht werden. Die „precision“ lag bei 88,0% und das „f-measure“ bei 87,1%.

2.5 Ansatz 3: AMUSE-Merkmale

Der dritte Ansatz wurde für diese Arbeit entwickelt. Er setzt auf den Random Forest als Klassifikator und nutzt die Merkmale, die sich mit dem *„Advanced MUSic Explorer (AMUSE)“* extrahieren lassen. Dieser wurde 2010 von Vatolkin et al. vorgestellt [18]. Das Tool zeichnet sich dadurch aus, dass es eine Vielzahl an anderen Tools aus dem Bereich *„Music Information Retrieval (MIR)“* kombiniert und Schnittstellen für diese bietet. Die in diesem Ansatz genutzten Merkmale lieferten die Tools *„jAudio“*[10], *„RapidMiner“*[1], *„Matlab“*[2], *„MIRtoolbox“*[3] und *„Sonic Annotator“*[4].

Neben MFCCs werden eine Reihe von Merkmalen verwendet, die nicht speziell für die Gesangserkennung entwickelt wurden, sich jedoch möglicherweise auch zur Gesangserkennung eignen. Dieser Ansatz soll zeigen, inwieweit sich diese Merkmale auch zur Gesangserkennung nutzen lassen. Eine Liste aller verwendeten Merkmale befindet sich in Anhang A.4.

[1] https://rapidminer.com/, Abrufdatum: 10.11.2016

[2] https://de.mathworks.com/products/matlab/, Abrufdatum: 10.11.2016

[3] https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox, Abrufdatum: 10.11.2016

[4] http://www.vamp-plugins.org/sonic-annotator/, Abrufdatum: 10.11.2016

Kapitel 3

Ziele der Arbeit

Das Hauptaugenmerk dieser Arbeit liegt auf dem Vergleich verschiedener Musikgenres bei der Gesangserkennung. Bisherige Arbeiten haben stets einen Datensatz genutzt, der eine breite Mischung aus verschiedenen Musikrichtungen und Stilen enthielt. Da es jedoch eine enorme Vielfalt an verschiedensten Musikrichtungen und Stilen gibt, die zum Teil einen völlig gegensätzlichen Charakter haben, soll in dieser Arbeit nun untersucht werden, ob sich Unterschiede erkennen lassen, wenn die Klassifikation in Abhängigkeit der Musikrichtung durchgeführt wird.

Gibt es eine Musikrichtung, bei der die Erkennung von Gesang generell besonders gut funktioniert? Oder gibt es eine Musikrichtung, bei der die Klassifikation ausgesprochen schlecht funktioniert? Zeigen sich bei den drei Ansätzen abhängig von der Musikrichtung Stärken und Schwächen, die bei einem gemischten Datensatz nicht auftreten? Ist es möglicherweise sinnvoll, zum Training ein anderes Genre zu nutzen, als zum Testen? All diese Fragen sollen in dieser Arbeit untersucht und geklärt werden.

3.1 Testdaten

Bisher wurde bei Vergleichen verschiedener Ansätze oft der von Ramona et al. in [12] eingeführte Jamendo-Datensatz verwendet. Er beinhaltet insgesamt 93 Lieder und ist in drei nicht überlappende Teile geteilt: 61 Stücke zum Trainieren, 16 zum Testen und weitere 16 zur Validierung. Die Lieder wurden zufällig aus der Jamendo-Datenbank entnommen. Sie sind frei verfügbar und stehen unter der „Creative Commons“-Lizenz. Der Datensatz steht inklusive Annotationen auf der Website von Ramona zur Verfügung[1].

Trotz der zufälligen Auswahl der Stücke enthält der gesamte Datensatz jedoch kein einziges elektronisches Musikstück. Somit ist ein großer Teil an Musikstilen nicht abgedeckt und es besteht die Möglichkeit der Überanpassung auf nicht elektronische Musikrichtungen. Mit Musik aus den elektronischen Stilen *Dubstep*, *Hardstyle* und *Trance* soll nun geprüft

[1] http://www.mathieuramona.com/wp/data/jamendo/, Abrufdatum: 10.11.2016

werden, ob die Testkandidaten auch mit elektronischer Musik brauchbare Ergebnisse liefern oder tatsächlich eine Überanpassung vorliegt.

Die drei Musikstile unterscheiden sich dabei grundlegend voneinander. Während Dubstep meist minimalistisch, aber bassbetont gehalten ist, lässt sich Hardstyle am ehesten mit den Attributen „laut“ und „wild“ versehen, wobei er jedoch auch meist von einem dominanten Bass geprägt ist. Trance wiederum zeichnet sich durch ruhige und harmonische Melodien aus, oft untermalt von einem Nachhall.

Die zweite Hälfte des Datensatzes besteht aus den nicht elektronischen Genres *Jazz*, *Klassik* und *Rock*. Sie wurden zum einen ausgewählt, um einen Kontrast zu den drei elektronischen Musikstilen zu bieten. Zum anderen enthalten die drei Genres besondere Charakteristika, welche die Algorithmen im Speziellen fordern. Laut [9] werden falsch-positiv Ergebnisse häufig durch Violinen und E-Gitarren verursacht. Um diesen Effekt genauer zu untersuchen, wurden die Genres Klassik und Rock gewählt.

Das Spektogramm in Abbildung 3.1 zeigt einen Ausschnitt aus einem Saxophon-Solo eines Jazz-Stücks. Zu erkennen ist, dass die erzeugten Muster denen des Gesangs in Abbildung 1.1 ähnlich sehen. Da der in [15] vorgestellte Ansatz zur Klassifikation Spektogramme auswertet, kann so untersucht werden, ob sich das neuronale Netz durch von Saxophonen geprägte Jazz-Stücke verwirren lässt.

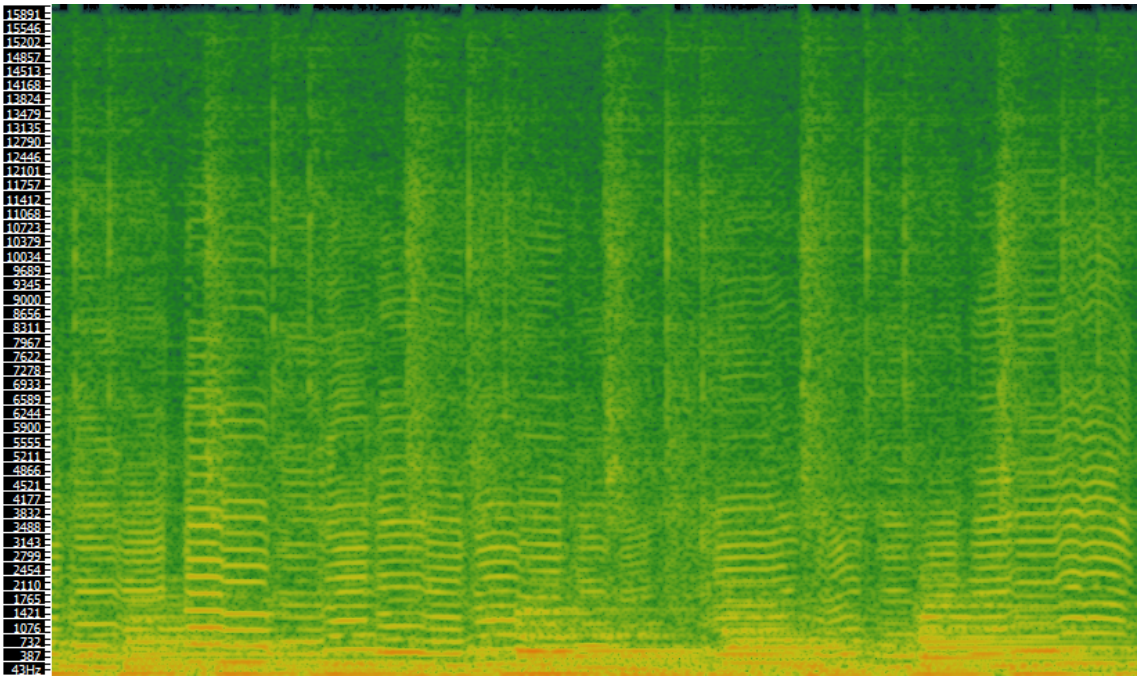

Abbildung 3.1: Spektogramm aus „John Coltrane - Giant Steps“: Saxophon-Solo, Zeitabschnitt von ca. 2 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse. Farbverlauf von Schwarz über grün und gelb bis hin zu rot. Schwarz entspricht fehlender Frequenz, rot zeigt eine stark vorhandene Frequenz an.

Insgesamt hat der Testdatensatz eine Länge von etwas mehr als zehn Stunden, davon ca. vier Stunden mit und sechs Stunden ohne Gesang. Ca. 20 Minuten wurden ignoriert (siehe Kapitel 4.1). Eine detaillierte Aufstellung zu allen Genres findet sich im Anhang A.1.

3.2 Parameterevaluation

Für den Ansatz aus [9] sowie den in Kapitel 2.5 beschriebenen Ansatz soll des Weiteren eine Parameterevaluation für das Training vorgenommen werden. Hierbei soll zum einen die Konfiguration des Random Forest variiert werden, um mögliches Optimierungspotenzial auf Seiten des Klassifikators zu zeigen. Zum Anderen soll die Größe des Bearbeitungsfensters evaluiert werden, um dessen Auswirkungen auf „recall“, „precision“ und „f-measure“ zu untersuchen.

Für den Ansatz auf Basis der AMUSE-Merkmale wird zusätzlich die Zusammensetzung der Merkmale variiert, um so die Auswirkungen von MFCC-Merkmalen auf das Klassifikationsergebnis genauer untersuchen zu können.

Kapitel 4

Experimente

Im folgenden Kapitel werden nun die für die Arbeit durchgeführten Experimente beschrieben. Das verwendete Testsystem bestand aus einem *Intel Xeon E3-1231v3, 3,4Ghz, 4 Kerne, 8 Threads*, *16GB DDR3-1600 RAM*, *Gigabyte GeForce GTX 970 G1 Gaming, 4GB GDDR5* sowie einem *Gigabyte GA-H97-D3H*. Es wurden keine Hardwaremodifikationen vorgenommen.

Die Tests für den Ansatz aus [15] wurden unter *Ubuntu 14.04 LTS x64* mit GPU-Beschleunigung über die Bibliothek *Theano*[1] durchgeführt. Die Tests auf dem vortrainierten Netz dauerten zwischen fünf und zehn Sekunden.

Die Tests der anderen Ansätze wurden unter *Windows 7 Professional x64* durchgeführt. Die Steuerung erfolgt über ein hierfür geschriebenes Java-Programm. Der Random Forest sowie die Evaluationsfunktionen wurden durch das Tool *Weka*[2] bereitgestellt. Das Erstellen der Modelle konnte mittels Parallelisierung auf acht Threads beschleunigt werden. Die benötigte Rechenzeit schwankte abhängig von der Testkonfiguration stark. Im „leave one out“ Verfahren pro Genre mit großem Bearbeitungsfenster und wenigen Bäumen benötigte das System für einen Fold circa zehn Sekunden. Bei der Kreuzvalidierung auf dem gesamten Testdatensatz mit kleinem Bearbeitungsfenster und vielen Bäumen nahm ein Fold circa eine Stunde in Anspruch.

4.1 Annotation der Testdaten

Zur Annotation der Musikstücke wurde die Software Sonic Visualizer[3] verwendet. Insgesamt 15 Stücke (6 Rock, 9 Klassik) wurden von Bernhard Lehner von der Johannes Kepler Universität Linz zur Verfügung gestellt. Die verbleibenden 135 Musikstücke wurden selbst annotiert. Bei der Annotation ist besondere Sorgfalt gefragt. Da bei der späteren Klassifikation mit Fenstergrößen von 200 Millisekunden und kleiner gearbeitet wird, muss die dafür

[1] https://github.com/Theano/Theano, Abrufdatum: 28.10.2016
[2] http://www.cs.waikato.ac.nz/ml/weka/index.html, Abrufdatum: 28.10.2016
[3] http://sonicvisualiser.org/, Abrufdatum: 28.10.2016

zugrunde liegende Annotation so genau wie möglich sein. Für ein genaues Ergebnis reicht es daher bei Weitem nicht aus, sich auf das Gehör zu verlassen. Der Sonic Visualizer bietet dazu einige hilfreiche Funktionen, die eine genaue Markierung ermöglichen. Eine erste grobe Verortung des Gesangs kann mittels Gehör und der Wellenform des Sonic Visualizer vorgenommen werden. Für die präzise Positionierung empfiehlt es sich, das Spektogramm zu betrachten. In Abbildung 4.1 lassen sich auf der rechten Bildhälfte gut die für Gesang typischen wellenartigen Muster erkennen, anhand derer genau annotiert werden kann. Hier gilt es jedoch zu beachten, dass der Gesang nicht immer optisch zu erkennen ist, wie es in Abbildung 4.2 der Fall ist. Daher ist ausschließliches Verwenden des Spektogramms nicht zu empfehlen. Eine Hörprobe sollte in jedem Fall erfolgen.

Damit ein Klassifikator korrekt arbeiten kann, benötigt er als Grundlage möglichst eindeutige Beispieldaten, anhand derer er sein Modell aufbauen kann. Wie in der Einleitung bereits beschrieben, kann der Begriff Gesang verschieden interpretiert werden. Um strittigen Punkten bei der Klassifikation aus dem Weg zu gehen, sind einige Passagen in den Musikstücken mit „vielleicht“ markiert. Diese Passagen werden weder zum Trainieren, noch zum Testen verwendet. Beispiele für solche Passagen sind ein langer Nachhall, wie im Lied „Wasted Penguinz & Chris One - I Miss You (Extended Version)“ ab ca. 1:44, elektronisch geloopte Wortsilben, wie zu Beginn des Liedes „Cosmic Gate - London Rain“ oder durch einen Synthesizer erzeugte Töne, die der menschlichen Stimme sehr nahe kommen, wie beispielsweise in „Skrux - Essence“ ab ca. 0:47.

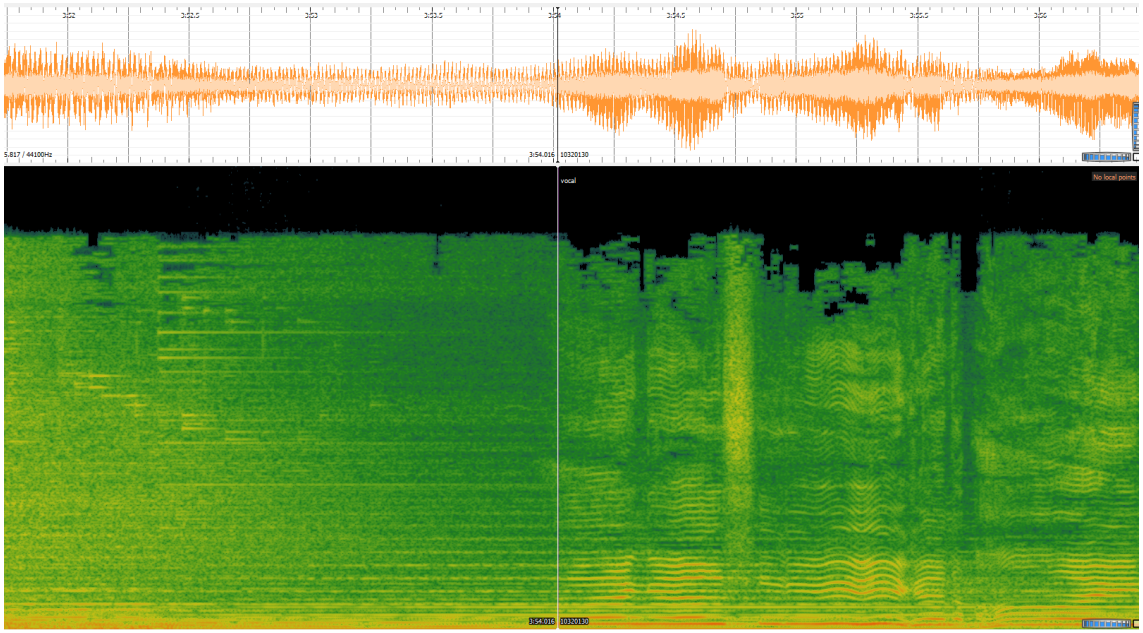

Abbildung 4.1: Wellenform mit Spektogramm aus: „Adoro - Engel fliegen einsam“, Zeitabschnitt von ca. 5 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse, links der Bildmitte ohne Gesang, rechts mit Gesang.

4.2 Testverfahren

Da für den Ansatz aus [15] nur ein mit dem Jamendo-Datensatz vortrainiertes neuronales Netz vorliegt, das nicht mehr nachbearbeitet oder trainiert werden kann, wird für diesen Ansatz jedes Lied einzeln getestet. Die Ergebnisse werden im Anschluss genreweise zusammen gefasst, sodass zum Schluss Messwerte für jedes Musikstück einzeln sowie ein Durchschnitt pro Genre vorliegen.

Bei den beiden anderen Ansätzen kann Einfluss auf das Training genommen werden, sodass hier umfangreichere Tests möglich sind. Neben einer Kreuzvalidierung nach dem Schema *„leave one out“* auf jedem Genre separat wird außerdem eine Konfusionsmatrix über alle Genres aufgebaut, indem jeweils mit dem einen Genre trainiert und mit dem anderen Genre getestet wird. In einem dritten Test wird des Weiteren eine zehnfache Kreuzvalidierung auf dem gesamten Datensatz vorgenommen. Die Aufteilung der Testdaten auf die zehn Testgruppen ist in Tabelle A.3 im Anhang A.1 zu finden. Die Gruppen wurden zufällig generiert. Da aber einige Interpreten mehrfach im Testdatensatz enthalten sind, wurde die generierte Aufteilung so modifiziert, dass sich Musikstücke des gleichen Interpreten auch in der gleichen Gruppe befinden.

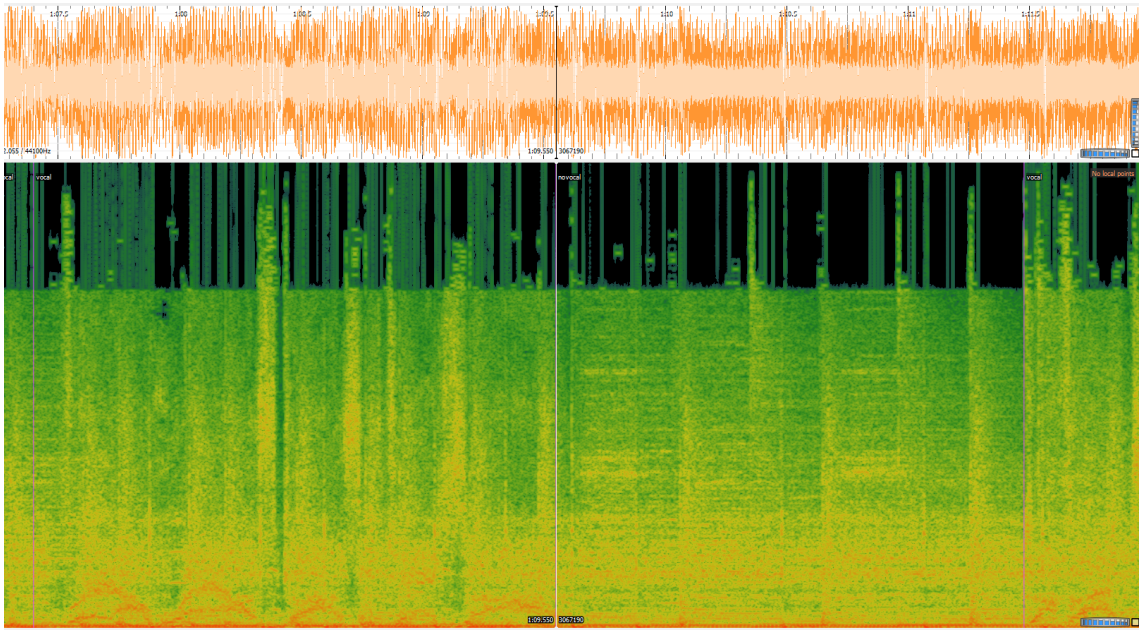

Abbildung 4.2: Wellenform mit Spektogramm aus: „Alexisonfire - A dagger through the heart of St. Angeles“, Zeitabschnitt von ca. 5 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse, links der Bildmitte mit Gesang, rechts ohne Gesang.

4.3 Ansatz 1: J. Schlüter und T. Grill

Für die Tests des Ansatzes aus [15] wurde ein auf dem Jamendo-Datensatz vortrainiertes neuronales Netz mit Augmentation der Trainingsdaten verwendet. Die Auswertung wurde für jedes Musikstück einzeln durchgeführt. Als Ausgabe lieferte das Programm für jedes Lied eine Liste mit Startzeitpunkten und einem Label zur Identifikation (0 für kein Gesang, 1 für Gesang). Diese Ergebnisse wurden mit den Annotationen verglichen und so Werte für „False Positive“, „True Positive“, „False Negative“ und „True Negative“ berechnet.

Für die Ergebnisse des jeweiligen Genres wurde der Durchschnitt über alle Lieder des betreffenden Genres gebildet. Die Ergebnisse sind in Tabelle 4.1 zu sehen. Ein direkter Vergleich der Genres anhand von „recall“, „precision“ und „f-measure“ wird in Abbildung 4.3 gezeigt.

Genre	FP	TN	FN	recall	precision	f-meas.	b. Fehler
Dubstep	12,23%	87,77%	29,26%	70,74%	88,21%	74,63%	21,12%
Hardstyle	7,32%	92,68%	16,89%	83,11%	72,06%	74,54%	12,15%
Jazz	12,13%	87,87%	8,59%	91,41%	90,01%	90,43%	11,76%
Klassik	11,60%	88,40%	16,46%	83,54%	93,49%	87,53%	15,86%
Rock	7,18%	92,82%	14,02%	85,98%	93,98%	89,06%	10,60%
Trance	4,29%	95,71%	15,36%	84,64%	95,11%	89,18%	10,27%
Gesamt	9,12%	90,88%	16,76%	83,24%	88,81%	84,23%	13,62%

Tabelle 4.1: Ergebnisse Ansatz 1 aufgeschlüsselt nach Genre: „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

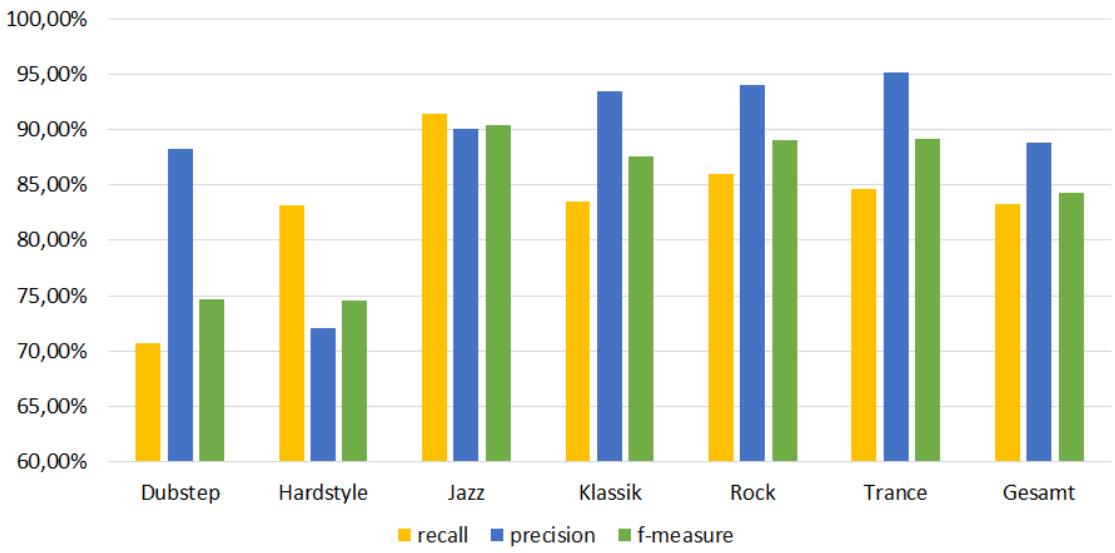

Abbildung 4.3: „Recall“, „precision“ und „f-measure“ für Ansatz 1 aufgeschlüsselt nach Genre und gesamt.

Das neuronale Netz liefert für alle sechs Genres deutlich bessere Ergebnisse als der Zufall. Es ist jedoch zu erkennen, dass die Genres Dubstep und Hardstyle sichtbar schlechter abschneiden, als die verbleibenden vier Genres. Das „f-measure“ weicht bei beiden um mehr als 10% vom Durchschnitt über alle Genres ab. Im Vergleich zum Genre Jazz, bei dem das „f-measure“ am größten ist, beträgt der Unterschied schon mehr als 15%. Während bei Hardstyle besonders die niedrige „precision“ auffällig ist, sticht bei Dubstep der sehr schlechte „recall“ hervor. Somit müssen überdurchschnittlich viele Gesangselemente als gesangslos markiert worden sein. Wie in Anhang A.1 zu sehen ist, wurden während der Annotation bereits mehr als fünf Minuten strittige Passagen, die solch hohe falsch-negativ Werte verhindern sollten, mit „vielleicht“ markiert und so von der Bewertung ausgeschlossen.

In Folge des niedrigen „recall“ im ersten Testdurchlauf wurde daher ein zweiter Test durchgeführt. Für diesen Durchlauf wurden die Annotationen für die Genres Dubstep und Hardstyle jedoch überarbeitet und weitere Passagen mit „vielleicht“ markiert, die zuvor als Gesang gewertet wurden. Dadurch erhöhte sich die ignorierte Zeit für das Genre Dubstep von ca. 5,6 auf ca. 7,5 Minuten, für Hardstyle leicht von 3,0 auf 3,6 Minuten. Tabelle 4.2 zeigt das Ergebnis für Dubstep und Hardstyle nach der Anpassung. Es ist zu sehen, dass auch weiteres Entfernen von möglicherweise strittigen Gesangspassagen nur eine kleine Verbesserung der Ergebnisse lieferte. Somit können diese Werte nicht auf eine mangelhafte Annotation zurück geführt werden, sondern sie müssen durch die Charakteristika des jeweiligen Genres verursacht worden sein.

Im Gegensatz zu den Genres Dubstep und Hardstyle lieferte das neuronale Netz für die verbleibenden vier Genres deutlich bessere Ergebnisse. Besonders die „precision“ von jeweils über 90% sticht hervor; für Trance lag die „precision“ sogar bei über 95%. Insgesamt schnitt das neuronale Netz auf dem Genre Jazz am besten ab. „recall“, „precision“ und „f-measure“ liegen alle auf einem Niveau knapp über 90%. Bei Klassik, Rock und Trance hingegen ist ein deutlicher Unterschied von jeweils ca. 10% zwischen „recall“ und „precision“ zu erkennen. Das lässt darauf schließen, dass das Modell des neuronalen Netzes zu sehr auf die zugrunde liegenden Trainingsdaten eingestellt ist. Der Jamendo-Datensatz enthält sowohl Klassik- als auch Rock-Titel. Mit dem Training auf diesem Datensatz sollten die Charakteristika

Genre	**FP**	**TN**	**FN**	**recall**	**precision**	**f-meas.**	**b. Fehler**
Dubstep	12,29%	87,71%	26,96%	73,04%	87,83%	75,72%	20,01%
Hardstyle	7,32%	92,68%	15,85%	84,15%	71,57%	74,75%	11,63%

Tabelle 4.2: Ergebnisse Ansatz 1 für Dubstep und Hardstyle nach Anpassung der Annotation: „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

der beiden Genres dem Netz also bekannt sein und trotzdem gibt es eine Abweichung von mehr als 5% von den Referenzwerten, die in [15] genannt wurden.

Da dem Trainingsdatensatz elektronische Musik fehlt, kann eine weitere interessante Beobachtung gemacht werden: Obwohl Dubstep, Hardstyle und Trance alle drei als elektronisch bezeichnet werden können, schneidet Trance im Vergleich zu Dubstep und Hardstyle deutlich besser ab. Somit hat der Ansatz kein generelles Problem mit elektronischer Musik an sich, aber sehr wohl mit zwei der drei gewählten elektronischen Stilen.

Insgesamt konnte dieser Ansatz mit den hier verwendeten Testdaten nicht an die Ergebnisse aus [15] heran reichen. Bei der Bewertung muss jedoch bedacht werden, dass für diesen Ansatz nur ein vortrainiertes Netz vorlag, sodass die Testdaten für das Netz als unbekannt einzustufen sind. Vor diesem Hintergrund liefert das neuronale Netz sehr gute Ergebnisse. Das Genre Jazz erreichte sogar annähernd die Ergebnisse aus der Veröffentlichung. Auf den Genres Dubstep und Hardstyle konnte das Netz nicht überzeugen.

4.4 Ansatz 2: B. Lehner, G. Widmer und R. Sonnleiter

Für die Tests des Ansatzes aus [9] wurde ein Random Forest mit zuvor extrahierten Merkmalen trainiert. Die Merkmale wurden von Bernhard Lehner extrahiert und für diese Arbeit zur Verfügung gestellt. Die Daten lagen in separaten Dateien vom Typ „arff“ ungelabelt für jedes Musikstück vor. Jedes Element repräsentierte einen Zeitabschnitt von 200ms und bestand aus insgesamt 116 Merkmalswerten. In einem Vorverarbeitungsschritt wurden die Elemente mit einem Label versehen. Dazu wurde die Position des jeweiligen Ausschnitts innerhalb des Musikstücks mit den Annotationen verglichen. Gesangsausschnitte wurden mit „vocal“, gesangslose Abschnitte mit „novocal“ markiert. Mit „vielleicht“ zu labelnde Elemente wurden entfernt und nicht zur Auswertung herangezogen. Konnte ein Element nicht eindeutig einer Klasse zugeordnet werden, da es sich genau auf einer Schnittstelle zwischen zwei Bereichen befand, so wurde es ebenfalls entfernt.

Bei allen Tests wurde eine Parameterevaluation durchgeführt. Zum Einen wurde die Fenstergröße von 200ms auf 400ms erhöht. Eine weitere Vergrößerung des Fensters schien wenig sinnvoll, da sich mit jeder weiteren Vergrößerung die Genauigkeit der Verortung innerhalb des Musikstücks reduziert. Da die standardmäßige Anzahl an Bäumen im Random Forest 100 beträgt, wurden mit 50 und 200 zum Anderen zwei weitere Konfigurationen unter- und oberhalb der Standardeinstellung getestet.

Um die für die Parameterevaluation benötigte zweite Fenstergröße von 400ms zu erreichen, wurden jeweils aus zwei Elementen der ungelabelten Daten das arithmetische Mittel gebildet. Danach wurde das zuvor beschriebene Verfahren zur Beschriftung der Elemente durchgeführt.

4.4.1 Kreuzvalidierung und Parameterevaluation

Bei der Kreuzvalidierung auf dem gesamten Datensatz zeigte sich, dass die von Lehner et al. extrahierten Merkmale unempfindlich auf veränderte Parameter reagieren. Tabelle 4.3 zeigt die Ergebnisse aus allen sechs Varianten der Parameterevaluation. Es ist zu sehen, dass sich das Klassifikationsergebnis verbessern ließ, indem das Klassifikationsfenster vergrößert wurde. Mit einer größeren Anzahl an Bäumen im Random Forest Klassifikator konnten zwar „precision“ und „f-measure“ erhöht werden, der „recall“ ging aber leicht zurück.

Insgesamt wird das beste Ergebnis mit einer Fenstergröße von 400ms und 200 Bäumen erreicht. Mit dieser Konfiguration liegt das „f-measure“ bei 81,95%. Im Vergleich dazu liegt das „f-measure“ der schlechtesten Konfiguration mit einem 200ms Fenster und 50 Bäumen bei 81,18%. Der Unterschied beträgt folglich weniger als 1%. Zur Bewertung der optimalen Konfiguration muss nun betrachtet werden, welche Folgen diese angepasste Konfiguration neben leicht höheren „f-measure“ hat.

Die Erhöhung der Baumanzahl führt zu einer erhöhten Rechenzeit während der Erstellung des Random Forest. Die vierfache Baummenge bedeutet eine etwa vierfache Rechenzeit. Dies lässt sich durch einen ausreichend starken Computer kompensieren. Die Vergrößerung des Klassifikationsfensters führt zu einem Verlust an Präzision was die Verortung von Gesangspassagen innerhalb des Musikstücks angeht. Dies lässt sich allerdings nicht kompensieren. Daher sollte die Fenstergröße nur so lange erhöht werden, wie sich auch deutliche Verbesserungen in den Ergebnissen zeigen. Bei einer gewählten Anzahl von 200 Bäumen liegt das „f-measure“ mit einem 200ms Fenster bei 81,54%. Mit einer Fenstergröße von 400ms kann nur ein kleiner Gewinn um 0,41% auf 81,95% erreicht werden. Daher ist die Vergrößerung des Klassifikationsfensters für diesen Ansatz nicht sinnvoll.

Betrachtet man den Gewinn von 50 zu 100 Bäumen, so ist dieser zwar klein, aber deutlich größer als von 100 zu 200 Bäumen. Anstatt die Laufzeit zu vervierfachen, muss sie für 100 Bäume nur verdoppelt werden. Für die meisten modernen Computer mit Mehrkern-

win.	trees	FP	TN	FN	recall	preci.	f-meas.	b. Feh.
200ms	50	10,18%	89,82%	20,48%	79,52%	83,24%	81,18%	15,33%
200ms	100	9,68%	90,32%	20,56%	79,44%	83,89%	81,45%	15,12%
200ms	200	9,43%	90,57%	20,66%	79,34%	84,19%	81,54%	15,05%
400ms	50	9,89%	90,11%	20,04%	79,96%	83,47%	81,50%	14,97%
400ms	100	9,14%	90,86%	20,19%	79,81%	84,48%	81,90%	14,67%
400ms	200	8,92%	91,08%	20,34%	79,66%	84,72%	81,95%	14,63%

Tabelle 4.3: Zehnfache Kreuzvalidierung mit Ansatz 2 auf dem gesamten Testdatensatz: Fenstergröße, Anzahl der Bäume im Random Forest, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

prozessoren sollten 100 Bäume kein Problem darstellen, sodass hier der Standardwert des Tools Weka zu empfehlen ist.

Vergleicht man die Ergebnisse, die auf dem hier verwendeten Datensatz erreicht wurden, mit denen aus [9] auf dem Jamendo-Datensatz, so lässt sich ein deutlicher Unterschied erkennen. Das „f-measure“ auf dem Jamendo-Datensatz liegt bei 87,1% und somit ca. 6% höher als auf dem hier getesteten Datensatz. Die „precision“ liegt mit 88,0% etwa 4% höher und der „recall“ liegt mit 86,2% ca. 7% über dem hier erreichten „recall“. Dieser Ansatz hat folglich mehr Probleme den Datensatz aus dieser Arbeit zu klassifizieren, als den Jamendo-Datensatz. Um zu untersuchen, ob einzelne Genres für das schlechtere Abschneiden verantwortlich sind, oder ob alle Genres gleichermaßen mehr Schwierigkeiten verursacht haben, wurden weitere Tests im „leave one out“ Verfahren durchgeführt.

4.4.2 „Leave one out“

Das „leave one out“ Verfahren wurde für jedes Genre separat durchgeführt, um die einzelnen Genres so voneinander zu trennen. Die Ergebnisse der 25 einzelnen Durchläufe pro Genre wurden nach Abschluss gemittelt. Die vollständige Tabelle mit Ergebnissen aus allen Parameterkonfigurationen findet sich im Anhang A.2 in Tabelle A.4. Der Tabelle ist zu entnehmen, dass die Parameterevaluation auch bei einer genreweisen Kreuzvalidierung keinen großen Einfluss auf das Klassifikationsergebnis genommen hat. Daher enthält Tabelle 4.4 nur einen Auszug aus der Gesamttabelle mit der im vorherigen Abschnitt gewählten Parameterkonfiguration.

In Abbildung 4.4 werden die Unterschiede zwischen den Genres besonders deutlich. Mit einem „f-measure“ von 94,48% schnitt Klassik am besten ab. Das zweitbeste Ergebnis konnte mit Rock erzielt werden. Das „f-measure“ lag dort bei 88,02%. Deutlich schlechter waren die Ergebnisse für das „f-measure“ bei den Genres Dubstep (78,02%), Hardstyle (69,01%) und Jazz (78,33%). Während das niedrige „f-measure“ bei Dubstep vor allem durch die Kombination aus vielen falsch Positiven und falsch Negativen hervorgerufen wird,

Genre	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-meas.**	**b. Feh.**
Dubstep	22,20%	77,80%	22,51%	77,49%	82,44%	78,02%	22,08%
Hardstyle	3,00%	97,00%	34,74%	65,26%	81,13%	69,01%	18,90%
Jazz	10,37%	89,63%	28,69%	71,31%	90,51%	78,33%	21,04%
Klassik	8,53%	91,47%	6,44%	93,56%	95,93%	94,48%	8,15%
Rock	14,73%	85,27%	12,07%	87,93%	89,73%	88,02%	13,40%
Trance	7,82%	92,18%	17,31%	82,69%	90,86%	85,27%	13,75%

Tabelle 4.4: „Leave one out“ mit Ansatz 2 auf jedem Genre getrennt mit einer Fenstergröße von 200ms und 100 Bäumen im Random Forest: Genre, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

erreicht Jazz eine hohe „precision“ von 90,51%, was mit wenigen falsch Positiven verbunden ist. Allerdings erreicht der „recall“ nur 71,31%. Das gleiche Problem ist bei Hardstyle zu beobachten. Hier liegen sowohl „recall“ als auch „precision“ mit 65,26% bzw. 81,13% noch deutlich unter den erreichten Werten von Jazz. Mit nur 3% kann Hardstyle jedoch bei den falsch Positiven mit Abstand am besten abschneiden.

Aus diesen Ergebnissen lässt sich schließen, dass der Klassifikator bei den Genres Hardstyle und Jazz ein zu restriktives Modell aufgebaut hat. Es gibt zwar vergleichsweise wenige falsch Positive, allerdings auch eine sehr hohe Anzahl an falsch Negativen. An dieser Stelle sei an das zweite Beispiel aus Kapitel 2.2 erinnert. Der dort beschriebene Klassifikator hat nahezu alle Elemente als gesangslos klassifiziert und so eine hohe „precision“ erreicht, den „recall“ dadurch aber extrem verschlechtert. Der zweite Ansatz hat genau das gleiche Problem auf den Genres Hardstyle und Dubstep. Es ist jedoch nicht so extrem ausgeprägt wie im Beispiel geschildert. Hier könnten zur Lösung des Problems weitere Trainingsbeispiele helfen, um das Modell allgemeiner auszurichten und so empfänglicher für andere Gesangselemente zu machen. Diese Beispiele sollten aber so gewählt sein, dass die „precision“ nicht unter dem weiteren Training zu leiden hat. Ansonsten bewegt sich das modifizierte Modell auf ein Problem zu, dass sich bei den Experimenten mit dem Genre Dubstep erkennen lässt.

Dort liegt das „f-measure“ unter 80% und da „recall“ und „precision“ in etwa gleich auf liegen, kann man weder sagen, dass das Modell zu restriktiv ist, noch kann man sagen, dass das Modell zu extensiv ist. Wäre das Modell restriktiv, dürften die falsch Positiven nicht so hoch sein; wäre es extensiv, gäbe es weniger falsch Negative. Um hier ein besseres Ergebnis

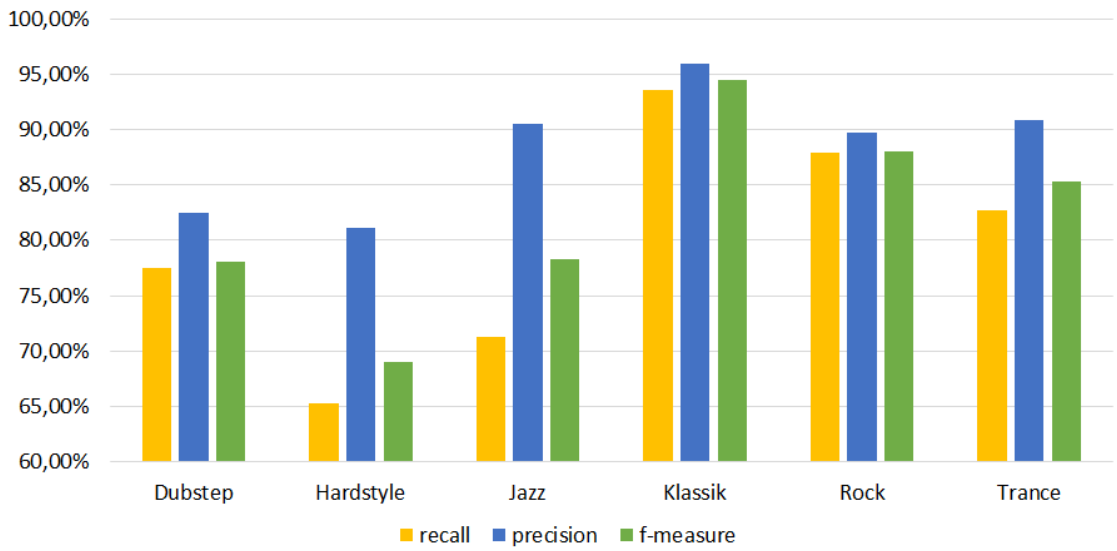

Abbildung 4.4: „Recall“, „precision“ und „f-measure“ im „leave one out“ Verfahren für Ansatz 2 mit 200ms Fenster und 100 Bäumen im Random Forest aufgeschlüsselt nach Genre und gesamt.

zu erzielen, muss das Modell beide Bedingungen erfüllen. Es muss extensiver gegenüber Gesang werden und gleichzeitig auch restriktiver gegenüber gesangslosen Passagen.

Dieser Test hat gezeigt, dass jedes Genre für sich genommen verschiedene Verhaltensweisen in der Klassifikation von Gesang zeigen. Klassik ließ sich mit Abstand am besten klassifizieren, gefolgt von Rock und Trance. Die größten Probleme verursachten Hardstyle und Dubstep, während Jazz besonders durch die hohe Differenz zwischen „recall“ und „precision“ auffiel.

4.4.3 Genre Konfusion

In einem dritten Experiment wurde Ansatz 2 dahingehend untersucht, wie sich die Ergebnisse verändern, wenn das Genre der Test- und Trainingsdaten nicht übereinstimmen. Hier sollte sich zeigen, ob es ein Genre gibt, dass eine besonders gute Grundlage für einen Trainingsdatensatz bietet und ob es ein Genre gibt, das sich besonders gut zum Testen eignet. In Tabelle 4.5 ist ein Auszug der Testreihe mit einer Fenstergröße von 200ms und 100 Bäumen abgebildet. Sie zeigt die Ergebnisse für ein Trainingsgenre. Die verbleibenden Genres wurden zum Testen verwendet. In Tabelle A.5 in Anhang A.2 sind alle Ergebnisse enthalten.

Tabelle 4.5 ist zu entnehmen, dass die Genres Dubstep und Trance mit einem „f-measure“ von 77,55% bzw. 77,01% am besten abgeschnitten haben. Beide liefern einen „recall“ von über 80% und eine „precision“ von über 73%. Vergleicht man diese Ergebnisse mit denen der Kreuzvalidierung auf dem gesamten Datensatz in Tabelle 4.3, so zeigt sich bei „f-measure“ lediglich ein Rückstand von ca. 4%. Der „recall“ liegt sogar ca. 1% höher. Bezieht man nun mit ein, dass bei der Kreuzvalidierung aus Tabelle 4.3 bei jedem Fold eine Trainingsmenge von 135 Liedern auf sechs verschiedenen Genres vorlag, so sind die Ergebnisse aus Tabelle 4.5 noch beachtlicher. Hier kamen jeweils 25 Lieder aus nur einem Genre zum Einsatz.

T. Genre	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-meas.**	**b. Feh.**
Dubstep	17,20%	82,80%	18,42%	81,58%	73,89%	77,55%	17,81%
Hardstyle	13,67%	86,33%	32,16%	67,84%	79,75%	73,32%	22,92%
Jazz	1,77%	98,23%	59,28%	40,72%	94,29%	56,88%	30,52%
Klassik	18,55%	81,45%	30,39%	69,61%	68,66%	69,13%	24,47%
Rock	35,12%	64,88%	17,48%	82,52%	58,34%	68,35%	26,30%
Trance	19,63%	80,37%	19,22%	80,78%	73,58%	77,01%	19,42%

Tabelle 4.5: Genre Konfusion mit Ansatz 2 mit einem 200ms Fenster, 100 Bäumen im Random Forest, einem Trainingsgenre und fünf Testgenres: Trainingsgenre, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Das drittbeste Ergebnis konnte mit Hardstyle erzielt werden. Dort liegt das „f-measure“ bei 73,32%. Die „precision“ liegt mit 79,75% sogar ca. 6% über der von Dubstep und Trance. Allerdings liegt der „recall“ deutlich niedriger bei nur 67,84%.

Am schlechtesten schnitt das Genre Jazz mit einem „f-measure“ von nur 56,88% ab. Zwar erreichte Jazz die mit Abstand höchste „precision“ von 94,29%, allerdings auch den mit Abstand schlechtesten „recall“ mit 40,72%. Somit konnten mit Jazz als Trainingsbasis nicht einmal die Hälfte aller Gesangspassagen korrekt klassifiziert werden. Ein Modell nur auf Basis von Jazz-Musik ist somit deutlich zu restriktiv.

Auffällig ist, dass sich Dubstep und Trance nicht nur durch ihr gutes Abschneiden von den verbleibenden Genres abheben, sondern hinzu kommt, dass nur bei diesen beiden Genres das Verhältnis von falsch Positiven zu falsch Negativen ausgeglichen ist. Die anderen vier Genres weisen alle eine deutliche Tendenz in die eine oder andere Richtung auf. Hardstyle hat mehr als das Doppelte an falsch Negativen als an falsch Positiven. Auch Jazz und Klassik weisen deutlich mehr falsch Negative als falsch Positive auf. Rock hingegen hat mit Abstand die meisten falsch Positiven. Das Modell auf Basis von Rock scheint also zu extensiv zu sein.

Tabelle 4.6 zeigt eine detaillierte Aufschlüsselung des „f-measure“ in Form einer Konfusionsmatrix. So lassen sich Stärken und Schwächen einzelner Genres genauer erkennen. Wird Dubstep zum Trainieren verwendet, so schneidet Jazz mit Abstand am schlechtesten ab. Im Gegensatz dazu schneidet Klassik mit einem „f-measure“ von 89,23% erheblich besser ab. Vergleichbare Werte liefert Trance als Trainingsgenre.

Bei Jazz, dem Genre, das insgesamt das schlechteste „f-measure“ liefert, ist zu beobachten, dass es für Dubstep als Testgenre mit 42,06% auch den schlechtesten Wert für das „f-measure“ bezogen auf alle Kombinationen aus Trainings- und Testgenre liefert. Das beste Testergebnis mit Jazz als Trainingsgenre liefert mit klarem Abstand Klassik mit

	Trainingsgenre					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	73,60%	42,06%	70,47%	72,97%	77,05%
Hardstyle	71,13%	–	46,06%	56,72%	45,74%	68,70%
Jazz	60,21%	65,36%	–	63,72%	59,19%	62,31%
Klassik	89,23%	83,03%	76,81%	–	81,63%	89,32%
Rock	78,41%	65,98%	51,44%	75,14%	–	81,42%
Trance	84,81%	75,30%	54,36%	76,33%	76,57%	–
Alle	77,55%	73,32%	56,88%	69,13%	68,35%	77,01%

Tabelle 4.6: „F-measure“ der Genre Konfusion von Ansatz 2 mit Fenstergröße 200ms und 100 Bäumen im Random Forest.

76,81%. Alle anderen Genres liegen im Bereich von ca. 50%, einem Niveau, das auch der Zufallsklassifikator erreichen würde.

Die durchweg besten Resultate als Testgenre liefert Klassik. Egal, welches der anderen fünf Genres zum Trainieren verwendet wurde, Klassik lieferte immer das beste „f-measure“. Auch das absolut höchste „f-measure“ konnte mit Klassik als Testgenre erreicht werden. Als Trainingsgrundlage diente hier Trance.

Am schlechtesten ließ sich Hardstyle testen. Während Dubstep und Trance als Trainingsgrundlage hier noch gute Ergebnisse um 70% liefern, war das „f-measure“ mit Jazz, Klassik und Rock mehr als 20% niedriger.

Tabelle 4.7 zeigt die detaillierte Aufschlüsselung des „recall“. Geht es nach dem „recall“, so schneidet Rock als Trainingsgenre mit 82,52% am besten ab. Dahinter folgen Dubstep und Trance mit 81,58% bzw. 80,78%. Am schlechtesten schneidet erneut Jazz mit 40,72% ab. Auf dem Testgenre Klassik schneidet Jazz mit 63,25% noch vergleichsweise gut ab. Bei allen vier anderen Genres liegt der „recall“ deutlich unter 40%, bei Dubstep wird sogar nur ein „recall“ von 27,38% erreicht, was auch dem insgesamt schlechtesten Wert entspricht.

Der absolut beste „recall“ wird mit einer Kombination aus Trance als Trainingsgenre und Jazz als Testgenre erreicht. Das insgesamt beste Testgenre im Bezug auf den „recall“ ist ebenfalls Jazz. Mit Ausnahme von Hardstyle kann jedes Trainingsgenre einen „recall“ größer als 90% auf Jazz erreichen. Bei Hardstyle liegt der „recall“ mit 87,6% nur knapp hinter den anderen.

Die Ergebnisse zur „precision“ in Tabelle 4.8 zeigen ein anderes Bild. Hier hat Jazz mit 94,29% einen deutlichen Vorsprung vor den anderen Genres. Wie bereits in Kapitel 4.4.2 festgestellt, wird auf Basis von Jazz ein sehr restriktives Modell aufgebaut. Dies erzeugt, wie in Tabelle 4.8 zu sehen, außerordentlich gute „precision“-Werte. Damit verbunden sind aber auch die außerordentlich schlechten „recall“-Werte aus Tabelle 4.7 und insgesamt das schlechte „f-measure“ in Tabelle 4.6.

	Trainingsgenre					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	65,18%	27,38%	63,75%	80,18%	75,59%
Hardstyle	74,72%	–	32,63%	60,85%	83,67%	73,32%
Jazz	91,83%	87,60%	–	93,51%	91,68%	94,03%
Klassik	86,44%	74,13%	63,25%	–	75,55%	86,77%
Rock	70,23%	52,85%	34,98%	65,22%	–	74,27%
Trance	83,10%	63,52%	37,79%	69,57%	88,92%	–
Alle	81,58%	67,84%	40,72%	69,61%	82,52%	80,78%

Tabelle 4.7: „Recall“ der Genre Konfusion von Ansatz 2 mit Fenstergröße 200ms und 100 Bäumen im Random Forest.

Die zweitbeste „precision“ liefert Hardstyle mit 79,75% gefolgt von Dubstep mit 73,89%. Am schlechtesten schneidet hier das Genre Rock mit nur 58,34% ab. Auffallend ist dort, dass Rock beim Testen des Genres Klassik zwar mit 88,79% sehr gut abgeschnitten, bei den Genres Jazz und Hardstyle mit 43,7% bzw. 31,47% hingegen versagt hat. Diese Werte stellen außerdem die beiden schlechtesten „precision“-Werte im gesamten Test dar.

Die beste „precision“ findet sich mit 97,79% bei der Kombination aus Jazz als Trainingsgenre und Klassik als Testgenre. Des Weiteren ist Klassik auch das insgesamt beste Testgenre im Bezug auf die „precision“, während Jazz das schlechteste Testgenre ist. Somit stellt Jazz das beste Trainingsgenre und das schlechteste Testgenre zugleich dar.

In einem nächsten Schritt kann diese Grundlage mit einem weiteren Genre erweitert werden. Dazu wurde die Genre Konfusion mit einer Trainingsmenge von zwei Genres erneut durchgeführt. Die verbleibenden vier Genres wurde als Testmenge genutzt. Die Ergebnisse dieses Versuchs sind für die Parameterkonfiguration 200ms Fenster und 100 Bäume in Tabelle 4.9 abgebildet. Die vollständige Tabelle befindet sich in Anhang A.2 in Tabelle A.6.

Wird das „f-measure“ betrachtet, so schneidet die Kombination aus Dubstep und Jazz mit 79,29% am besten ab. Mit 77,53% folgt die Paarung Hardstyle und Trance, liegt damit aber schon knapp hinter dem Ergebnis, dass Dubstep im vorherigen Test alleine erzielen konnte. Doch auch der Zugewinn, den Dubstep und Jazz gegenüber Dubstep alleine erreichen, fällt mit ca. 2% gering aus. Deutlich interessanter ist, zu beobachten, wie sich „recall“ und „precision“ verändern, wenn zwei Genres kombiniert werden.

Dubstep alleine erreichte einen „recall“ von 81,58%, Jazz kam auf 40,72%. Zusammen wird ein „recall“ von 69,58% erreicht. Dieser ist zwar mehr als 10% schlechter als bei Dubstep, aber auch fast 30% besser als bei Jazz. Auch bei der „precision“ gibt es deutliche Veränderungen. Dubstep erreichte zuvor 73,89%, Jazz kam auf 94,29%. Die „precision“ der Kombination liegt mit 92,15% nur ca. 2% hinter dem Einzelergebnis des Genres Jazz, aber

	Trainingsgenre					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	84,51%	90,68%	78,77%	66,94%	78,58%
Hardstyle	67,88%	–	78,29%	53,11%	31,47%	64,62%
Jazz	44,79%	52,13%	–	48,33%	43,70%	46,59%
Klassik	92,21%	94,34%	97,79%	–	88,79%	92,03%
Rock	88,74%	87,80%	97,16%	88,62%	–	90,10%
Trance	86,59%	92,43%	96,80%	84,55%	67,23%	–
Alle	73,89%	79,75%	94,29%	68,66%	58,34%	73,58%

Tabelle 4.8: „Precision“ der Genre Konfusion von Ansatz 2 mit Fenstergröße 200ms und 100 Bäumen im Random Forest.

knapp 20% vor der von Dubstep erreichten „precision“. Der Verlust im „recall“ kann also durch die stark erhöhte „precision“ ausgeglichen werden, wodurch sich der Zugewinn im „f-measure“ ergibt.

Auch die anderen Kombinationen, an den Jazz beteiligt ist, haben sich durch eine besonders hohe „precision“ und sehr wenige falsch Positive von den anderen Genrekombinationen ab. Eine Ausnahme bildet hier allerdings die Kombination aus Jazz und Rock. Mit 13,31% falsch Positiven und einer „precision“ von 75,21% beträgt der Abstand mehr als 10% zu den anderen Jazz-Kombinationen. Dies ist auf die Eigenarten des Genres Rock zurückzuführen. Im vorherigen Einzeltest gab es bei Rock die mit Abstand meisten falsch Positiven sowie die schlechteste „precision“ mit 58,34%. Beide Werte konnten durch das Hinzufügen von Jazz zwar erheblich verbessert werden, an die Ergebnisse der anderen Kombinationen kommt dieses Paar jedoch nicht heran. Insgesamt profitieren aber beide Genres stark voneinander, da sie ihre Stärken und Schwächen gegenseitig ausgleichen können. Anhand der vorherigen Ergebnisse war zu sehen, dass auf Basis von Jazz ein sehr restriktives Modell entsteht, was mit einem niedrigen „recall“ und einer hohen „precision“ einher geht. Bei Rock war genau das Gegenteil der Fall. Das Modell auf Basis von Rock war sehr extensiv, der „recall“ war hoch, die „precision“ sehr niedrig. Das Modell auf Basis der Kombination bildet somit einen Kompromiss aus diesen beiden Extremen.

T. Genre	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-Meas.**	**b. Feh.**
Dubs, Hard	18,43%	81,57%	21,42%	78,58%	76,04%	77,29%	19,92%
Dubs, Jazz	3,87%	96,13%	30,42%	69,58%	92,15%	79,29%	17,15%
Dubs, Klas	19,73%	80,27%	19,08%	80,92%	67,05%	73,34%	19,41%
Dubs, Rock	21,73%	78,27%	12,68%	87,32%	67,48%	76,13%	17,20%
Dubs, Tran	18,58%	81,42%	18,02%	81,98%	72,77%	77,10%	18,30%
Hard, Jazz	2,04%	97,96%	51,88%	48,12%	95,64%	64,03%	26,96%
Hard, Klas	16,82%	83,18%	30,31%	69,69%	74,95%	72,22%	23,57%
Hard, Rock	18,78%	81,22%	21,52%	78,48%	75,30%	76,86%	20,15%
Hard, Tran	21,23%	78,77%	21,14%	78,86%	76,23%	77,53%	21,18%
Jazz, Klas	3,81%	96,19%	59,97%	40,03%	86,97%	54,83%	31,89%
Jazz, Rock	13,31%	86,69%	37,62%	62,38%	75,21%	68,20%	25,47%
Jazz, Tran	5,93%	94,07%	37,07%	62,93%	88,97%	73,72%	21,50%
Klas, Rock	33,87%	66,13%	14,20%	85,80%	56,07%	67,82%	24,03%
Klas, Tran	22,03%	77,97%	22,35%	77,65%	67,52%	72,23%	22,19%
Rock, Tran	21,87%	78,13%	18,34%	81,66%	69,29%	74,97%	20,11%

Tabelle 4.9: Genre Konfusion mit Ansatz 2 mit einem 200ms Fenster, 100 Bäumen im Random Forest, zwei Trainingsgenres und vier Testgenres: Trainingsgenres, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Das schlechteste „f-measure“ erzielen mit klarem Abstand Jazz und Klassik gemeinsam. Es liegt bei 54,83%. Die hohe „precision“ von 86,97% reicht in dieser Kombination nicht aus, um den extrem schlechten „recall“ von nur 40,03% zu kompensieren.

Eine Kombination aus den zuvor am besten abschneidenden Genres Dubstep und Trance ergibt ein „f-measure“ von 77,1%, womit es 0,45% hinter dem Einzelergebnis von Dubstep und 0,09% vor dem Einzelergebnis von Trance liegt. Somit kann durch die Kombination von zwei Genres, die einzeln sehr ähnliche Ergebnisse in der Klassifikation liefern, keine Verbesserung erreicht werden.

Den höchsten „recall“ erreicht das Paar aus Dubstep und Rock mit 87,32%, gefolgt von der Kombination aus Klassik und Rock mit 85,8%.

Bezieht man alle Erkenntnisse der Genre Konfusion ein, so kann man festhalten, dass Dubstep mit einem „f-measure“ von 77,55% in jedem Fall eine solide Grundlage für einen Trainingsdatensatz darstellt. Das daraus resultierende Modell ist weder zu restriktiv noch zu extensiv. Trance bietet eine minimal schlechtere alternative.

Des Weiteren hat der Test gezeigt, dass das Genre Klassik sich außerordentlich gut Klassifizieren lässt. Obwohl in diesem Test Trainings- und Testgenre gänzlich verschieden waren und die Trainingsmenge nur aus 25 Liedern bestand, konnte Klassik mit jedem Trainingsgenre mit Ausnahme von Jazz ein besseres „f-measure“ erreichen, als die Kreuzvalidierung auf dem gesamten Datensatz mit einer erheblich größeren und gemischten Trainingsmenge.

Die Tests mit Kombinationen aus zwei Genres zeigten, dass die Kombination von mehreren Genres zu einem verbesserten Ergebnis führen kann, wenn die zwei gewählten Genres ihr jeweiligen Schwächen gegenseitig ausgleichen. Dies war beispielsweise bei den Kombinationen Dubstep und Jazz sowie Jazz und Rock der Fall. Nahezu ohne Veränderung im Vergleich zu den Einzelergebnissen blieb hingegen die Paarung Dubstep und Trance. Dies ist auf deren ähnliches Verhalten in der Klassifikation zurückzuführen. Alle in den Tabellen aufgeführten Kennzahlen bewegen sich auf dem gleichen Niveau. Keines der Genres hat dort eine Stärke, wo das andere eine Schwäche hat, sodass sich weder der „recall“ noch „precision“ und „f-measure“ nennenswert steigern ließen.

4.5 Ansatz 3: AMUSE-Merkmale

Die Experimente für den dritten Ansatz auf Basis der AMUSE-Merkmale wurden, wie bei Ansatz 2, mit einem Random Forest durchgeführt. Die Merkmale wurden von Dr. Igor Vatolkin extrahiert und für die Tests zur Verfügung gestellt. Sie lagen nach Musikstück und Merkmal getrennt in Dateien des Typs „arff“ ungelabelt vor. Da die Merkmale zum Teil auf unterschiedlichen Fenstergrößen berechnet wurden, war eine Trennung nach Merkmal nötig. Eine Liste mit allen eingesetzten Merkmalen findet sich in Tabelle A.11 in Anhang A.4.

Für die Tests wurden drei verschiedene Merkmalskombination eingesetzt. In der ersten Konfiguration, genannt „all“, kamen alle Merkmale mit insgesamt 225 Dimensionen gemeinsam zum Einsatz. Die zweite Konfiguration mit dem Namen „mfcc“ beinhaltete nur die MFCC-Merkmale mit insgesamt 13 Dimensionen, während die dritte Konfiguration „o mfcc“ alle Merkmale mit Ausnahme der MFCC-Merkmale enthielt, was eine Größe von 212 Dimensionen ergab.

Auch für diesen Ansatz wurde die Anzahl der Bäume des Random Forest Klassifikators variiert. Wie zuvor kamen die gleichen Größen (50, 100, 200) zum Einsatz. Alle Tests wurden außerdem mit den Fenstergrößen 512, 2048 und 4096 Frames durchgeführt. Diese entsprechen etwa 23ms, 93ms und 186ms.

Bevor die Merkmale nun mit einem Label versehen werden konnten, mussten diese zuerst miteinander kombiniert und an die Zielfenstergröße angepasst werden. Stimmte die Zielgröße mit der Größe des Merkmals überein, so konnten die Werte direkt übernommen werden. War die Zielgröße größer als die Größe des Merkmals, so wurden mehrere Merkmalswerte über das arithmetische Mittel zusammengefasst. Eine zu kleine Zielgröße wurde durch mehrmaliges Wiederholen des gleichen Merkmalwertes kompensiert. Bei einer Zielgröße von 512 Frames kommen alle Merkmalswerte mit einer Größe von ebenfalls 512 Frames genau einmal vor, während jedes Element eines Merkmals mit Größe 2048 genau vier mal nacheinander vorkommt.

Nach diesem ersten Schritt lagen nun für jedes Musikstück drei „arff“-Dateien vor. Jede enthielt alle Merkmale angepasst auf je eine der drei Zielgrößen. Diese Dateien konnten nun, nach dem gleichen Verfahren wie für Ansatz 2 geschildert gelabelt werden.

win.	**type**	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-meas.**	**b. Feh.**
512	all	14,51%	85,49%	40,22%	59,78%	72,89%	65,18%	27,37%
2048	all	14,47%	85,53%	36,18%	63,82%	74,04%	68,07%	25,33%
4096	all	14,34%	85,66%	33,51%	66,49%	74,80%	69,89%	23,93%
512	mfcc	16,16%	83,84%	49,03%	50,97%	67,46%	57,74%	32,60%
2048	mfcc	17,09%	82,91%	42,64%	57,36%	68,85%	62,31%	29,86%
4096	mfcc	17,52%	82,48%	40,50%	59,50%	69,07%	63,62%	29,01%
512	o mfcc	15,07%	84,93%	40,71%	59,29%	72,03%	64,52%	27,89%
2048	o mfcc	14,81%	85,19%	37,12%	62,88%	73,34%	67,19%	25,96%
4096	o mfcc	14,88%	85,12%	34,37%	65,63%	73,92%	68,99%	24,63%

Tabelle 4.10: Zehnfache Kreuzvalidierung mit Ansatz 3 auf dem gesamten Testdatensatz: Fenstergröße, Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

4.5.1 Kreuzvalidierung und Parameterevaluation

Der erste durchgeführte Test mit Ansatz 3 war die zehnfache Kreuzvalidierung auf dem gesamten Testdatensatz. Die hierbei ebenfalls durchgeführte Parameterevaluation ergab, dass sowohl die Fenstergröße als auch die Merkmalszusammensetzung einen spürbaren Einfluss auf das Ergebnis nehmen. Auf Veränderungen an der Anzahl der Bäume im Random Forest Klassifikator reagierte das Modell kaum. Daher wurde, ähnlich wie im vorherigen Ansatz, die Standardeinstellung aus Weka mit 100 Bäumen gewählt. In Tabelle 4.10 sind die Ergebnisse der Kreuzvalidierung mit 100 Bäumen zu sehen. Die vollständige Ergebnisliste mit allen Parameterkonfigurationen findet sich in Anhang A.3 in Tabelle A.7. Das höchste „f-measure“ kann mit allen Merkmalen und einem Fenster von 4096 Frames erreicht werden. Mit dieser Konfiguration sind „recall“ und „precision“ ebenfalls am größten. Die „precision“ erreicht 74,8%, der „recall“ 66,49%.

In Abbildung 4.5 werden „recall“, „precision“ und „f-measure“ in den neun Variationen aus Fenstergröße und Merkmalskombination gegenüber gestellt. Auch dort kam ein Random Forest mit 100 Bäumen zu Einsatz. In der Grafik lassen sich die Effekte der Parameterevaluation gut ablesen.

Eine Vergrößerung des Klassifikationsfensters wirkt sich besonders positiv auf den „recall“ aus, während die „precision“ nur leicht vom größeren Fenster profitiert. In der Summe steigt so in jedem Fall das „f-measure“ an.

Bei der Variation der Merkmale ist zu beobachten, dass alle Merkmale zusammen sowohl beim „recall“, als auch bei „precision“ und „f-measure“ am besten abschneiden. Dieses Ergebnis war auch zu erwarten, da dem Klassifikator in dieser Konfiguration die meisten

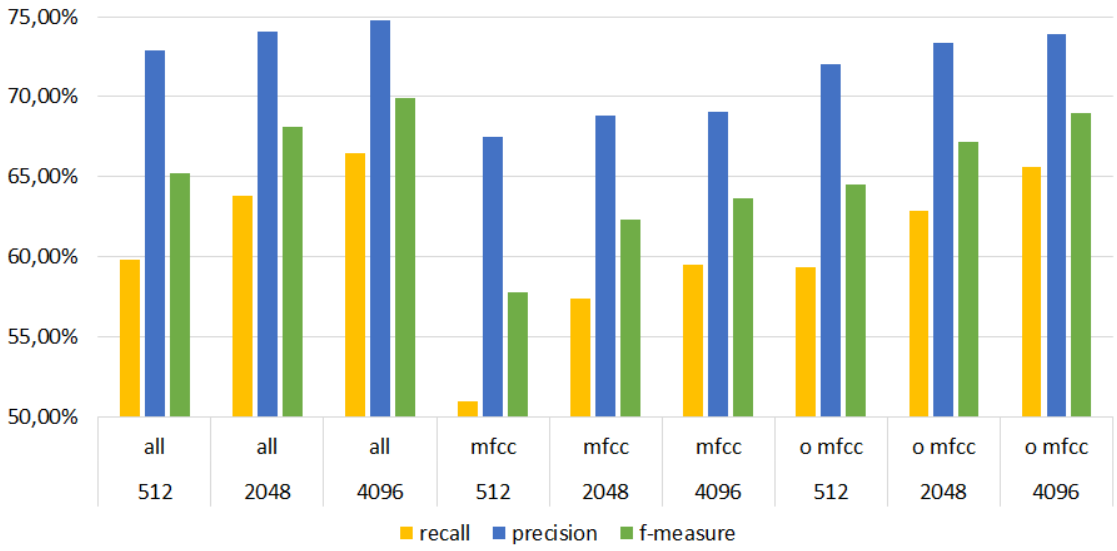

Abbildung 4.5: „Recall“, „precision“ und „f-measure“ in der Kreuzvalidierung für Ansatz 3 mit 100 Bäumen im Random Forest aufgeschlüsselt nach Fenstergröße und Merkmalen.

Informationen zur Verfügung stehen. Erstaunlicher ist allerdings, dass alle Merkmale ohne MFCCs nur minimal schlechter abgeschnitten haben. Hier wäre zu erwarten gewesen, dass die Klassifikation ohne die speziell für Sprach- und Gesangserkennung entwickelten MFCC-Merkmale deutlich schlechter durchzuführen ist. Somit muss es in der Menge der AMUSE-Merkmale neben MFCCs weitere Merkmale geben, die sich gut zur Klassifikation von Gesang eignen.

Werden ausschließlich MFCC-Merkmale verwendet, so schneidet Ansatz 3 am schlechtesten ab. Der Rückstand auf die zwei anderen Merkmalskombinationen beträgt ca. 5%. Des Weiteren reagieren MFCC-Merkmale alleine deutlicher auf eine Anpassung der Fenstergröße. Von 512 Frames zu 2048 Frames beträgt der Unterschied im „recall“ ca. 6%. Mit dem kleinsten Fenster erreichen die MFCCs nur einen „recall“ von 50,97%. Damit liegen sie nur minimal vor dem Zufallsklassifikator.

Insgesamt sollten für die optimale Konfiguration, neben den zuvor gesetzten 100 Bäumen im Random Forest, alle AMUSE-Merkmale sowie eine Fenstergröße von 4096 Frames bzw. 186ms verwendet werden. Zwar gibt es zwischen den Gruppen „alle Merkmale“ und „alle Merkmale ohne MFCCs“ nur einen kleinen Unterschied, dennoch sollte nicht auf MFCCs verzichtet werden, da die Hinzunahme nur 13 Dimensionen hinzufügt und, wie in 2 bereits erwähnt, diese speziell zur Spracherkennung entwickelt wurden.

Die Wahl der Fenstergröße von 4096 Frames begründet sich auf drei wesentliche Aspekte: den „recall“, die Datenmenge und die Rechenzeit. Unter der Verwendung von allen Merkmalen erhöht sich der „recall“ um ca. 7% im Vergleich zum kleineren Fenster mit 512 Frames, sodass auch das „f-measure“ insgesamt auf 69,89% steigen kann. Zusätzlich reduzieren sich durch diesen Schritt auch die Datenmenge sowie die benötigte Rechenzeit auf ca. $\frac{1}{8}$ der Menge und Zeit, die bei einem 512 Frames großen Fenster benötigt wird.

Das in dieser Arbeit verwendete Testsystem, beschrieben zu Beginn dieses Kapitels, erreichte mit allem Merkmalen und der kleinsten Fenstergröße bereits sein Ressourcenlimit. Die Erstellung des Modells für einen einzigen Fold aus diesem Experiment benötigte unter voller Auslastung aller 8 Threads knapp eine Stunde, wodurch der Test dieser einen Parameterkonfiguration insgesamt mehr als 10 Stunden in Anspruch nahm. Außerdem war der Systemspeicher während des gesamten Testes vollständig belegt.

Zuletzt empfiehlt sich ein großes Bearbeitungsfenster auch im Hinblick auf die zugrunde liegenden Annotationen. Alle Annotationen, die für den Computer später die „Ground Truth“ oder Grundwahrheit bilden, werden von einem Menschen erstellt. Um die Klassifikation auf Ausschnitten mit einer Größe von 512 Frames bzw. 23ms vorzunehmen, müssen auch die zugrunde liegenden Annotationen eine ebenso hohe Präzision aufweisen. Genau an dieser Stelle liegt das Problem. Für einen Menschen ist es kaum möglich bis auf 23ms genau zu entscheiden, wann eine Gesangspassage beginnt oder endet. Selbst mit grafischer Unterstützung, wie in Kapitel 4.1 beschrieben, stellt dies immer noch eine enorme Heraus-

forderung dar. Daher ist eine Klassifikation auf solchen kleinen Abschnitten nicht sinnvoll, da die dazu benötigte Grundlage fehlt.

Insgesamt hat dieser erste Test gezeigt, dass sich mit Ansatz 3 solide Ergebnisse mit einem „f-measure“ von ca. 70% erreichen lassen. Damit arbeitet dieser Ansatz klar besser als der Zufallsklassifikator. Die optimale Parametereinstellung besteht aus einem 4096 Frames großen Fenster sowie 100 Bäumen im Random Forest. Des Weiteren sollten alle Merkmale verwendet werden. Der Vorsprung vor den anderen Merkmalskombinationen ist aber eher gering. Hier überrascht besonders das vergleichsweise gute Abschneiden der Merkmalskombination ohne MFCC-Merkmale, da sie, wie bereits zuvor erwähnt, speziell für die Sprach- und Gesangserkennung entwickelt wurden.

4.5.2 „Leave one out“

Auch für diesen dritten Ansatz wurde das „leave one out“ Verfahren für jedes Genre separat durchgeführt. Die vollständigen Ergebnisse sind im Anhang A.3 in Tabelle A.8. Eine gekürzte Fassung mit fester Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest ist in Tabelle 4.11 zu sehen.

Dort lässt sich ablesen, dass die sechs Genres recht unterschiedlich auf die drei Merkmalskombinationen reagieren. Generell lässt sich sagen, dass es zwischen der Variante mit allen Merkmalen und der Variante ohne MFCCs nur kleine Unterschiede gibt, die das „f-measure“ kaum verändern. Deutliche Veränderungen sind jedoch bei ausschließlicher Verwendung von MFCCs zu erkennen.

Während die alleinige Verwendung von MFCCs beim Genre Dubstep kaum Auswirkungen auf die falsch Positiven hat, erhöhen sich die falsch Positiven bei den Genres Hardstyle, Jazz, Rock und Trance. Beim Genre Rock ist der Zuwachs mit 17% am größten. Ein ganz anderes Verhalten zeigt das Genre Klassik. Hier ist die Zahl der falsch Positiven geringer, wenn nur MFCC-Merkmale verwendet werden. In Kombination mit den anderen Merkmalen oder ganz ohne MFCCs scheidet Klassik schlechter ab.

Betrachtet man die falsch Negativen, so reagieren die Genres auch dort verschieden. Mit MFCC-Merkmalen allein erhöhen sich die falsch Negativen auf dem Genre Dubstep um ca. 7%. Auch auf den Genres Hardstyle, Klassik, Rock und Trance geht die Tendenz nach oben. Der deutlichste Anstieg findet bei Trance mit ca. 10% statt. Mit nur 2% ist die Erhöhung der falsch Negativen auf Rock am kleinsten. Ausreißer ist hier das Genre Jazz. Dort gehen die falsch Negativen um ca. 5% zurück, wenn nur MFCCs genutzt werden.

In Abbildung 4.6 ist eine Gegenüberstellung von „recall“, „precision“ und „f-measure“ zu sehen. Da sich die Merkmalskombinationen „all“ und „ohne mfcc“ nicht bedeutend unterscheiden, wurde letztere in dieser Abbildung entfernt.

Das höchste „f-measure“ von 85,16% kann das Genre Klassik mit allen Merkmalen erzielen. Unter ausschließlicher Verwendung von MFCCs geht das „f-measure“ leicht auf

82,55% zurück. Setzt man diese Ergebnisse mit der jeweiligen Anzahl an Dimensionen in Verbindung, so ist das gute Abschneiden von MFCCs alleine noch beachtlicher, da die MFCC-Merkmale insgesamt 13 Dimensionen haben, während es bei allen Merkmalen zusammen 225 Dimensionen sind. Bei den Genres Dubstep, Rock und Trance liegen die MFCC-Merkmale ca. 5% hinter der Gruppe aus allen Merkmalen und auch hier muss der enorme Unterschied in den Dimensionen beachtet werden.

Deutlich größer ist der Abstand im Genre Hardstyle. Dort liegen zwischen allen Merkmalen gemeinsam und MFCCs alleine ca. 10% im „f-measure“. Des Weiteren schneidet Hardstyle zusammen mit Jazz mit Abstand am schlechtesten ab. Während das „f-measure“ auf dem Genre Hardstyle mit allem Merkmalen knapp die 50% übersteigt, so liegt es unter ausschließlicher Verwendung der MFCC-Merkmale nur noch knapp über 40%. Auch beide Varianten auf dem Genre Jazz erreichen nicht die 50%-Marke.

Weitere Auffälligkeiten zeigen sich außerdem an den Daten, die für Jazz ermittelt wurden. Jazz ist das einzige Genre, bei dem die alleinige Verwendung von MFCCs ein besseres Ergebnis lieferte als die Verwendung aller Merkmale gemeinsam. Eine zweite Abweichung

Genre	type	FP	TN	FN	recall	preci.	f-Mea.	b. F.
Dubs	all	40,45%	59,55%	25,35%	74,65%	71,20%	71,12%	31,61%
Dubs	mfcc	39,56%	60,44%	33,49%	66,51%	69,34%	66,23%	35,75%
Dubs	o mfcc	40,82%	59,18%	25,49%	74,51%	71,01%	70,87%	31,88%
Hard	all	2,51%	97,49%	57,09%	42,91%	76,43%	51,02%	29,76%
Hard	mfcc	6,10%	93,90%	63,23%	36,77%	59,37%	41,25%	34,55%
Hard	o mfcc	2,56%	97,44%	56,95%	43,05%	76,73%	51,01%	29,73%
Jazz	all	7,83%	92,17%	68,21%	31,79%	88,63%	44,10%	36,56%
Jazz	mfcc	12,27%	87,73%	63,35%	36,65%	79,24%	47,31%	38,02%
Jazz	o mfcc	8,49%	91,51%	69,35%	30,65%	87,61%	42,51%	37,37%
Klas	all	25,86%	74,14%	8,85%	86,15%	85,43%	85,16%	18,04%
Klas	mfcc	18,14%	81,86%	14,59%	80,41%	86,32%	82,55%	17,24%
Klas	o mfcc	27,72%	72,28%	9,25%	85,75%	84,70%	84,52%	19,12%
Rock	all	33,75%	66,25%	13,60%	86,40%	77,58%	80,43%	23,68%
Rock	mfcc	50,75%	49,25%	15,55%	84,45%	69,03%	74,55%	33,15%
Rock	o mfcc	33,91%	66,09%	13,57%	86,43%	77,61%	80,52%	23,74%
Tran	all	15,48%	84,52%	33,49%	66,51%	82,84%	71,68%	24,85%
Tran	mfcc	20,56%	79,44%	43,10%	56,90%	75,31%	63,23%	31,82%
Tran	o mfcc	15,45%	84,55%	33,85%	66,15%	83,35%	71,61%	24,91%

Tabelle 4.11: „Leave one out“ mit Ansatz 3 auf jedem Genre getrennt mit einer Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest: Genre, Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

wird bei einem Blick auf den „recall“ deutlich. Die Kombination aus MFCCs und den anderen Merkmalen führt zu einer Verminderung des „recall“. Bei den fünf übrigen Genres ist das genaue Gegenteil zu beobachten. Dort ist der „recall“ bei der Gruppe mit allen Merkmalen größer, als bei den MFCCs allein. Weiterhin sticht beim Genre Jazz die außerordentlich hohe „precision“ hervor. Mit allen Merkmalen erreicht sie mit 88,63% den Höchstwert in diesem Testlauf.

Vergleicht man die Ergebnisse aus Abbildung 4.11 mit denen aus Abbildung 4.10, so ist zu erkennen, das sich das Verhältnis von „recall“ zu „precision“ bei manchen Genres gedreht hat. Bei der Kreuzvalidierung auf dem gesamten Datensatz lag stets die „precision“ deutlich vor dem „recall“. Sind die Genres voneinander getrennt, so übertrifft der „recall“ in einigen Fällen die „precision“. Dies ist bei Dubstep in Verbindung mit allen Merkmalen, Klassik mit allen Merkmalen und bei beiden Variationen mit Rock der Fall. Doch da vor allem Hardstyle und Jazz eine enorme Differenz zwischen „recall“ und „precision“ zugunsten der „precision“ aufweisen, übertrifft auch auf dem ganzen Datensatz die „precision“ den „recall“.

4.5.3 Genre Konfusion

Der dritte Versuch widmete sich der Genre Konfusion. Die Tests wurden nach dem gleichen Schema wie bei Ansatz 2 durchgeführt: mit einem Genre trainieren und mit allen anderen Testen. Wie auch bei den vorherigen Experimenten mit diesem Ansatz, wurde eine Para-

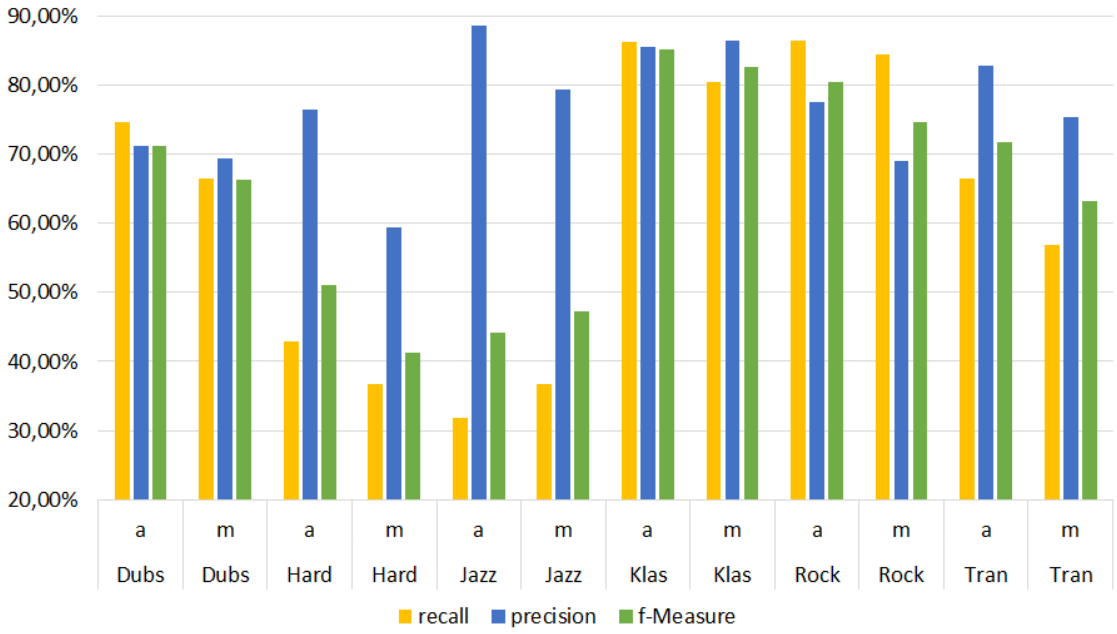

Abbildung 4.6: „Recall“, „precision“ und „f-measure“ im „leave one out“ Verfahren für Ansatz 3 mit 4096 Frames Fenster und 100 Bäumen im Random Forest aufgeschlüsselt nach Genre und Merkmalskombination.

meterevaluation durchgeführt. Die vollständige Liste mit allen Parameterkonfigurationen befindet sich im Anhang A.3 in Tabelle A.9. In Tabelle 4.12 sind die Ergebnisse mit festem Bearbeiungsfenster von 4096 Frames sowie 100 Bäumen im Random Forest zu sehen.

Insgesamt kann man beobachten, dass sich die Merkmalskombinationen bestehend aus allen Merkmalen bzw. allen Merkmalen ohne MFCCs gleich verhalten. Der Unterschied bewegt sich bei jedem Maß im Bereich von unter 1%. Die Ergebnisse, die nur auf Basis von MFCCs entstanden, heben sich jedoch sichtbar von den anderen ab.

Das höchste „f-measure“ in diesem Testdurchlauf erreichte das Genre Trance unter ausschließlicher Verwendung von MFCC-Merkmalen mit 62,78%. Auch die anderen beiden elektronischen Genres Dubstep und Hardstyle erzielen jeweils ihr bestes Ergebnis auf Basis der MFCC-Merkmale. Die beiden anderen Merkmalskombinationen schneiden sichtbar schlechter ab. Gegenteiliges ist bei den verbleibenden drei Genres zu beobachten. Dort schneiden die MFCC-Merkmale alleine am schlechtesten ab.

Gen.	**type**	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-meas.**	**b. Feh.**
Dubs	a	39,30%	60,70%	38,28%	61,72%	48,42%	54,27%	38,79%
Dubs	m	38,63%	61,37%	33,61%	66,39%	50,63%	57,45%	36,12%
Dubs	o	39,41%	60,59%	38,02%	61,98%	48,46%	54,39%	38,71%
Hard	a	14,80%	85,20%	62,01%	37,99%	67,12%	48,52%	38,40%
Hard	m	19,41%	80,59%	54,25%	45,75%	65,16%	53,75%	36,83%
Hard	o	15,15%	84,85%	61,24%	38,76%	67,04%	49,12%	38,20%
Jazz	a	34,86%	65,14%	51,28%	48,72%	50,11%	49,41%	43,07%
Jazz	m	9,94%	90,06%	64,59%	35,41%	71,87%	47,45%	37,26%
Jazz	o	34,01%	65,99%	52,79%	47,21%	49,95%	48,54%	43,40%
Klas	a	84,35%	15,65%	4,11%	95,89%	39,94%	56,39%	44,23%
Klas	m	15,33%	84,67%	67,97%	32,03%	54,94%	40,47%	41,65%
Klas	o	84,28%	15,72%	3,85%	96,15%	40,03%	56,53%	44,06%
Rock	a	46,72%	53,28%	35,03%	64,97%	45,38%	53,43%	40,87%
Rock	m	67,56%	32,44%	29,26%	70,74%	38,43%	49,81%	48,41%
Rock	o	45,80%	54,20%	35,84%	64,16%	45,56%	53,28%	40,82%
Tran	a	19,53%	80,47%	50,59%	49,41%	63,19%	55,45%	35,06%
Tran	m	31,92%	68,08%	32,67%	67,33%	58,81%	62,78%	32,30%
Tran	o	19,58%	80,42%	51,16%	48,84%	62,86%	54,97%	35,37%

Tabelle 4.12: Genre Konfusion mit Ansatz 3 mit einem 4096 Frames Fenster, 100 Bäumen im Random Forest, einem Trainingsgenre und fünf Testgenres: Trainingsgenre, Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Das niedrigste „f-measure“ wird auf dem Genre Klassik mit MFCC-Merkmalen erreicht. Es liegt dort bei 40,47%. Verursacht wird es durch den ebenfalls niedrigsten „recall“ im gesamten Test von nur 32,03%. Den höchsten „recall“ liefert überraschender Weise auch Klassik. Er kommt mit der Merkmalskombination ohne MFCCs zustande und beträgt 96,15%. Damit gehen aber auch die enorm hohen falsch Positiven einher. Sie betragen 84,28%. Knapp darüber liegt nur Klassik in Verbindung mit allen Merkmalen mit 84,35%.

Während MFCCs alleine auf Klassik folglich ein sehr restriktives Modell verursachen, entsteht mit den zwei anderen Merkmalskombinationen ein sehr extensives Modell mit sehr hohem „recall“ und sehr vielen falsch Positiven.

Die höchste „precision“ von 71,87% kann mit MFCC-Merkmalen auf dem Genre Jazz erzielt werden. Mit nur 9,94% falsch Positiven wird auch dort der niedrigste Wert erreicht. Die geringste „precision“ mit nur 38,43% wird auf dem Genre Rock unter ausschließlicher Verwendung von MFCCs erreicht.

Vergleicht man alle sechs Genres im Bezug auf das beste Ergebnis und die dazu benötigte Merkmalskombination, so lässt sich kein Muster erkennen. Wie zuvor erwähnt, erreichen die elektronischen Genres ihr bestes Ergebnis für das „f-measure“, wenn nur MFCCs verwendet werden. Damit sind auch „recall“ und „precision“ bei Dubstep am höchsten. Bei Hardstyle hingegen ist dann zwar auch der „recall“ am höchsten, die „precision“ erreicht ihr Maximum allerdings erst, wenn alle Merkmale verwendet werden.

Jazz erreicht das maximale „f-measure“ sowie den maximalen „recall“ mit allen Merkmalen, die maximale „precision“, wenn nur MFCCs genutzt werden. Klassik wiederum liefert mit allem Merkmalen außer MFCCs das größte „f-measure“ und den höchsten „recall“. Die maximale „precision“ wird nur mit MFCCs erzielt. Auf dem Genre Rock kann mit jeder Merkmalskombination ein anderen Maß maximiert werden. Das „f-measure“ ist mit allem Merkmalen maximal, der „recall“ nur mit MFCCs und die „precision“ ist mit allem Merkmalen ohne MFCCs am größten. Trance hat das Maximum bei „recall“ und „f-measure“, wenn MFCCs alleine verwendet werden. Die höchste „precision“ kann mit allen Merkmalen erreicht werden.

Um trotz dieser sehr unterschiedlichen Ergebnisse eine „optimale“ Kombination zu finden, wurde in Tabelle 4.13 der Durchschnitt über alle Ergebnisse einer Kombination ge-

type	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-meas.**	**b. Feh.**
a	39,93%	60,07%	40,22%	59,78%	52,36%	52,91%	40,07%
m	30,46%	69,54%	47,06%	52,94%	56,64%	51,95%	38,76%
o	39,71%	60,29%	40,48%	59,52%	52,32%	52,80%	40,09%

Tabelle 4.13: Genre Konfusion mit Ansatz 3 mit einem 4096 Frames Fenster, 100 Bäumen im Random Forest, Durchschnitt nach Merkmalskombination: Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

bildet. Das „f-measure“ ist mit 52,91% am größten, wenn alle Merkmale genutzt werden. Knapp dahinter befindet sich die Kombination ohne MFCC-Merkmale. Mit 51,95% schneiden MFCCs alleine am schlechtesten ab. Der Abstand zu allem Merkmalen liegt jedoch unter einem Prozent. Auch der „recall“ ist mit allem Merkmalen am größten, die „precision“ hat ihr Maximum, wenn nur MFCCs zum Einsatz kommen. Somit ist die Merkmalskombination bestehend aus allen Merkmalen knapp die beste Kombination.

In Tabelle 4.14 sind zwei Konfusionsmatrizen für das „f-measure“ abgebildet. Die obere Matrix enthält alle Werte, die auf Grundlage aller Merkmale erzielt wurden. Die untere Matrix enthält die Werte, die anhand der MFCC-Merkmale erreicht wurden. Das höchste „f-measure“ mit allen Merkmalen wird mit der Paarung aus Jazz als Trainingsgenre und Rock als Testgenre erreicht. Es beträgt 73,46%. Die gleiche Genrekombination schneidet auf den MFCC-Merkmalen aber bei weitem schlechter ab. Dort liegt das „f-measure“ nur bei 46,31%. Das niedrigste „f-measure“ mit allen Merkmalen von nur 29,02% ergibt das Training auf Jazz mit einem Test auf Hardstyle. Werden nur MFCCs verwendet, steigt es auf 40,17% an. In der Menge aus allen Merkmalen müssen folglich Merkmale enthalten sein, die den Klassifikator verwirren, anstatt eine klare Trennung zu ermöglichen.

	Trainingsgenre mit allen Merkmalen					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	51,39%	45,61%	70,54%	68,01%	67,53%
Hardstyle	40,79%	–	29,02%	34,14%	37,58%	44,98%
Jazz	41,02%	41,58%	–	47,09%	45,93%	30,90%
Klassik	55,65%	48,22%	49,38%	–	47,06%	46,67%
Rock	69,39%	47,39%	73,46%	72,24%	–	70,19%
Trance	61,31%	53,98%	47,27%	55,38%	61,17%	–
Alle	54,27%	48,52%	49,41%	56,39%	53,43%	55,45%
	Trainingsgenre mit MFCC-Merkmalen					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	48,02%	28,28%	40,39%	64,96%	62,61%
Hardstyle	43,39%	–	40,17%	43,64%	31,10%	50,35%
Jazz	43,25%	46,58%	–	50,33%	42,30%	49,11%
Klassik	70,69%	73,15%	70,42%	–	52,66%	80,29%
Rock	64,22%	41,76%	46,31%	27,93%	–	58,52%
Trance	59,80%	50,74%	38,39%	40,45%	52,81%	–
Alle	57,45%	53,75%	47,45%	40,47%	49,81%	62,78%

Tabelle 4.14: „F-measure“ der Genre Konfusion von Ansatz 3 mit Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest, oben mit allen Merkmalen, unten nur mit MFCC-Merkmalen.

Auf den MFCC-Merkmalen kann mit der Kombination aus Trance als Trainingsgenre und Klassik als Testgenre das maximale „f-measure“ von 80,29% erreicht werden. Mit allen Merkmalen halbiert sich das „f-measure“ fast auf 46,67%. Mit nur 27,93% wird mit dem Paar aus Klassik als Trainingsgenre und Rock als Testgenre das schlechteste „f-measure“ auf Basis der MFCC-Merkmale erreicht. Nur minimal besser ist das „f-measure“ wenn mit Jazz trainiert und mit Hardstyle getestet wird.

Das beste Trainingsgenre auf allen Merkmalen ist Klassik. Es erreicht ein „f-measure“ von 56,39%. Wird die Merkmalsmenge nur auf MFCCs reduziert, so schneidet Klassik mit 40,47% am schlechtesten ab. Die fehlenden Merkmale wirken sich, auf alle Testgenres betrachtet, negativ aus. Zwar schneidet Klassik mit MFCCs auf den Testgenres Hardstyle und Jazz besser ab als mit allen Merkmalen, in den anderen drei Genres ist das Ergebnis aber erheblich schlechter, sodass die Zugewinne auf Hardstyle und Jazz kaum ins Gewicht fallen.

Das schlechteste Trainingsgenre auf allen Merkmalen ist mit 48,52% Hardstyle. Mit MFCC-Merkmalen allein erhöht sich das „f-measure“ um ca. 5% auf 53,75%. In dieser Variation verschlechtern sich zwar die Ergebnisse der Testgenres Dubstep, Rock und Trance, die Zugewinne bei Jazz (ca. 5%) und Klassik (ca. 25%) sind allerdings auch hier so groß, dass sie die Verluste übertreffen.

Wenn nur MFCC-Merkmale zur Verfügung stehen, kann Trance mit 62,78% das höchste „f-measure“ erreichen. Durch das Hinzufügen der anderen Merkmale reduziert sich das „f-measure“ dann wieder um ca. 7% auf 55,45%. Somit sind auch hier mehr Merkmale eher hinderlich, als hilfreich.

Das beste Testgenre auf Basis aller Merkmale ist Rock. Mit Ausnahme des Trainingsgenres Hardstyle kann beim Testen auf Rock immer ein „f-measure“ von ca. 70% erzielt werden. Mit Hardstyle liegt das „f-measure“ aber nur bei 47,39%. Das schlechteste Testgenre mit allem Merkmalen ist Hardstyle. Dort liegt es „f-measure“ im Durchschnitt über alle Trainingsgenres bei nur ca. 37%.

Stehen nur MFCC-Merkmale zur Verfügung, so lässt sich Klassik am besten testen. Das „f-measure“ liegt dort im Durchschnitt bei ca. 69%. Am schlechtesten lässt sich auch hier Hardstyle testen, dessen durchschnittliches „f-measure“ bei ca. 41% liegt. Damit liegt zumindest eine kleine Steigerung gegenüber der anderen Merkmalskombination vor.

In Tabelle 4.15 sind die Konfusionsmatrizen für den „recall“ für die beiden Merkmalskombinationen „alle Merkmale“ und „nur MFCC“ zu sehen. Den höchsten „recall“ liefert das Trainingsgenre Klassik mit Trance als Testgenre. Der „recall“ erreicht dort 98,98%. Auch insgesamt ist Klassik das beste Trainingsgenre im Bezug auf den „recall“. Auf allen Testgenres zusammen liegt er im Durchschnitt bei 95,89%. Vergleicht man diese enorm hohen Werte mit den Gegenstücken auf Basis der MFCC-Merkmale, so ist zu erkennen, dass der „recall“ dort nur noch ca. ein Drittel beträgt. Das ist zugleich auch der schlechteste „recall“ auf Basis der MFCCs. Daraus folgt, dass mit allen Merkmalen gemeinsam ein sehr

extensives Modell aufgebaut wird, während die Reduktion auf MFCC-Merkmale eine sehr restriktives Modell erzeugt.

Das beste Testgenre im Bezug auf den „recall“ mit allen Merkmalen ist Hardstyle. Im Durchschnitt wird ein Ergebnis von über 72% erreicht. Knapp hinter Hardstyle befindet sich das Testgenre Trance mit einem durchschnittlichen „recall“ von ca. 71%. Das schlechteste Testgenre ist mit einem Durchschnitt von ca. 34% Klassik. Werden die Merkmale auf MFCCs reduziert, so wird Klassik zum besten Testgenre mit ca. 65%. Am schlechtesten lässt sich nun Rock testen. Dort wird im Durchschnitt nur noch ein „recall“ von ca. 42% erreicht.

Der kleinste „recall“ mit allen Merkmalen entsteht aus der Kombination von Trance als Trainingsgenre und Jazz als Testgenre und beträgt 26,78%. Im Durchschnitt über alle Testgenres liefert Hardstyle als Trainingsgenre mit 37,99% den niedrigsten „recall“. Durch ausschließliches Verwenden der MFCCs können diese beiden Werte jedoch gesteigert werden. Für das Paar aus Trance und Jazz ist der „recall“ mit 66,64% sogar mehr als doppelt so groß, als mit allen Merkmalen.

Mit nur 19,5% ist der „recall“ der Paarung bestehend aus Klassik als Trainingsgenre und Rock als Testgenre der kleinste bei alleiniger Nutzung von MFCCs. Den maximalen „recall“

	Trainingsgenre mit allen Merkmalen					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	37,70%	38,82%	98,28%	79,02%	64,75%
Hardstyle	77,66%	–	54,01%	96,03%	80,31%	54,94%
Jazz	52,07%	45,52%	–	92,06%	66,58%	26,78%
Klassik	40,05%	32,63%	34,11%	–	33,71%	31,19%
Rock	71,36%	36,37%	78,34%	92,72%	–	66,32%
Trance	82,79%	42,63%	48,90%	98,98%	81,71%	–
Alle	61,72%	37,99%	48,72%	95,89%	64,97%	49,41%
	Trainingsgenre mit MFCC-Merkmalen					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	36,05%	17,88%	29,29%	81,16%	60,46%
Hardstyle	67,90%	–	34,07%	37,93%	79,93%	62,22%
Jazz	63,66%	54,58%	–	54,38%	64,73%	66,64%
Klassik	65,67%	64,96%	58,34%	–	52,86%	86,06%
Rock	69,22%	33,05%	35,99%	19,50%	–	54,54%
Trance	65,51%	38,34%	26,74%	29,12%	82,06%	–
Alle	66,39%	45,75%	35,41%	32,03%	70,74%	67,33%

Tabelle 4.15: „Recall“ der Genre Konfusion von Ansatz 3 mit Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest, oben mit allen Merkmalen, unten nur mit MFCC-Merkmalen.

auf MFCC-Basis in Höhe von 86,06% erzielt die Kombination aus dem Trainingsgenre Trance und dem Testgenre Klassik.

Auffällig ist außerdem, dass das Training mit Dubstep und MFCC-Merkmalen sehr stabile Ergebnisse auf den Testgenres liefert. Der „recall“ bewegt sich zwischen 63,66% bei Jazz und 69,22% bei Rock, womit die Schwankung zwischen den Ergebnissen nur knapp 6% beträgt. Bei allen anderen Trainingsgenres ist die Differenz zwischen minimalem und maximalem „recall“ erheblich größer.

In Tabelle 4.16 ist die „precision“ der sechs Genres in der Genre Konfusion aufgeführt. Auch hier wird zwischen „alle Merkmale“ und „nur MFCCs“ unterschieden. Mit 92,63% erreicht die Kombination aus Trance als Trainingsgenre und Klassik als Testgenre die höchste „precision“ mit allen Merkmalen. Interessant ist hier die Tatsache, dass diese beiden Genres in getauschten Positionen den höchsten „recall“ erreichen konnten. Am schwächsten schneidet die Paarung Jazz und Hardstyle ab, wobei Jazz zum Trainieren verwendet wurde. Deren „precision“ liegt nur bei 19,84%.

Ein Blick auf die untere Matrix zeigt, dass dort die maximale „precision“ von 88,8% vom Trainingsgenre Jazz zusammen mit Klassik als Testgenre erzielt wird. Die niedrigste „pre-

	Trainingsgenre mit allen Merkmalen					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	80,71%	55,28%	55,02%	59,70%	70,57%
Hardstyle	27,66%	–	19,84%	20,76%	24,53%	38,08%
Jazz	33,84%	38,27%	–	31,63%	35,06%	36,51%
Klassik	91,17%	92,38%	89,37%	–	77,94%	92,63%
Rock	67,53%	67,98%	69,15%	59,17%	–	74,54%
Trance	48,68%	73,56%	45,74%	38,45%	48,88%	–
Alle	48,42%	67,12%	50,11%	39,94%	45,38%	63,19%
	Trainingsgenre mit MFCC-Merkmalen					
Testgenre ↓	Dubstep	Hardstyle	Jazz	Klassik	Rock	Trance
Dubstep	–	71,89%	67,59%	65,03%	54,16%	64,91%
Hardstyle	31,89%	–	48,93%	51,37%	19,30%	42,29%
Jazz	32,74%	40,62%	–	46,84%	31,41%	38,88%
Klassik	76,55%	83,72%	88,80%	–	52,47%	75,25%
Rock	59,90%	56,68%	64,94%	49,21%	–	63,13%
Trance	55,00%	75,01%	68,06%	66,23%	38,94%	–
Alle	50,63%	65,16%	71,87%	54,94%	38,43%	58,81%

Tabelle 4.16: „Precision“ der Genre Konfusion von Ansatz 3 mit Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest, oben mit allen Merkmalen, unten nur mit MFCC-Merkmalen.

cision“ erreichen Rock und Hardstyle mit Rock als Trainingsgenre. Bei dieser Kombination liegt die „precision“ nur bei 19,3%.

Die höchste „precision“ auf allen Genres mit allen Merkmalen erreicht Hardstyle mit 67,12%. Kommen nur MFCCs zum Einsatz, so liegt die „precision“ mit 65,16% ca. 2% niedriger. Die maximale „precision“ auf MFCCs kann das Trainingsgenre Jazz mit 71,78% erzielen, womit auch das beste Trainingsgenre auf allen Merkmalen übertroffen wird.

Im Schnitt über alle Testgenres mit allen Merkmalen schneidet Klassik als Trainingsgenre mit 39,94% am schlechtesten ab. Mit einem Wechsel auf ausschließlich MFCC-Merkmale steigt die „precision“ im Durchschnitt um 15%. Auf Basis dieser Merkmale schneidet nun Rock mit durchschnittlich 38,43% am schlechtesten ab.

Das Genre, das sich mit der höchsten Präzision Testen lässt, ist auf beiden Merkmalskombination Klassik. Im Durchschnitt kann beim testen von Klassik eine „precision“ von über 88% mit allen Merkmalen bzw. über 75% nur mit MFCC-Merkmalen erreicht werden. Am unpräzisesten lässt sich Hardstyle mit allen Merkmalen testen. Die „precision“ beträgt durchschnittlich nur 26,1%. Werden die Merkmale auf MFCCs beschränkt, erhöht sich die „precision“ auf 38,7%. Damit liegt Hardstyle knapp vor Jazz, das mit 38,1% am schlechtesten unter alleiniger Verwendung von MFCCs abschnitt.

In einem letzten Test wurde schließlich untersucht, ob sich die, teils sehr durchwachsenen, Ergebnisse verbessern ließen, indem bei der Genre Konfusion zwei Genres zum Trainieren verwendet werden. Eine vollständige Ergebnisliste findet sich im Anhang A.3 in Tabelle A.10. Tabelle 4.17 enthält eine gekürzte Liste mit fester Fenstergröße von 4096 Frames und fester Größe des Random Forest von 100 Bäumen. Da auch in diesem Versuch nur kleine Unterschiede zwischen den Merkmalskombinationen „alle Merkmale“ und „alle Merkmale ohne MFCC“ zu erkennen waren, sind in dieser kurzen Tabelle außerdem nur die Kombinationen „alle Merkmale“ und „nur MFCC“ aufgeführt.

Die wichtigste Erkenntnis, die dieser Test geliefert hat, ist, dass ein zweites Trainingsgenre in der Genre Konfusion auf Ansatz 3 für keine Verbesserung sorgt. Das maximale „f-measure“ erreicht die Kombination aus Dubstep und Trance auf Basis von MFCC-Merkmalen mit 61,89%. Im vorherigen Test konnte das Genre Trance alleine schon ein „f-measure“ von 62,67% erreichen. Somit liegt bei zwei Genres eine Verschlechterung von ca. 1% gegenüber einem Genre vor. Das kleinste „f-measure“ erzielt die Paarung aus Jazz und Klassik auf MFCC-Merkmalen mit nur 38,23%, womit auch das schlechteste Ergebnis aus der einfachen Genre Konfusion untertroffen wurde.

Die höchste „precision“ kann das Paar Hardstyle und Jazz auf allen Merkmalen mit 86,42% erreichen, womit das Maximum von 71,87% aus der einfachen Konfusion auf dem Genre Jazz mit MFCC-Merkmalen übertroffen wird. Die schlechteste „precision“ erzielen Klassik und Rock auf Basis der MFCC-Merkmale mit nur 38,11%. Das Miminum der einfachen Konfusion lag bei 38,43% und wurde vom Genre Rock auf MFCCs erreicht.

T. Genre	ty.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Dubs, Hard	a	20,87%	79,13%	51,42%	48,58%	63,44%	55,02%	36,15%
Dubs, Hard	m	29,14%	70,86%	41,52%	58,48%	59,88%	59,17%	35,33%
Dubs, Jazz	a	33,57%	66,43%	36,64%	63,36%	55,28%	59,05%	35,10%
Dubs, Jazz	m	23,62%	76,38%	43,04%	56,96%	61,18%	58,99%	33,33%
Dubs, Klas	a	50,83%	49,17%	20,52%	79,48%	43,74%	56,43%	35,68%
Dubs, Klas	m	38,90%	61,10%	33,30%	66,70%	45,98%	54,43%	36,10%
Dubs, Rock	a	39,99%	60,01%	38,82%	61,18%	44,18%	51,31%	39,41%
Dubs, Rock	m	41,96%	58,04%	33,27%	66,73%	45,10%	53,82%	37,62%
Dubs, Tran	a	20,99%	79,01%	50,41%	49,59%	58,93%	53,86%	35,70%
Dubs, Tran	m	32,98%	67,02%	30,82%	69,18%	55,98%	61,89%	31,90%
Hard, Jazz	a	4,56%	95,44%	68,85%	31,15%	86,42%	45,79%	36,71%
Hard, Jazz	m	9,40%	90,60%	63,05%	36,95%	78,51%	50,25%	36,22%
Hard, Klas	a	30,15%	69,85%	48,77%	51,23%	55,15%	53,12%	39,46%
Hard, Klas	m	18,92%	81,08%	62,13%	37,87%	59,09%	46,16%	40,53%
Hard, Rock	a	18,82%	81,18%	54,00%	46,00%	64,11%	53,56%	36,41%
Hard, Rock	m	23,48%	76,52%	46,87%	53,13%	62,27%	57,34%	35,18%
Hard, Tran	a	15,44%	84,56%	55,58%	44,42%	71,34%	54,75%	35,51%
Hard, Tran	m	27,90%	72,10%	41,51%	58,49%	64,41%	61,31%	34,70%
Jazz, Klas	a	48,05%	51,95%	29,89%	70,11%	48,17%	57,10%	38,97%
Jazz, Klas	m	9,98%	90,02%	72,66%	27,34%	63,53%	38,23%	41,32%
Jazz, Rock	a	32,01%	67,99%	51,98%	48,02%	49,33%	48,67%	42,00%
Jazz, Rock	m	34,60%	65,40%	51,18%	48,82%	47,76%	48,28%	42,89%
Jazz, Tran	a	18,46%	81,54%	45,06%	54,94%	69,41%	61,33%	31,76%
Jazz, Tran	m	15,33%	84,67%	49,12%	50,88%	71,63%	59,50%	32,22%
Klas, Rock	a	59,51%	40,49%	13,29%	86,71%	42,38%	56,93%	36,40%
Klas, Rock	m	59,62%	40,38%	27,14%	72,86%	38,11%	50,05%	43,38%
Klas, Tran	a	36,22%	63,78%	28,63%	71,37%	53,81%	61,36%	32,42%
Klas, Tran	m	27,79%	72,21%	43,79%	56,21%	54,40%	55,29%	35,79%
Rock, Tran	a	25,23%	74,77%	47,94%	52,06%	55,56%	53,75%	36,59%
Rock, Tran	m	34,49%	65,51%	32,17%	67,83%	54,33%	60,33%	33,33%

Tabelle 4.17: Genre Konfusion mit Ansatz 3 mit zwei Trainingsgenres, vier Testgenres, 4096 Frames Fenster und 100 Bäumen im Random Forest: Trainingsgenres, Merkmalstyp, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Der maximale „recall“ wird von einer Kombination aus Klassik und Rock erreicht und es kamen alle Merkmale zum Einsatz. Er betrug 86,71%, womit er jedoch nicht an das Maximum aus der einfachen Genre Konfusion von 96,15% heran reicht. Den minimalen „recall“ von 27,34% erzielen Jazz und Klassik gemeinsam auf MFCC-Merkmalen.

Das Paar aus Hardstyle und Jazz hebt sich von den anderen Kombinationen besonders durch einen extrem niedrigen Prozentsatz von nur 4,56% an falsch Positiven. Für dieses Ergebnis wurden alle Merkmale verwendet. Kommen nur MFCC-Merkmale mit dieser Trainingskombination zum Einsatz, so gibt es 9,4% falsch Positive. Mit 9,98% falsch Positiven auf MFCC-Merkmalen erreicht auch die Kombination aus Jazz und Klassik einen sehr guten Wert. Werden alle Merkmale mit diesen Trainingsgenres verwendet, so liefert der Random Forest erheblich mehr falsch Positive in Höhe von 48,05%.

Ähnlich wie bei der einfachen Genre Konfusion sind auch bei der doppelten Genre Konfusion die Ergebnisse sehr durchwachsen. Es sind keine klaren Tendenzen zu einer Merkmalskombination zu erkennen. Abhängig von den Trainingsgenres schneidet mal die eine Kombination, mal die andere besser ab. Auch sind keine deutlichen Verbesserung in den Bewertungsmaßen zu erkennen. Die meisten Ergebnisse sind sogar schlechter,n als bei der einfachen Konfusion.

Insgesamt fällt eine Bewertung der Genre Konfusion auf Ansatz 3 schwierig. Die Parameterevaluation für den Random Forest und die Fenstergröße lieferte mit 100 Bäumen bzw. 4096 Frames zwar deutliche Ergebnisse, die auch mit der Evaluation von Ansatz 2 übereinstimmen, doch über die optimale Merkmalskombination lässt sich nur schwer eine Aussage treffen. Nach Tabelle 4.13 schneiden alle Merkmale gemeinsam am besten ab. Der Vorsprung auf die zwei weiteren Kombinationen ist allerdings nur sehr gering. Des Weiteren geben auch die Konfusionsmatrizen kein klares Bild. Die Verknüpfung von zwei Trainingsgenres, die im vorherigen Ansatz die Ergebnisse im „f-measure“ noch leicht steigern konnte, stellte sich für die AMUSE-Merkmale als kontraproduktiv heraus.

4.6 Vergleich der Ansätze

In dieser Arbeit wurden drei Ansätze zur Gesangserkennung getestet. Jeder einzelne Ansatz hatte sein ganz eigenes System, um Gesangspassagen von gesangslosen Passagen zu trennen. Es kamen verschiedene Klassifikatoren und Merkmale zum Einsatz. Trotz aller Punkte, in den sich die drei Ansätze und vor allem auch ihre Ergebnisse unterscheiden, gibt es doch auch eine Menge Gemeinsamkeiten, die in dieser Arbeit gefunden wurden.

Ein Blick auf die Genres, die sich am erfolgreichsten bzw. am erfolglosesten klassifizieren ließen, zeigt, dass die Genres Klassik und Rock auf allen drei Ansätzen mit zu den besten Genres zählen. Des Weiteren hatten alle drei Ansätze Probleme mit dem Genre Hardstyle - es ließ sich stets mit am schlechtesten testen.

Doch es gibt auch Genres, auf denen die drei Ansätze widersprüchliche Ergebnisse produzierten. So war das absolut beste Genre auf dem ersten Ansatz Jazz, während es auf Ansatz 3 das schlechteste war. Auch der Test auf Dubstep brachte sehr unterschiedliche Ergebnisse hervor. Während die Ansätze eins und zwei verglichen mit den anderen Genres ein schlechtes Ergebnis auf Dubstep lieferten, so hatte Ansatz drei im Vergleich zu den anderen Genres weniger Problem mit Dubstep.

Die Genre Konfusion auf den Ansätzen zwei und drei brachte Gemeinsamkeiten und Widersprüche zugleich. Gemein haben beide Ansätze, dass Trance jeweils gut abgeschnitten hat. Auf Ansatz drei war Trance das beste Genre, auf Ansatz zwei mit knappem Rückstand zweitbestes Genre. Beim Genre Dubstep herrscht allerdings keine Einigkeit. Auf Ansatz zwei war es das beste Genre in der Konfusion, auf Ansatz drei lag es deutlich hinter dem besten Genre Trance. Die Genre Konfusion mit zwei Genres erbrachte auf Ansatz zwei einen kleinen Zugewinn. Auf dem dritten Ansatz verschlechterte sich das Ergebnis jedoch.

Die Parameterevaluation auf den beiden letzten Ansätzen lieferte wiederum stimmige Ergebnisse. Tendenziell profitiert das „f-measure“ auf diesen Ansätzen, wenn die Anzahl der Bäume im Random Forest gesteigert wird. Das beste Ergebnis im Verhältnis zum Rechenaufwand liefern beide Ansätze mit 100 Bäumen - dem Standardwert des Tools Weka. Die optimale Fenstergröße der beiden letzten Ansätze liegt bei einem 200ms Fenster für Ansatz zwei bzw. einem 186ms Fenster für Ansatz drei. Beide Werte liegen folglich auf einem Niveau. Die Evaluation auf Ansatz zwei zeigt, dass eine weitere Vergrößerung mehr Nachteile als Vorteile bringt. Ebenso zeigte die Evaluation auf Ansatz drei, dass eine Verkleinerung des Fensters das „f-measure“ reduziert und gleichzeitig einen höheren Rechenaufwand bedeutet.

Insgesamt schnitt der erste Ansatz in dieser Arbeit am besten ab mit einem durchschnittlichen „f-measure“ über alle Genres von 84,23%. Es war zwar keine Parameterevaluation möglich und auch ein Training auf dem in dieser Arbeit genutzten Datensatz konnte nicht durchgeführt werden, umso bemerkenswerter ist, dass dieser Ansatz trotz fehlendem Trainings am besten abgeschnitten hat. Ein gezieltes Training auf den für das neuronale Netz unbekannten Genres Dubstep und Hardstyle könnte sich jedoch positiv auswirken und ein noch höheres Ergebnis ermöglichen.

Knapp hinter dem ersten Ansatz liegt der zweite Ansatz mit einem „f-measure“ von 81,45% in der Kreuzvalidierung. Auffallend bei diesem Ansatz war, dass die Merkmale äußerst gering auf die Parameterevaluation reagiert haben, denn alle Parametervariationen lieferten Ergebnisse, die eine Toleranz von nur einem Prozent lieferten. Die zugrunde liegenden Merkmale müssen daher extrem optimiert sein, sodass sie auch bei einer schlechten Parameterkonfiguration noch solch hohe Ergebnisse liefern.

Am schlechtesten Schnitt der dritte Ansatz ab. Doch auch dessen „f-measure“ in der Kreuzvalidierung liegt mit 69,89% noch deutlich über dem des Zufallsklassifikators. Bemerkenswert an dessen Ergebnissen war, dass die Merkmalskombination ohne MFCC-

Merkmale beinahe so gut abgeschnitten hat wie die Kombination aus allen Merkmalen. Es ist also möglich, auch ohne MFCC-Merkmale einen brauchbaren Ansatz zu erstellen. Durch Optimierungen an den Merkmalen und ein gezielteres Auswählen von Merkmalen ließe sich auch dieser Ansatz noch weiter verbessern.

Kapitel 5

Zusammenfassung

5.1 Fazit

Das Fazit dieser Arbeit fällt durchaus positiv aus. Neben einigen zu erwartenden Ergebnissen konnten auch einige neue Erkenntnisse gewonnen werden. Besonders über das Verhalten der drei Ansätze auf verschiedenen Genres wurden umfangreiche Beobachtungen gemacht. Die Genres Klassik und Rock lassen sich allgemein recht gut klassifizieren. Keiner der drei Ansätze hat dort große Schwierigkeiten. Problematischer waren allerdings die Genres Dubstep und Hardstyle. Dort schnitten die drei Ansätze im Vergleich zu den anderen Genres deutlich schlechter ab.

Der erste Ansatz aus [15] überzeugte vor allem durch die Tatsache, dass er trotz des Trainings auf dem Jamendo-Datensatz ein Ergebnis nahe an den Werten aus der Veröffentlichung liefern konnte. Die Stärken dieses Ansatzes auf den Genres Jazz und Rock haben gezeigt, dass er sich nicht, wie anfangs vermutet, von Saxophonen oder E-Gitarren verwirren lässt. Die Schwäche dieses Ansatzes auf den Genres Dubstep und Hardstyle kann man auf fehlendes Training zurückführen, da der Jamendo-Datensatz keine elektronische Musik enthält. Ein generelles Problem mit elektronischer Musik scheint es jedoch nicht zu geben, da das dritte elektronische Genre Trance deutlich besser abschnitt.

Ansatz zwei aus [9] erreicht in den Kreuzvalidierung ebenfalls Ergebnisse nahe an seiner Veröffentlichung. Seine Stärken lagen auf den Genres Klassik und Rock, womit sich auch hier die anfängliche Vermutung, dass E-Gitarren und Violinen den Klassifikator verwirren, widerlegen lies. Ähnlich wie der erste Ansatz, hatte auch dieser Ansatz Probleme mit den Genres Dubstep und Hardstyle, obwohl in diesem Fall ein Training auf beiden Genres erfolgte. Da die anderen Genres jedoch deutlich besser abschnitten, scheint das Problem tatsächlich durch die Charakteristika dieser beiden Genres verursacht zu werden.

Auch wenn der Ansatz auf Basis der AMUSE-Merkmale nicht an die hohen Werte der anderen Ansätze im „f-measure“ heran kam, so konnte auch dieser Ansatz überzeugen. Ohne Optimierungen und mit einigen Merkmalen, die nicht für die Gesangserkennung

ausgelegt wurden, konnte trotzdem ein System entworfen werden, das deutlich besser als der Zufallsklassifikator arbeitet. Die Stärken dieses Ansatzes lagen in den Genres Klassik und Rock, die Schwächen bei Hardstyle und Jazz.

Die Tests zur Genre Konfusion haben interessante Ergebnisse über die sechs Testgenres geliefert. So wird mit Jazz als Trainingsgrundlage stets ein sehr restriktives Modell aufgebaut, das die falsch Positiven sehr gering, die falsch Negativen jedoch sehr hoch hält. Rock hingegen erzeugt ein extensives Modell, was viele falsch Positive ergibt. Als alleiniges Testgenre eignen sich Dubstep und Trance am besten.

Die in dieser Arbeit durchgeführte Parameterevaluation auf den Ansätze zwei und drei lieferte Übereinstimmende Ergebnisse für die Fenstergröße sowie die Anzahl der Bäume im Random Forest. Das Klassifikationsfenster sollte eine Größe von ca. 200ms haben, der Random Forest eine Größe von 100 Bäumen.

Abschließend lässt sich sagen, dass diese Arbeit gezeigt hat, dass das Ergebnis der Gesangsklassifikation nicht nur von den verwendeten Merkmalen und dem gewählten Klassifikator abhängt, sondern auch von den zugrunde liegenden Testdaten. Je nach Genre kann es zu starken Schwankungen bei „recall“, „precision“ und „f-measure“ kommen.

5.2 Ausblick

Neben einigen Antworten hat diese Arbeit auch eine neue Fragen aufgeworfen. Auf die Frage, ob es Unterschiede in den Musikgenres bei der Gesangserkennung gibt, folgt nun die Frage nach dem Warum. Welche Charakteristika sorgen dafür, dass sich beispielsweise Klassik so gut testen lässt, während Dubstep und Hardstyle große Probleme verursachen? Hierzu müssen die einzelnen Musikstück genau untersucht werden und die Passagen exakt identifiziert werden, welche der Klassifikator falsch zugeordnet hat.

Weiterhin bleibt die Frage nach dem optimalen Trainingsdatensatz offen. Ein potenzielles perfektes System zur Gesangserkennung, das in jeder Form von Musik problemlos Gesang erkennen kann, besteht neben einem Klassifikator und Merkmalen eben auch aus Beispieldaten, mit denen der Klassifikator trainiert wurde. Welch großen Einfluss diese Trainingsdaten auf die Klassifikation nehmen, hat diese Arbeit gezeigt.

Speziell bleibt zu klären, ob der oft verwendete Jamendo-Datensatz weiterhin für Vergleiche heran gezogen werden sollte. Seine durch Zufall bestimmte Zusammensetzung deckt zwar ein breites Spektrum an Genres und Musikstilen ab, der Mangel an elektronischer Musik stellt jedoch eine deutliche Schwäche dar. Es besteht die Gefahr, dass auf diesem Datensatz optimierte Systeme durch eine Überanpassung nicht mehr auf elektronischer Musik klassifizieren können. Dieser Effekt war bereits bei dem ersten Ansatz zu beobachten: die Klassifikation auf Dubstep und Hardstyle verlief mangels Trainingsbeispielen schlecht.

Bei der Findung eines optimalen Trainingsdatensatz könnten die Ergebnisse aus der Genre Konfusion hilfreich sein, denn dort hat sich gezeigt, dass genau die im Jamdeno-

Datensatz fehlenden elektronischen Genres Dubstep und Trance als alleinige Trainingsgrundlage am besten abschneiden. Außerdem lassen sich die Auswirkungen auf „recall“, „precision“ bei Kombination zweier Genres gut beobachten.

Im Bezug auf die AMUSE-Merkmale bleibt zu klären, weshalb sie in einigen Fällen solch durchwachsene Ergebnisse lieferten. Es stellt sich die Frage, weshalb der Unterschied zwischen den Gruppen „alle Merkmale“ und „ohne MFCC“ so gering ausfiel. Gab es Merkmale, die das Ergebnis negativ beeinflussten? Oder gab es vielleicht Merkmale, die sich mit MFCCs vergleichbar gut zur Gesangserkennung eignen? Eine detaillierte Untersuchung der einzelnen Merkmale scheint angebracht.

Anhang A

Anhang

A.1 Verwendeter Testdatensatz

Tabelle A.1: Liste aller verwendeten Genres, Zeitangaben in Minuten

Genre	Länge	ignorierte Zeit	vocal Zeit	nonVocal Zeit
Dubstep	114,3	5,61	54,03	54,66
Hardstyle	111,44	3,06	21,43	86,95
Jazz	98,48	0,04	29,03	69,41
Klassik	104,6	0	55,02	49,58
Rock	75,2	0	41,75	33,45
Trance	97,92	2,56	34,37	60,99
Gesamt	601,94	11,27	235,63	355,04

Tabelle A.2: Liste aller verwendeten Teststücke, Zeitangaben in Sekunden

ID	Genre	Track	gesamt	igno.	vocal	nVocal
1	Dubstep	Above & Beyond - Tri-State (5vel Remix)	391,47	56,53	67,69	267,25
2	Dubstep	Adventure Club feat. Delaney Jane - Limitless	243,49	14,81	196,59	32,08
3	Dubstep	Azedia - Something (Rameses B Remix)	266,50	55,48	156,63	54,40
4	Dubstep	Crywolf & Ianborg - Stay	289,52	6,69	236,37	46,45
5	Dubstep	Crywolf & Skrux - Tides	272,48	18,33	143,52	110,63
6	Dubstep	Flux Pavilion - Freeway (Flux Pavilion & Kill The Noise Remix)	163,53	0,00	99,52	64,01

ID	Genre	Track	gesamt	igno.	vocal	nVocal
7	Dubstep	Gemini - Blue	314,10	3,01	162,21	148,87
8	Dubstep	Jakwob feat. Maiday - Fade	206,52	0,00	165,09	41,44
9	Dubstep	L.a.o.s - Healing	226,53	0,00	173,20	53,34
10	Dubstep	Lana Del Rey - Ultraviolence (Datsik Remix)	236,46	0,00	131,55	104,91
11	Dubstep	Metric - Artificial Nocturne (Love Thy Brother Remix)	271,52	0,00	108,28	163,23
12	Dubstep	Mr Fijiwiji feat. Danyka Nadeau - Yours Truly	281,52	0,00	186,52	95,00
13	Dubstep	Myon & Shane 54 feat. Aruna - Lights (5vel Remix)	465,53	37,18	198,21	230,14
14	Dubstep	O.a.e - Starstruck	250,62	0,00	0,00	250,62
15	Dubstep	Oceanlab - Satellite (Seven Lions Remix)	312,48	0,00	141,97	170,50
16	Dubstep	Phaeleh feat. Soundmouse - Afterglow (Akira Kiteshi Remix)	199,00	39,44	133,32	26,24
17	Dubstep	Roeyksopp feat. Susanne Sundfor - Running To The Sea (Seven Lions Remix)	341,47	27,39	170,31	143,77
18	Dubstep	Rogue - Goodbye (Winter Chords)	221,52	0,00	0,00	221,52
19	Dubstep	Salm - Endless Stairs	251,51	0,00	0,00	251,51
20	Dubstep	Skrux - Essence	217,47	15,60	55,32	146,54
21	Dubstep	Sound Remedy & Illenium feat. King Deco - Spirals	362,53	61,98	156,32	144,23
22	Dubstep	Superbus - All Alone (Seven Lions Remix)	264,52	0,00	89,00	175,52
23	Dubstep	Synthetic Epiphany & Coma - Beauty In Errors	261,51	0,00	174,40	87,11
24	Dubstep	Tut Tut Child feat. Isabel Higuero - Gravity	334,52	0,00	138,56	195,96
25	Dubstep	Zhu - Paradise Awaits	211,54	0,00	157,31	54,23
26	Hardstyle	A-wak - F.N.W.	151,07	4,17	2,21	144,69
27	Hardstyle	A. Spencer, DJ Gollum - In The Shadows (Mikesh's Fucking Cocaine Remix)	385,02	6,27	84,06	294,69

ID	Genre	Track	gesamt	igno.	vocal	nVocal
28	Hardstyle	Activator & Zatox - Freedom	278,20	9,75	57,74	210,71
29	Hardstyle	Builder - Her Voice (Headhunterz Remix Edit)	274,86	0,00	69,39	205,47
30	Hardstyle	Coone - Musical Notez (Original)	315,40	38,92	26,49	249,99
31	Hardstyle	D-block & S-te-fan - Kingdom (The Pitcher Re-amp)	259,53	14,88	63,49	181,15
32	Hardstyle	DJ Desnar, Distractor - Back 2 Back	361,85	7,20	41,47	313,18
33	Hardstyle	Donkey Rollers - Chaos	207,96	0,00	0,00	207,96
34	Hardstyle	E-fect - Glorification (Activator & Zatox Mix)	326,06	11,54	87,37	227,15
35	Hardstyle	Headhunterz - Psychedelic	267,18	0,00	88,73	178,45
36	Hardstyle	Jan Wayne Vs. Raindropz! - Numb (The Illuminatorz Remix)	315,40	2,36	61,14	251,91
37	Hardstyle	Killer Clown - Silentium	226,32	2,62	14,98	208,72
38	Hardstyle	Max Enforcer - Damned	208,56	0,00	77,47	131,09
39	Hardstyle	Noisecontrollers - Yellow Minute (Alpha Twins Remix)	268,54	0,00	33,46	235,08
40	Hardstyle	Pavo - Communicate	190,96	0,00	26,23	164,73
41	Hardstyle	Rephex & Abw - Dangerous Danes (Abw's Slasher Mix)	203,52	0,00	63,42	140,10
42	Hardstyle	Smd - Just Like You	294,58	1,96	88,58	204,04
43	Hardstyle	The Pitcher - Sleeping	230,30	28,24	16,86	185,19
44	Hardstyle	The Pitcher and Noisecontrollers - Cherish	206,00	4,97	26,83	174,20
45	Hardstyle	Wasted Penguinz - Extension (Extended Version)	330,37	0,00	54,73	275,64
46	Hardstyle	Wasted Penguinz feat. Mattias Welin - Out Of Reach (Extended Version)	333,64	11,37	76,29	245,97
47	Hardstyle	Wasted Penguinz and Chris One - I Miss You (Extended Version)	301,51	8,28	115,89	177,34
48	Hardstyle	Wildstylez & DJ Isaac - Lost In Music	246,28	28,22	41,56	176,50

ID	Genre	Track	gesamt	igno.	vocal	nVocal
49	Hardstyle	Zairon - Stand Up	260,28	0,00	19,21	241,08
50	Hardstyle	Zatox - A New Dimension	243,28	2,99	48,03	192,26
51	Jazz	John Coltrane - Giant Steps	286,22	0,00	0,00	286,22
52	Jazz	Jacques Loussier Trio - Spring - Concerto No. 1 in E Major - Allegro	418,14	0,00	0,00	418,14
53	Jazz	Herbie Mann - Blue Dip	566,52	0,00	0,00	566,52
54	Jazz	Miles Davis - Blue In Green	337,63	0,00	0,00	337,63
55	Jazz	Michael Brecker - Song for Bilbao	345,03	0,00	0,00	345,03
56	Jazz	Duke Ellington - Very Special	266,21	0,00	0,00	266,21
57	Jazz	David Benoit - Snow Dancing	245,84	0,00	0,00	245,84
58	Jazz	Sonny Rollis - The Stopper	178,02	0,00	0,00	178,02
59	Jazz	Charlie Parker - Segment	200,65	0,00	0,00	200,65
60	Jazz	Count Basie - Every Tub	192,34	0,00	94,46	97,88
61	Jazz	Chet Baker - There Is No Greater Love	155,27	0,00	121,56	33,71
62	Jazz	Count Basie Orchestra - Cottontail	196,05	0,00	154,71	41,34
63	Jazz	David Benoit - When The Winter's Gone	251,51	0,00	133,78	117,72
64	Jazz	Glenn Miller - The Jumpin Jive	153,34	0,00	28,53	124,81
65	Jazz	Heavy Metal Horns - Hey Ramon	181,55	0,00	100,48	81,08
66	Jazz	Heavy Metal Horns - Inferno	268,25	2,14	126,05	140,06
67	Jazz	Louis Armstrong - Hello Dolly	148,30	0,00	70,53	77,77
68	Jazz	Michael Brecker - The Nearness Of You - Nearness Of You	278,05	0,00	143,22	134,83
69	Jazz	Nina Simone - I Need A Little Sugar In My Bowl	217,86	0,00	117,98	99,88
70	Jazz	Ray Charles - Busted	124,34	0,00	66,44	57,90
71	Jazz	Ray Charles - Crying Time	174,34	0,00	136,98	37,36
72	Jazz	Supercharge - Blue Monday	175,18	0,00	99,88	75,30
73	Jazz	Tower Of Power - Down To The Nightclub (Bump City)	166,01	0,00	118,14	47,87

ID	Genre	Track	gesamt	igno.	vocal	nVocal
74	Jazz	Tower Of Power - Gone (In Memory Of Jacqueline Mesquite)	224,03	0,00	130,76	93,27
75	Jazz	Wil Salden And His Orchestra - Well Alright, Okay, You Win - Glenn Miller	158,01	0,00	98,15	59,86
76	Klassik	Wagner - Tristan und Isolde Akt 3 - Liebestod	427,55	0,00	284,36	143,19
77	Klassik	Carl Orff - O Fortuna	157,62	0,00	127,34	30,29
78	Klassik	Beethoven - Sonata quasi una Fantasia No.14 in C# minor, Op 27 No.2 Moonlight - Adagio sostenuto"	328,80	0,00	0,00	328,80
79	Klassik	Alfred Brendel - Italian Concerto in F [BWV 971] 2 Andante	341,37	0,00	0,00	341,37
80	Klassik	Johann Sebastian Bach - Triosonate No.1 BWV525 es-dur - 1. Allegro	193,07	0,00	0,00	193,07
81	Klassik	Carl Orff - Omnia Sol Temerat	114,10	0,00	75,99	38,11
82	Klassik	Hector Berlioz - Damnation de Faust, 2. Marche Hongroise	274,60	0,00	0,00	274,60
83	Klassik	Frederic Chopin - Mazurka in e-Moll Op.41 No. 2	129,80	0,00	0,00	129,80
84	Klassik	Vladimir Ashkenazy - E-Dur Op. Posth.	131,13	0,00	0,00	131,13
85	Klassik	David Hansen, countertenor - Finals Concert of the Australian Singing Competition	498,09	0,00	260,89	237,20
86	Klassik	Adoro - Engel Fliegen Einsam	255,82	0,00	211,26	44,55
87	Klassik	Amaury Vassili - Nella Fantasia	264,39	0,00	177,03	87,36
88	Klassik	Andrea Bocelli - Romanza	225,36	0,00	156,81	68,55
89	Klassik	ARIA - Che Faro Senza Euridice (Orfeo Ed Euridice)	233,29	0,00	156,39	76,90
90	Klassik	Beverly Sills - Una Voce Poco Fa!!	373,24	0,00	233,14	140,10

ID	Genre	Track	gesamt	igno.	vocal	nVocal
91	Klassik	Cosi Fan tutte trio - Soave il vento	159,78	0,00	127,71	32,07
92	Klassik	Denza - Funiculi, Funicula	145,10	0,00	114,68	30,42
93	Klassik	Il Divo - Everytime I Look At You	210,02	0,00	161,58	48,45
94	Klassik	Il Divo - Nella Fantasia	266,89	0,00	175,40	91,50
95	Klassik	Luciano Pavarotti - Chanson D'adieu	165,30	0,00	119,36	45,94
96	Klassik	Paul Potts - Nessun Dorma	172,62	0,00	140,18	32,43
97	Klassik	Sarah Brightman & The London Symphony Orchestra - Who Wants To Live Forever	236,04	0,00	172,04	64,00
98	Klassik	Sarah Brightman, Andrea Bocelli - Time To Say Goodbye	245,05	0,00	193,11	51,94
99	Klassik	Verdi - Rigoletto - La Donna E Mobile	134,10	0,00	91,44	42,66
100	Klassik	Vivaldi - Gelido in ogni vena - Max Emanuel Cencic, countertenor	592,90	0,00	322,75	270,15
101	Rock	Alexisonfire - A Dagger Through The Heart of St. Angeles	252,05	0,00	89,50	162,55
102	Rock	Armor For Sleep - Dream To Make Believe	181,34	0,00	84,06	97,28
103	Rock	Atreyu - Bleeding Mascara	146,23	0,00	104,53	41,71
104	Rock	Band of Horses - Funeral	322,27	0,00	146,83	175,45
105	Rock	Bayside - Masterpiece	224,71	0,00	142,53	82,18
106	Rock	Cheap Trick - Hello There	99,85	0,00	60,88	38,97
107	Rock	Coheed and Cambria - A Favor House Atlantic	234,76	0,00	158,98	75,78
108	Rock	Cursive - The Martyr	237,58	0,00	112,91	124,68
109	Rock	Death from Above - Blood On Our Hands	179,96	0,00	107,70	72,26
110	Rock	Descendents - Cool To Be You	144,90	0,00	105,98	38,92
111	Rock	Devo - Uncontrollable Urge	198,39	0,00	75,17	123,22
112	Rock	Dreamwalk - Simple	172,88	0,00	97,27	75,60
113	Rock	Ex-Girl - Rocket Keronian	269,64	0,00	150,13	119,50

ID	Genre	Track	gesamt	igno.	vocal	nVocal
114	Rock	Fall Out Boy - Dead On Arrival	194,56	0,00	132,34	62,22
115	Rock	Konks - 29 Fingers	178,00	0,00	88,30	89,70
116	Rock	Libyans - Welcome To The Neighborhood	141,64	0,00	83,00	58,64
117	Rock	Motion City Soundtrack - Make Out Kids	184,89	0,00	121,26	63,63
118	Rock	NOFX - The Rationality Of Irrationality	150,60	0,00	71,40	79,19
119	Rock	Paint It Black - Exit Wounds	97,96	0,00	57,03	40,93
120	Rock	Ramones - Blitzkrieg Bop	137,59	0,00	68,40	69,19
121	Rock	The Clash - I Fought The Law	156,20	0,00	77,80	78,56
122	Rock	The Faint - Call Call	146,63	0,00	57,54	89,08
123	Rock	The Hong Kong - Mazerati	194,43	0,00	115,59	78,84
124	Rock	The Soviettes - Tonight	143,99	0,00	93,12	50,87
125	Rock	Yeah Yeah Yeahs - Art Star	120,76	0,00	102,96	17,80
126	Trance	4 Strings feat. Samantha Fox - Forever	195,60	0,00	129,66	65,95
127	Trance	8 Wonders - The Return	361,82	0,00	0,00	361,82
128	Trance	Airbeat One Project - Airbeat Army (Arena Edit)	213,42	7,87	159,56	46,00
129	Trance	Akesson - Perfect Blue	140,93	0,00	0,00	140,93
130	Trance	Armin Van Buuren & Sophie Ellis-bextor - Not Giving Up On Love	173,95	10,89	125,57	37,49
131	Trance	ATB - Ecstasy	248,71	0,00	120,00	128,71
132	Trance	Cosmic Gate - London Rain	240,12	36,38	100,68	103,06
133	Trance	Cosmic Gate feat. Jes - Flying Blind	240,48	21,78	154,42	64,29
134	Trance	Dash Berlin feat. Jonathan Medelsohn - World Falls Apart	196,57	0,00	153,58	43,00
135	Trance	DJ Tatana - Children	369,92	0,00	0,00	369,92
136	Trance	Emma Hewitt - Miss You Paradise (Shogun Remix)	209,19	0,00	146,59	62,60

ID	Genre	Track	gesamt	igno.	vocal	nVocal
137	Trance	Emotional Horizons & X-plorations feat. Aneym - Nobody Knows	196,34	0,00	70,31	126,03
138	Trance	Eric De Koning - Dream Flight	169,93	0,00	0,00	169,93
139	Trance	Johmar feat. Carie - Let Go (Trance Arts Remix Edit)	233,40	0,00	144,17	89,24
140	Trance	Kai Tracid - 4 Just 1 Day (Club Version)	223,74	13,68	43,07	166,98
141	Trance	Nadia Ali - Rapture (Avicii Remix)	217,97	2,25	83,31	132,41
142	Trance	Neo Cortex - Don't You (Trance Radio Mix)	214,99	0,00	116,97	98,02
143	Trance	Nic Chagall - Sky Blue	229,59	0,00	0,00	229,59
144	Trance	Paul Van Dyk - Home	365,71	0,00	122,49	243,23
145	Trance	Peter Heppner - Alleinesein	203,36	0,00	142,99	60,37
146	Trance	Primer - Everlast (Original Edit)	419,40	0,00	0,00	419,40
147	Trance	Sander Van Doorn Vs. Robbie Williams - Close My Eyes (Radio Edit)	180,38	44,89	36,72	98,77
148	Trance	Tiesto - Adagio For Strings	206,13	0,00	0,00	206,13
149	Trance	Tiesto feat. Kirsty Hawkshaw - Just Be	192,03	0,00	102,17	89,85
150	Trance	Will Holland feat. Jeza - Start Again (Juventa Remix Edit)	231,50	15,63	109,93	105,94

Tabelle A.3: Einteilung des Testdatensatzes in die zehn Folds für die Kreuzvalidierung.

Fold	IDs der Musikstücke
0	119, 120, 0, 31, 39, 98, 103, 110, 68, 130, 131, 16, 29, 121, 24
1	84, 4, 45, 32, 11, 102, 122, 113, 112, 127, 52, 97, 28, 63, 109
2	18, 6, 105, 126, 85, 15, 94, 41, 118, 138, 26, 9, 44, 60, 82
3	91, 133, 88, 92, 124, 80, 111, 2, 71, 7, 106, 23, 59, 73, 145
4	53, 93, 87, 37, 20, 30, 89, 27, 14, 50, 125, 79, 12, 51, 104
5	42, 58, 96, 49, 101, 146, 5, 43, 25, 17, 34, 134, 128, 115, 81
6	33, 99, 75, 76, 55, 148, 22, 57, 77, 78, 135, 136, 143, 72, 56
7	140, 141, 142, 147, 149, 117, 21, 132, 137, 90, 62, 139, 144, 107, 66
8	47, 67, 65, 40, 1, 64, 46, 61, 8, 100, 70, 38, 74, 10, 108
9	3, 13, 19, 35, 36, 48, 54, 69, 83, 86, 95, 114, 116, 123, 129

A.2 Auswertungstabellen Ansatz 2

Tabelle A.4: „leave one out“ mit Ansatz 2 auf jedem Genre getrennt: Genre, Fenstergröße, Anzahl der Bäume im Random Forest, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Genre	win.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Dubstep	200	50	0,230	0,770	0,223	0,777	0,817	0,778	0,224
Dubstep	200	100	0,222	0,778	0,225	0,775	0,824	0,780	0,221
Dubstep	200	200	0,221	0,780	0,230	0,771	0,824	0,778	0,222
Dubstep	400	50	0,226	0,774	0,228	0,772	0,821	0,775	0,226
Dubstep	400	100	0,216	0,784	0,233	0,767	0,826	0,775	0,222
Dubstep	400	200	0,210	0,790	0,234	0,766	0,832	0,777	0,220
Hardstyle	200	50	0,032	0,968	0,343	0,657	0,788	0,685	0,188
Hardstyle	200	100	0,030	0,970	0,347	0,653	0,811	0,690	0,189
Hardstyle	200	200	0,030	0,970	0,352	0,648	0,810	0,687	0,191
Hardstyle	400	50	0,030	0,970	0,340	0,660	0,790	0,678	0,185
Hardstyle	400	100	0,027	0,974	0,348	0,652	0,815	0,688	0,187
Hardstyle	400	200	0,025	0,975	0,347	0,653	0,822	0,692	0,186
Jazz	200	50	0,108	0,892	0,287	0,713	0,897	0,781	0,212
Jazz	200	100	0,104	0,896	0,287	0,713	0,905	0,783	0,210
Jazz	200	200	0,102	0,899	0,287	0,713	0,905	0,783	0,209
Jazz	400	50	0,096	0,904	0,283	0,717	0,917	0,788	0,201
Jazz	400	100	0,094	0,907	0,278	0,722	0,919	0,793	0,199
Jazz	400	200	0,094	0,906	0,277	0,723	0,920	0,794	0,199

Genre	win.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Klassik	200	50	0,093	0,908	0,062	0,938	0,956	0,945	0,085
Klassik	200	100	0,085	0,915	0,064	0,936	0,959	0,945	0,082
Klassik	200	200	0,084	0,916	0,064	0,937	0,960	0,946	0,080
Klassik	400	50	0,077	0,923	0,059	0,941	0,964	0,950	0,073
Klassik	400	100	0,073	0,927	0,058	0,942	0,966	0,951	0,070
Klassik	400	200	0,073	0,927	0,058	0,942	0,966	0,952	0,070
Rock	200	50	0,155	0,845	0,121	0,879	0,893	0,878	0,138
Rock	200	100	0,147	0,853	0,121	0,879	0,897	0,880	0,134
Rock	200	200	0,141	0,859	0,122	0,878	0,902	0,881	0,132
Rock	400	50	0,142	0,858	0,125	0,875	0,899	0,877	0,133
Rock	400	100	0,127	0,873	0,127	0,873	0,908	0,882	0,127
Rock	400	200	0,123	0,877	0,121	0,879	0,912	0,886	0,122
Trance	200	50	0,086	0,915	0,175	0,825	0,901	0,848	0,142
Trance	200	100	0,078	0,922	0,173	0,827	0,909	0,853	0,138
Trance	200	200	0,078	0,922	0,174	0,826	0,909	0,853	0,138
Trance	400	50	0,075	0,925	0,171	0,829	0,912	0,857	0,134
Trance	400	100	0,070	0,930	0,172	0,828	0,916	0,857	0,132
Trance	400	200	0,069	0,931	0,171	0,829	0,918	0,858	0,130

Tabelle A.5: Genre Konfusion mit Ansatz 2 mit einem Trainingsgenre und fünf Testgenres: Trainingsgenre, Fenstergröße, Anzahl der Bäume im Random Forest, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

TGenre	wi.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Dubstep	200	50	0,180	0,820	0,183	0,817	0,731	0,772	0,181
Hardstyle	200	50	0,145	0,855	0,319	0,681	0,788	0,731	0,232
Jazz	200	50	0,020	0,980	0,590	0,410	0,937	0,570	0,305
Klassik	200	50	0,193	0,807	0,294	0,706	0,681	0,693	0,244
Rock	200	50	0,351	0,649	0,172	0,828	0,584	0,685	0,262
Trance	200	50	0,207	0,793	0,187	0,813	0,727	0,767	0,197
Dubstep	200	100	0,172	0,828	0,184	0,816	0,739	0,775	0,178
Hardstyle	200	100	0,137	0,863	0,322	0,678	0,798	0,733	0,229
Jazz	200	100	0,018	0,982	0,593	0,407	0,943	0,569	0,305
Klassik	200	100	0,185	0,815	0,304	0,696	0,687	0,691	0,245
Rock	200	100	0,351	0,649	0,175	0,825	0,583	0,684	0,263
Trance	200	100	0,196	0,804	0,192	0,808	0,736	0,770	0,194
Dubstep	200	200	0,169	0,831	0,185	0,815	0,743	0,777	0,177

TGenre	wi.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Hardstyle	200	200	0,132	0,868	0,324	0,676	0,802	0,734	0,228
Jazz	200	200	0,016	0,984	0,595	0,405	0,947	0,568	0,305
Klassik	200	200	0,181	0,819	0,304	0,696	0,691	0,693	0,243
Rock	200	200	0,341	0,659	0,182	0,818	0,588	0,685	0,261
Trance	200	200	0,192	0,808	0,196	0,804	0,740	0,770	0,194
Dubstep	400	50	0,171	0,829	0,175	0,825	0,739	0,780	0,173
Hardstyle	400	50	0,132	0,868	0,318	0,682	0,802	0,737	0,225
Jazz	400	50	0,017	0,983	0,598	0,402	0,945	0,564	0,307
Klassik	400	50	0,190	0,810	0,291	0,709	0,681	0,695	0,241
Rock	400	50	0,350	0,650	0,167	0,833	0,583	0,686	0,258
Trance	400	50	0,195	0,805	0,193	0,807	0,734	0,769	0,194
Dubstep	400	100	0,163	0,837	0,177	0,823	0,747	0,783	0,170
Hardstyle	400	100	0,123	0,877	0,325	0,675	0,811	0,737	0,224
Jazz	400	100	0,015	0,985	0,593	0,407	0,952	0,570	0,304
Klassik	400	100	0,191	0,809	0,279	0,721	0,684	0,702	0,235
Rock	400	100	0,341	0,659	0,177	0,823	0,587	0,685	0,259
Trance	400	100	0,189	0,811	0,196	0,804	0,739	0,770	0,193
Dubstep	400	200	0,162	0,838	0,177	0,823	0,748	0,784	0,170
Hardstyle	400	200	0,122	0,878	0,329	0,671	0,812	0,735	0,225
Jazz	400	200	0,014	0,986	0,594	0,406	0,954	0,569	0,304
Klassik	400	200	0,185	0,815	0,288	0,712	0,689	0,700	0,236
Rock	400	200	0,334	0,666	0,183	0,817	0,589	0,685	0,258
Trance	400	200	0,185	0,815	0,198	0,802	0,742	0,771	0,192

Tabelle A.6: Genre Konfusion mit Ansatz 2 mit zwei Trainingsgenres und vier Testgenres: Trainingsgenres, Fenstergröße, Anzahl der Bäume im Random Forest, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

T. Genre	wi.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Dubs, Hard	200	50	0,192	0,808	0,211	0,789	0,753	0,771	0,202
Dubs, Hard	200	100	0,184	0,816	0,214	0,786	0,760	0,773	0,199
Dubs, Hard	200	200	0,181	0,819	0,216	0,784	0,763	0,774	0,198
Dubs, Hard	400	50	0,188	0,812	0,208	0,792	0,756	0,774	0,198
Dubs, Hard	400	100	0,177	0,823	0,210	0,790	0,766	0,778	0,194
Dubs, Hard	400	200	0,175	0,825	0,210	0,790	0,768	0,779	0,192
Dubs, Jazz	200	50	0,044	0,956	0,297	0,703	0,913	0,794	0,171
Dubs, Jazz	200	100	0,039	0,961	0,304	0,696	0,921	0,793	0,171

T. Genre	wi.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Dubs, Jazz	200	200	0,037	0,963	0,304	0,696	0,924	0,794	0,171
Dubs, Jazz	400	50	0,040	0,960	0,297	0,703	0,918	0,796	0,169
Dubs, Jazz	400	100	0,034	0,966	0,302	0,698	0,929	0,797	0,168
Dubs, Jazz	400	200	0,032	0,968	0,302	0,698	0,934	0,799	0,167
Dubs, Klas	200	50	0,203	0,797	0,184	0,816	0,666	0,733	0,193
Dubs, Klas	200	100	0,197	0,803	0,191	0,809	0,671	0,733	0,194
Dubs, Klas	200	200	0,193	0,807	0,190	0,810	0,676	0,737	0,191
Dubs, Klas	400	50	0,200	0,800	0,181	0,819	0,665	0,734	0,191
Dubs, Klas	400	100	0,193	0,807	0,182	0,818	0,673	0,738	0,188
Dubs, Klas	400	200	0,190	0,810	0,185	0,815	0,676	0,739	0,187
Dubs, Rock	200	50	0,225	0,775	0,128	0,872	0,666	0,755	0,177
Dubs, Rock	200	100	0,217	0,783	0,127	0,873	0,675	0,761	0,172
Dubs, Rock	200	200	0,210	0,790	0,128	0,872	0,682	0,765	0,169
Dubs, Rock	400	50	0,220	0,780	0,127	0,873	0,668	0,756	0,174
Dubs, Rock	400	100	0,212	0,788	0,126	0,874	0,677	0,763	0,169
Dubs, Rock	400	200	0,206	0,794	0,123	0,877	0,683	0,768	0,165
Dubs, Tran	200	50	0,191	0,809	0,177	0,823	0,723	0,770	0,184
Dubs, Tran	200	100	0,186	0,814	0,180	0,820	0,728	0,771	0,183
Dubs, Tran	200	200	0,183	0,817	0,181	0,819	0,731	0,772	0,182
Dubs, Tran	400	50	0,189	0,811	0,170	0,830	0,722	0,772	0,180
Dubs, Tran	400	100	0,183	0,817	0,170	0,830	0,729	0,776	0,176
Dubs, Tran	400	200	0,179	0,821	0,172	0,828	0,732	0,777	0,176
Hard, Jazz	200	50	0,022	0,978	0,513	0,487	0,954	0,644	0,268
Hard, Jazz	200	100	0,020	0,980	0,519	0,481	0,956	0,640	0,270
Hard, Jazz	200	200	0,020	0,980	0,519	0,481	0,958	0,640	0,269
Hard, Jazz	400	50	0,018	0,982	0,505	0,495	0,963	0,654	0,261
Hard, Jazz	400	100	0,016	0,984	0,518	0,482	0,965	0,643	0,267
Hard, Jazz	400	200	0,015	0,985	0,522	0,478	0,967	0,640	0,268
Hard, Klas	200	50	0,176	0,824	0,298	0,702	0,742	0,721	0,237
Hard, Klas	200	100	0,168	0,832	0,303	0,697	0,750	0,722	0,236
Hard, Klas	200	200	0,166	0,834	0,303	0,697	0,752	0,723	0,235
Hard, Klas	400	50	0,173	0,827	0,297	0,703	0,743	0,722	0,235
Hard, Klas	400	100	0,166	0,834	0,302	0,698	0,750	0,723	0,234
Hard, Klas	400	200	0,162	0,838	0,306	0,694	0,754	0,723	0,234
Hard, Rock	200	50	0,199	0,801	0,216	0,784	0,742	0,763	0,207
Hard, Rock	200	100	0,188	0,812	0,215	0,785	0,753	0,769	0,201

T. Genre	wi.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Hard, Rock	200	200	0,185	0,815	0,217	0,783	0,755	0,769	0,201
Hard, Rock	400	50	0,190	0,810	0,213	0,787	0,749	0,767	0,202
Hard, Rock	400	100	0,182	0,818	0,220	0,780	0,756	0,768	0,201
Hard, Rock	400	200	0,180	0,820	0,220	0,780	0,758	0,768	0,200
Hard, Tran	200	50	0,218	0,782	0,209	0,791	0,758	0,774	0,213
Hard, Tran	200	100	0,212	0,788	0,211	0,789	0,762	0,775	0,212
Hard, Tran	200	200	0,210	0,790	0,213	0,787	0,764	0,775	0,211
Hard, Tran	400	50	0,222	0,778	0,201	0,799	0,754	0,776	0,212
Hard, Tran	400	100	0,212	0,788	0,207	0,793	0,762	0,777	0,210
Hard, Tran	400	200	0,206	0,794	0,209	0,791	0,767	0,779	0,207
Jazz, Klas	200	50	0,041	0,959	0,596	0,404	0,863	0,550	0,319
Jazz, Klas	200	100	0,038	0,962	0,600	0,400	0,870	0,548	0,319
Jazz, Klas	200	200	0,036	0,964	0,603	0,397	0,875	0,546	0,320
Jazz, Klas	400	50	0,036	0,964	0,589	0,411	0,876	0,559	0,313
Jazz, Klas	400	100	0,034	0,966	0,601	0,399	0,880	0,549	0,317
Jazz, Klas	400	200	0,033	0,967	0,603	0,397	0,882	0,548	0,318
Jazz, Rock	200	50	0,143	0,857	0,371	0,629	0,740	0,680	0,257
Jazz, Rock	200	100	0,133	0,867	0,376	0,624	0,752	0,682	0,255
Jazz, Rock	200	200	0,128	0,872	0,380	0,620	0,759	0,682	0,254
Jazz, Rock	400	50	0,138	0,862	0,374	0,626	0,743	0,679	0,256
Jazz, Rock	400	100	0,125	0,875	0,376	0,624	0,761	0,686	0,251
Jazz, Rock	400	200	0,116	0,884	0,384	0,616	0,772	0,685	0,250
Jazz, Tran	200	50	0,064	0,936	0,367	0,633	0,882	0,737	0,216
Jazz, Tran	200	100	0,059	0,941	0,371	0,629	0,890	0,737	0,215
Jazz, Tran	200	200	0,058	0,942	0,374	0,626	0,892	0,736	0,216
Jazz, Tran	400	50	0,058	0,942	0,365	0,635	0,891	0,742	0,211
Jazz, Tran	400	100	0,055	0,945	0,370	0,630	0,896	0,740	0,212
Jazz, Tran	400	200	0,052	0,948	0,373	0,627	0,900	0,739	0,212
Klas, Rock	200	50	0,349	0,651	0,141	0,859	0,553	0,673	0,245
Klas, Rock	200	100	0,339	0,661	0,142	0,858	0,561	0,678	0,240
Klas, Rock	200	200	0,335	0,665	0,143	0,857	0,563	0,679	0,239
Klas, Rock	400	50	0,344	0,656	0,141	0,859	0,553	0,673	0,243
Klas, Rock	400	100	0,338	0,662	0,141	0,859	0,558	0,676	0,240
Klas, Rock	400	200	0,333	0,667	0,145	0,855	0,560	0,677	0,239
Klas, Tran	200	50	0,226	0,774	0,219	0,781	0,671	0,722	0,222
Klas, Tran	200	100	0,220	0,780	0,223	0,777	0,675	0,722	0,222

T. Genre	wi.	tr.	FP	TN	FN	recall	preci.	f-mea.	b. Feh.
Klas, Tran	200	200	0,217	0,783	0,226	0,774	0,678	0,723	0,221
Klas, Tran	400	50	0,220	0,780	0,222	0,778	0,672	0,721	0,221
Klas, Tran	400	100	0,214	0,786	0,226	0,774	0,677	0,722	0,220
Klas, Tran	400	200	0,211	0,789	0,230	0,770	0,678	0,721	0,221
Rock, Tran	200	50	0,224	0,776	0,183	0,817	0,688	0,747	0,204
Rock, Tran	200	100	0,219	0,781	0,183	0,817	0,693	0,750	0,201
Rock, Tran	200	200	0,213	0,787	0,184	0,816	0,698	0,753	0,198
Rock, Tran	400	50	0,219	0,781	0,187	0,813	0,688	0,745	0,203
Rock, Tran	400	100	0,212	0,788	0,187	0,813	0,695	0,749	0,200
Rock, Tran	400	200	0,210	0,790	0,187	0,813	0,697	0,750	0,198

A.3 Auswertungstabellen Ansatz 3

Tabelle A.7: Zehnfache Kreuzvalidierung mit Ansatz 3 auf dem gesamten Testdatensatz: Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalskonfiguration, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

win.	trees	type	FP	TN	FN	recall	preci.	f-meas.	b. Feh.
512	50	all	0,157	0,843	0,397	0,603	0,715	0,649	0,277
512	100	all	0,145	0,855	0,402	0,598	0,729	0,652	0,274
512	200	all	0,139	0,861	0,403	0,597	0,736	0,654	0,271
2048	50	all	0,156	0,844	0,358	0,642	0,728	0,678	0,257
2048	100	all	0,145	0,855	0,362	0,638	0,740	0,681	0,253
2048	200	all	0,138	0,862	0,363	0,637	0,748	0,683	0,251
4096	50	all	0,155	0,845	0,332	0,668	0,735	0,695	0,243
4096	100	all	0,143	0,857	0,335	0,665	0,748	0,699	0,239
4096	200	all	0,139	0,861	0,337	0,663	0,754	0,701	0,238
512	50	mfcc	0,175	0,825	0,480	0,520	0,662	0,579	0,327
512	100	mfcc	0,162	0,838	0,490	0,510	0,675	0,577	0,326
512	200	mfcc	0,155	0,845	0,495	0,505	0,682	0,577	0,325
2048	50	mfcc	0,182	0,818	0,419	0,581	0,678	0,623	0,301
2048	100	mfcc	0,171	0,829	0,426	0,574	0,688	0,623	0,299
2048	200	mfcc	0,165	0,835	0,430	0,570	0,694	0,623	0,298
4096	50	mfcc	0,187	0,813	0,398	0,602	0,679	0,635	0,293
4096	100	mfcc	0,175	0,825	0,405	0,595	0,691	0,636	0,290
4096	200	mfcc	0,170	0,830	0,407	0,593	0,697	0,638	0,288
512	50	o mfcc	0,162	0,838	0,403	0,597	0,707	0,642	0,283

win.	trees	type	FP	TN	FN	recall	preci.	f-meas.	b. Feh.
512	100	o mfcc	0,151	0,849	0,407	0,593	0,720	0,645	0,279
512	200	o mfcc	0,145	0,855	0,408	0,592	0,728	0,647	0,277
2048	50	o mfcc	0,160	0,840	0,367	0,633	0,720	0,668	0,264
2048	100	o mfcc	0,148	0,852	0,371	0,629	0,733	0,672	0,260
2048	200	o mfcc	0,142	0,858	0,372	0,628	0,741	0,674	0,257
4096	50	o mfcc	0,160	0,840	0,340	0,660	0,727	0,686	0,250
4096	100	o mfcc	0,149	0,851	0,344	0,656	0,739	0,690	0,246
4096	200	o mfcc	0,143	0,857	0,345	0,655	0,746	0,692	0,244

Tabelle A.8: „leave one out“ mit Ansatz 3 auf jedem Genre getrennt: Genre, Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalskombination, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Gen.	win.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Dubs	512	50	a	0,426	0,574	0,264	0,736	0,698	0,702	0,332
Dubs	512	50	m	0,407	0,593	0,363	0,637	0,677	0,642	0,378
Dubs	512	50	o	0,430	0,570	0,272	0,729	0,693	0,695	0,338
Dubs	512	100	a	0,421	0,579	0,266	0,734	0,701	0,701	0,330
Dubs	512	100	m	0,394	0,606	0,369	0,631	0,682	0,640	0,375
Dubs	512	100	o	0,427	0,573	0,271	0,729	0,697	0,696	0,335
Dubs	512	200	a	0,419	0,581	0,266	0,734	0,702	0,701	0,329
Dubs	512	200	m	0,385	0,615	0,373	0,627	0,685	0,640	0,372
Dubs	512	200	o	0,423	0,577	0,272	0,728	0,699	0,696	0,334
Dubs	2.048	50	a	0,414	0,586	0,254	0,746	0,706	0,709	0,321
Dubs	2.048	50	m	0,403	0,597	0,340	0,660	0,686	0,658	0,365
Dubs	2.048	50	o	0,421	0,579	0,263	0,737	0,699	0,702	0,329
Dubs	2.048	100	a	0,410	0,590	0,256	0,744	0,707	0,708	0,321
Dubs	2.048	100	m	0,393	0,607	0,345	0,655	0,689	0,657	0,363
Dubs	2.048	100	o	0,413	0,587	0,266	0,734	0,703	0,702	0,326
Dubs	2.048	200	a	0,407	0,594	0,256	0,744	0,710	0,709	0,319
Dubs	2.048	200	m	0,386	0,614	0,348	0,652	0,693	0,657	0,361
Dubs	2.048	200	o	0,410	0,590	0,264	0,736	0,707	0,704	0,324
Dubs	4.096	50	a	0,413	0,588	0,249	0,751	0,708	0,712	0,318
Dubs	4.096	50	m	0,407	0,593	0,333	0,667	0,689	0,661	0,362
Dubs	4.096	50	o	0,413	0,587	0,254	0,746	0,705	0,707	0,322
Dubs	4.096	100	a	0,405	0,596	0,254	0,747	0,712	0,711	0,316
Dubs	4.096	100	m	0,396	0,604	0,335	0,665	0,693	0,662	0,358

Gen.	win.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Dubs	4.096	100	o	0,408	0,592	0,255	0,745	0,710	0,709	0,319
Dubs	4.096	200	a	0,403	0,597	0,251	0,749	0,713	0,713	0,314
Dubs	4.096	200	m	0,389	0,611	0,341	0,659	0,696	0,660	0,357
Dubs	4.096	200	o	0,401	0,599	0,259	0,741	0,712	0,707	0,317
Hard	512	50	a	0,027	0,973	0,613	0,387	0,731	0,479	0,320
Hard	512	50	m	0,049	0,951	0,664	0,336	0,609	0,400	0,355
Hard	512	50	o	0,027	0,973	0,611	0,389	0,732	0,480	0,319
Hard	512	100	a	0,024	0,976	0,615	0,386	0,753	0,484	0,319
Hard	512	100	m	0,045	0,955	0,674	0,326	0,624	0,396	0,359
Hard	512	100	o	0,024	0,976	0,619	0,381	0,746	0,476	0,322
Hard	512	200	a	0,023	0,977	0,615	0,385	0,765	0,485	0,319
Hard	512	200	m	0,043	0,957	0,678	0,322	0,629	0,394	0,360
Hard	512	200	o	0,023	0,977	0,621	0,379	0,759	0,478	0,322
Hard	2048	50	a	0,029	0,971	0,571	0,429	0,740	0,509	0,300
Hard	2048	50	m	0,060	0,940	0,617	0,384	0,609	0,432	0,337
Hard	2048	50	o	0,030	0,970	0,577	0,423	0,735	0,503	0,303
Hard	2048	100	a	0,026	0,974	0,576	0,424	0,759	0,510	0,300
Hard	2048	100	m	0,057	0,943	0,630	0,370	0,613	0,426	0,342
Hard	2048	100	o	0,026	0,974	0,582	0,418	0,759	0,502	0,304
Hard	2048	200	a	0,024	0,976	0,580	0,420	0,775	0,508	0,302
Hard	2048	200	m	0,055	0,945	0,626	0,374	0,622	0,428	0,339
Hard	2048	200	o	0,025	0,975	0,589	0,411	0,763	0,499	0,307
Hard	4096	50	a	0,028	0,972	0,569	0,431	0,761	0,516	0,298
Hard	4096	50	m	0,064	0,936	0,620	0,380	0,589	0,417	0,341
Hard	4096	50	o	0,028	0,972	0,575	0,425	0,744	0,497	0,301
Hard	4096	100	a	0,025	0,975	0,571	0,429	0,764	0,510	0,298
Hard	4096	100	m	0,061	0,939	0,632	0,368	0,594	0,413	0,346
Hard	4096	100	o	0,026	0,974	0,570	0,431	0,767	0,510	0,297
Hard	4096	200	a	0,024	0,976	0,578	0,422	0,768	0,505	0,301
Hard	4096	200	m	0,060	0,940	0,633	0,367	0,597	0,413	0,345
Hard	4096	200	o	0,025	0,975	0,582	0,418	0,772	0,503	0,303
Jazz	512	50	a	0,086	0,915	0,750	0,250	0,853	0,367	0,402
Jazz	512	50	m	0,117	0,883	0,661	0,339	0,783	0,452	0,390
Jazz	512	50	o	0,090	0,911	0,769	0,231	0,839	0,336	0,413
Jazz	512	100	a	0,074	0,926	0,763	0,237	0,871	0,352	0,405
Jazz	512	100	m	0,106	0,894	0,671	0,329	0,795	0,444	0,389

Gen.	win.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Jazz	512	100	o	0,075	0,925	0,788	0,212	0,855	0,315	0,416
Jazz	512	200	a	0,066	0,935	0,772	0,228	0,888	0,342	0,406
Jazz	512	200	m	0,100	0,900	0,674	0,326	0,802	0,441	0,388
Jazz	512	200	o	0,068	0,932	0,796	0,204	0,870	0,307	0,418
Jazz	2048	50	a	0,093	0,907	0,691	0,309	0,857	0,430	0,378
Jazz	2048	50	m	0,132	0,868	0,627	0,373	0,784	0,480	0,381
Jazz	2048	50	o	0,096	0,904	0,707	0,293	0,852	0,409	0,385
Jazz	2048	100	a	0,078	0,922	0,709	0,291	0,875	0,415	0,380
Jazz	2048	100	m	0,121	0,879	0,634	0,366	0,794	0,475	0,380
Jazz	2048	100	o	0,083	0,917	0,725	0,275	0,866	0,395	0,389
Jazz	2048	200	a	0,071	0,929	0,713	0,287	0,886	0,413	0,379
Jazz	2048	200	m	0,114	0,886	0,639	0,361	0,802	0,471	0,379
Jazz	2048	200	o	0,074	0,926	0,734	0,267	0,880	0,387	0,389
Jazz	4096	50	a	0,091	0,909	0,674	0,326	0,870	0,449	0,366
Jazz	4096	50	m	0,132	0,868	0,617	0,383	0,792	0,487	0,376
Jazz	4096	50	o	0,098	0,902	0,681	0,319	0,854	0,435	0,374
Jazz	4096	100	a	0,078	0,922	0,682	0,318	0,886	0,441	0,366
Jazz	4096	100	m	0,123	0,877	0,634	0,367	0,792	0,473	0,380
Jazz	4096	100	o	0,085	0,915	0,694	0,307	0,876	0,425	0,374
Jazz	4096	200	a	0,070	0,930	0,689	0,311	0,898	0,437	0,366
Jazz	4096	200	m	0,119	0,881	0,635	0,365	0,799	0,470	0,380
Jazz	4096	200	o	0,077	0,923	0,703	0,297	0,880	0,417	0,375
Klas	512	50	a	0,301	0,699	0,104	0,847	0,834	0,833	0,209
Klas	512	50	m	0,231	0,769	0,178	0,772	0,834	0,795	0,216
Klas	512	50	o	0,323	0,677	0,105	0,845	0,826	0,828	0,220
Klas	512	100	a	0,297	0,703	0,102	0,848	0,836	0,835	0,206
Klas	512	100	m	0,223	0,777	0,181	0,770	0,838	0,795	0,213
Klas	512	100	o	0,320	0,680	0,105	0,845	0,827	0,828	0,219
Klas	512	200	a	0,297	0,703	0,102	0,848	0,836	0,835	0,207
Klas	512	200	m	0,219	0,781	0,181	0,769	0,839	0,795	0,211
Klas	512	200	o	0,320	0,680	0,104	0,846	0,828	0,829	0,218
Klas	2048	50	a	0,274	0,726	0,095	0,855	0,845	0,844	0,191
Klas	2048	50	m	0,203	0,797	0,156	0,794	0,851	0,815	0,190
Klas	2048	50	o	0,302	0,698	0,096	0,854	0,836	0,838	0,206
Klas	2048	100	a	0,270	0,730	0,096	0,854	0,848	0,844	0,190
Klas	2048	100	m	0,197	0,803	0,158	0,792	0,854	0,815	0,187

Gen.	win.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Klas	2048	100	o	0,295	0,705	0,097	0,853	0,838	0,839	0,203
Klas	2048	200	a	0,268	0,732	0,094	0,856	0,848	0,846	0,188
Klas	2048	200	m	0,196	0,804	0,159	0,792	0,854	0,815	0,187
Klas	2048	200	o	0,291	0,709	0,097	0,853	0,840	0,839	0,201
Klas	4096	50	a	0,260	0,740	0,090	0,860	0,853	0,851	0,181
Klas	4096	50	m	0,190	0,810	0,145	0,805	0,860	0,824	0,177
Klas	4096	50	o	0,283	0,717	0,092	0,858	0,844	0,844	0,195
Klas	4096	100	a	0,259	0,741	0,089	0,862	0,854	0,852	0,180
Klas	4096	100	m	0,181	0,819	0,146	0,804	0,863	0,826	0,172
Klas	4096	100	o	0,277	0,723	0,093	0,858	0,847	0,845	0,191
Klas	4096	200	a	0,256	0,744	0,087	0,863	0,856	0,853	0,178
Klas	4096	200	m	0,179	0,822	0,147	0,803	0,865	0,826	0,171
Klas	4096	200	o	0,276	0,724	0,092	0,858	0,847	0,845	0,191
Rock	512	50	a	0,395	0,605	0,170	0,830	0,734	0,766	0,282
Rock	512	50	m	0,510	0,490	0,206	0,794	0,671	0,716	0,358
Rock	512	50	o	0,385	0,615	0,174	0,826	0,737	0,768	0,279
Rock	512	100	a	0,386	0,614	0,170	0,831	0,739	0,769	0,278
Rock	512	100	m	0,503	0,497	0,205	0,795	0,675	0,718	0,354
Rock	512	100	o	0,380	0,620	0,174	0,827	0,741	0,769	0,277
Rock	512	200	a	0,381	0,619	0,169	0,831	0,742	0,771	0,275
Rock	512	200	m	0,499	0,501	0,206	0,794	0,676	0,718	0,353
Rock	512	200	o	0,379	0,621	0,171	0,829	0,742	0,770	0,275
Rock	2048	50	a	0,363	0,637	0,159	0,841	0,756	0,784	0,261
Rock	2048	50	m	0,500	0,500	0,180	0,820	0,685	0,734	0,340
Rock	2048	50	o	0,358	0,642	0,162	0,838	0,757	0,782	0,260
Rock	2048	100	a	0,354	0,646	0,160	0,840	0,761	0,785	0,257
Rock	2048	100	m	0,493	0,507	0,178	0,822	0,689	0,737	0,335
Rock	2048	100	o	0,349	0,651	0,165	0,835	0,762	0,783	0,257
Rock	2048	200	a	0,350	0,650	0,159	0,841	0,764	0,787	0,254
Rock	2048	200	m	0,492	0,508	0,178	0,822	0,690	0,736	0,335
Rock	2048	200	o	0,347	0,653	0,162	0,838	0,765	0,786	0,255
Rock	4096	50	a	0,352	0,648	0,135	0,865	0,767	0,800	0,243
Rock	4096	50	m	0,519	0,481	0,155	0,845	0,685	0,743	0,337
Rock	4096	50	o	0,352	0,648	0,133	0,867	0,769	0,803	0,243
Rock	4096	100	a	0,338	0,663	0,136	0,864	0,776	0,804	0,237
Rock	4096	100	m	0,508	0,493	0,156	0,845	0,690	0,746	0,332

Gen.	win.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Rock	4096	100	o	0,339	0,661	0,136	0,864	0,776	0,805	0,237
Rock	4096	200	a	0,335	0,666	0,136	0,864	0,779	0,806	0,235
Rock	4096	200	m	0,509	0,491	0,158	0,842	0,690	0,744	0,334
Rock	4096	200	o	0,331	0,669	0,140	0,860	0,780	0,804	0,236
Tran	512	50	a	0,147	0,853	0,422	0,578	0,813	0,655	0,288
Tran	512	50	m	0,181	0,819	0,543	0,457	0,737	0,549	0,361
Tran	512	50	o	0,149	0,851	0,424	0,576	0,813	0,653	0,290
Tran	512	100	a	0,136	0,864	0,427	0,573	0,821	0,654	0,286
Tran	512	100	m	0,169	0,831	0,552	0,448	0,746	0,545	0,359
Tran	512	100	o	0,138	0,862	0,430	0,571	0,821	0,652	0,288
Tran	512	200	a	0,131	0,869	0,428	0,572	0,824	0,655	0,284
Tran	512	200	m	0,163	0,837	0,559	0,441	0,749	0,541	0,360
Tran	512	200	o	0,132	0,868	0,432	0,568	0,825	0,652	0,286
Tran	2048	50	a	0,151	0,849	0,360	0,640	0,825	0,701	0,259
Tran	2048	50	m	0,201	0,799	0,452	0,548	0,750	0,618	0,326
Tran	2048	50	o	0,158	0,843	0,365	0,635	0,819	0,696	0,265
Tran	2048	100	a	0,143	0,857	0,362	0,638	0,831	0,702	0,256
Tran	2048	100	m	0,191	0,809	0,457	0,543	0,757	0,618	0,323
Tran	2048	100	o	0,146	0,854	0,369	0,631	0,829	0,696	0,261
Tran	2048	200	a	0,140	0,860	0,365	0,635	0,834	0,700	0,257
Tran	2048	200	m	0,186	0,814	0,458	0,542	0,762	0,618	0,321
Tran	2048	200	o	0,141	0,859	0,371	0,629	0,837	0,697	0,259
Tran	4096	50	a	0,161	0,839	0,329	0,672	0,825	0,719	0,248
Tran	4096	50	m	0,215	0,785	0,428	0,572	0,746	0,631	0,322
Tran	4096	50	o	0,162	0,838	0,336	0,664	0,825	0,714	0,252
Tran	4096	100	a	0,155	0,845	0,335	0,665	0,828	0,717	0,249
Tran	4096	100	m	0,206	0,794	0,431	0,569	0,753	0,632	0,318
Tran	4096	100	o	0,155	0,846	0,339	0,662	0,834	0,716	0,249
Tran	4096	200	a	0,151	0,849	0,337	0,663	0,833	0,718	0,248
Tran	4096	200	m	0,198	0,802	0,433	0,567	0,760	0,633	0,316
Tran	4096	200	o	0,151	0,849	0,338	0,662	0,838	0,718	0,248

Tabelle A.9: Genre Konfusion mit Ansatz 3 mit einem Trainingsgenre und fünf Testgenres: Trainingsgenre, Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalskonfiguration, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

Gen.	**wi.**	**tr.**	**type**	**FP**	**TN**	**FN**	**recall**	**preci.**	**f-m.**	**b. F.**
Dubs	512	50	a	0,458	0,542	0,376	0,624	0,452	0,524	0,417
Dubs	512	50	m	0,407	0,593	0,362	0,638	0,486	0,552	0,385
Dubs	512	50	o	0,462	0,538	0,375	0,625	0,450	0,523	0,419
Dubs	512	100	a	0,457	0,543	0,377	0,623	0,452	0,524	0,417
Dubs	512	100	m	0,394	0,606	0,367	0,633	0,492	0,554	0,380
Dubs	512	100	o	0,460	0,540	0,376	0,624	0,451	0,523	0,418
Dubs	512	200	a	0,455	0,545	0,377	0,623	0,453	0,524	0,416
Dubs	512	200	m	0,387	0,613	0,370	0,630	0,496	0,555	0,378
Dubs	512	200	o	0,459	0,541	0,375	0,625	0,451	0,524	0,417
Dubs	2048	50	a	0,418	0,582	0,378	0,622	0,472	0,537	0,398
Dubs	2048	50	m	0,392	0,608	0,333	0,667	0,506	0,575	0,362
Dubs	2048	50	o	0,418	0,582	0,381	0,619	0,472	0,536	0,399
Dubs	2048	100	a	0,410	0,590	0,383	0,617	0,475	0,537	0,397
Dubs	2048	100	m	0,381	0,619	0,337	0,663	0,511	0,577	0,359
Dubs	2048	100	o	0,412	0,588	0,380	0,620	0,475	0,538	0,396
Dubs	2048	200	a	0,411	0,589	0,381	0,619	0,475	0,538	0,396
Dubs	2048	200	m	0,374	0,626	0,340	0,660	0,515	0,579	0,357
Dubs	2048	200	o	0,413	0,587	0,381	0,619	0,474	0,537	0,397
Dubs	4096	50	a	0,400	0,600	0,378	0,622	0,482	0,543	0,389
Dubs	4096	50	m	0,398	0,602	0,329	0,671	0,501	0,574	0,364
Dubs	4096	50	o	0,401	0,599	0,381	0,619	0,480	0,540	0,391
Dubs	4096	100	a	0,393	0,607	0,383	0,617	0,484	0,543	0,388
Dubs	4096	100	m	0,386	0,614	0,336	0,664	0,506	0,574	0,361
Dubs	4096	100	o	0,394	0,606	0,380	0,620	0,485	0,544	0,387
Dubs	4096	200	a	0,388	0,612	0,384	0,616	0,487	0,544	0,386
Dubs	4096	200	m	0,383	0,617	0,336	0,664	0,508	0,576	0,360
Dubs	4096	200	o	0,391	0,609	0,382	0,618	0,486	0,544	0,387
Hard	512	50	a	0,145	0,855	0,642	0,358	0,663	0,465	0,394
Hard	512	50	m	0,158	0,842	0,613	0,387	0,661	0,488	0,386
Hard	512	50	o	0,150	0,850	0,641	0,359	0,657	0,464	0,395
Hard	512	100	a	0,139	0,861	0,648	0,352	0,670	0,462	0,393
Hard	512	100	m	0,150	0,850	0,621	0,379	0,669	0,484	0,385
Hard	512	100	o	0,142	0,858	0,649	0,351	0,664	0,459	0,395

Gen.	wi.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Hard	512	200	a	0,136	0,864	0,650	0,350	0,673	0,460	0,393
Hard	512	200	m	0,144	0,856	0,625	0,375	0,675	0,482	0,385
Hard	512	200	o	0,139	0,861	0,652	0,348	0,666	0,457	0,396
Hard	2048	50	a	0,156	0,844	0,610	0,390	0,667	0,493	0,383
Hard	2048	50	m	0,196	0,804	0,539	0,461	0,652	0,540	0,367
Hard	2048	50	o	0,159	0,841	0,614	0,386	0,659	0,487	0,387
Hard	2048	100	a	0,149	0,851	0,618	0,382	0,672	0,487	0,383
Hard	2048	100	m	0,189	0,811	0,545	0,455	0,657	0,538	0,367
Hard	2048	100	o	0,152	0,848	0,624	0,376	0,664	0,480	0,388
Hard	2048	200	a	0,144	0,856	0,621	0,379	0,678	0,486	0,382
Hard	2048	200	m	0,185	0,815	0,549	0,451	0,660	0,536	0,367
Hard	2048	200	o	0,147	0,853	0,628	0,372	0,668	0,478	0,388
Hard	4096	50	a	0,152	0,848	0,614	0,386	0,668	0,489	0,383
Hard	4096	50	m	0,206	0,794	0,534	0,466	0,642	0,540	0,370
Hard	4096	50	o	0,161	0,839	0,601	0,399	0,663	0,498	0,381
Hard	4096	100	a	0,148	0,852	0,620	0,380	0,671	0,485	0,384
Hard	4096	100	m	0,194	0,806	0,543	0,457	0,652	0,538	0,368
Hard	4096	100	o	0,152	0,848	0,612	0,388	0,670	0,491	0,382
Hard	4096	200	a	0,146	0,854	0,618	0,382	0,676	0,488	0,382
Hard	4096	200	m	0,189	0,811	0,546	0,454	0,656	0,537	0,368
Hard	4096	200	o	0,148	0,852	0,619	0,381	0,672	0,487	0,383
Jazz	512	50	a	0,285	0,715	0,616	0,384	0,494	0,432	0,450
Jazz	512	50	m	0,107	0,893	0,662	0,338	0,695	0,455	0,385
Jazz	512	50	o	0,303	0,697	0,616	0,384	0,479	0,426	0,459
Jazz	512	100	a	0,278	0,722	0,617	0,383	0,499	0,433	0,448
Jazz	512	100	m	0,095	0,905	0,673	0,327	0,713	0,449	0,384
Jazz	512	100	o	0,298	0,702	0,621	0,379	0,480	0,424	0,459
Jazz	512	200	a	0,280	0,720	0,614	0,386	0,500	0,436	0,447
Jazz	512	200	m	0,089	0,911	0,679	0,321	0,723	0,444	0,384
Jazz	512	200	o	0,287	0,713	0,631	0,369	0,481	0,417	0,459
Jazz	2048	50	a	0,326	0,674	0,542	0,458	0,503	0,479	0,434
Jazz	2048	50	m	0,105	0,895	0,643	0,357	0,710	0,475	0,374
Jazz	2048	50	o	0,324	0,676	0,559	0,441	0,496	0,467	0,441
Jazz	2048	100	a	0,321	0,679	0,543	0,457	0,507	0,481	0,432
Jazz	2048	100	m	0,094	0,906	0,651	0,349	0,728	0,472	0,372
Jazz	2048	100	o	0,321	0,679	0,563	0,437	0,495	0,464	0,442

Gen.	wi.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Jazz	2048	200	a	0,319	0,681	0,545	0,455	0,507	0,480	0,432
Jazz	2048	200	m	0,089	0,911	0,655	0,345	0,737	0,470	0,372
Jazz	2048	200	o	0,309	0,691	0,566	0,434	0,503	0,466	0,438
Jazz	4096	50	a	0,347	0,653	0,513	0,487	0,502	0,494	0,430
Jazz	4096	50	m	0,110	0,890	0,632	0,368	0,705	0,483	0,371
Jazz	4096	50	o	0,353	0,647	0,518	0,482	0,496	0,489	0,435
Jazz	4096	100	a	0,349	0,651	0,513	0,487	0,501	0,494	0,431
Jazz	4096	100	m	0,099	0,901	0,646	0,354	0,719	0,474	0,373
Jazz	4096	100	o	0,340	0,660	0,528	0,472	0,499	0,485	0,434
Jazz	4096	200	a	0,345	0,655	0,515	0,485	0,503	0,494	0,430
Jazz	4096	200	m	0,094	0,906	0,650	0,350	0,729	0,473	0,372
Jazz	4096	200	o	0,332	0,668	0,536	0,464	0,501	0,482	0,434
Klas	512	50	a	0,844	0,156	0,049	0,951	0,400	0,563	0,447
Klas	512	50	m	0,198	0,802	0,638	0,362	0,520	0,427	0,418
Klas	512	50	o	0,849	0,151	0,048	0,952	0,399	0,562	0,448
Klas	512	100	a	0,847	0,153	0,048	0,952	0,399	0,562	0,448
Klas	512	100	m	0,187	0,813	0,650	0,350	0,525	0,420	0,418
Klas	512	100	o	0,854	0,146	0,046	0,954	0,398	0,562	0,450
Klas	512	200	a	0,848	0,152	0,049	0,951	0,399	0,562	0,448
Klas	512	200	m	0,181	0,819	0,656	0,344	0,529	0,417	0,419
Klas	512	200	o	0,855	0,145	0,045	0,955	0,398	0,562	0,450
Klas	2048	50	a	0,843	0,157	0,044	0,956	0,400	0,564	0,443
Klas	2048	50	m	0,183	0,817	0,643	0,357	0,535	0,428	0,413
Klas	2048	50	o	0,853	0,147	0,039	0,961	0,398	0,563	0,446
Klas	2048	100	a	0,849	0,151	0,043	0,957	0,399	0,563	0,446
Klas	2048	100	m	0,174	0,826	0,649	0,351	0,542	0,426	0,412
Klas	2048	100	o	0,857	0,143	0,037	0,963	0,398	0,563	0,447
Klas	2048	200	a	0,850	0,150	0,043	0,957	0,399	0,563	0,446
Klas	2048	200	m	0,168	0,832	0,657	0,343	0,545	0,421	0,413
Klas	2048	200	o	0,859	0,141	0,037	0,963	0,397	0,563	0,448
Klas	4096	50	a	0,838	0,162	0,044	0,956	0,400	0,564	0,441
Klas	4096	50	m	0,164	0,836	0,665	0,335	0,544	0,415	0,414
Klas	4096	50	o	0,833	0,167	0,043	0,957	0,402	0,566	0,438
Klas	4096	100	a	0,844	0,156	0,041	0,959	0,399	0,564	0,442
Klas	4096	100	m	0,153	0,847	0,680	0,320	0,549	0,405	0,417
Klas	4096	100	o	0,843	0,157	0,038	0,962	0,400	0,565	0,441

Gen.	wi.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Klas	4096	200	a	0,844	0,156	0,040	0,960	0,399	0,564	0,442
Klas	4096	200	m	0,149	0,851	0,684	0,316	0,553	0,402	0,417
Klas	4096	200	o	0,852	0,148	0,037	0,963	0,398	0,563	0,445
Rock	512	50	a	0,513	0,487	0,334	0,666	0,439	0,529	0,424
Rock	512	50	m	0,649	0,351	0,294	0,706	0,396	0,507	0,472
Rock	512	50	o	0,503	0,497	0,340	0,660	0,442	0,529	0,421
Rock	512	100	a	0,510	0,490	0,335	0,665	0,440	0,530	0,423
Rock	512	100	m	0,646	0,354	0,295	0,705	0,397	0,508	0,470
Rock	512	100	o	0,502	0,498	0,341	0,659	0,442	0,529	0,421
Rock	512	200	a	0,510	0,490	0,334	0,666	0,441	0,531	0,422
Rock	512	200	m	0,644	0,356	0,296	0,704	0,397	0,507	0,470
Rock	512	200	o	0,501	0,499	0,340	0,660	0,443	0,530	0,421
Rock	2048	50	a	0,496	0,504	0,337	0,663	0,445	0,533	0,417
Rock	2048	50	m	0,651	0,349	0,278	0,722	0,400	0,515	0,464
Rock	2048	50	o	0,478	0,522	0,352	0,648	0,449	0,530	0,415
Rock	2048	100	a	0,493	0,507	0,337	0,663	0,447	0,534	0,415
Rock	2048	100	m	0,646	0,354	0,282	0,718	0,400	0,514	0,464
Rock	2048	100	o	0,480	0,520	0,350	0,650	0,448	0,531	0,415
Rock	2048	200	a	0,491	0,509	0,339	0,661	0,447	0,533	0,415
Rock	2048	200	m	0,646	0,354	0,284	0,716	0,399	0,513	0,465
Rock	2048	200	o	0,477	0,523	0,350	0,650	0,450	0,532	0,414
Rock	4096	50	a	0,464	0,536	0,350	0,650	0,455	0,535	0,407
Rock	4096	50	m	0,675	0,325	0,282	0,718	0,388	0,504	0,479
Rock	4096	50	o	0,464	0,536	0,356	0,644	0,453	0,532	0,410
Rock	4096	100	a	0,467	0,533	0,350	0,650	0,454	0,534	0,409
Rock	4096	100	m	0,676	0,324	0,293	0,707	0,384	0,498	0,484
Rock	4096	100	o	0,458	0,542	0,358	0,642	0,456	0,533	0,408
Rock	4096	200	a	0,470	0,530	0,348	0,652	0,453	0,534	0,409
Rock	4096	200	m	0,672	0,328	0,293	0,707	0,385	0,499	0,483
Rock	4096	200	o	0,456	0,544	0,357	0,643	0,457	0,534	0,406
Tran	512	50	a	0,160	0,840	0,587	0,413	0,638	0,502	0,374
Tran	512	50	m	0,242	0,758	0,439	0,561	0,613	0,586	0,341
Tran	512	50	o	0,165	0,835	0,583	0,417	0,633	0,503	0,374
Tran	512	100	a	0,154	0,846	0,590	0,410	0,646	0,502	0,372
Tran	512	100	m	0,230	0,770	0,447	0,553	0,622	0,585	0,338
Tran	512	100	o	0,158	0,842	0,586	0,414	0,642	0,503	0,372

Gen.	wi.	tr.	type	FP	TN	FN	recall	preci.	f-m.	b. F.
Tran	512	200	a	0,150	0,850	0,593	0,407	0,650	0,501	0,371
Tran	512	200	m	0,223	0,777	0,451	0,549	0,627	0,585	0,337
Tran	512	200	o	0,154	0,846	0,588	0,412	0,647	0,503	0,371
Tran	2048	50	a	0,191	0,809	0,525	0,475	0,629	0,541	0,358
Tran	2048	50	m	0,306	0,694	0,348	0,652	0,592	0,621	0,327
Tran	2048	50	o	0,190	0,810	0,531	0,469	0,627	0,537	0,361
Tran	2048	100	a	0,185	0,815	0,528	0,472	0,635	0,541	0,357
Tran	2048	100	m	0,294	0,706	0,353	0,647	0,599	0,622	0,324
Tran	2048	100	o	0,184	0,816	0,533	0,467	0,634	0,538	0,358
Tran	2048	200	a	0,182	0,818	0,532	0,468	0,637	0,539	0,357
Tran	2048	200	m	0,289	0,711	0,356	0,644	0,603	0,623	0,322
Tran	2048	200	o	0,181	0,819	0,536	0,464	0,636	0,537	0,359
Tran	4096	50	a	0,202	0,798	0,503	0,497	0,625	0,554	0,353
Tran	4096	50	m	0,331	0,669	0,323	0,677	0,581	0,625	0,327
Tran	4096	50	o	0,204	0,796	0,512	0,488	0,618	0,546	0,358
Tran	4096	100	a	0,195	0,805	0,506	0,494	0,632	0,555	0,351
Tran	4096	100	m	0,319	0,681	0,327	0,673	0,588	0,628	0,323
Tran	4096	100	o	0,196	0,804	0,512	0,488	0,629	0,550	0,354
Tran	4096	200	a	0,193	0,807	0,508	0,492	0,634	0,554	0,350
Tran	4096	200	m	0,316	0,684	0,328	0,672	0,590	0,628	0,322
Tran	4096	200	o	0,192	0,808	0,513	0,487	0,632	0,550	0,353

Tabelle A.10: Genre Konfusion mit Ansatz 3 mit zwei Trainingsgenres und vier Testgenres: Trainingsgenres, Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalstyp, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler.

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Dubs, Hard	512	50	a	0,214	0,786	0,557	0,443	0,609	0,513	0,385
Dubs, Hard	512	50	m	0,258	0,742	0,489	0,511	0,597	0,551	0,374
Dubs, Hard	512	50	o	0,219	0,781	0,554	0,446	0,604	0,513	0,387
Dubs, Hard	512	100	a	0,208	0,792	0,561	0,439	0,613	0,512	0,385
Dubs, Hard	512	100	m	0,247	0,753	0,499	0,501	0,604	0,548	0,373
Dubs, Hard	512	100	o	0,210	0,790	0,564	0,436	0,609	0,508	0,387
Dubs, Hard	512	200	a	0,204	0,796	0,565	0,435	0,616	0,510	0,384
Dubs, Hard	512	200	m	0,240	0,760	0,503	0,497	0,608	0,547	0,372
Dubs, Hard	512	200	o	0,206	0,794	0,567	0,433	0,612	0,507	0,386
Dubs, Hard	2048	50	a	0,217	0,783	0,523	0,477	0,622	0,540	0,370

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Dubs, Hard	2048	50	m	0,288	0,712	0,424	0,576	0,599	0,587	0,356
Dubs, Hard	2048	50	o	0,218	0,782	0,523	0,477	0,621	0,540	0,370
Dubs, Hard	2048	100	a	0,208	0,792	0,527	0,473	0,630	0,541	0,367
Dubs, Hard	2048	100	m	0,278	0,722	0,429	0,571	0,606	0,588	0,353
Dubs, Hard	2048	100	o	0,209	0,791	0,529	0,471	0,628	0,538	0,369
Dubs, Hard	2048	200	a	0,204	0,796	0,529	0,471	0,634	0,541	0,366
Dubs, Hard	2048	200	m	0,273	0,727	0,431	0,569	0,609	0,588	0,352
Dubs, Hard	2048	200	o	0,206	0,794	0,534	0,466	0,629	0,535	0,370
Dubs, Hard	4096	50	a	0,216	0,784	0,508	0,492	0,630	0,553	0,362
Dubs, Hard	4096	50	m	0,300	0,700	0,407	0,593	0,595	0,594	0,354
Dubs, Hard	4096	50	o	0,221	0,779	0,511	0,489	0,622	0,547	0,366
Dubs, Hard	4096	100	a	0,209	0,791	0,514	0,486	0,634	0,550	0,361
Dubs, Hard	4096	100	m	0,291	0,709	0,415	0,585	0,599	0,592	0,353
Dubs, Hard	4096	100	o	0,212	0,788	0,518	0,482	0,629	0,546	0,365
Dubs, Hard	4096	200	a	0,207	0,793	0,518	0,482	0,634	0,548	0,363
Dubs, Hard	4096	200	m	0,287	0,713	0,418	0,582	0,601	0,592	0,352
Dubs, Hard	4096	200	o	0,210	0,790	0,520	0,480	0,631	0,545	0,365
Dubs, Jazz	512	50	a	0,407	0,593	0,403	0,597	0,492	0,539	0,405
Dubs, Jazz	512	50	m	0,236	0,764	0,471	0,529	0,596	0,561	0,354
Dubs, Jazz	512	50	o	0,402	0,598	0,421	0,579	0,488	0,529	0,411
Dubs, Jazz	512	100	a	0,397	0,603	0,404	0,596	0,498	0,543	0,400
Dubs, Jazz	512	100	m	0,219	0,781	0,481	0,519	0,610	0,561	0,350
Dubs, Jazz	512	100	o	0,396	0,604	0,422	0,578	0,491	0,531	0,409
Dubs, Jazz	512	200	a	0,393	0,607	0,405	0,595	0,500	0,543	0,399
Dubs, Jazz	512	200	m	0,209	0,791	0,488	0,512	0,618	0,560	0,348
Dubs, Jazz	512	200	o	0,392	0,608	0,422	0,578	0,493	0,532	0,407
Dubs, Jazz	2048	50	a	0,364	0,636	0,381	0,619	0,528	0,570	0,373
Dubs, Jazz	2048	50	m	0,239	0,761	0,432	0,568	0,610	0,589	0,335
Dubs, Jazz	2048	50	o	0,374	0,626	0,394	0,606	0,516	0,558	0,384
Dubs, Jazz	2048	100	a	0,362	0,638	0,384	0,616	0,528	0,569	0,373
Dubs, Jazz	2048	100	m	0,224	0,776	0,442	0,558	0,621	0,588	0,333
Dubs, Jazz	2048	100	o	0,359	0,641	0,397	0,603	0,525	0,561	0,378
Dubs, Jazz	2048	200	a	0,362	0,638	0,385	0,615	0,528	0,568	0,374
Dubs, Jazz	2048	200	m	0,216	0,784	0,446	0,554	0,628	0,588	0,331
Dubs, Jazz	2048	200	o	0,355	0,645	0,397	0,603	0,528	0,563	0,376
Dubs, Jazz	4096	50	a	0,354	0,646	0,368	0,632	0,539	0,582	0,361

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Dubs, Jazz	4096	50	m	0,252	0,748	0,424	0,576	0,599	0,588	0,338
Dubs, Jazz	4096	50	o	0,350	0,650	0,381	0,619	0,536	0,575	0,366
Dubs, Jazz	4096	100	a	0,336	0,664	0,366	0,634	0,553	0,590	0,351
Dubs, Jazz	4096	100	m	0,236	0,764	0,430	0,570	0,612	0,590	0,333
Dubs, Jazz	4096	100	o	0,333	0,667	0,385	0,615	0,548	0,580	0,359
Dubs, Jazz	4096	200	a	0,332	0,668	0,371	0,629	0,554	0,589	0,351
Dubs, Jazz	4096	200	m	0,229	0,771	0,435	0,565	0,617	0,590	0,332
Dubs, Jazz	4096	200	o	0,325	0,675	0,386	0,614	0,553	0,582	0,355
Dubs, Klas	512	50	a	0,591	0,409	0,205	0,795	0,404	0,536	0,398
Dubs, Klas	512	50	m	0,403	0,597	0,377	0,623	0,438	0,515	0,390
Dubs, Klas	512	50	o	0,594	0,406	0,203	0,797	0,403	0,536	0,399
Dubs, Klas	512	100	a	0,591	0,409	0,201	0,799	0,406	0,538	0,396
Dubs, Klas	512	100	m	0,386	0,614	0,386	0,614	0,445	0,516	0,386
Dubs, Klas	512	100	o	0,592	0,408	0,202	0,798	0,405	0,537	0,397
Dubs, Klas	512	200	a	0,592	0,408	0,198	0,802	0,406	0,539	0,395
Dubs, Klas	512	200	m	0,376	0,624	0,390	0,610	0,449	0,517	0,383
Dubs, Klas	512	200	o	0,594	0,406	0,198	0,802	0,405	0,538	0,396
Dubs, Klas	2048	50	a	0,545	0,455	0,206	0,794	0,422	0,551	0,375
Dubs, Klas	2048	50	m	0,397	0,603	0,344	0,656	0,453	0,536	0,370
Dubs, Klas	2048	50	o	0,547	0,453	0,208	0,792	0,420	0,549	0,378
Dubs, Klas	2048	100	a	0,540	0,460	0,202	0,798	0,426	0,555	0,371
Dubs, Klas	2048	100	m	0,381	0,619	0,349	0,651	0,461	0,540	0,365
Dubs, Klas	2048	100	o	0,546	0,454	0,205	0,795	0,422	0,551	0,375
Dubs, Klas	2048	200	a	0,537	0,463	0,203	0,797	0,427	0,556	0,370
Dubs, Klas	2048	200	m	0,373	0,627	0,353	0,647	0,465	0,541	0,363
Dubs, Klas	2048	200	o	0,543	0,457	0,202	0,798	0,424	0,554	0,373
Dubs, Klas	4096	50	a	0,516	0,484	0,205	0,795	0,434	0,561	0,360
Dubs, Klas	4096	50	m	0,400	0,600	0,333	0,667	0,453	0,540	0,366
Dubs, Klas	4096	50	o	0,528	0,472	0,197	0,803	0,431	0,561	0,362
Dubs, Klas	4096	100	a	0,508	0,492	0,205	0,795	0,437	0,564	0,357
Dubs, Klas	4096	100	m	0,389	0,611	0,333	0,667	0,460	0,544	0,361
Dubs, Klas	4096	100	o	0,514	0,486	0,195	0,805	0,438	0,567	0,355
Dubs, Klas	4096	200	a	0,506	0,494	0,203	0,797	0,439	0,566	0,354
Dubs, Klas	4096	200	m	0,381	0,619	0,338	0,662	0,463	0,545	0,359
Dubs, Klas	4096	200	o	0,511	0,489	0,196	0,804	0,439	0,568	0,353
Dubs, Rock	512	50	a	0,456	0,544	0,394	0,606	0,410	0,489	0,425

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Dubs, Rock	512	50	m	0,451	0,549	0,351	0,649	0,430	0,517	0,401
Dubs, Rock	512	50	o	0,452	0,548	0,399	0,601	0,410	0,488	0,426
Dubs, Rock	512	100	a	0,454	0,546	0,393	0,607	0,412	0,491	0,424
Dubs, Rock	512	100	m	0,437	0,563	0,355	0,645	0,436	0,520	0,396
Dubs, Rock	512	100	o	0,454	0,546	0,394	0,606	0,412	0,490	0,424
Dubs, Rock	512	200	a	0,455	0,545	0,393	0,607	0,412	0,491	0,424
Dubs, Rock	512	200	m	0,429	0,571	0,357	0,643	0,439	0,522	0,393
Dubs, Rock	512	200	o	0,454	0,546	0,393	0,607	0,412	0,490	0,424
Dubs, Rock	2048	50	a	0,423	0,577	0,396	0,604	0,427	0,500	0,410
Dubs, Rock	2048	50	m	0,423	0,577	0,332	0,668	0,451	0,539	0,378
Dubs, Rock	2048	50	o	0,429	0,571	0,389	0,611	0,426	0,502	0,409
Dubs, Rock	2048	100	a	0,421	0,579	0,396	0,604	0,428	0,501	0,408
Dubs, Rock	2048	100	m	0,408	0,592	0,337	0,663	0,458	0,542	0,373
Dubs, Rock	2048	100	o	0,419	0,581	0,393	0,607	0,430	0,503	0,406
Dubs, Rock	2048	200	a	0,419	0,581	0,395	0,605	0,430	0,503	0,407
Dubs, Rock	2048	200	m	0,402	0,598	0,339	0,661	0,461	0,543	0,370
Dubs, Rock	2048	200	o	0,417	0,583	0,393	0,607	0,431	0,504	0,405
Dubs, Rock	4096	50	a	0,403	0,597	0,387	0,613	0,440	0,512	0,395
Dubs, Rock	4096	50	m	0,430	0,570	0,326	0,674	0,448	0,538	0,378
Dubs, Rock	4096	50	o	0,401	0,599	0,386	0,614	0,442	0,514	0,394
Dubs, Rock	4096	100	a	0,400	0,600	0,388	0,612	0,442	0,513	0,394
Dubs, Rock	4096	100	m	0,420	0,580	0,333	0,667	0,451	0,538	0,376
Dubs, Rock	4096	100	o	0,395	0,605	0,388	0,612	0,445	0,515	0,392
Dubs, Rock	4096	200	a	0,397	0,603	0,388	0,612	0,444	0,514	0,393
Dubs, Rock	4096	200	m	0,412	0,588	0,335	0,665	0,455	0,540	0,373
Dubs, Rock	4096	200	o	0,394	0,606	0,390	0,610	0,445	0,514	0,392
Dubs, Tran	512	50	a	0,209	0,791	0,550	0,450	0,570	0,503	0,379
Dubs, Tran	512	50	m	0,294	0,706	0,388	0,612	0,561	0,585	0,341
Dubs, Tran	512	50	o	0,210	0,790	0,556	0,444	0,566	0,498	0,383
Dubs, Tran	512	100	a	0,201	0,799	0,555	0,445	0,577	0,503	0,378
Dubs, Tran	512	100	m	0,279	0,721	0,394	0,606	0,571	0,588	0,337
Dubs, Tran	512	100	o	0,202	0,798	0,560	0,440	0,573	0,497	0,381
Dubs, Tran	512	200	a	0,197	0,803	0,559	0,441	0,579	0,501	0,378
Dubs, Tran	512	200	m	0,271	0,729	0,399	0,601	0,576	0,589	0,335
Dubs, Tran	512	200	o	0,198	0,802	0,562	0,438	0,576	0,498	0,380
Dubs, Tran	2048	50	a	0,213	0,787	0,516	0,484	0,581	0,528	0,365

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Dubs, Tran	2048	50	m	0,320	0,680	0,323	0,677	0,564	0,615	0,321
Dubs, Tran	2048	50	o	0,215	0,785	0,516	0,484	0,580	0,528	0,365
Dubs, Tran	2048	100	a	0,207	0,793	0,520	0,480	0,587	0,528	0,363
Dubs, Tran	2048	100	m	0,309	0,691	0,326	0,674	0,571	0,618	0,317
Dubs, Tran	2048	100	o	0,207	0,793	0,521	0,479	0,586	0,528	0,364
Dubs, Tran	2048	200	a	0,203	0,797	0,522	0,478	0,589	0,528	0,363
Dubs, Tran	2048	200	m	0,304	0,696	0,327	0,673	0,574	0,620	0,316
Dubs, Tran	2048	200	o	0,204	0,796	0,525	0,475	0,587	0,525	0,365
Dubs, Tran	4096	50	a	0,218	0,782	0,501	0,499	0,582	0,537	0,360
Dubs, Tran	4096	50	m	0,341	0,659	0,307	0,693	0,552	0,614	0,324
Dubs, Tran	4096	50	o	0,220	0,780	0,502	0,498	0,579	0,535	0,361
Dubs, Tran	4096	100	a	0,210	0,790	0,504	0,496	0,589	0,539	0,357
Dubs, Tran	4096	100	m	0,330	0,670	0,308	0,692	0,560	0,619	0,319
Dubs, Tran	4096	100	o	0,213	0,787	0,503	0,497	0,587	0,538	0,358
Dubs, Tran	4096	200	a	0,206	0,794	0,508	0,492	0,592	0,538	0,357
Dubs, Tran	4096	200	m	0,324	0,676	0,309	0,691	0,564	0,621	0,316
Dubs, Tran	4096	200	o	0,206	0,794	0,508	0,492	0,592	0,537	0,357
Hard, Jazz	512	50	a	0,054	0,946	0,720	0,280	0,828	0,418	0,387
Hard, Jazz	512	50	m	0,090	0,910	0,669	0,331	0,774	0,463	0,380
Hard, Jazz	512	50	o	0,051	0,949	0,734	0,266	0,830	0,403	0,392
Hard, Jazz	512	100	a	0,046	0,954	0,732	0,268	0,845	0,407	0,389
Hard, Jazz	512	100	m	0,083	0,917	0,677	0,323	0,784	0,457	0,380
Hard, Jazz	512	100	o	0,042	0,958	0,746	0,254	0,849	0,391	0,394
Hard, Jazz	512	200	a	0,042	0,958	0,737	0,263	0,853	0,402	0,390
Hard, Jazz	512	200	m	0,079	0,921	0,682	0,318	0,790	0,454	0,380
Hard, Jazz	512	200	o	0,038	0,962	0,753	0,247	0,859	0,384	0,395
Hard, Jazz	2048	50	a	0,058	0,942	0,682	0,318	0,837	0,461	0,370
Hard, Jazz	2048	50	m	0,098	0,902	0,627	0,373	0,779	0,505	0,362
Hard, Jazz	2048	50	o	0,056	0,944	0,705	0,295	0,831	0,435	0,381
Hard, Jazz	2048	100	a	0,046	0,954	0,698	0,302	0,858	0,447	0,372
Hard, Jazz	2048	100	m	0,091	0,909	0,636	0,364	0,788	0,498	0,364
Hard, Jazz	2048	100	o	0,045	0,955	0,712	0,288	0,857	0,431	0,378
Hard, Jazz	2048	200	a	0,043	0,957	0,703	0,297	0,867	0,442	0,373
Hard, Jazz	2048	200	m	0,087	0,913	0,642	0,358	0,793	0,493	0,364
Hard, Jazz	2048	200	o	0,041	0,959	0,719	0,281	0,866	0,425	0,380
Hard, Jazz	4096	50	a	0,052	0,948	0,683	0,317	0,852	0,462	0,367

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Hard, Jazz	4096	50	m	0,101	0,899	0,623	0,377	0,776	0,508	0,362
Hard, Jazz	4096	50	o	0,052	0,948	0,689	0,311	0,847	0,455	0,370
Hard, Jazz	4096	100	a	0,046	0,954	0,689	0,311	0,864	0,458	0,367
Hard, Jazz	4096	100	m	0,094	0,906	0,630	0,370	0,785	0,503	0,362
Hard, Jazz	4096	100	o	0,043	0,957	0,700	0,300	0,866	0,445	0,372
Hard, Jazz	4096	200	a	0,043	0,957	0,694	0,306	0,868	0,453	0,369
Hard, Jazz	4096	200	m	0,090	0,910	0,636	0,364	0,791	0,499	0,363
Hard, Jazz	4096	200	o	0,039	0,961	0,706	0,294	0,874	0,440	0,373
Hard, Klas	512	50	a	0,298	0,702	0,521	0,479	0,540	0,508	0,409
Hard, Klas	512	50	m	0,174	0,826	0,681	0,319	0,572	0,410	0,427
Hard, Klas	512	50	o	0,314	0,686	0,513	0,487	0,531	0,508	0,413
Hard, Klas	512	100	a	0,294	0,706	0,524	0,476	0,541	0,506	0,409
Hard, Klas	512	100	m	0,166	0,834	0,689	0,311	0,577	0,404	0,427
Hard, Klas	512	100	o	0,308	0,692	0,519	0,481	0,532	0,505	0,414
Hard, Klas	512	200	a	0,293	0,707	0,524	0,476	0,542	0,506	0,409
Hard, Klas	512	200	m	0,161	0,839	0,694	0,306	0,581	0,401	0,427
Hard, Klas	512	200	o	0,306	0,694	0,521	0,479	0,533	0,504	0,413
Hard, Klas	2048	50	a	0,306	0,694	0,491	0,509	0,547	0,528	0,398
Hard, Klas	2048	50	m	0,192	0,808	0,621	0,379	0,588	0,461	0,407
Hard, Klas	2048	50	o	0,313	0,687	0,495	0,505	0,539	0,522	0,404
Hard, Klas	2048	100	a	0,302	0,698	0,495	0,505	0,549	0,526	0,398
Hard, Klas	2048	100	m	0,183	0,817	0,629	0,371	0,595	0,457	0,406
Hard, Klas	2048	100	o	0,310	0,690	0,498	0,502	0,540	0,521	0,404
Hard, Klas	2048	200	a	0,300	0,700	0,497	0,503	0,549	0,525	0,399
Hard, Klas	2048	200	m	0,180	0,820	0,634	0,366	0,596	0,453	0,407
Hard, Klas	2048	200	o	0,311	0,689	0,498	0,502	0,540	0,520	0,404
Hard, Klas	4096	50	a	0,307	0,693	0,478	0,522	0,551	0,536	0,393
Hard, Klas	4096	50	m	0,200	0,800	0,613	0,387	0,583	0,465	0,406
Hard, Klas	4096	50	o	0,314	0,686	0,482	0,518	0,545	0,531	0,398
Hard, Klas	4096	100	a	0,302	0,698	0,488	0,512	0,551	0,531	0,395
Hard, Klas	4096	100	m	0,189	0,811	0,621	0,379	0,591	0,462	0,405
Hard, Klas	4096	100	o	0,309	0,691	0,482	0,518	0,548	0,532	0,396
Hard, Klas	4096	200	a	0,297	0,703	0,491	0,509	0,554	0,531	0,394
Hard, Klas	4096	200	m	0,184	0,816	0,624	0,376	0,596	0,461	0,404
Hard, Klas	4096	200	o	0,310	0,690	0,482	0,518	0,547	0,532	0,396
Hard, Rock	512	50	a	0,195	0,805	0,558	0,442	0,625	0,517	0,377

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Hard, Rock	512	50	m	0,213	0,787	0,524	0,476	0,621	0,539	0,369
Hard, Rock	512	50	o	0,200	0,800	0,558	0,442	0,619	0,516	0,379
Hard, Rock	512	100	a	0,188	0,812	0,564	0,436	0,631	0,516	0,376
Hard, Rock	512	100	m	0,201	0,799	0,533	0,467	0,631	0,537	0,367
Hard, Rock	512	100	o	0,193	0,807	0,563	0,437	0,625	0,514	0,378
Hard, Rock	512	200	a	0,185	0,815	0,565	0,435	0,634	0,516	0,375
Hard, Rock	512	200	m	0,194	0,806	0,538	0,462	0,637	0,535	0,366
Hard, Rock	512	200	o	0,189	0,811	0,566	0,434	0,628	0,513	0,377
Hard, Rock	2048	50	a	0,201	0,799	0,533	0,467	0,630	0,536	0,367
Hard, Rock	2048	50	m	0,240	0,760	0,468	0,532	0,618	0,572	0,354
Hard, Rock	2048	50	o	0,199	0,801	0,542	0,458	0,628	0,530	0,371
Hard, Rock	2048	100	a	0,194	0,806	0,537	0,463	0,636	0,536	0,366
Hard, Rock	2048	100	m	0,228	0,772	0,475	0,525	0,628	0,572	0,351
Hard, Rock	2048	100	o	0,194	0,806	0,545	0,455	0,632	0,529	0,370
Hard, Rock	2048	200	a	0,190	0,810	0,540	0,460	0,639	0,535	0,365
Hard, Rock	2048	200	m	0,222	0,778	0,480	0,520	0,632	0,571	0,351
Hard, Rock	2048	200	o	0,192	0,808	0,545	0,455	0,634	0,530	0,369
Hard, Rock	4096	50	a	0,194	0,806	0,539	0,461	0,635	0,534	0,366
Hard, Rock	4096	50	m	0,244	0,756	0,466	0,534	0,615	0,572	0,355
Hard, Rock	4096	50	o	0,200	0,800	0,536	0,464	0,629	0,534	0,368
Hard, Rock	4096	100	a	0,188	0,812	0,540	0,460	0,641	0,536	0,364
Hard, Rock	4096	100	m	0,235	0,765	0,469	0,531	0,623	0,573	0,352
Hard, Rock	4096	100	o	0,190	0,810	0,541	0,459	0,638	0,534	0,365
Hard, Rock	4096	200	a	0,184	0,816	0,541	0,459	0,646	0,536	0,363
Hard, Rock	4096	200	m	0,226	0,774	0,472	0,528	0,630	0,574	0,349
Hard, Rock	4096	200	o	0,189	0,811	0,544	0,456	0,639	0,532	0,366
Hard, Tran	512	50	a	0,134	0,866	0,620	0,380	0,712	0,496	0,377
Hard, Tran	512	50	m	0,218	0,782	0,512	0,488	0,660	0,561	0,365
Hard, Tran	512	50	o	0,135	0,865	0,626	0,374	0,707	0,490	0,380
Hard, Tran	512	100	a	0,126	0,874	0,629	0,371	0,719	0,490	0,377
Hard, Tran	512	100	m	0,206	0,794	0,518	0,482	0,670	0,560	0,362
Hard, Tran	512	100	o	0,127	0,873	0,633	0,367	0,715	0,485	0,380
Hard, Tran	512	200	a	0,122	0,878	0,632	0,368	0,723	0,488	0,377
Hard, Tran	512	200	m	0,200	0,800	0,523	0,477	0,674	0,559	0,361
Hard, Tran	512	200	o	0,124	0,876	0,635	0,365	0,719	0,484	0,380
Hard, Tran	2048	50	a	0,152	0,848	0,577	0,423	0,707	0,530	0,364

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Hard, Tran	2048	50	m	0,277	0,723	0,423	0,577	0,644	0,609	0,350
Hard, Tran	2048	50	o	0,153	0,847	0,575	0,425	0,706	0,530	0,364
Hard, Tran	2048	100	a	0,144	0,856	0,582	0,418	0,716	0,528	0,363
Hard, Tran	2048	100	m	0,264	0,736	0,429	0,571	0,652	0,609	0,346
Hard, Tran	2048	100	o	0,145	0,855	0,587	0,413	0,712	0,522	0,366
Hard, Tran	2048	200	a	0,140	0,860	0,583	0,417	0,720	0,528	0,362
Hard, Tran	2048	200	m	0,259	0,741	0,431	0,569	0,655	0,609	0,345
Hard, Tran	2048	200	o	0,141	0,859	0,590	0,410	0,717	0,522	0,365
Hard, Tran	4096	50	a	0,162	0,838	0,551	0,449	0,705	0,549	0,357
Hard, Tran	4096	50	m	0,291	0,709	0,410	0,590	0,636	0,612	0,351
Hard, Tran	4096	50	o	0,161	0,839	0,551	0,449	0,708	0,549	0,356
Hard, Tran	4096	100	a	0,154	0,846	0,556	0,444	0,713	0,548	0,355
Hard, Tran	4096	100	m	0,279	0,721	0,415	0,585	0,644	0,613	0,347
Hard, Tran	4096	100	o	0,153	0,847	0,562	0,438	0,713	0,543	0,357
Hard, Tran	4096	200	a	0,147	0,853	0,561	0,439	0,721	0,545	0,354
Hard, Tran	4096	200	m	0,275	0,725	0,414	0,586	0,647	0,615	0,345
Hard, Tran	4096	200	o	0,150	0,850	0,565	0,435	0,715	0,541	0,357
Jazz, Klas	512	50	a	0,452	0,548	0,363	0,637	0,475	0,544	0,408
Jazz, Klas	512	50	m	0,111	0,889	0,730	0,270	0,610	0,375	0,420
Jazz, Klas	512	50	o	0,469	0,531	0,357	0,643	0,468	0,542	0,413
Jazz, Klas	512	100	a	0,442	0,558	0,368	0,632	0,479	0,545	0,405
Jazz, Klas	512	100	m	0,101	0,899	0,742	0,258	0,622	0,365	0,421
Jazz, Klas	512	100	o	0,457	0,543	0,364	0,636	0,472	0,541	0,411
Jazz, Klas	512	200	a	0,438	0,562	0,369	0,631	0,480	0,545	0,404
Jazz, Klas	512	200	m	0,095	0,905	0,749	0,251	0,628	0,358	0,422
Jazz, Klas	512	200	o	0,448	0,552	0,366	0,634	0,476	0,544	0,407
Jazz, Klas	2048	50	a	0,461	0,539	0,338	0,662	0,479	0,556	0,399
Jazz, Klas	2048	50	m	0,112	0,888	0,709	0,291	0,624	0,397	0,410
Jazz, Klas	2048	50	o	0,476	0,524	0,328	0,672	0,475	0,556	0,402
Jazz, Klas	2048	100	a	0,457	0,543	0,334	0,666	0,482	0,559	0,396
Jazz, Klas	2048	100	m	0,100	0,900	0,724	0,276	0,638	0,386	0,412
Jazz, Klas	2048	100	o	0,470	0,530	0,329	0,671	0,478	0,558	0,399
Jazz, Klas	2048	200	a	0,452	0,548	0,334	0,666	0,485	0,561	0,393
Jazz, Klas	2048	200	m	0,094	0,906	0,730	0,270	0,646	0,381	0,412
Jazz, Klas	2048	200	o	0,461	0,539	0,333	0,667	0,481	0,559	0,397
Jazz, Klas	4096	50	a	0,489	0,511	0,300	0,700	0,477	0,567	0,395

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Jazz, Klas	4096	50	m	0,108	0,892	0,715	0,285	0,626	0,392	0,411
Jazz, Klas	4096	50	o	0,491	0,509	0,300	0,700	0,476	0,567	0,396
Jazz, Klas	4096	100	a	0,481	0,519	0,299	0,701	0,482	0,571	0,390
Jazz, Klas	4096	100	m	0,100	0,900	0,727	0,273	0,635	0,382	0,413
Jazz, Klas	4096	100	o	0,487	0,513	0,296	0,704	0,479	0,570	0,391
Jazz, Klas	4096	200	a	0,475	0,525	0,300	0,700	0,484	0,572	0,388
Jazz, Klas	4096	200	m	0,095	0,905	0,735	0,265	0,640	0,374	0,415
Jazz, Klas	4096	200	o	0,490	0,510	0,295	0,705	0,478	0,570	0,393
Jazz, Rock	512	50	a	0,339	0,661	0,559	0,441	0,459	0,450	0,449
Jazz, Rock	512	50	m	0,336	0,664	0,520	0,480	0,482	0,481	0,428
Jazz, Rock	512	50	o	0,332	0,668	0,569	0,431	0,459	0,445	0,450
Jazz, Rock	512	100	a	0,328	0,672	0,565	0,435	0,464	0,449	0,447
Jazz, Rock	512	100	m	0,323	0,677	0,530	0,470	0,487	0,478	0,426
Jazz, Rock	512	100	o	0,322	0,678	0,579	0,421	0,461	0,440	0,451
Jazz, Rock	512	200	a	0,324	0,676	0,566	0,434	0,467	0,450	0,445
Jazz, Rock	512	200	m	0,314	0,686	0,536	0,464	0,491	0,477	0,425
Jazz, Rock	512	200	o	0,320	0,680	0,582	0,418	0,461	0,439	0,451
Jazz, Rock	2048	50	a	0,333	0,667	0,522	0,478	0,484	0,481	0,427
Jazz, Rock	2048	50	m	0,350	0,650	0,498	0,502	0,483	0,492	0,424
Jazz, Rock	2048	50	o	0,327	0,673	0,551	0,449	0,473	0,461	0,439
Jazz, Rock	2048	100	a	0,322	0,678	0,535	0,465	0,485	0,475	0,428
Jazz, Rock	2048	100	m	0,336	0,664	0,509	0,491	0,487	0,489	0,423
Jazz, Rock	2048	100	o	0,319	0,681	0,556	0,444	0,476	0,460	0,437
Jazz, Rock	2048	200	a	0,320	0,680	0,540	0,460	0,484	0,471	0,430
Jazz, Rock	2048	200	m	0,330	0,670	0,516	0,484	0,489	0,486	0,423
Jazz, Rock	2048	200	o	0,314	0,686	0,563	0,437	0,476	0,456	0,438
Jazz, Rock	4096	50	a	0,331	0,669	0,514	0,486	0,488	0,487	0,423
Jazz, Rock	4096	50	m	0,357	0,643	0,505	0,495	0,473	0,484	0,431
Jazz, Rock	4096	50	o	0,317	0,683	0,539	0,461	0,485	0,473	0,428
Jazz, Rock	4096	100	a	0,320	0,680	0,520	0,480	0,493	0,487	0,420
Jazz, Rock	4096	100	m	0,346	0,654	0,512	0,488	0,478	0,483	0,429
Jazz, Rock	4096	100	o	0,303	0,697	0,549	0,451	0,492	0,470	0,426
Jazz, Rock	4096	200	a	0,312	0,688	0,531	0,469	0,494	0,481	0,421
Jazz, Rock	4096	200	m	0,339	0,661	0,516	0,484	0,480	0,482	0,428
Jazz, Rock	4096	200	o	0,303	0,697	0,550	0,450	0,491	0,470	0,426
Jazz, Tran	512	50	a	0,153	0,847	0,540	0,460	0,697	0,554	0,347

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Jazz, Tran	512	50	m	0,132	0,868	0,560	0,440	0,718	0,546	0,346
Jazz, Tran	512	50	o	0,151	0,849	0,562	0,438	0,689	0,536	0,357
Jazz, Tran	512	100	a	0,141	0,859	0,550	0,450	0,710	0,551	0,346
Jazz, Tran	512	100	m	0,121	0,879	0,570	0,430	0,731	0,542	0,345
Jazz, Tran	512	100	o	0,139	0,861	0,568	0,432	0,705	0,536	0,353
Jazz, Tran	512	200	a	0,136	0,864	0,555	0,445	0,715	0,549	0,346
Jazz, Tran	512	200	m	0,116	0,884	0,575	0,425	0,738	0,540	0,345
Jazz, Tran	512	200	o	0,134	0,866	0,570	0,430	0,712	0,537	0,352
Jazz, Tran	2048	50	a	0,185	0,815	0,472	0,528	0,685	0,596	0,329
Jazz, Tran	2048	50	m	0,156	0,844	0,495	0,505	0,712	0,591	0,326
Jazz, Tran	2048	50	o	0,181	0,819	0,495	0,505	0,681	0,580	0,338
Jazz, Tran	2048	100	a	0,175	0,825	0,479	0,521	0,695	0,595	0,327
Jazz, Tran	2048	100	m	0,145	0,855	0,503	0,497	0,723	0,589	0,324
Jazz, Tran	2048	100	o	0,170	0,830	0,501	0,499	0,692	0,580	0,335
Jazz, Tran	2048	200	a	0,168	0,832	0,482	0,518	0,703	0,597	0,325
Jazz, Tran	2048	200	m	0,141	0,859	0,509	0,491	0,727	0,586	0,325
Jazz, Tran	2048	200	o	0,167	0,833	0,501	0,499	0,696	0,581	0,334
Jazz, Tran	4096	50	a	0,199	0,801	0,444	0,556	0,680	0,612	0,321
Jazz, Tran	4096	50	m	0,164	0,836	0,482	0,518	0,706	0,597	0,323
Jazz, Tran	4096	50	o	0,194	0,806	0,457	0,543	0,681	0,604	0,325
Jazz, Tran	4096	100	a	0,185	0,815	0,451	0,549	0,694	0,613	0,318
Jazz, Tran	4096	100	m	0,153	0,847	0,491	0,509	0,716	0,595	0,322
Jazz, Tran	4096	100	o	0,181	0,819	0,463	0,537	0,693	0,605	0,322
Jazz, Tran	4096	200	a	0,180	0,820	0,453	0,547	0,698	0,613	0,317
Jazz, Tran	4096	200	m	0,148	0,852	0,494	0,506	0,722	0,595	0,321
Jazz, Tran	4096	200	o	0,174	0,826	0,468	0,532	0,700	0,605	0,321
Klas, Rock	512	50	a	0,645	0,355	0,141	0,859	0,405	0,550	0,393
Klas, Rock	512	50	m	0,589	0,411	0,299	0,701	0,378	0,491	0,444
Klas, Rock	512	50	o	0,643	0,357	0,144	0,856	0,405	0,549	0,394
Klas, Rock	512	100	a	0,648	0,352	0,136	0,864	0,405	0,551	0,392
Klas, Rock	512	100	m	0,584	0,416	0,302	0,698	0,379	0,491	0,443
Klas, Rock	512	100	o	0,645	0,355	0,141	0,859	0,405	0,550	0,393
Klas, Rock	512	200	a	0,647	0,353	0,134	0,866	0,406	0,553	0,391
Klas, Rock	512	200	m	0,581	0,419	0,304	0,696	0,379	0,491	0,442
Klas, Rock	512	200	o	0,645	0,355	0,139	0,861	0,405	0,551	0,392
Klas, Rock	2048	50	a	0,620	0,380	0,140	0,860	0,414	0,559	0,380

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Klas, Rock	2048	50	m	0,588	0,412	0,276	0,724	0,385	0,502	0,432
Klas, Rock	2048	50	o	0,614	0,386	0,145	0,855	0,414	0,558	0,380
Klas, Rock	2048	100	a	0,616	0,384	0,138	0,862	0,415	0,560	0,377
Klas, Rock	2048	100	m	0,583	0,417	0,278	0,722	0,386	0,503	0,430
Klas, Rock	2048	100	o	0,614	0,386	0,142	0,858	0,415	0,560	0,378
Klas, Rock	2048	200	a	0,620	0,380	0,135	0,865	0,415	0,561	0,378
Klas, Rock	2048	200	m	0,579	0,421	0,281	0,719	0,386	0,503	0,430
Klas, Rock	2048	200	o	0,612	0,388	0,142	0,858	0,416	0,560	0,377
Klas, Rock	4096	50	a	0,597	0,403	0,135	0,865	0,422	0,568	0,366
Klas, Rock	4096	50	m	0,600	0,400	0,269	0,731	0,380	0,500	0,435
Klas, Rock	4096	50	o	0,584	0,416	0,143	0,857	0,425	0,569	0,364
Klas, Rock	4096	100	a	0,595	0,405	0,133	0,867	0,424	0,569	0,364
Klas, Rock	4096	100	m	0,596	0,404	0,271	0,729	0,381	0,500	0,434
Klas, Rock	4096	100	o	0,582	0,418	0,143	0,857	0,426	0,569	0,363
Klas, Rock	4096	200	a	0,591	0,409	0,133	0,867	0,425	0,571	0,362
Klas, Rock	4096	200	m	0,594	0,406	0,272	0,728	0,382	0,501	0,433
Klas, Rock	4096	200	o	0,582	0,418	0,142	0,858	0,427	0,570	0,362
Klas, Tran	512	50	a	0,312	0,688	0,394	0,606	0,537	0,570	0,353
Klas, Tran	512	50	m	0,221	0,779	0,555	0,445	0,546	0,491	0,388
Klas, Tran	512	50	o	0,324	0,676	0,388	0,612	0,531	0,569	0,356
Klas, Tran	512	100	a	0,303	0,697	0,397	0,603	0,543	0,572	0,350
Klas, Tran	512	100	m	0,208	0,792	0,565	0,435	0,555	0,488	0,386
Klas, Tran	512	100	o	0,315	0,685	0,391	0,609	0,536	0,570	0,353
Klas, Tran	512	200	a	0,299	0,701	0,398	0,602	0,546	0,573	0,348
Klas, Tran	512	200	m	0,202	0,798	0,570	0,430	0,560	0,486	0,386
Klas, Tran	512	200	o	0,311	0,689	0,393	0,607	0,539	0,571	0,352
Klas, Tran	2048	50	a	0,356	0,644	0,322	0,678	0,531	0,596	0,339
Klas, Tran	2048	50	m	0,266	0,734	0,457	0,543	0,548	0,546	0,361
Klas, Tran	2048	50	o	0,361	0,639	0,320	0,680	0,529	0,595	0,340
Klas, Tran	2048	100	a	0,344	0,656	0,324	0,676	0,539	0,600	0,334
Klas, Tran	2048	100	m	0,257	0,743	0,464	0,536	0,553	0,544	0,361
Klas, Tran	2048	100	o	0,352	0,648	0,324	0,676	0,533	0,596	0,338
Klas, Tran	2048	200	a	0,343	0,657	0,323	0,677	0,540	0,601	0,333
Klas, Tran	2048	200	m	0,251	0,749	0,469	0,531	0,557	0,543	0,360
Klas, Tran	2048	200	o	0,351	0,649	0,324	0,676	0,534	0,597	0,337
Klas, Tran	4096	50	a	0,367	0,633	0,284	0,716	0,536	0,613	0,325

T. Genre	wi.	tr.	ty.	FP	TN	FN	rec.	pre.	f-m.	b.F.
Klas, Tran	4096	50	m	0,291	0,709	0,430	0,570	0,536	0,552	0,361
Klas, Tran	4096	50	o	0,379	0,621	0,283	0,717	0,528	0,609	0,331
Klas, Tran	4096	100	a	0,362	0,638	0,286	0,714	0,538	0,614	0,324
Klas, Tran	4096	100	m	0,278	0,722	0,438	0,562	0,544	0,553	0,358
Klas, Tran	4096	100	o	0,366	0,634	0,287	0,713	0,535	0,611	0,327
Klas, Tran	4096	200	a	0,357	0,643	0,287	0,713	0,541	0,615	0,322
Klas, Tran	4096	200	m	0,271	0,729	0,442	0,558	0,548	0,553	0,357
Klas, Tran	4096	200	o	0,364	0,636	0,290	0,710	0,536	0,611	0,327
Rock, Tran	512	50	a	0,232	0,768	0,521	0,479	0,558	0,515	0,377
Rock, Tran	512	50	m	0,303	0,697	0,388	0,612	0,553	0,581	0,345
Rock, Tran	512	50	o	0,232	0,768	0,525	0,475	0,556	0,513	0,378
Rock, Tran	512	100	a	0,223	0,777	0,525	0,475	0,566	0,516	0,374
Rock, Tran	512	100	m	0,290	0,710	0,394	0,606	0,561	0,582	0,342
Rock, Tran	512	100	o	0,224	0,776	0,527	0,473	0,564	0,515	0,375
Rock, Tran	512	200	a	0,219	0,781	0,526	0,474	0,570	0,517	0,373
Rock, Tran	512	200	m	0,282	0,718	0,398	0,602	0,566	0,583	0,340
Rock, Tran	512	200	o	0,220	0,780	0,528	0,472	0,568	0,516	0,374
Rock, Tran	2048	50	a	0,252	0,748	0,479	0,521	0,557	0,538	0,366
Rock, Tran	2048	50	m	0,333	0,667	0,335	0,665	0,549	0,601	0,334
Rock, Tran	2048	50	o	0,255	0,745	0,487	0,513	0,551	0,532	0,371
Rock, Tran	2048	100	a	0,242	0,758	0,486	0,514	0,564	0,538	0,364
Rock, Tran	2048	100	m	0,323	0,677	0,338	0,662	0,555	0,604	0,330
Rock, Tran	2048	100	o	0,244	0,756	0,493	0,507	0,558	0,531	0,369
Rock, Tran	2048	200	a	0,237	0,763	0,490	0,510	0,567	0,537	0,364
Rock, Tran	2048	200	m	0,317	0,683	0,339	0,661	0,559	0,606	0,328
Rock, Tran	2048	200	o	0,239	0,761	0,496	0,504	0,563	0,532	0,367
Rock, Tran	4096	50	a	0,258	0,742	0,476	0,524	0,552	0,538	0,367
Rock, Tran	4096	50	m	0,353	0,647	0,318	0,682	0,539	0,602	0,336
Rock, Tran	4096	50	o	0,253	0,747	0,484	0,516	0,553	0,533	0,369
Rock, Tran	4096	100	a	0,252	0,748	0,479	0,521	0,556	0,538	0,366
Rock, Tran	4096	100	m	0,345	0,655	0,322	0,678	0,543	0,603	0,333
Rock, Tran	4096	100	o	0,246	0,754	0,488	0,512	0,558	0,534	0,367
Rock, Tran	4096	200	a	0,248	0,752	0,479	0,521	0,560	0,540	0,363
Rock, Tran	4096	200	m	0,339	0,661	0,324	0,676	0,547	0,604	0,332
Rock, Tran	4096	200	o	0,244	0,756	0,487	0,513	0,560	0,535	0,366

A.4 AMUSE-Merkmale

Tabelle A.11: Liste aller verwendeten AMUSE-Merkmale: Merkmals-ID im AMUSE-Tool, Beschreibung des Merkmals, Fenstergröße des Merkmals in Frames, Anzahl der Dimensionen

ID	Merkmal Beschreibung	Fenstergröße	Dim.
0	Zero-crossing rate	512	1
1	Linear prediction coefficients	512	10
2	Average distance between extremal spectral values and its variance	512	2
3	Average distance between zero-crossings of the time-domain signal and its variance	512	2
4	Root mean square	512	1
6	Low energy	512	1
7	Normalized energy of harmonic components	512	1
10	Tristimulus	512	2
14	Spectral centroid	512	1
15	Spectral irregularity	512	1
16	Spectral bandwidth	512	1
17	Spectral skewness	512	1
18	Spectral kurtosis	512	1
19	Spectral crest factor	512	4
20	Spectral flatness measure	512	4
21	Spectral extent	512	1
22	Spectral flux	512	1
23	Spectral brightness	512	1
24	Sensory roughness	1024	1
25	Sub-band energy ratio	512	4
29	Spectral slope	512	1
30	y-axis intercept	512	1
31	Discrepancy	512	1
32	Angles in phase domain	512	1
33	Distances in phase domain	512	1
39	Mel frequency cepstral coefficients - MIR Toolbox	512	13
40	Bark scale magnitudes	512	23
200	Fundamental frequency	512	1
201	Inharmonicity	512	1
202	Key and its clarity	512	2

ID	Merkmal Beschreibung	Fenstergröße	Dim.
203	Major/minor alignment	512	1
204	Chroma vector	512	24
205	Amplitude of maximum in the chromagram	512	1
206	Chroma vector - MIR Toolbox	512	12
207	Tone with maximum strength in the chromagram	512	1
209	Strengths of major keys	512	12
210	Strengths of minor keys	512	12
211	Amplitude, position and width of 1st spectral peak	512	3
212	Amplitude, position and width of 2nd spectral peak	512	3
213	Amplitude, position and width of 3rd spectral peak	512	3
214	Amplitude, position and width of 4th spectral peak	512	3
215	Amplitude, position and width of 5th spectral peak	512	3
216	Tonal centroid vector	512	6
250	Chroma - NNLS	2048	12
251	Bass chroma - NNLS	2048	12
253	Local tuning - NNLS	8192	1
254	Harmonic change - NNLS	2048	1
255	Consonance - NNLS	2048	1
260	Interval strengths estimated from 10 highest semitone peaks	2048	12
261	Interval strengths estimated from the semitone peaks above 3/4 of the maximum peak	2048	12
402	First relative periodicity amplitude peak	512	1
405	First periodicity peak in bpm	512	1
406	Second periodicity peak in bpm	512	1
407	Sum of correlated components	512	1

Abbildungsverzeichnis

1.1 Spektogramm aus „Carl Orff - O Fortuna“: Ein Sänger, Zeitabschnitt von ca. 2 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse. Farbverlauf von Schwarz über grün und gelb bis hin zu rot. Schwarz entspricht fehlender Frequenz, rot zeigt eine stark vorhandene Frequenz an. 2

1.2 Spektogramm aus „Il Divo - Everytime I Look At You“: Chorgesang, Zeitabschnitt von ca. 2 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse. Farbverlauf von Schwarz über grün und gelb bis hin zu rot. Schwarz entspricht fehlender Frequenz, rot zeigt eine stark vorhandene Frequenz an. 3

2.1 Unterteilung des Klassifikationsergebnises in „True Positive“, „False Positive“, „True Negative“ und „False Negative“. 10

2.2 Schaubild eines einzelnen Entscheidungsbaumes bestehend aus sieben Knoten in drei Schichten. Zur Entscheidungsfindung werden die Merkmale A-E verwendet. Durch die zufallsbedingte Auswahl der Merkmale werden die Merkmale A und B mehrfach verwendet. 12

2.3 Aufbau eines Neurons mit vier Eingaben X1 bis X4. Sie werden mit den vier Gewichten W1 bis W4 verrechnet und aufsummiert. Anschließend wird die Summe mit WS erneut gewichtet, bevor die Aktivierungsfunktion die Ausgabe des Neurons berechnet. Die grün gefärbten Elemente zeigen Eingabe- und Ausgabefelder, die rot gefärbten Elemente sind Variablen, welche im Lernprozess angepasst werden können. 14

2.4 Aufbau eines neuronalen Netzes mit drei Eingaben und zwei Ausgaben: In rot die Eingabeschicht, in blau zwei verdeckte Schichten und in grün die Ausgabeschicht. 15

3.1 Spektogramm aus „John Coltrane - Giant Steps“: Saxophon-Solo, Zeitabschnitt von ca. 2 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse. Farbverlauf von Schwarz über grün und gelb bis hin zu rot. Schwarz entspricht fehlender Frequenz, rot zeigt eine stark vorhandene Frequenz an. 20

4.1 Wellenform mit Spektogramm aus: „Adoro - Engel fliegen einsam“, Zeitabschnitt von ca. 5 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse, links der Bildmitte ohne Gesang, rechts mit Gesang. . . . 24

4.2 Wellenform mit Spektogramm aus: „Alexisonfire - A dagger through the heart of St. Angeles“, Zeitabschnitt von ca. 5 Sekunden, Frequenzbereich auf der Y-Achse, Zeitbereich auf der X-Achse, links der Bildmitte mit Gesang, rechts ohne Gesang. 25

4.3 „Recall“, „precision“ und „f-measure“ für Ansatz 1 aufgeschlüsselt nach Genre und gesamt. 26

4.4 „Recall“, „precision“ und „f-measure“ im „leave one out“ Verfahren für Ansatz 2 mit 200ms Fenster und 100 Bäumen im Random Forest aufgeschlüsselt nach Genre und gesamt. 31

4.5 „Recall“, „precision“ und „f-measure“ in der Kreuzvalidierung für Ansatz 3 mit 100 Bäumen im Random Forest aufgeschlüsselt nach Fenstergröße und Merkmalen. 39

4.6 „Recall“, „precision“ und „f-measure“ im „leave one out“ Verfahren für Ansatz 3 mit 4096 Frames Fenster und 100 Bäumen im Random Forest aufgeschlüsselt nach Genre und Merkmalskombination. 43

Tabellenverzeichnis

4.1 Ergebnisse Ansatz 1 aufgeschlüsselt nach Genre: „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 26

4.2 Ergebnisse Ansatz 1 für Dubstep und Hardstyle nach Anpassung der Annotation: „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 27

4.3 Zehnfache Kreuzvalidierung mit Ansatz 2 auf dem gesamten Testdatensatz: Fenstergröße, Anzahl der Bäume im Random Forest, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 29

4.4 „Leave one out“ mit Ansatz 2 auf jedem Genre getrennt mit einer Fenstergröße von 200ms und 100 Bäumen im Random Forest: Genre, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 30

4.5 Genre Konfusion mit Ansatz 2 mit einem 200ms Fenster, 100 Bäumen im Random Forest, einem Trainingsgenre und fünf Testgenres: Trainingsgenre, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 32

4.6 „F-measure“ der Genre Konfusion von Ansatz 2 mit Fenstergröße 200ms und 100 Bäumen im Random Forest. 33

4.7 „Recall“ der Genre Konfusion von Ansatz 2 mit Fenstergröße 200ms und 100 Bäumen im Random Forest. 34

4.8 „Precision“ der Genre Konfusion von Ansatz 2 mit Fenstergröße 200ms und 100 Bäumen im Random Forest. 35

4.9 Genre Konfusion mit Ansatz 2 mit einem 200ms Fenster, 100 Bäumen im Random Forest, zwei Trainingsgenres und vier Testgenres: Trainingsgenres, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 36

4.10 Zehnfache Kreuzvalidierung mit Ansatz 3 auf dem gesamten Testdatensatz: Fenstergröße, Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 38
4.11 „Leave one out“ mit Ansatz 3 auf jedem Genre getrennt mit einer Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest: Genre, Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 42
4.12 Genre Konfusion mit Ansatz 3 mit einem 4096 Frames Fenster, 100 Bäumen im Random Forest, einem Trainingsgenre und fünf Testgenres: Trainingsgenre, Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 44
4.13 Genre Konfusion mit Ansatz 3 mit einem 4096 Frames Fenster, 100 Bäumen im Random Forest, Durchschnitt nach Merkmalskombination: Merkmalskombination, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 45
4.14 „F-measure“ der Genre Konfusion von Ansatz 3 mit Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest, oben mit allen Merkmalen, unten nur mit MFCC-Merkmalen. 46
4.15 „Recall“ der Genre Konfusion von Ansatz 3 mit Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest, oben mit allen Merkmalen, unten nur mit MFCC-Merkmalen. 48
4.16 „Precision“ der Genre Konfusion von Ansatz 3 mit Fenstergröße von 4096 Frames und 100 Bäumen im Random Forest, oben mit allen Merkmalen, unten nur mit MFCC-Merkmalen. 49
4.17 Genre Konfusion mit Ansatz 3 mit zwei Trainingsgenres, vier Testgenres, 4096 Frames Fenster und 100 Bäumen im Random Forest: Trainingsgenres, Merkmalstyp, „False Positive“, „True Negative“, „False Negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 51

A.1 Liste aller verwendeten Genres, Zeitangaben in Minuten 59
A.2 Liste aller verwendeten Teststücke, Zeitangaben in Sekunden 59
A.3 Einteilung des Testdatensatzes in die zehn Folds für die Kreuzvalidierung. . 67
A.4 „leave one out“ mit Ansatz 2 auf jedem Genre getrennt: Genre, Fenstergröße, Anzahl der Bäume im Random Forest, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 67
A.5 Genre Konfusion mit Ansatz 2 mit einem Trainingsgenre und fünf Testgenres: Trainingsgenre, Fenstergröße, Anzahl der Bäume im Random Forest, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 68

A.6 Genre Konfusion mit Ansatz 2 mit zwei Trainingsgenres und vier Testgenres: Trainingsgenres, Fenstergröße, Anzahl der Bäume im Random Forest, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 69
A.7 Zehnfache Kreuzvalidierung mit Ansatz 3 auf dem gesamten Testdatensatz: Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalskonfiguration, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 72
A.8 „leave one out“ mit Ansatz 3 auf jedem Genre getrennt: Genre, Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalskombination, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 73
A.9 Genre Konfusion mit Ansatz 3 mit einem Trainingsgenre und fünf Testgenres: Trainingsgenre, Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalskonfiguration, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 78
A.10 Genre Konfusion mit Ansatz 3 mit zwei Trainingsgenres und vier Testgenres: Trainingsgenres, Fenstergröße, Anzahl der Bäume im Random Forest, Merkmalstyp, „false positive“, „true negative“, „false negative“, „recall“, „precision“, „f-measure“ und balancierter Fehler. 82
A.11 Liste aller verwendeten AMUSE-Merkmale: Merkmals-ID im AMUSE-Tool, Beschreibung des Merkmals, Fenstergröße des Merkmals in Frames, Anzahl der Dimensionen 94

Literaturverzeichnis

[1] *Brockhaus, Band 10.* F. A. Brockhaus, Leipzig, 21. Auflage, 2006.

[2] Berenzweig, A. L. und D. P. W. Ellis: *Locating singing voice segments within music signals.* In: *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, Seiten 119–122, 2001.

[3] Breiman, L.: *Random Forests.* Machine Learning, 45(1):5–32, 2001.

[4] Dittmar, C., B. Lehner, T. Prätzlich, M. Müller und G. Widmer: *Cross-Version Singing Voice Detection in Classical Opera Recordings.* In: *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Seiten 618–624, 2015.

[5] Goto, M., H. Hashiguchi, T. Nishimura und R. Oka: *RWC music database: Popular, classical and jazz databases.* In: *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Seiten 287–288, 2002.

[6] Hastie, T. J., R. J. Tibshirani und J. H. Friedman: *The elements of statistical learning: Data mining, inference, and prediction.* Springer, New York, 2. Auflage, 2013.

[7] Herzfeld, F.: *Das neue Ullstein-Lexikon der Musik.* Ullstein, Frankfurt/M. and Berlin, 8. Auflage, 1989.

[8] Leglaive, S., R. Hennequin und R. Badeau: *Singing voice detection with deep recurrent neural networks.* In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seiten 121–125, 2015.

[9] Lehner, B., G. Widmer und R. Sonnleitner: *On the reduction of false positives in singing voice detection.* In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seiten 7480–7484, 2014.

[10] McEnnis, D., C. McKay, I. Fujinaga und P. Depalle: *jAudio: An Feature Extraction Library.* In: *Proceedings of the 6th International Society for Music Information Retrieval Conference ISMIR*, Seiten 600–603, 2005.

[11] Niemann, H.: *Klassifikation von Mustern.* Springer, Berlin, 2. Auflage, 2003.

[12] Ramona, M., G. Richard und B. David: *Vocal detection in music with support vector machines.* In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Seiten 1885–1888, 2008.

[13] Regnier, L. und G. Peeters: *Singing Voice Detection in Music Tracks Using Direct Voice Vibrato Detection.* In: *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seiten 1685–1688, 2009.

[14] Rossing, T. D., F. R. Moore und P. A. Wheeler.: *The science of sound.* Addison Wesley, San Francisco, Third edition Auflage, 2002.

[15] Schlüter, J. und T. Grill: *Exploring Data Augmentation for Improved Singing Voice Detection with Neural Networks.* In: *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Seiten 121–126, 2015.

[16] Urbano, J., D. Bogdanov, P. Herrera, E. Gómez und X. Serra: *What is the Effect of Audio Quality on the Robustness of MFCCs and Chroma Features?* In: *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR)*, Seiten 573–578, 2014.

[17] Vatolkin, I.: *Improving supervised music classification by means of multi-objective evolutionary feature selection.* Doktorarbeit, Lehrstuhl 11 Informatik, Technische Universität Dortmund, 2013.

[18] Vatolkin, I., W. M. Theimer und M. Botteck: *AMUSE (Advanced MUSic Explorer) - A Multitool Framework for Music Data Analysis.* In: *Proceedings of the 11th International Society for Music Information Retrieval Conference ISMIR*, Seiten 33–38, 2010.

[19] You, S. D., Y.-C. Wu und S.-H. Peng: *Comparative study of singing voice detection methods.* Multimedia Tools and Applications, Seiten 1–16, 2015.

[20] Ziegenrücker, W.: *ABC Musik - Allgemeine Musiklehre: 446 Lehr- und Lernsätze.* Breitkopf & Härtel, Wiesbaden, 1997.

Danksagungen

An dieser Stelle möchte ich all denen danken, die mich bei der Anfertigung meiner Bachelorarbeit unterstützt haben.

Zuerst gilt mein Dank Prof. Dr. Günter Rudolph und Dr. Igor Vatolkin, die mich während meiner Bachelorarbeit begleitet haben und mir ermöglicht haben, an diesem sehr interessanten Thema zu arbeiten. Besonderer Dank gilt Dr. Vatolkin, der alle benötigten Merkmale für den dritten Ansatz mit dem AMUSE-Tool extrahiert hat und mir bei Fragen immer mit einem guten Rat zur Verfügung stand.

Ich danke Jan Schlüter für die Bereitstellung seines vortrainierten neuronalen Netzes zur Durchführung meiner Experimente sowie für die Unterstützung beim Setup der Bibliothek Theano.

Mein Dank gilt ebenfalls Bernhard Lehner, der mir alle nötigen Merkmale extrahiert hat, um seinen Ansatz zwei zu testen. Hervorheben möchte ich außerdem seine vielen hilfreichen Hinweise zur Annotation und zum Umgang mit dem Sonic Visualizer.

Ich danke Stefanie Rahder für das Korrekturlesen meiner Arbeit und die vielen Hinweise, die sie mir zur Gestaltung der Arbeit gegeben hat.

Meinen Eltern Hiltrud und Reiner danke ich für die Unterstützung, die sie mir während meiner gesamten Ausbildung gegeben haben. Ohne euch wäre ich nie so weit gekommen!

Meinem Bruder Daniel danke ich dafür, dass er mir stets das geschulte Ohr eines Musikers zur Verfügung gestellt hat, wenn ich während der Annotation an ein Hindernis geraten bin.

Zuletzt gilt mein besonderer Dank meiner Verlobten Veronika für ihre Unterstützung und ihre Geduld, die sie mir während der Bearbeitungszeit entgegen gebracht hat. Ich danke ihr dafür, dass sie mir immer wieder Mut gemacht hat, mich aufgebaut hat und sie mir stets mit Optimismus zur Seite stand, wenn meine eigene Zuversicht verflogen war.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet sowie Zitate kenntlich gemacht habe.

Dortmund, den 28. November 2016

Florian Scholz