

# Design-Space von Audio Sample Browsern

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieur/in**

im Rahmen des Studiums

**Medieninformatik**

eingereicht von

**Georg Julius Peter Gruber**

Matrikelnummer 09008390

an der  
Fakultät für Informatik der Technischen Universität Wien

Betreuung  
Betreuer/in: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Andreas Rauber

Wien, 30.09.2017

\_\_\_\_\_  
(Unterschrift Verfasser/in)

\_\_\_\_\_  
(Unterschrift Betreuer/in)

Georg Julius Peter Gruber, Neubaugasse 18/2/6, 1070 Wien

„Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit –einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.“

Wien, 30.09.2017

---

(Unterschrift Verfasser/in)

## Kurzfassung

Für Musiker, Gamesound-Designer und Toningenieure, die sich mit der Vertonung von Filmen und Videos beschäftigen, ist Sample-Browsing ein Bestandteil des täglichen Arbeitsablaufes. Während sich die Methoden und Benutzeroberflächen mit denen die genannten Berufsgruppen ihre Samplesuche durchführen kaum verändert haben, ist der Umfang der zur Verfügung stehenden Samples in den letzten 15 Jahren exponentiell gestiegen. Das kreative Potential derartig großer Sample-Mengen kann weder mit listenbasierten Interfaces noch mit einer Stichwortsuche ausgeschöpft werden.

Als Alternative zu listenbasierten Interfaces und Stichwortsuchen sind in den letzten 15 Jahren mehrere Forschungsprototypen entwickelt worden, die die psychoakustische Ähnlichkeit von Samples algorithmisch berechnen. In den meisten Fällen stellen diese Anwendungen einzelne Samples als grafische Objekte auf einer 2D Karte dar. Die Ähnlichkeit der Samples wird durch den Abstand der grafischen Objekte visualisiert. Nach 15 Jahren universitärer Forschung und einer Vielzahl an vorgestellten Forschungsprototypen bleibt der Befund, dass sich dieses Designkonzept (=ähnlichkeitsbasierter Sample-Browser) in der Praxis nicht durchgesetzt hat. Die zentrale Fragestellung dieser Arbeit ist daher die Identifikation jener Faktoren, welche die Usability dieses Ansatzes beeinträchtigen.

Die bisher vorgestellten Forschungsprototypen wurden häufig durch relative kurze, pro Anwenderin einmalige User-Tests evaluiert. Ähnlichkeitsbasierte Sample-Browser unterscheiden sich in vielen Aspekten deutlich von listenbasierten Interfaces und erfordern eine Lernphase um das volle Potential dieser Interfaces auszuschöpfen. In einem mehrmonatigen, partizipativen Designprozess wurde daher im Rahmen dieser Diplomarbeit ein ähnlichkeitsbasierter Sample-Browser entwickelt und zwei Anwender über einen längeren Zeitraum mit dem Designkonzept „ähnlichkeitsbasierter Sample-Browser“ vertraut gemacht. Im Laufe des Designprozesses wurden jene Faktoren identifiziert, die eine weitreichende Verwendung von ähnlichkeitsbasierten Sample-Browsern durch die genannten Berufsgruppen verhindern.

Drei Faktoren beeinträchtigen die Usability von ähnlichkeitsbasierten Sample-Browsern wesentlich. Einer dieser Faktoren ist die Diskrepanz zwischen der algorithmisch ermittelten und durch Abstand visualisierten Ähnlichkeit der Samples und dem subjektiven Ähnlichkeitsempfinden der Anwenderinnen. Weiters ergeben die Berechnungen und Visualisierungen von ähnlichkeitsbasierten Sample-Browsern für in der Praxis verwendete Sample-Sammlungen zumeist nur eine einzige, diffuse Punktwolke ohne erkennbare Gruppen ähnlicher Samples. Damit sind für den realen Anwendungsfall die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der 2D Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind, stark beeinträchtigt. Drittens wurde die für jede ähnlichkeitsbasierte 2D Karte unterschiedliche semantische Struktur als gravierende Schwäche identifiziert. Anwenderinnen gelingt es grundsätzlich die semantische Struktur der Karten durch Anhören der Samples zu ermitteln. Ein Memorisieren der semantischen Struktur gelingt den Anwenderinnen aber in der Regel nicht. Eine zielgerichtete Suche in Karten, deren semantische Struktur bereits einmal ermittelt wurde, ist daher nicht möglich. Jeder erneute Zugriff auf eine Karte gleicht (-wie bei einem listenbasierten Interface-) eher einem zufallsgesteuerten „Try and Error“ Prozess.

## Abstract

Sample-Browsing is a part of musicians, gamesound-designers and filmsound-designers daily workflow. While the methods and user interfaces that the named above professions use while searching for samples hardly changed, the number of available samples increased exponentially during the last 15 years. The full, creative potential of such a huge number of samples cannot be tapped with list-based interfaces or a keyword search.

As an alternative to list-based interfaces and keyword searches several research prototypes, that calculate the psychoacoustic similarity of samples algorithmically, have been developed during the last 15 years. In most of cases these applications depict single samples as graphical objects on a 2D map. The similarity of samples is visualized by the distance between these graphical objects. After 15 years of scientific research and a multitude of proposed research prototypes the finding is that this design concept (= similarity-based sample-browser) is not widely used in practice. Therefore, the central issue of this thesis is the identification of factors that impair the usability of this approach.

In many cases the research prototypes proposed so far have been evaluated with comparatively short and for each user one-off user tests. Similarity-based sample-browsers differ significantly in many aspects from list-based interfaces and a learning phase is needed to tap the full potential of these user interfaces. Therefore, in a participatory design process lasting several months a similaritybased sample-browser was developed as part of this thesis and two users were acquainted with the design concept of „similaritybased sample-browsing “. During the design process factors that forestall the wide-ranging use by the named above professional guilds were identified.

Three factors impair the usability of similarity-based sample-browsers significantly. One of these factors is the discrepancy between the algorithmically calculated similarity (-visualized by distance) and the resemblance of samples subjectively perceived by users. Furthermore, the calculations and visualizations of similarity based sample-browsers show only a single, diffuse cloud of points without any recognizable groups of similar samples for real world sample-collections. Thereby the visual perception of similarity, the systematic navigation as well as the creation and memorizing of a mental picture where on the map samples with specific acoustic und tonal properties are to be found are severely jeopardized for the realistic case of application. Thirdly the semantic structure differing for each similaritybased 2D map was identified as a severe weakness. Principally users manage to identify the semantic structure while listening to the samples on the map but generally fail to memorize the semantic structure. Therefore, a targeted search on maps, the semantic structure of which was already identified beforehand is not possible. Each access on the 2D maps is (-similar to a list-based interface-) a try and error process.

# Inhalt

1	Einleitung.....	8
1.1	Motivation und Problemstellung .....	8
1.2	Praxisnahes Anwendungsszenario für Sample-Browser .....	10
1.3	Umfangreiche Stichwortverzeichnisse als Sample-Browsing Werkzeuge.....	11
1.4	Forschungsfragen .....	12
1.5	Methodisches Vorgehen .....	13
1.6	Struktur der Arbeit .....	13
2	Stand der Technik.....	15
2.1	Ähnlichkeitsbasierte Sample-Browser.....	15
2.1.1	Sonic Browser [12].....	16
2.1.2	MARSYAS3D [13] .....	17
2.1.3	ARB – Audio Retrieval Browser [15].....	18
2.1.4	SoundTorch [16], [9].....	19
2.1.5	Sound Navigator [7], [17] .....	21
2.1.6	MediaCycle [19], AudioCycle [18] .....	23
2.1.7	Grill und Flexer [20], [29], [30] .....	25
2.1.8	AudioQuilt, Snare Drum Navigator [21] .....	29
2.1.9	AudioMetro [2].....	33
2.1.10	Drumspace [6] .....	36
2.1.11	Konzeptioneller Aufbau von ähnlichkeitsbasierten Sample-Browsern.....	39
2.2	Inhaltsbasierte Merkmalsextraktion .....	41
2.2.1	Konzeptioneller Aufbau von inhaltsbasierten Merkmalsextraktoren.....	42
2.2.2	Inhaltsbasierte Features für die Modellierung von Klangfarben .....	45
2.2.3	Klangfarbe – Psychoakustik und MIR Forschung.....	49
2.3	Dimensionsreduktion .....	58
2.3.1	PCA - Principal Component Analysis.....	59
2.3.2	MDS (Multidimensional Scaling) .....	60
2.3.3	SOM (Self-organizing Map).....	60
2.3.4	t-SNE - t-Distributed Stochastic Neighbor Embedding.....	60
2.3.5	Quantitative und qualitative Evaluierung von DRvis Methoden.....	62
2.4	Zusammenfassung.....	65
3	Analyse der ähnlichkeitsbasierten Sample-Browser und resultierende Forschungsfragen.....	66
3.1	Usability von ähnlichkeitsbasierten 2D Karten als Sample-Browsing Werkzeuge .....	66
3.1.1	Forschungsfragen - Usability von ähnlichkeitsbasierten 2D Karten.....	71
3.2	Algorithmische Ähnlichkeit und subjektives Ähnlichkeitsempfinden .....	71

3.2.1	Forschungsfragen - Algorithmisch berechnete Ähnlichkeit und subjektives Ähnlichkeitsempfinden.....	73
3.3	Beschleunigtes Sample-Browsing durch selektive, auditive Aufmerksamkeit.....	73
3.3.1	Forschungsfragen - Sonifikation.....	74
3.4	Zusammenfassung.....	75
4	Implementierung und Evaluierung des MM-Browsers .....	76
4.1	Usability der ähnlichkeitsbasierten 2D Karten des MM-Browsers.....	77
4.1.1	Feature-Kombinationen und optische Clustering .....	77
4.1.2	Evaluierungsmethodik: Gruppen ähnlicher Samples .....	78
4.1.3	Feature-Kombinationen der Timbre Toolbox.....	80
4.1.4	Feature-Kombinationen der MIR Toolbox.....	83
4.1.5	Feature-Kombinationen des Essentia Frameworks.....	88
4.1.6	Auswahl von Schlagzeug-Samples .....	89
4.1.7	Farbkodierung der Sample-Punkte.....	94
4.1.8	Semantische Struktur von ähnlichkeitsbasierten 2D Karten.....	95
4.2	Algorithmische Ähnlichkeit und subjektives Ähnlichkeitsempfinden .....	96
4.3	Sonifikation.....	97
4.4	Zusammenfassung.....	99
5	Zusammenfassung und Ausblick .....	100
6	Literaturverzeichnis.....	103
7	Abbildungsverzeichnis.....	110
8	Anhang.....	111
8.1	Auswertungstabellen der getesteten Feature-Kombinationen .....	111
8.1.1	Tabelle „Drums“ .....	111
8.1.2	Tabelle „ESS_feat_vs_mfcc“ .....	112
8.1.3	Tabelle „MIR_deltaMFCC_vs_MFCC“ .....	114
8.1.4	Tabelle „MIR_mfcc_sc_lat_allStats“.....	115
8.1.5	Tabelle „MIR_timbre_allstats“ .....	116
8.1.6	Tabelle „TT_independent_vs_mfcc“ .....	117
8.1.7	Tabelle „TT_lat_sc_vs_mfcc“ .....	118
8.1.8	Tabelle „TT_max_pca_vs_mfcc“.....	119
8.1.9	Tabelle „Vergleich MIR_max vs MIR_max_pca“ .....	120
8.1.10	Tabelle „Vergleich TT_max vs TT_max_pca“ .....	121
8.2	GäS -Sammlungen .....	122
8.2.1	BD 1 .....	122
8.2.2	BD 2 .....	123

8.2.3	FSD 1 .....	124
8.2.4	GSD 1 .....	125
8.2.5	HH 1 .....	126
8.2.6	M 1.....	127
8.2.7	M 2.....	128
8.2.8	P 1.....	129
8.2.9	SD 1.....	130
8.2.10	SD 2.....	131

# 1 Einleitung

## 1.1 Motivation und Problemstellung

Mit dem Begriff „Sample“ werden in dieser Arbeit digitale, maximal mehrere Sekunden lange Audio-Aufnahmen von Klängen und Geräuschen bezeichnet. Für das Durchsuchen großer Sample-Sammlungen nach Samples, die sich für einen bestimmten Einsatzzweck eignen, wird der Begriff „Sample-Browsing“ verwendet.

Für Musiker, DJs, Remix-Künstler, Gamesound-Designer sowie für Toningenieure, die sich mit der Vertonung von Filmen und Videos beschäftigen, ist Sample-Browsing ein Bestandteil des täglichen Arbeitsablaufes.

Neben der Klangsynthese ist das Verwenden von Samples das derzeit vorherrschende Paradigma in der Musikproduktion. Samples sind die Grundbausteine vieler Musikgenres. Einige zeitgenössische Musikgenres (z.B. HipHop, Techno) bestehen fast ausschließlich aus Samples. Musikerinnen setzen Rhythmen aus einzelnen Drum-Samples zusammen. Sie verwenden Sample-Loops um Melodien und Rhythmen zu erschaffen. Anstatt einzelne Instrumente selbst aufzunehmen setzen sie Samples des jeweiligen Instruments ein. Samples werden auch für die Vertonung von Filmen oder Videos verwendet. Steht eine Toningenieurin vor der Aufgabe eine Szene zu vertonen stehen ihr für jedes erdenkliche Geräusch eigene Sample-Sammlungen zur Verfügung. Mit der globalen Verbreitung von Videospiele ist „Gamesound-Design“ zu einem eigenen Berufsbild geworden. Viele Universitäten und private Bildungseinrichtungen bieten bereits Lehrgänge für die Ausbildung zum „Gamesound-Designer“ an. Einige Konferenzen beschäftigen sich ausschließlich mit diesem Thema. Sample-Browsing ist ein unumgänglicher Teil des täglichen Arbeitsablaufs dieser relativ neuartigen Berufsgruppe da Gamesound-Designer den Soundtrack eines Videospiele aus vielen verschiedenen Samples zusammensetzen.

In der Praxis arbeiten die genannten Berufsgruppen mit sogenannten DAWs (Digital Audio Workstation). Die Bereiche der graphischen Benutzeroberfläche von DAWs, die die Anwenderin bei ihrer Samplesuche unterstützen sollen, haben sich seit zwei Jahrzehnten kaum verändert. Abbildung 1 zeigt exemplarisch einen solchen Sample-Browsing Bereich einer DAW, den Avid Pro Tools Workspace Browser. Fast alle professionellen DAWs bieten der Anwenderin nur eine listenbasierte Benutzeroberfläche und eine Stichwortsuche um nach geeigneten Samples zu suchen. In den meisten Fällen befindet sich am linken oder rechten Rand der DAW-Benutzeroberfläche eine alphabetisch sortierte Liste der Samples. Der Umfang der Liste kann durch eine Stichwortsuche eingeschränkt werden. Für die Stichwortsuche wird entweder der Dateiname oder sogenannte „Tags“ (als Metadaten gespeicherte, semantische Beschreibungen der einzelnen Samples) verwendet. (vgl. die Beschreibung von DAWs in [1])

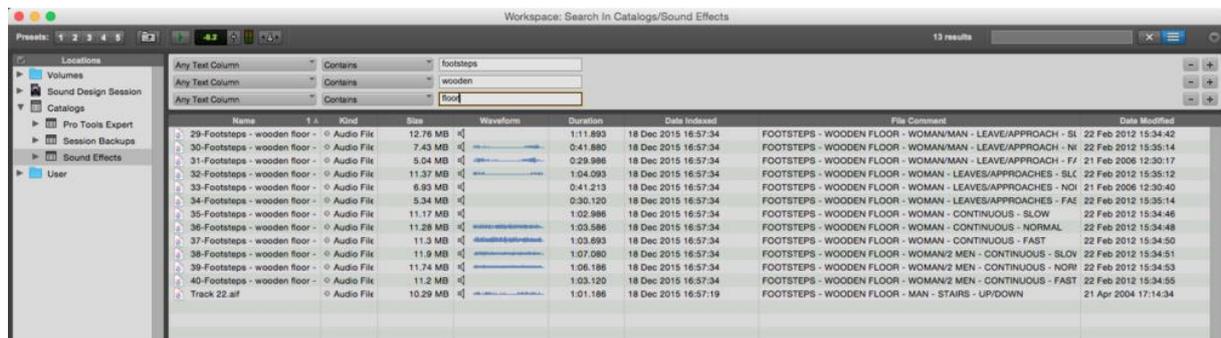


Abbildung 1: Avid Pro Tools Workspace Browser, als Beispiel für ein listenbasiertes Interface einer DAW zur Samplesuche

Neben den Sample-Browsing Interfaces von DAWs verwenden die genannten Berufsgruppen bei ihrer Sample-Suche traditioneller Weise den Dateibrowser des jeweiligen Betriebssystems. Die Suche im Dateibrowser bedeutet, dass jedes Sample einzeln angeklickt und auch einzeln angehört werden muss. Eine kleine Anzahl von Samples auf diese Art und Weise auf ihre Eignung zu prüfen ist vielleicht noch praktikabel. Bei hunderten oder tausenden von Samples wird das „Durchklicken“ von Dateilisten aber zu langwierig und der Vergleich zwischen unterschiedlichen Samples schwierig. (vgl. [2])

Während sich die Methoden und Benutzeroberflächen mit denen die genannten Berufsgruppen ihre Samplesuche durchführen kaum verändert haben, ist der Umfang der zur Verfügung stehenden Sample-Datenbanken in den letzten 15 Jahren exponentiell gestiegen.(vgl. [3]) Die Suche nach geeigneten Samples ist damit von der (von Woods et al. in [4] beschriebenen) „Data Overload“ Problematik betroffen: „... more and more data is available in principle, but our ability to interpret what is available has not increased. On one hand, all participants in a field of activity recognise that having greater access to data is a benefit in principle. On the other hand, these same participants recognise how the flood of available data challenges their ability to find what is informative or meaningful for their goals and tasks.“ [4]

Am deutlichsten zeigt sich diese Problematik im Bereich der Musikproduktion. Das von der EU geförderte Forschungsprogramm Giant Steps „aims to create musical tools that provide intuitive and meaningful interfaces to complex selections of sounds and musical data through expert agents, music analysis algorithms and new interfaces.“ [5] Im Rahmen des Projekts wurden mit 16 „expert users“ [5] semistrukturierte Interviews durchgeführt. Die Interviewpartner wurden unter anderem nach ihren Wünschen und Erwartungen bezüglich zukünftiger, digitaler Musikproduktionstools befragt. Die meisten Interviewpartner wünschen sich neuartige Werkzeuge die sie bei ihrer Samplesuche unterstützen. (vgl. [5]). Ihre derzeitige Strategie bei der Suche nach Samples beschreiben die befragten Musiker als langwierigen und eher zufallsgesteuerten Suchprozess: „... I am just going through all these sound files, and I am just waiting for the sound which I had in mind to suddenly appear...“ [5], „You just click randomly and just scrolling, it takes for ever!“ [5]

Musikerinnen stehen immer größere Sample-Sammlungen, sogenannte „Sample-Packs“ zur Verfügung. Sample-Packs werden zumeist als digitaler Download vertrieben. Das Angebot an frei verfügbaren und kommerziellen Sample-Packs ist unüberschaubar groß geworden. Das Online Gratis-Angebot an Sample-Packs ermöglicht es Musikerinnen sich innerhalb von 1-2 Stunden Sample-Datenbanken zu erstellen, die pro Instrument weit über 1000 Samples beinhalten (z.B. 1000 Snaredrum-Samples). Professionelle, kommerzielle Sample-Packs beinhalten teilweise 10 000 Samples eines einzelnen Instruments (z.B. 10 000 Bassdrum-Samples). Audiosuchmaschinen und Sample-Sharingplattformen erweitern das Sample-Angebot zusätzlich.

Die moderne Musikerin erlebt einen „information overflow“. Ihr stehen für ein einzelnes Instrument tausende Samples zur Verfügung. Sie steht vor der Aufgabe aus über z.B. 10 000 Bassdrum-Samples, das Sample auszuwählen das ihren künstlerischen Vorstellungen am ehesten entspricht. Welches der 10 000 Samples erzielt den besten ästhetischen Effekt? Das kreative Potential derartig großer Sample-Packs kann weder mit einem listenbasierten Interface noch mit einer Stichwortsuche ausgeschöpft werden. Sich 10 000 Aufnahmen eines einzelnen Instruments hintereinander anzuhören erscheint aufgrund der dazu benötigten Zeit unmöglich. Stichwortverzeichnisse, die alle relevanten, klanglichen Dimensionen von Samples beschreiben, existieren derzeit nicht. Es ist also nicht übertrieben wenn Gomez et al., 2016 feststellen: „Nowadays, one of the most well identified bottlenecks to creativity in digital music production is the difficulty of finding suitable sound samples.“ [6]

Sound-Designer, die sich mit der Vertonung von Filmen und Videos beschäftigen, stehen vor ähnlichen Herausforderungen. Auch dieser Berufsgruppe steht ein noch nie dagewesenes Angebot an frei

verfügbaren und kommerziellen Sample-Packs zur Verfügung. „The number and size of commercially available sound effects databases, such as Hollywood Edge, Sound Ideas, ..., and of community-driven on-line collections like freesound are growing steadily, with rising network bandwidth, growing harddisk capacity, and falling prices for storage and distribution media accelerating this growth even further.” [7]

Will eine Sound-Designerin eine Szene vertonen, in der die Meeresbrandung zusehen ist, sucht sie Samples mit den Geräuschen von Wellen. Das Gratis- Online Angebot ermöglichen der Sound-Designerin innerhalb kürzester Zeit eine Sample-Sammlung mit hunderten Aufnahmen von Wellen zusammenzustellen. Kommerzielle Sample-Packs bieten ihr ebenfalls hunderte Wellen-Samples. Dasselbe gilt für die meisten, in Filmen und Videos vorkommenden Geräusche. Für Wind- und Wettergeräusche, Aufnahmen von Tieren, Fahrzeuge jeder Art, Boote, Haushaltsgeräusche, Waffen, Explosionen, Schritte, Lachen, Applaus, Alarmsignale, usw. stehen der Sound-Designerin Sample-Packs mit mehreren hundert Aufnahmen pro Geräuschkategorie (z.B. Wellen, Türgeräusche, Schreie) zur Verfügung. Auch für dieses Anwendungsszenario sind listenbasierte Benutzeroberflächen und eine Stichwortsuche keine idealen Werkzeuge um eine Sample-Suche durchzuführen.

Professionelle, kommerzielle Sample-Packs für die Film-Vertonung werden häufig mit Stichwortverzeichnissen geliefert, die pro Samples ca. 2 -10 Tags (Stichwörter) und eine textuelle Beschreibung enthalten. Diese Sample-Packs bestehen zumeist aus tausenden von Samples aus einer Geräuschkategorie z.B. Explosionen, Wassergeräusche, Schritte usw. Die Tags geben meistens Auskunft über das Objekt, das das Geräusch erzeugt, die Quelle des Geräuschs, das verwendete Material um das Geräusch zu erzeugen, die Handlung, die das Geräusch erzeugt oder den Ort der Aufnahme. Hinweise auf die akustischen Eigenschaften oder die emotionale Wirkung eines Samples bieten diese Verzeichnisse aber nicht. Diese Stichwortverzeichnisse ermöglichen nur ein Einschränken der in Frage kommenden Samples, eine zielgerichtete Auswahl eines Samples nur mit einer Stichwortsuche ist nicht möglich. Mit Bezug auf solche, mit Tags versehenen Sample-Sammlungen, halten Dupont et al. in [2] fest: „When it comes to looking for sounds in such collections, for each query, once successive keywords helped the user to filter down the results, but attained a limit, hundreds of sounds are left to be reviewed.“ [2] Verwendet die Anwenderin gratis angebotene Geräusch-Samplesammlungen, die in den meisten Fällen nicht mit einem Stichwortverzeichnis versehen sind, steht sie vor der gleichen Situation. Nachdem die Filmsound-Designerin Online nach geeigneten, gratis Samples gesucht hat, hat sie hunderte Aufnahmen von z.B. Wellen zur Verfügung. Welches der Samples am besten für die Vertonung geeignet ist kann die Sound-Designerin nur feststellen indem sie sich alle Samples anhört.

Auch im Bereich der Film- und Videovertonung gilt der Befund, dass sich das kreative Potential der zur Verfügung stehenden Sample-Sammlungen sich mit der derzeit gebräuchlichen Kombination aus Stichwortsuche und Ergebnisliste nicht ausschöpfen lässt. Eine zielgerichtete Suche nach Samples mit bestimmten akustischen Eigenschaften ist mit den verfügbaren Tags nicht möglich. Sich alle in Frage kommenden Samples der Reihe nach anzuhören ist zu zeitaufwendig.

Gamesound-Designer stehen bei der Auswahl von Samples für ein Videospiel vor den gleichen Problemen. Sie verwenden dieselben Sample-Sammlungen wie Sound-Designer für die Film- und Videoproduktion. Zusätzlich gibt es ein riesiges Angebot an Sample-packs mit Gamesound-Samples.

## 1.2 Praxisnahes Anwendungsszenario für Sample-Browser

Die im vorherigen Abschnitt skizzierten Anwendungsszenarien (Auswahl eines Bass Drum Samples aus tausenden Bassdrum-Samples, Auswahl eines Wellengeräusches aus hunderten Wellengeräuschen) sind typische Anwendungsszenarien für Sample-Browser in der Musikproduktion bzw. der Film -und Videovertonung. Beide Beispiele sind dadurch gekennzeichnet dass die Anwenderin ein passendes Sample aus hunderten oder tausenden, ähnlicher Samples auswählen muss.

Das Erstellen und Verwalten von Stichwortverzeichnissen erfreut sich unter Musikerinnen keiner großen Beliebtheit. Typischerweise speichern Musikerinnen ihre Sample-Sammlungen auf Festplatten in Ordner und Unterordnern. Als Ordnernamen werden dabei häufig Begriffe verwendet die eine Klasse oder Kategorie von Samples bezeichnen. (Zur Verdeutlichung: Eine Musikerin speichert alle Schlagzeug-Samples in einem Ordner „Drums“, die Unterordner „Snaredrums“, „Basdrums“, „HighHats“, „Toms“, usw. dienen als Speicherorte für Samples des jeweiligen Schlaginstruments) In diesen Unterordnern liegen hunderte bis tausende ähnlicher Samples, für die im günstigsten Fall ein Stichwortverzeichnis mit einigen, wenigen Tags existiert. Meistens sind aber keine Metadaten für die Samples vorhanden und der semantische Gehalt der Dateinamen reicht nicht aus um eine Filterung innerhalb des Unterordners durchzuführen.

Mit dieser Ordner- und Unterordnerstruktur nimmt die Musikerin eine grobe Klassifizierung aller Samples selbst vor. Musikerinnen benötigen Sample-Browser daher weniger um Samples aus unterschiedlichen Kategorien voneinander zu unterscheiden. Neuartige Sample-Browser müssen vor allem die Auswahl aus derselben Sample-Kategorie unterstützen. Die Musikerin benötigt einen Sample-Browser höchstwahrscheinlich nicht um verschiedene Schlaginstrumente voneinander zu unterscheiden, sondern um eine Snare-Drum aus tausenden von Snare-Drums auszuwählen.

Das selbe Argument gilt für die Auswahl von Samples in der Film- und Videoproduktion und im Bereich Gamesound-Design. Für professionelle Sample-Packs existieren Stichwortverzeichnisse die eine grobe Vorauswahl der Samples ermöglichen. Für ihre eigenen Sample-Sammlungen verwenden Filmsound-Designer oder Gamesound-Designer ähnliche Ordnerstrukturen wie Musikerinnen oder verwalten und erstellen kleinere Stichwortverzeichnisse in Eigenregie. Diese beiden Berufsgruppen benötigen einen Sample Browser nicht dazu um Windgeräusche von Wellengeräuschen zu unterscheiden. Sie benötigen ein Programm das ihnen hilft das passende Wellengeräusch aus hunderten von Wellengeräuschen auszuwählen. Sample-Browser müssen daher vor allem die Auswahl eines Samples aus hunderten bis tausenden Samples derselben Kategorie unterstützen.

### 1.3 Umfangreiche Stichwortverzeichnisse als Sample-Browsing Werkzeuge

Derzeit existieren im günstigsten Fall für professionelle Sample-Packs Stichwortverzeichnisse mit 2- 10 Tags, die ein Einschränken der in Frage kommenden Samples ermöglichen. Ein Durchsuchen einer Sample-Sammlung nach klanglichen oder akustischen Eigenschaften ermöglichen diese Stichwortverzeichnisse aber nicht. Sind umfangreichere Stichwortverzeichnisse, die die wesentlichen klanglichen und akustischen Eigenschaften der Samples mit einer Vielzahl von Tags beschreiben, eine Lösung?

In [8] gehen Cano et al. der Frage nach wie umfangreich ein Stichwortverzeichnis sein muss, das eine Suche in professionellen Sample-Datenbanken für die Film- und Videoproduktion ermöglicht. Die Autoren schätzen das mehrere tausend verschiedene Stichwörter notwendig sind „...that account the level of detail needed in a production-size sound effect management system...“ [8]. Die Verwendung von domainspezifischen Ontologien oder Taxonomien ermöglicht vielleicht die Reduzierung der benötigten Tags für andere Anwendungsgebiete. Es ist aber unwahrscheinlich das die wesentlichen klanglichen und akustischen Eigenschaften eines Samples mit weniger als hundert unterschiedlichen Tags ausreichend genau beschrieben werden können.

Die manuelle Erstellung und Verwaltung derartiger Stichwortverzeichnisse scheitert damit aus mehreren Gründen aufgrund der Größe der verfügbaren Sample-Sammlungen. Für die Anbieter von Sample-Packs bedeutet die manuelle Erstellung eines solchen Verzeichnisses einen enormen Aufwand. Bestens geschultes Personal müsste pro Sample hunderte Stichwörter in eine Datenbank eingeben. Es ist daher aus wirtschaftlichen Gründen unwahrscheinlich das die Anbieter von Sample Packs derartige Stichwortverzeichnisse erstellen. Ebenfalls unwahrscheinlich ist, dass die genannten Berufsgruppen

sich die Mühe machen ihre persönliche Sample-Sammlung mit einem derartigen Stichwortverzeichnis zu verwalten. Das Hinzufügen eines einzelnen Samples würde die Eingabe hunderter Tags bedeuten. Das Hinzufügen ganzer Sample Packs mit tausenden von Samples hat einen tage- und wochenlang dauernden Verwaltungsaufwand zur Folge um das Stichwortverzeichnis in einen konsistenten Zustand zu bringen.

Fraglich ist auch ob derartige Stichwortverzeichnisse auch effizient genutzt werden können. „However, all available solutions that come with pre-tagged sounds use a static semantic structure, which cannot adapt to the users individual understanding of the sounds. This is especially problematic when the user has a specific target sound in mind, but does not know how this would be described in the pre-tagged semantic structure of the database.“ [5] Um von einem solchen Stichwortverzeichnis effektiv Gebrauch zu machen müsste die Anwenderin genau wissen welcher Tag welche klangliche oder akustische Eigenschaft beschreibt. Die Auswahl eines Snaredrum-Samples würde die Eingabe einer langen Wortkette bedeuten: „snare, rimshot, acoustic, sharp, crisp, short attack, bright, dry,....“ Die Anwenderin müsste sich durch einen langwierigen Lernprozess klarmachen welche akustische oder klangliche Eigenschaft mit welchem Tag beschrieben wird. Sie muss die semantische Struktur des Stichwortverzeichnisses erlernen.

Selbst wenn die Sample-Datenbank mit der größten Sorgfalt mit einem umfangreichen Stichwortkatalog versehen wurde, ist noch nicht garantiert das bei der Suche dieselben Stichwörter wie beim Erstellen der Datenbank verwendet werden. Das Zusammenknüllen einer Plastikfolie klingt wie ein Schritt im Schnee. Ob der Ersteller der Datenbank bei den Stichwörtern dieses Samples „Schnee“ oder „Schritte“ vermerkt ist aber fraglich. Er vermerkt als Stichwort wahrscheinlich „Plastik“, und „Zusammenknüllen“. Der Benützer der Datenbank sucht aber nach „Schnee“ und „Schritte“. (vgl. [9])

#### 1.4 Forschungsfragen

Als Alternative zu listenbasierten Interfaces und Stichwortsuchen sind in den letzten 15 Jahren mehrere Forschungsprototypen entwickelt worden, die die psychoakustische Ähnlichkeit von Samples algorithmisch berechnen. In den meisten Fällen stellen diese Anwendungen einzelne Samples als grafische Objekte auf einer 2D Karte dar. Die Ähnlichkeit der Samples wird durch den Abstand dieser grafischen Objekte visualisiert.

Nach 15 Jahren universitärer Forschung und einer Vielzahl an vorgestellten Forschungsprototypen bleibt der Befund, dass sich dieser Ansatz in der Praxis nicht durchgesetzt hat. Kapitel 4 dieser Arbeit beinhaltet daher eine kritische Analyse dieser Forschungsprototypen. Im selben Kapitel werden die aus dieser Analyse resultierenden Forschungsfragen aufgelistet. In der Analyse wurden drei Problemfelder identifiziert, welche die Benutzerfreundlichkeit dieses Ansatzes beeinflussen könnten:

- Faktoren, die die Benutzerfreundlichkeit von ähnlichkeitsbasierten 2D Karten als Sample-Browsing Werkzeuge beeinträchtigen
- Subjektives, individuelles Ähnlichkeitsempfinden und algorithmisch berechnete Ähnlichkeit
- Sonifikation (Verklanglichung) der 2D Karte

Die Analyse der Forschungsprototypen zeigt, dass die ähnlichkeitsbasierten 2D Karten die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind, unterstützen sollten. Mit einer einzigen Ausnahme erstellen alle in dieser Arbeit behandelten Forschungsprototypen diese Karten mit Hilfe einer algorithmischen Verarbeitungskette aus inhaltsbasierter Merkmalsextraktion und einer Dimensionsreduktion. Die Forschungsfragen des ersten Problemfeldes thematisieren welche optischen Ergebnisse mit dieser

Verarbeitungskette erzielt werden sollten um die drei genannten Interaktionen möglichst optimal zu unterstützen.

Das zweite Problemfeld beinhaltet Forschungsfragen, die thematisieren wie weit die algorithmisch ermittelte Ähnlichkeit der Samples mit dem subjektiven Ähnlichkeitsempfinden der Anwenderinnen übereinstimmt. Eine zentrale Fragestellung dieser Arbeit ist wie genau das individuelle, subjektive Ähnlichkeitsempfinden von Anwenderinnen algorithmisch berechnet werden kann.

Das dritte Problemfeld beinhaltet Fragen zu einem Sonifikations-Konzept, das in einigen Forschungsprototypen verwendet wird. Das Aura Sonifikations-Konzept beruht auf der Vorstellung, dass Anwenderinnen in der Lage sind ihre auditive Aufmerksamkeit gezielt auf einzelne Samples zu lenken, während gleichzeitig mehrere andere Samples abgespielt werden. In einem Mix aus mehreren Samples treten aber Maskierungseffekte auf und die Fähigkeit von Anwenderinnen einzelne Samples in einem Mix aus mehreren Samples überhaupt zu erkennen nimmt mit einer relativ geringen Steigerung der gleichzeitig abgespielten Samples dramatisch ab. Die zentrale Fragestellung des dritten Problemfeldes ist daher ob das Aura Sonifikations-Konzept trotz sinkender Erkennungsraten und Maskierungseffekten, die beim Abspielen mehrerer Samples auftreten, einen praktischen Nutzen hat.

## 1.5 Methodisches Vorgehen

Ziel dieser Arbeit ist die Entwicklung eines Sample-Browsers. Der erste Schritt im Designprozess des Sample-Browsers war die Erstellung einer Literaturübersicht über einige Forschungsprototypen und deren Analyse. Semistrukturierte Experteninterviews und Beobachtungen lieferten zusätzliche Hinweise für das Design des Sample-Browsers.

Die Anwendung wurde in einem partizipativen Designprozess mit zwei Personen aus den genannten Berufsgruppen entwickelt. Als Entwicklungsmethodik wurde „Rapid Application Development“ [10] verwendet. Nach der gemeinsamen Anforderungserhebung wurden mit Skizzen und „low fidelity“ mock ups erste Designkonzepte entworfen. Danach wurden mehrere vertikale, „high fidelity“ Prototypen entwickelt und qualitativ evaluiert. Diese Prototypen waren entweder die Umsetzung der Designvorschläge der beiden Anwender oder es wurden Prototypen entwickelt, die dazu dienten die Forschungsfragen zu beantworten, die sich aus der Analyse der Forschungsprototypen ergaben. Nach mehreren Designiterationen (=Prototypenentwicklung und Evaluierung durch die beiden Anwender) wurden die gewonnen Erkenntnisse in ein Gesamtsystem integriert.

Die Evaluierung des Sample-Browsers erfolgte durch eine qualitative Beobachtung unter Anwendung der „think aloud“ [11] Methode. Beobachtet wurden 2 Personen aus den genannten Berufsgruppen während einer explorativen Suche.

## 1.6 Struktur der Arbeit

Im Kapitel „*Stand der Technik*“ werden zuerst einige Sample-Browser Forschungsprototypen beschrieben. Anschließend erfolgt im selben Kapitel eine Darstellung der technologischen Grundlagen dieser Sample-Browser. Alle vorgestellten Sample-Browser verwenden Algorithmen aus den Forschungsbereichen Music Information Retrieval(MIR) und Machine Learning. Die verwendeten MIR Algorithmen werden im Abschnitt „*Inhaltsbasierte Merkmalsextraktion*“ besprochen. Eine Darstellung der verwendeten Machine Learning Algorithmen erfolgt im Abschnitt „*Dimensionsreduktion*“

Im Kapitel „*Analyse der ähnlichkeitsbasierten Sample-Browser*“ werden in getrennten Abschnitten drei Problemfelder identifiziert, die die Benutzerfreundlichkeit der im vorherigen Kapitel beschriebenen Sample-Browser beeinträchtigen könnten. Zu jedem dieser drei Problemfelder werden in den jeweiligen Abschnitten Forschungsfragen formuliert, die durch einen partizipativen Designprozess geklärt werden sollen.

Im Kapitel „*Implementierung und Evaluierung des MM-Browsers*“ werden Erkenntnisse zu den im vorhergehenden Kapitel formulierten Forschungsfragen dargestellt. Neben einer Analyse welche inhaltsbasierten Features für eine bestimmte Art von Samples geeignet sind und einem qualitativen Vergleich von drei Sonifikations-Konzepten wird in diesem Kapitel die Frage erörtert in wie weit die algorithmisch ermittelte Ähnlichkeit von Audio-Samples mit dem Ähnlichkeitsempfinden von Anwenderinnen übereinstimmt.

## 2 Stand der Technik

### 2.1 Ähnlichkeitsbasierte Sample-Browser

Seit ca. 17 Jahren sind immer wieder Alternativen zu listenbasierten Benutzeroberflächen und einer Stichwortsuche entwickelt worden. Bereits im Jahr 1998 wird der „Sonic Browser“ [12] vorgestellt. „MARSYAS3D“ [13] wird 2001 veröffentlicht. Die grundlegenden Designkonzepte beider Anwendungen sind in einer Vielzahl von später entwickelten Sample-Browsern erkennbar. 2002 wird der „SonicBrowser“ mit dem „Marsyas“ Framework [14] erweitert zum „Audio Retrieval Browser“ [15]. Vom 2008 vorgestellten Sample-Browser „Soundtorch“ [16] wurde eine (nicht mehr erhältliche) Public Beta Version zum Download angeboten. Im selben Jahr präsentieren Schwarz et al. den „Sound Navigator“ [17]. 2009 wird mit AudioCycle [18] ein Sample-Browser für Sample-Loops vorgestellt. Dupont et al. [19] entwickelten 2012 einen Sample-Browser der zusätzlich zu einem ähnlichkeitsbasierten Interface eine „Query by Example“ Samplesuche ermöglicht. Ebenfalls 2012 zeigen Grill und Flexer [20] mit ihrem Prototypen neue Wege der Visualisierung von klanglichen Eigenschaften auf. „AudioQuilt“ [21] und „AudioMetro“ [2] wurden 2014 präsentiert. „Drumspace“ [6], ein 2016 vorgestellter Samplebrowser, zeigt wie ähnlichkeitsbasierte 2D Karten verwendet werden können um Musiker bei der Auswahl von Schlagzeugsamples zu unterstützen.

Die genannten Prototypen weisen Gemeinsamkeiten auf, sie haben konzeptionell den selben Aufbau. Man kann von einem Designtypus sprechen, der im restlichen Text mit „ähnlichkeitsbasierter Sample-Browser“ bezeichnet wird.

Die im Folgenden beschriebenen, ähnlichkeitsbasierten Sample-Browser stellen einzelne Samples als grafische Objekte (z.B. Kreis) in einer 2-dimensionalen Karte dar. Auf dieser Karte entspricht der Abstand der grafischen Objekte der psychoakustischen Ähnlichkeit der Samples. Nahe beieinanderliegende Samples weisen eine große akustische Ähnlichkeit auf, je grösser die Distanz desto unähnlicher sind die Samples.

Als ersten Verarbeitungsschritt führen ähnlichkeitsbasierte Sample-Browser eine inhaltsbasierte Merkmalsextraktion (Contentbased Feature Extraction) durch, d.h. es werden einige Kennzahlen (=Features) direkt aus den digitalen Daten eines Audiofiles (=Contentbased) berechnet. Die Anzahl der verwendeten Features variiert dabei von Browser zu Browser. Üblicherweise werden pro Audio-Sample zwischen wenigen Dutzend und einigen Hundert Kennzahlen berechnet. Die Anzahl der berechneten Kennzahlen wird im weiteren Text mit  $n$  bezeichnet. Ist die Feature Extraction für alle Samples einer Sample-Sammlung abgeschlossen steht pro Sample ein  $n$ -dimensionaler Feature-Vektor zur Verfügung, der bestimmte akustische und klangliche Eigenschaften des Samples in  $n$  Zahlen abbildet.

Danach wenden ähnlichkeitsbasierte Samplebrowser verschiedene „machine learning“ Algorithmen auf die Menge der Feature-Vektoren an. Diese Algorithmen ermöglichen die ähnlichkeitsbasierten 2D Karten. Sie reduzieren den  $n$ -dimensionalen Raum der Feature-Vektoren auf einen 2-dimensionalen Raum. Bei dieser Dimensionsreduktion wird die Ähnlichkeitsstruktur des  $n$ -dimensionalen Feature - Raums auf der 2D Karte abgebildet. Ist die Distanz im  $n$ -dimensionalen Feature-Raum zwischen zwei Samples gering so ist die Distanz zwischen diesen beiden Samples in der 2-dimensionalen Darstellung klein. Umgekehrt gilt, dass eine große Distanz zwischen zwei Samples im  $n$ -dimensionalen Raum auch zu einer großen Distanz in der 2D Darstellung führt.

Nach der Dimensionsreduktion stellen ähnlichkeitsbasierte Sample-Browser einzelne Samples als grafische Objekte auf 2D Karten dar. Der Abstand der Sample-Objekte visualisiert die Ähnlichkeit der Samples. Farbe, Größe, Form der Sample-Objekte visualisieren häufig akustische Eigenschaften oder eine Gruppenzugehörigkeit der Samples.

Ein weiteres gemeinsames Merkmal aller ähnlichkeitsbasierten Sample-Browser ist die Sonifikation (Verklanglichung) der 2D Karte. In den meisten Fällen wird das zu einem grafischen Objekt auf der Karte gehörige Sample abgespielt sobald die Anwenderin den Mauscursor über das Objekt oder in die Nähe des Objekts bewegt.

### 2.1.1 Sonic Browser [12]

Im Userinterface des Sonic Browsers (siehe Abbildung 2) werden Samples als kleine, grafische Objekte in einem 2 Achsendiagramm („star-field display“ [12]) dargestellt. Größe, Farbe, Form oder Position dieser grafischen Objekte zeigen bestimmte Eigenschaften der Samples an (z.B. die Dateigröße, die Samplingrate, den Dateityp, die Dauer oder das Aufnahmedatum). Die Anwenderin kann frei entscheiden welche Sample-Eigenschaft mit welchem grafischem Parameter visualisiert wird. „As an example, file size can be mapped to size of visual symbols, sampling rates to colour, symbol shape to file type and horizontal and vertical location to date and time.“ [12]

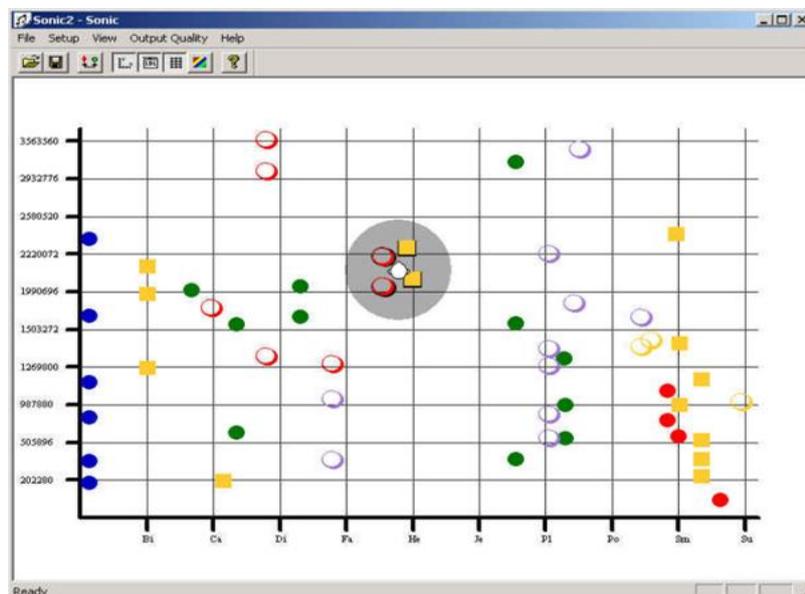


Abbildung 2: Sonic Browser, Aura als grauer, semitransparenter Kreis dargestellt, alle Samples innerhalb der Aura werden gleichzeitig abgespielt

Die Evaluierung des Sonic Browsers als Sample-Browser für Umgebungsgeräusche wird in [12] beschrieben. Die Autoren vergleichen in dieser Studie den Windows Explorer mit dem Sonic Browser. Das Ziel der Studie ist die Unterschiede zwischen „single versus multiple stream audio browsing.“ [12] zu erfassen. Der Windows Explorer ermöglicht Anwenderinnen nur ein „single stream audio browsing“. Jede Audiodatei kann nur einzeln angehört werden und muss mit einem Doppelklick gestartet werden. Der Sonic Browser ermöglicht den Anwenderinnen durch sein „aura“ [12] Design-Konzept mehrere Sounds ohne Anklicken gleichzeitig abzuspielen (= „multiple stream audio Browsing“)

Grafisch wird die Aura als ein grauer, semitransparenter Kreis rund um den Mauscursor dargestellt. Die Anwenderin bewegt die Aura über das 2 Achsen Diagramm in dem alle Samples durch kleine grafische Objekte dargestellt werden. Die Größe der Aura ist frei wählbar. Alle Audiosamples die innerhalb der Aura liegen werden auf den Lautsprechern in einem Stereopanorama ausgegeben. Die Lautstärke der einzelnen Samples wird durch den Abstand zum Mittelpunkt der Aura bestimmt. Je näher ein Sample zum Mittelpunkt liegt desto lauter wird es abgespielt.

Die Entwickler des Sonic Browsers gehen davon aus, dass die Anwenderin in der Lage ist mehrere gleichzeitige abgespielte Samples zu unterscheiden und sich selektiv auf einen dieser Sounds zu konzentrieren. Sie berufen sich dabei auf den „cocktail party effect“ [12] , „our ability to switch our

attention between different sounds in the auditory scene" [12] und vergleichen das Anhören von mehreren Sounds im Diagramm des Sonic Browsers mit der akustischen Wahrnehmung in einer Alltagssituation: „In everyday listening one is often exposed to hundreds of different sounds simultaneously and is still able to pick out important parts of the auditory scene" [12]

Für die Evaluierung wurden insgesamt sechs Anwenderinnen auf zwei Gruppen aufgeteilt. Die erste Gruppe verwendet den Windows Explorer, die zweite Gruppe den Sonic Browser. Aufgabe der Studienteilnehmer war es Sounds zu finden die einer kurzen textlichen Beschreibung am besten entsprechen. Für die Studie wurde eine relative kleine Datenbank mit 244 Aufnahmen verwendet, die einmal mit langen deskriptiven Dateinamen (z.B. "dog barking", "car starting", "ball bouncing" [12]) und einmal mit kurzen, semantisch nicht aufschlussreichen Dateinamen versehen wurde. Für die Beobachtungen wurde die „Thinking Aloud“ Methode verwendet. Im Anschluss an die Beobachtung wurde ein Likert Skala Fragebogen verwendet um die Beurteilungen der beiden getesteten Interfaces zu erfassen. Die wesentlichen Ergebnisse der Beobachtung und der Befragung werden im Folgenden zusammengefasst.

Die Beobachtung während der Tests zeigen, dass die 2D Karte und das Aura Konzept einen rascheren Zugriff als der listenbasierte Windows Explorer ermöglichen, bei dem die Sounds einzelnen angeklickt werden müssen.

Die kumulierten Ergebnisse der Fragebögen zeigen, dass falls lange, deskriptive Dateinamen vorhanden sind, es den Teilnehmerinnen mit dem Windows Explorer leichter fällt geeignete Sounds zu finden. Fehlen deskriptive Dateinamen fällt es den Anwenderinnen leichter mit dem Sonic Browser passende Sounds zu finden. Die Autoren weisen darauf hin das lange, deskriptive Dateinamen in der Praxis selten vorhanden sind: "Commercial data sets with sound resources most often use cryptic filenames..." [12]

Für den Sonic Browser wünschen sich die Anwenderinnen eine Zoomfunktionalität und eine „display by category“ [12] Funktionalität. Der Wunsch nach einer Zoomfunktionalität entsteht durch Cluster im Diagramm des Sonic Browsers. Befinden sich zu viele grafische Objekte auf zu engem Raum wird das Ansteuern der einzelnen Samples schwierig. Die Zoomfunktionalität soll die Navigation in Clustern erleichtern. Mit der „display by category“ [12] Funktionalität wünschen sich die Userinnen die Möglichkeit Samples nach bestimmten Kriterien auf der Karte ein- und auszublenden.

Für beide getesteten Interfaces gilt, dass sich die Anwenderinnen eine Markierungsmöglichkeit für interessante bzw. passende Sounds wünschen. Diese „files of interest“ [12] hören sich die Userinnen meistens mehrmals hintereinander an und würden sie gerne im Anschluss als „interessant“ oder als „bereits gehört“ markieren.

### 2.1.2 MARSYAS3D [13]

MARSYAS3D ist ein an der Universität von Princeton entwickelter "... prototype audio browser and editor designed for working with large collections of audio files." [13] und bietet der Anwenderin mehrere, unterschiedliche Benutzeroberflächen für eine Samplesuche. Die Entwickler nehmen explizit Bezug auf den Sonic Browser als Inspirationsquelle für ihr Design. Wesentliche Innovation gegenüber dem Sonic Browser ist, dass einige dieser Benutzeroberflächen die Ergebnisse einer inhaltsbasierten Merkmalsextraktion und Dimensionsreduktion anzeigen.

Ähnlich wie der Sonic Browser stellen zwei dieser Oberflächen, TimbreSpace2D und TimbreSpace3D (siehe Abbildung 3), jedes Sample als grafisches Objekt auf einer 2D bzw. 3D Karte dar. TimbreSpace3D unterstützt Skalierungs-, Rotations- und Zoominteraktionen. Das „Marsyas“ [14] Audioanalyse Framework wird verwendet um für jedes Sample einen multidimensionalen Feature Vektor zu berechnen, für die Dimensionsreduktion wird eine PCA (Principal Component Analysis) verwendet. Als

Features werden unter anderem Spectral Centroid, Spectral Flux, Spectrol Rolloff, RMS Energie, Linear Prediction Coefficients(LPC) und Mel Frequency Cepstral Coefficients(MFCC) berechnet. Diese Features werden auch für ein algorithmisches Clustering aller Samples (= jeder Feature Vektor wird einer bestimmten Gruppe zugeordnet) verwendet. Jedes Sample auf den 2D bzw. 3D Karten kann dann je nach Gruppenzugehörigkeit farbcodiert dargestellt werden. Die Autoren stellen fest, dass die Kombination aus Feature Extraction, PCA und Clustering der Anwenderin eine ähnlichkeitsbasierte Sample-Suche ermöglicht: „The TimbreSpace reveals similarity and clustering of sounds.“ [22]

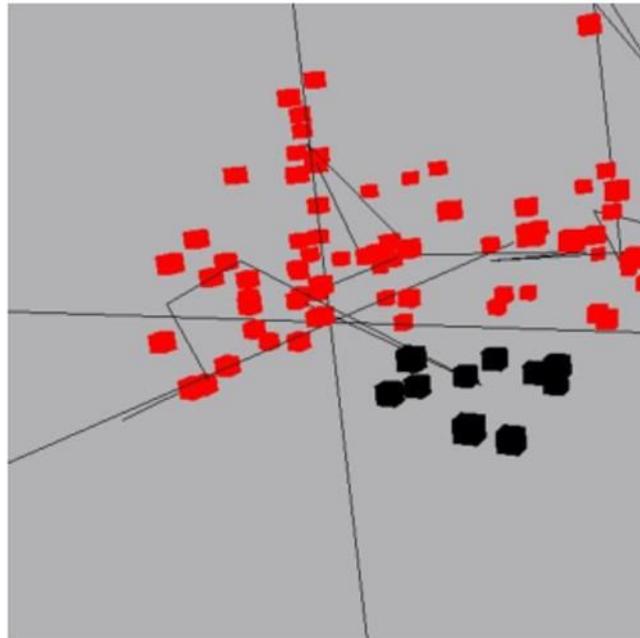


Abbildung 3: TimbreSpace3D, Farbkodierung je nach Clusterzugehörigkeit

Soundspace, eine weitere graphische Benutzeroberfläche von MARSYAS3D, bietet der Anwenderin eine Sonifikation, die der „aura“ Funktionalität des Sonic Browser ähnelt: „As the user browses, a neighborhood (aura) around the currently selected item is auralized by playing all the corresponding files simultaneously“ [13] Auch die Entwickler von MARSYAS3D verweisen dabei auch auf den Cocktail Party Effekt. Sie halten aber fest, dass die selektive Wahrnehmung eines einzelnen Samples in einem Mix mehrerer, gleichzeitig abgespielter Samples nur begrenzt möglich ist: „...we have found that at most 8 simultaneous audio streams are useful“ [13].

Eine formale Evaluierung oder eine qualitative Beschreibung wie Userinnen MARSYAS3D für eine Samplesuche einsetzen ist nicht veröffentlicht worden.

### 2.1.3 ARB – Audio Retrieval Browser [15]

Der Audio Retrieval Browser ist eine Kombination aus dem Sonic Browser [12] und dem Marsyas Framework [14]. „These two systems were presented at ICAD 2001, where the authors discovered the common goals and complementary approaches of the two projects and decided to combine them in one system“ [15]. Die erklärte Intension der Entwickler des ARBs ist, die zum damaligen Zeitpunkt erst kürzlich entwickelten Algorithmen des Forschungsbereichs „audio information retrieval“ [15] für das Durchsuchen von großen Audio-Datenbanken zu nützen.

Der ARB stellt wie der Sonic Browser jedes Sample als grafisches Objekt auf einer 2D Karte dar. Als Sonifikation wird das „aura“ Konzept des Sonic Browsers verwendet. Das Marsyas Framework wird für die Contentbased Feature Extraction eingesetzt um pro Sample einen multidimensionalen Feature

Vektor zu berechnen. Auf diese Feature Vektoren werden dann „standard pattern recognition techniques“ [15] wie eine automatische Klassifikation (Clustering) oder eine Dimensionsreduktion angewandt.

Der ARB ermöglicht (wie der Sonic Browser) der Anwenderin eine Zuweisung von Sample Eigenschaften auf verschiedene graphische Elemente der Benutzeroberfläche. Zusätzlich zu den bereits im Abschnitt über den Sonic Browser erwähnten Zuweisungen hat die Anwenderin die Möglichkeit die Ergebnisse mehrerer MIR (Music Information Retrieval) Algorithmen den grafischen Eigenschaften der Sample-Objekte oder den beiden Achsen einer 2D Karte zuzuweisen. Neben dieser frei konfigurierbaren 2 Achsen Darstellung bietet der ARB als Visualisierung eine ähnlichkeitsbasierte 2D oder 3D Karte, die durch die Anwendung einer PCA (= Dimensionsreduktion) auf die Feature Vektoren erstellt wird. Zusätzlich wurden im Rahmen des ARB Projekts eine Treemap [23] und Hyperbolic Tree [24] als alternative Visualisierungen implementiert. Abbildung 4 zeigt diese beiden Benutzeroberflächen. Eine Evaluierung des ARB ist nicht veröffentlicht worden.

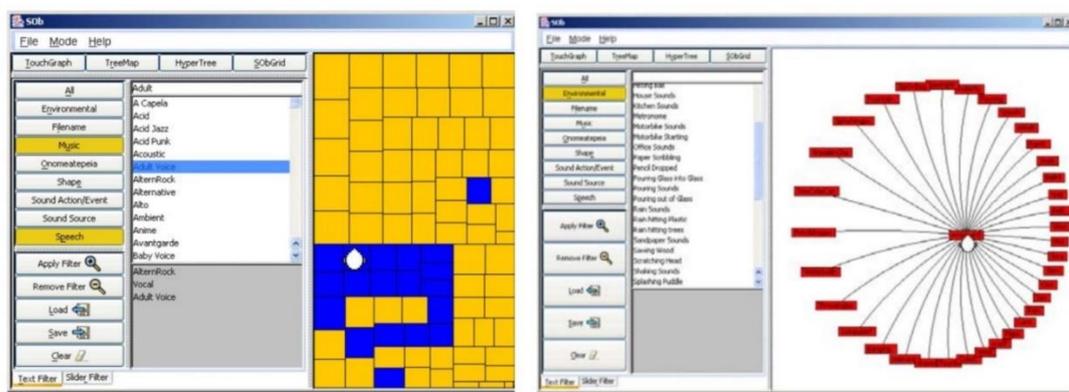


Abbildung 4: ARB, Treemap Visualisierung links, HyperbolicTree Visualisierung rechts

#### 2.1.4 SoundTorch [16], [9]

Der in [16] und [9] vorgestellte Sample-Browser SoundTorch visualisiert ebenfalls einzelne Samples als grafische Objekte auf einer 2D Karte. Abbildung 5 zeigt die Benutzeroberfläche von SoundTorch. Wie auch die Entwickler von MARSYAS3D [13] gehen die Entwickler von SoundTorch davon aus, dass eine Kombination aus MIR(Music Information Retrieval) Algorithmen und einer Dimensionsreduktion eine Visualisierung der akustischen Ähnlichkeiten der Samples ermöglicht: “The layout in 2D is based on methods from MIR, so that acoustically similar sounds are placed next to each other.” [9]

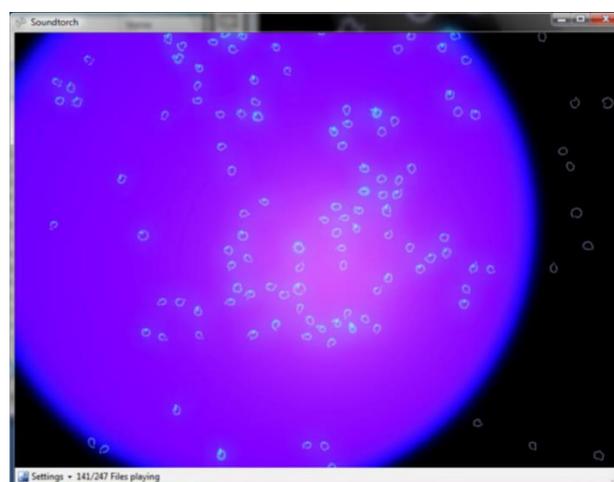


Abbildung 5: Benutzeroberfläche von SoundTorch

Die Sonifikation der einzelnen Samples in SoundTorch ähnelt dem "aura" Konzept des SonicBrowsers. Die Entwickler von SoundTorch verwenden aber für die Beschreibung ihrer Sonifikation die Metapher einer Taschenlampe. Die Anwenderin bewegt mit der Maus den Lichtkegel einer Taschenlampe über die ähnlichkeitsbasierte 2D Karte. Alle Samples innerhalb des Lichtkegels werden gleichzeitig abgespielt. Die Lichtintensität die auf ein Sample fällt bestimmt dabei die Lautstärke mit der ein Sample abgespielt wird. Die Größe der Taschenlampe kann jederzeit durch die Anwenderin geändert werden. Zur Auflösung von Clustern unterstützt die 2D Karte von SoundTorch Zoomoperationen der Anwenderin.

Als contentbased Features werden für SoundTorch nur MFCCs verwendet, die mit sich halbüberlagernden, 50 ms langen Fenstern extrahiert werden. Für die Erstellung der ähnlichkeitsbasierten 2D Karte verwendet SoundTorch eine Self-organizing Map(SOM). Typischerweise werden SOM Algorithmen mit einer Zufallsinitialisierung der Gewichtsvektoren gestartet. Die Folge einer Zufallsinitialisierung ist, dass den beiden Achsen der 2D Karte für unterschiedliche Sample-Sammlungen nicht immer dieselben klanglichen oder akustischen Eigenschaften zugeordnet sind. Um mehrere Karten unterschiedlicher Sample-Sammlungen miteinander vergleichen zu können wird der SOM Algorithmus in SoundTorch nicht mit einer Zufallsinitialisierung gestartet sondern es erfolgt eine Initialisierung der Gewichtsvektoren die eine gleichbleibende, von der Samplesammlung unabhängige, semantische Struktur der Karte ermöglichen soll: „This very roughly amounts to placing noisy and dull sounds at large x and small y; toward small x, the sound gets more tonal; toward large y, pitch goes up“ [9]

In [9] evaluieren die Entwickler SoundTorch mit Hilfe von qualitativen Beobachtungen und Anwendung der „Think Aloud“ Methode. Insgesamt 15 Anwenderinnen mussten drei verschiedene Aufgaben sowohl mit SoundTorch als auch mit einem listenbasierten Interface bewältigen. Für jeden dieser Tests wurde eine andere Sammlung von 100 Soundeffekten zur Filmvertonung verwendet. Die Dateinamen der Files wurden durch Nummern ersetzt. Die Reihenfolge der Files im listenbasierten Interface wurde für jeden Test per Zufall bestimmt.

Im ersten Test sollten die Anwenderinnen sich einen Überblick über eine unbekannte Sammlung von Samples verschaffen. Die Aufgabe bestand darin eine passende, mentale Gruppierung der Samples vorzunehmen. Mit SoundTorch gelang es 12 der 15 Versuchsteilnehmer die Gruppenzugehörigkeit der Samples zu erkennen. Mit dem listenbasierten Interface gelang dies nur einer Versuchsperson.

Die Aufgabe für den zweiten Test bestand darin in der Sammlung ein bestimmtes Sample zu finden. Mit SoundTorch gelang es 14 der 15 Teilnehmer das Sample zu finden, mit dem listenbasierten Interface nur 8 Teilnehmern. Mit SoundTorch benötigten die Teilnehmer wesentlich weniger Zeit das Sample zu finden.

Im dritten Test sollten die Teilnehmer passende Samples finden aus denen sich der Schrei eines Dinosauriers und das Alarmsignal eines Raumschiffs zusammensetzen lässt. Mit SoundTorch vergleichen die Teilnehmerinnen mehrere Samples miteinander und nahmen sich durchschnittlich 3 Minuten Zeit um geeignete Samples zu finden. Alle Teilnehmer waren mit der von Ihnen getroffene Auswahl zufrieden und sind der Meinung „that they enjoyed exploring the set“ [9] Mit dem listenbasierten Interface nahmen sich die Teilnehmer nur eine Minute Zeit und „most users claimed to be bored by having to listen to all that stuff“ [9] Nur 10 von 15 Teilnehmerinnen waren mit ihrer Auswahl zufrieden. Mit dem listenbasierten Interface wählten die Teilnehmerinnen immer das erste Sample das ihren Anforderungen genügte, ein Vergleich mehrerer Samples miteinander erfolgte nicht.

In Interviews bewerten die Anwenderinnen SoundTorch im Vergleich mit dem listenbasierten Interface folgendermaßen: „...working with SoundTorch is much more pleasing than stepping through a list“ [9]

Ob sich die positiven Resultate der Usertests auf den schnelleren Zugriff den SoundTorch gegenüber einer listenbasierten Benützeroberfläche bietet oder eher auf die ähnlichkeitsbasierte 2D Anordnung der Samples zurückführen lassen geht aus den qualitativen Beschreibungen in [16] und [9] nicht hervor. Hinweise ob die User die Semantik der ähnlichkeitsbasierten 2D Karte verstehen sind in beiden Veröffentlichungen nicht zu finden. Diesbezüglich wesentlich wäre eine Beurteilung der Testteilnehmer gewesen ob sich die ähnlichkeitsbasierte Anordnung der Samples in SoundTorch tatsächlich mit ihrem eigenem Ähnlichkeitsempfinden deckt.

Erklärtes Ziel des Designs von SoundTorch ist die Samplesuche durch das gleichzeitige Abspielen mehrerer Samples zu beschleunigen. „The acoustic presentation of several audio files in parallel leverages the potential of the human brain to home in on specific target patterns inside a mixture“ [16] Auch die Entwickler von SoundTorch berufen sich dabei auf den cocktail party effect: “The proposed quick auditioning method leverages the “Cocktail-Party Effect,” the human capability for complex auditory scene analysis“ [9] Die Tests zeigen aber auch hier das die menschliche Fähigkeit gleichzeitig mehrere Samples zu verarbeiten Grenzen hat. Zumindest für jene Anwender die nicht beruflich im Audiobereich arbeiteten zeigt sich, dass das gleichzeitige Anhören mehrerer Samples eine kognitive Belastung ist: „However, users with no professional audio expertise reported that they felt overtaxed with the focused listening to many simultaneous sounds for several minutes.“ [9] Auch die Anzahl der gleichzeitig abgespielten Samples wird von dieser Usergruppe während der Test eher gering gehalten: „On top of that, the non-professionals hesitated to engage the full power of the system by using a large light cone to play back a dozen of sounds at the same time.“ [9] Implizit wird mit dieser Formulierung angenommen, dass ein gut trainierter Hörer in der Lage ist mehrere Duzend gleichzeitig abgespielter Samples zu unterscheiden und einen oder mehrere als Kandidaten für einen bestimmten Einsatzzweck auszuwählen.

#### 2.1.5 Sound Navigator [7], [17]

Der Sound Navigator [17] bietet der Anwenderin insgesamt 229 verschiedene, inhaltsbasierte Features. Neben den im MPEG 7 Standard spezifizierten, inhaltsbasierten Audio-Features kann die Anwenderin Features auswählen, die im Rahmen des CUIDADO Projekts [25] neu entwickelt wurden. Die vollständige Auflistung dieser neu entwickelten Features ist in [26] zu finden. Zusätzlich werden Features angeboten, die für eine inhaltsbasierte Verschlagwortung von Soundeffekten entwickelt wurden. Die Details dieser Features werden in [27] beschrieben.

Jedes dieser Features kann von der Anwenderin entweder der X oder Y-Achse eines 2D Diagramms zugeordnet werden. Anstatt jedes Sample als ein grafisches Objekt auf diesem Diagramm darzustellen wird jedes Samples einer Sample-Sammlung zuerst in kurze, 200 bis 500 Millisekunden lange Segmente aufgeteilt. Für jedes dieser Segmente werden dann die verschiedenen Features berechnet. Anschließend wird jedes Segment als ein Punkt im XY Diagramm eingefügt. Die XY Koordinaten des Segments ergeben sich dabei aus den Werten der Features die für die jeweilige Achse gewählt wurden. Abbildung 6 zeigt die Benützer Oberfläche des Sound Navigators nach der Zuweisung von 2 Features auf die X bzw. Y-Achse und der automatischen Segmentierung der einzelnen Samples.

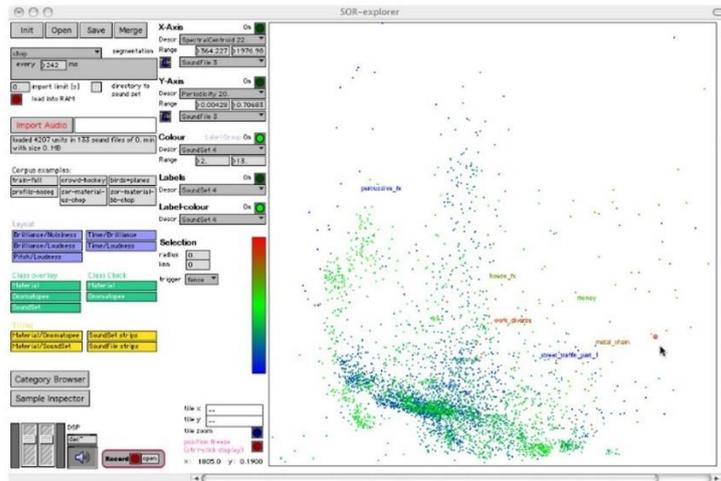


Abbildung 6: Benutzeroberfläche des Sound Navigator Prototyps

Neben der Zuweisung von zwei Features auf die X oder Y-Achse des Diagramms bietet der Sound Navigator die Möglichkeit mehrere Features auszuwählen. In diesem Fall wird ein mehrdimensionaler Feature Vektor pro Sample erstellt und mittels einer MDS (Multidimensional Scaling) Variante [7] oder einer PCA (Principal Component Analysis) auf zwei Dimensionen reduziert.

Bewegt die Anwenderin den Maus Cursor über das 2 Achsen Diagramm wird das Segment abgespielt das der Mausposition am nächsten ist. Welche klanglichen Eigenschaften die Segmente in einem bestimmten Bereich des Diagramms haben bringt die Anwenderin In Erfahrung in dem sie die Maus im Diagramm bewegt. Erfolgt eine rasche Mausbewegung werden viele Segmente in rascher Abfolge getriggert und teilweise gleichzeitig auf den Lautsprechern ausgegeben. „This strong interactivity enables the user to quickly understand the dimensions and areas of the presented space by an initial traversal, probing sound snippets that are played as they are passed by.“ [7] Wird der Maus Cursor nicht bewegt werden alle Segmente aus denen sich ein einzelnes Sample zusammensetzt abgespielt.

Die Entwickler des Sound Navigators bezeichnen dieses Design als „search-by-interaction paradigm“ [7] und betonen die Vorteile dieses Ansatzes im Vergleich zu listenbasierten Interfaces und einer Stichwortsuche: „This approach greatly speeds up the usual workflow of hierarchical menu or search mask, result list, and play/stop buttons that put many mouseclicks between the user’s idea of the sound and listening to appropriate contents of the database.“ [17]

Um das „search-by-interaction“ [7] Paradigma auch auf große Samplesammlungen anwenden zu können wurden im Rahmen des Projekts die Effizienz oder Funktionsweise von drei Algorithmen verbessert bzw. geändert: „Fast similarity-based search by a kD-tree in the high-dimensional descriptor space, a mass– spring model for layout, efficient dimensionality reduction for visualisation by hybrid multi-dimensional scaling...“ [17] Die Details der Implementierung dieser modifizierten Algorithmen wird in [7] beschrieben.

Die Segmentierung der Samples und die Darstellung aller Segmente als Punkte im Diagramm des Sound Navigators führt zwangsläufig dazu, dass sich viele Punkte überlagern und verdecken. Um diese Überlagerungen zu vermeiden verwenden die Entwickler ein physikalisches Model aus Masseteilchen, die über Federn miteinander verbunden sind. Sie erweitern dieses physikalische Model mit Abstoßungskräften zwischen den einzelnen Teilchen und wende diesen Repulsionsalgorithmus auf die Punkte des Diagramms an um damit Überlagerungen zu vermeiden. Abbildung 7 zeigt einen Cluster im 2D Diagramm des Sound Navigators vor und nach der Anwendung dieses Algorithmus.



Abbildung 7: Vermeidung von sich überlagernden Samplesegmenten im Sound Navigator, links sich überlagernde Segmente, rechts nach Anwendung des Repulsionsalgorithmus

Für die Dimensionsreduktion (= Projektion aller mehr dimensional Feature Vektoren auf eine XY Position im Diagramm des Sound Navigator) wurde eine neue Variante des in [28] beschriebenen „Hybrid MDS with Pivot-Based Searching“ [28] entwickelt.

Eine qualitative Beschreibung wie Anwenderinnen eine Samplesuche mit dem Sound Navigator durchführen oder eine Evaluierung des Sound Navigators sind nicht veröffentlicht worden.

#### 2.1.6 MediaCycle [19], AudioCycle [18]

MediaCycle [19] bietet 2 verschiedene Ansätze um eine Samplesuche durchzuführen. Neben einer interaktiven, ähnlichkeitsbasierten 2D Visualisierung einer Samplesammlung hat die Anwenderin auch die Möglichkeit eine „Query by Example“ Funktion zu nutzen. Beide Ansätze verwenden inhaltsbasierte Features um Ähnlichkeiten zwischen verschiedenen Samples zu berechnen. AudioCycle [18] bietet die gleiche ähnlichkeitsbasierte 2D Visualisierung wie MediaCycle, ist aber für das „browsing through music loop libraries“ [18] konzipiert. Die Anwenderin kann mit AudioCycle mehrere Sample-Loops taktsynchron und in derselben Tonart abspielen.

Die Entwickler von MediaCycle betonen, dass Audiosignale anhand mehrerer Aspekte wie z.B. Klangfarbe, Harmonie oder Rhythmus analysiert und interpretiert werden können. Dementsprechend verwendet MediaCycle Features, die die Klangfarbe (Timbre) sowie die harmonischen und rhythmischen Eigenschaften von Samples quantifizieren. Die Entwickler halten zusätzlich fest, dass die Anwenderin möglicherweise ihre Samplesammlung nur anhand bestimmter Aspekte durchsuchen oder gruppieren möchte.

Die Timbre Features werden mit 21 Millisekunden großen Fenstern, die sich zu 75% überlagern, extrahiert. Durchschnitt und Varianz aller Fenster werden dazu verwendet um einen Feature Vektor mit gleicher Dimensionalität für jedes Sample zu erstellen. MediaCycle verwendet die folgenden Timbre Features: „total energy, zero-crossing rate, spectral shape descriptors (spectral centroid, spectral spread, spectral skewness and spectral kurtosis), loudness, sharpness, spread, spectral flatness, spectral crest, spectral slope, spectral decrease, spectral rolloff, spectral variation, as well as mel-frequency cepstral coefficients and their first temporal derivate“ [19] Die in Media Cycle verwendeten Rhythmus Features ermitteln Tempo (BPM -Beats per Minute) und den Taktanfang, die harmonischen Features geben Aufschluss darüber welche Töne oder Akkorde im jeweiligen Sample verwendet wurden.

In der 2D Visualisierung von MediaCycle werden einzelne Samples als Rechtecke dargestellt, die je nach Gruppenzugehörigkeit farbkodiert sind. Abbildung 8 zeigt diese Visualisierung. Die Anwenderin kann sich einzelne Samples anhören indem sie eines der Rechtecke anklickt oder den Maus Cursor über

ein Rechteck platziert. Welche der beiden Sonifikationen verwendet wird bestimmt die Anwenderin. In beiden Fällen wird die Waveform des jeweiligen Samples direkt in der 2D Karte angezeigt.

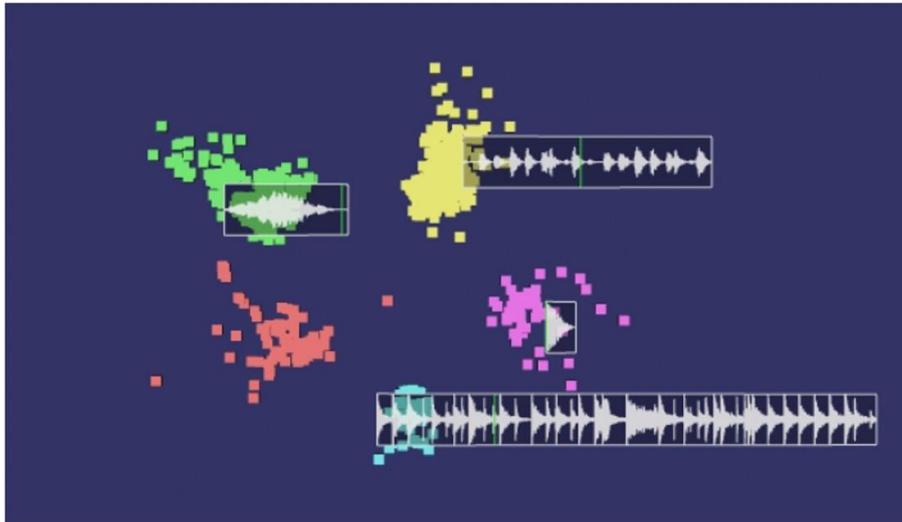


Abbildung 8: Ähnlichkeitsbasierte Gruppierung von Samples in MediaCycle. Das orangefarbene Referenzsample befindet sich in der Mitte der Visualisierung

Um eine Gruppierung der Samples nach Ähnlichkeit zu ermöglichen wird in MediaCycle der k-Means Algorithmus verwendet. Als Distanzmaß wird die euklidische Distanz der Feature Vektoren eingesetzt. Die Anwenderin bestimmt dabei die Anzahl der Gruppen in die eine Samplesammlung aufgeteilt wird.

Der k-Means Algorithmus liefert pro Sample eine Clusterzugehörigkeit und pro Cluster einen Clusterschwerpunkt, berechnet aber keine 2D Koordinaten für das jeweilige Sample. (Ein Clusterschwerpunkt ist ein Vektor mit gleicher Dimensionalität wie die Feature Vektoren der Samples) Die 2D Koordinaten aller Samples wird daher wie folgt bestimmt. Nach der Gruppierung durch den k-Means Algorithmus wählt MediaCycle ein Referenz-Sample und platziert es in der Mitte der Visualisierung (siehe Abbildung 8). Dieses Referenz-Sample wird als Pol eines Polarkoordinatensystem festgelegt. Die 2D Koordinaten jedes Samples bestehen in diesem Polarkoordinatensystem aus dem Abstand zum Pol und einer Winkelkoordinate. Der Abstand zwischen dem Referenz-Sample und einem bestimmten Sample entspricht proportional dem euklidischen Abstand der beiden Feature Vektoren dieser beiden Samples. (Deshalb gilt, dass je näher ein Sample dem Referenzsample ist desto ähnlicher ist das Sample dem Referenzsample. Je größer der Abstand desto unähnlicher sind die beiden Samples.) Die Winkelkoordinate wird aus der euklidischen Distanz zwischen dem Feature Vektors des jeweiligen Samples und dem Clusterschwerpunkt -Vektors des Clusters, dem das jeweilige Sample zugeordnet wurde, berechnet. Diese Vorgehensweise führt zu einer Anordnung aller Samples in Gruppen rund um das Referenzsample wie sie in Abbildung 8 zu sehen ist.

Um der Anwenderin die Möglichkeit zu bieten ihre Samplesammlung nur anhand bestimmter Aspekte zu durchsuchen oder zu gruppieren wurden in MediaCycle drei Schieberegler implementiert. Die drei Schieberegler sind im graphischen Userinterface von MediaCycle mit „Timbre“, „Harmony“ und „Rhythm“ beschriftet. Jeder dieser drei Aspekte wird durch mehrere Features quantifiziert. Die Anwenderin kann mit diesen Schieberegler bestimmen welche Features für die Ähnlichkeitsberechnung verwendet werden. Eine Nullstellung einer dieser Schieberegler hat zu Folge das eine der drei Featuregruppen nicht für die Ähnlichkeitsberechnung verwendet wird. Jede andere Stellung der Schieberegler hat eine entsprechende Gewichtung der jeweiligen Featuregruppe bei der

Ähnlichkeitsberechnung zur Folge. Die 2D Visualisierung reagiert unmittelbar auf eine Bewegung der Schieberegler und zeigt sofort eine neue Gruppierung der Samples.

Nachdem die erste Gruppierung erstellt wurde kann die Anwenderin einen der Cluster auswählen und diesen wiederum in Untergruppen aufteilen lassen. Das Referenzsample kann durch die Anwenderin geändert werden. Auch in diesem Fall reagiert die 2D Visualisierung unmittelbar und zeigt eine neue Gruppierung der Samples.

Neben der ähnlichkeitsbasierten 2D Karte bietet MediaCycle die Möglichkeit eine "Query by Example" durchzuführen. Die Anwenderin wählt ein Sample, MediaCycle sucht automatisch ähnliche Samples. In vielen „Query by Example“ Applikationen wird die euklidische oder Cosinus Distanz der Feature Vektoren als Ähnlichkeitsmaß verwendet. Den Nachteil dieser Methode identifizieren die Entwickler von Media Cycle wie folgt: „The underlying assumption is that all features are equally important for the perceived similarity, thereby discarding content that is very similar for only a subset of features.“ [19] Erklärtes Ziel der Entwickler ist aber, dass das Ergebnis der „Query by Example“ sowohl Samples enthält die sich in allen Features ähnlich sind, aber auch Samples enthält die sich nur in bestimmten Features ähnlich sind. Um dieses Ziel zu erreichen verwenden die Entwickler „... a mixture of Pareto rank and of euclidean distance to retrieve sounds that are close for all the features and sounds that are close considering a subset of features“ [19]

Eine qualitative Beschreibung wie Anwenderinnen eine Samplesuche mit MediaCycle durchführen oder eine Evaluierung von MediaCycle sind nicht veröffentlicht worden. Dasselbe gilt für AudioCycle.

#### 2.1.7 Grill und Flexer [20], [29], [30]

In [29] versuchen Flexer und Grill herauszufinden nach welchen Aspekten, welchen Wahrnehmungsqualitäten, welchen „personal constructs“ [29] Anwenderinnen ihre Sample-Sammlungen durchsuchen oder organisieren wollen. Ihr erklärtes Ziel: "Knowledge about those personal constructs shall eventually lead to more intuitive interfaces for browsing large sound libraries." [29] Die gesamte Studie zur Ermittlung dieser Konstrukte, die nachfolgende Implementierung neuer Features, die diese Konstrukte quantifizieren (vgl. [30]) und die Visualisierung der Konstrukte in einem Sample-Browser (vgl. [20]) beschränkt sich auf „textural sounds“ [20], d.h. Samples deren klangliche Eigenschaften während ihrer gesamten Dauer gleich bleiben.

Die Autoren führen Hörtest mit 16 Studienteilnehmerinnen durch. „By and large, those persons are all used to talking about sound and music and able to formulate their listening experiences, being either professional artists or researchers in sound“ [29]. Zur Erhebung der „personal constructs“ während der Hörtests wird die repertory grid Methode (siehe [31]) verwendet. Den Teilnehmerinnen wurden insgesamt 10 Sample-Paare vorgespielt. Nach jedem vorgespielten Sample-Paar wurden die Teilnehmerinnen gefragt wie sich die zwei Samples unterscheiden. Die Wahrnehmungsqualität in der sich die beiden Samples unterscheiden wurde dabei von den Teilnehmerinnen selbst formuliert. Sie wurden aufgefordert verbale Gegensatzpaare (z.B. tonal–noisy) zu bilden. Die zwei Wörter aus denen sich das Gegensatzpaar zusammensetzt soll sich dazu eignen die Enden einer mehrstufigen, zweipoligen Skala zu beschriften. Nachdem eine Teilnehmerin mehrere dieser „bi-polar personal constructs“ [29] definiert hatte wurde sie aufgefordert alle 20 Samples auf den Skalen der jeweiligen Konstrukte einzuordnen. Aus den insgesamt 202, von den Teilnehmern definierten, bipolaren Konstrukten ermitteln die Autoren sowohl mit einer statistischen als auch qualitativ, semantischen Analyse 10 synonym verwendete bipolare Konstrukte wie z.B. high–low, ordered–chaotic, smooth–coarse und tonal–noisy.

Da die Konstrukte mit einer relativ geringen Anzahl an Anwenderinnen und Samples ermittelt wurden ist fraglich ob sich die Ergebnisse auf eine große Anzahl an Anwenderinnen und Samples

verallgemeinern lassen. Welche der Konstrukte ist für eine möglichst große Zahl an Anwenderinnen relevant und eindeutig? Um diese Frage zu beantworten führen Flexer und Grill eine Online Umfrage durch. Die Teilnehmerinnen der Onlineumfrage spielen ein Sample ab und bewerten das Sample auf der jeweiligen Skala der 10 bipolaren Konstrukte. Abbildung 9 zeigt die Benutzeroberfläche der Online Umfrage. Ist den Teilnehmerinnen eines der Konstrukte unklar oder auf das jeweilige Sample nicht anwendbar haben sie Möglichkeit keine Wertung auf der jeweiligen Skala vorzunehmen. Durchschnittlich wurden pro Sample 9,4 von 10 Wertungen vorgenommen. Von der Möglichkeit für ein Konstrukt keine Wertung abzugeben wurde also in den seltensten Fällen Gebrauch gemacht. Man kann daher davon ausgehen, dass die Konstrukte von einer größeren Anzahl an Anwenderinnen semantisch verstanden werden. Wesentlich ist auch die Frage ob ein „inter-rater agreement“ der Teilnehmer erkennbar ist. Ordnen die Teilnehmer einem bestimmten Sample ähnliche Werte auf den jeweiligen Skalen zu? „...the ranking of the constructs in respect to inter-rater agreement is consistent, indicating the constructs high–low, ordered– chaotic, natural–artificial, smooth–coarse, tonal–noisy and homogeneous–heterogeneous to be shared among subjects of both groups with sufficient agreement.“ [29]

Please listen to the following sound and try to rate it according to the criteria listed below.

So far, you have rated 0 sounds (out of 20) in set 1.

For each criterion try to find a grade that reflects a position between the notions on the left and right sides.

In the case that a specific criterion doesn't seem to be applicable, the grading can be omitted resp. de-selected by double-clicking.

high bright	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	low dull
ordered coherent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	chaotic erratic
static rigid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	dynamic eventful
natural analog	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	artificial digital
smooth soft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	coarse raspy
near clear	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	far blurred
edgy disjointed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	flowing continuous
dense expansive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	sparse selective
homogeneous uniform	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	heterogeneous differentiated
tonal	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	noisy

Remarks (optional):

submit

Abbildung 9: Online Umfrage für die Evaluierung der zehn bipolaren Konstrukte

In [29] testen Flexer et al. welche bereits entwickelten Standard Audiofeatures geeignet sind die ermittelten, bipolaren Konstrukte zu quantifizieren. Die Autoren testen dabei alle skalaren Features des YAFFE [32] Feature Extractors. Die höchsten Korrelationen zeigt dabei das Konstrukt „high - low“ mit den Features „perceptual sharpness“, „spectral centroid“ und „spectral slope“. Drei weitere Konstrukte weisen eine hohe Korrelation zu einzelnen Features auf. Die Konstrukte „smooth–coarse“, „near–far“ und „edgy–flowing“ lassen sich jeweils durch „spectral skewness“, „spectral centroid“ und „spectral slope“ beziehungsweise „spectral flatness“ quantifizieren. Die hohe Korrelation der Features „spectral centroid“ und „spectral slope“ mit mehreren Konstrukten „... motivates the construction of specifically sensitive descriptors (or combinations thereof) in order to be able to discriminate those perceptual qualities.“ [29] Für einige andere Konstrukte lassen sich keine entsprechenden Standard Features ermitteln, was die Autoren dazu veranlasst neue Features zu entwickeln „... that correlate well with each (and exactly one) of the constructs, thereby allowing automatic classification of sounds according to perceptually oriented constructs.“ [29]

Die Entwicklung, Funktionsweise und Evaluierung dieser neuen Features für die Konstrukte „high–low“, „ordered–chaotic“, „smooth–coarse“, „tonal–noisy“, und „homogeneous–heterogeneous“ wird in [30] beschrieben. „Evaluation has yielded Pearson correlations between the audio descriptors and human ratings obtained from listening tests of above 0.74 for the constructs ordered–chaotic, smooth–coarse, tonal–noisy, homogeneous–heterogeneous, and up to 0.90 for the construct high–low.“ [30]

Einen mehrdimensionalen Feature Vektor pro Sample zu berechnen um anschließend eine Ähnlichkeitserhaltende Projektion auf eine 2D Position mittels einer Dimensionsreduktion durchzuführen ist zum Zeitpunkt der Veröffentlichungen von Grill und Flexer bereits ein MIR (Music Information Retrieval) Standardverfahren. Den wesentlichen Nachteil dieses Ansatzes identifizieren Grill und Flexer in [29]: „A drawback of this method is that the projection axes of the low-dimensional space are not pre-defined and/or not interpretable. When used as a visualization as part of a user interface, the systematics of organization consequently have to be learned by the user by means of exploration“ [29] Der Abstand der Samples zueinander zeigt der Anwenderin nur wie ähnlich sich bestimmte Samples sind. Welche klanglichen Eigenschaften die Samples an bestimmten Positionen in der 2D Visualisierung haben, kann die Anwenderin aber nicht erkennen ohne sich die Samples anzuhören. In [20] versuchen die Autoren daher eine Visualisierung zu entwickeln, die den klanglichen Eigenschaften grafische Eigenschaften zuweist. Die Anwenderin soll optisch erkennen können welche akustischen oder klanglichen Eigenschaften Samples in bestimmten Bereichen der 2D Karte haben.

Der in [20] vorgestellte Prototyp von Grill und Flexer verwendet für die Dimensionsreduktion ( von fünf, die ermittelten Konstrukte quantifizierenden Features, auf zwei Dimensionen) den t-SNE [33] Algorithmus. Damit wird eine Ähnlichkeitserhaltende Projektion des mehrdimensionalen Feature Raums auf 2D Koordinaten erreicht. Einzelne Samples werden in der 2D Karte als weiße Punkte dargestellt. Abbildung 10 zeigt den Prototypen von Grill und Flexer.

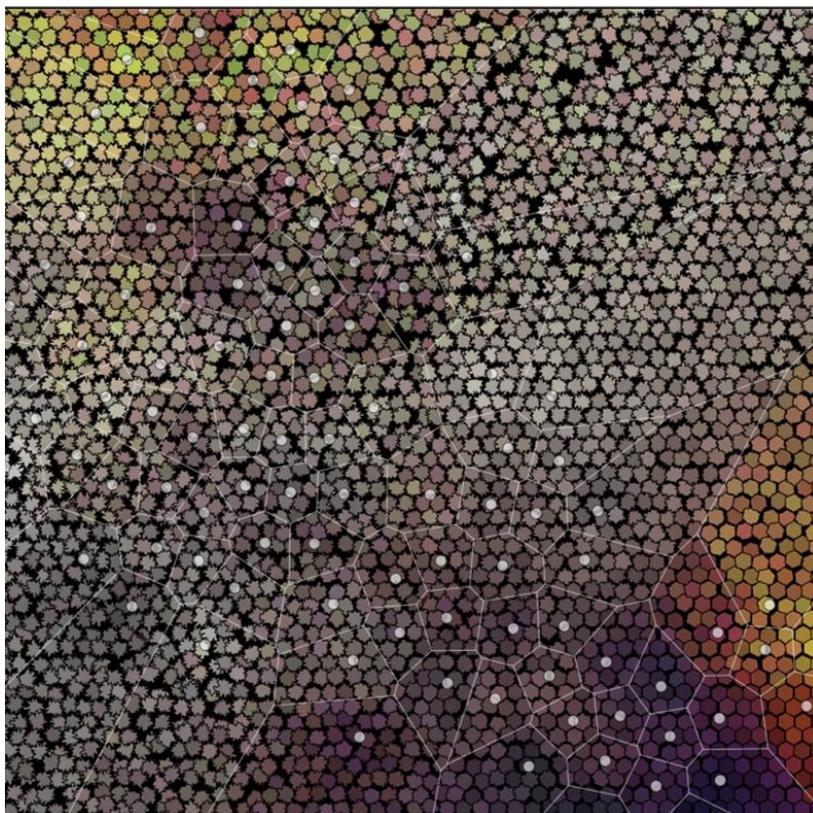


Abbildung 10: Prototyp des Sample-Browsers von Grill und Flexer

Um der Anwenderin eine Vorstellung davon zu vermitteln welche klanglichen Eigenschaften in welchen Bereichen der 2D Karte vorherrschen wird die gesamte 2D Karte mit einem dichten Gitter von ca. 20 Pixel großen, grafischen Elementen („tiles“ [20] - Kacheln ) ausgefüllt. Jedem der fünf bipolaren Konstrukte mit dem höchsten „inter-rater agreement“ ordnen die Entwickler bestimmte grafische Eigenschaften dieser Elemente zu. Abbildung 11 zeigt diese Zuordnung. Dem Konstrukt „tonhaltig – geräuschhaft“ wird z.B. die Farbsättigung zugeordnet. Tonhaltige Bereiche sind daher färbig, eher geräuschhafte Bereiche sind in Graustufen gehalten.

Auditory construct	Visual parameter
high→low	brightness + color hue (bright yellow→dark red)
ordered→chaotic	regularity of elements (on grid→deviating from grid)
smooth→coarse	jaggedness of element outline (smooth outline→jagged outline)
tonal→noisy	saturation (colorful→gray)
homogeneous→heterogeneous	variability in color parameters, especially hue (no variation→much variation)

Abbildung 11: Zuordnung der bipolaren Konstrukte auf grafische Eigenschaften der Tiles

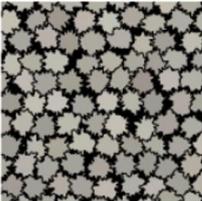
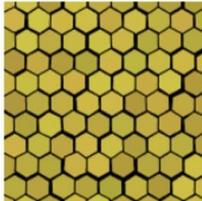
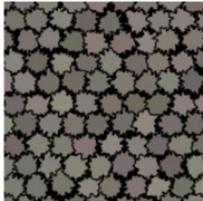
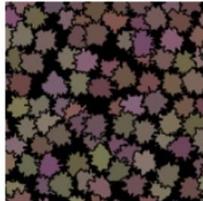
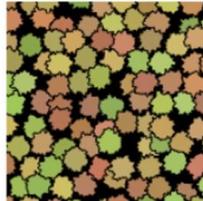
Um die Zuordnung der bipolaren Konstrukte auf visuelle Parameter zu evaluieren führen Grill und Flexer eine Online Umfrage durch. Die Teilnehmerinnen spielen ein Sample ab und wählen aus 5 Visualisierungen jene die dem gehörten Sample entspricht. Die Teilnehmerinnen erhielten nach einer Auswahl keine Rückmeldung ob ihre Auswahl richtig oder falsch war. Abbildung 12 zeigt die Benutzeroberfläche, die für die Online Umfrage verwendet wurde.

**Please listen to the following sound and try to associate it to one of the graphics below:**



So far, you have rated 0 sounds (out of 100 available).

**Choose the representation that to your opinion fits best to the sound.  
Click on the respective image and then 'submit'.**

Difficulty of the association:  straightforward/unambiguous  difficult/ambiguous  impossible

Remarks (optional):

Abbildung 12: Online Umfrage zur Evaluierung der Visualisierung der fünf bipolaren Konstrukte mit dem höchsten „inter-rater agreement“

Die Ergebnisse zeigen „...that subjects were able to successfully associate sounds with graphical representations, with RMS errors and rates of correct assignment better than random with a very high significance“ [20] Besonders die Zuordnung des Konstrukts „high – low“ auf die Helligkeit und den

Farbton scheint den Teilnehmerinnen intuitiv klar zu sein. Auch die Zuordnungen der Konstrukte „tonal-noisy“ und „smooth-coarse“ auf die visuelle Parameter „färbig-graustufen“ bzw. „geradlinige Umrandung – gezackte Umrandung“ scheinen die Teilnehmerinnen intuitiv erfassen zu können. Für jene Teilnehmer die eine größere Anzahl an Samples bewerten lassen sich statistisch signifikante Lerneffekte nachweisen. Die Anzahl der richtig ausgewählten Visualisierungen steigt mit der Anzahl der bewerteten Samples. Die Autoren folgern daraus „...that there is a correct implicit understanding of the employed mapping“ [20]

### 2.1.8 AudioQuilt, Snare Drum Navigator [21]

Der Snare Drum Navigator visualisiert einzelne Samples als gleich große, farbkodierte Rechtecke. In jedem Rechteck wird die Waveform des Samples dargestellt. Bewegt die Anwenderin den Maus Cursor über eines der Rechtecke wird das dazugehörige Sample abgespielt. Die Rechtecke sind in einem regelmäßigen Gitter angeordnet. Die Ähnlichkeit der Samples wird sowohl durch die Farbkodierung als auch durch den Abstand der Samples visualisiert. Je geringer der Abstand zweier Sample-Rechtecke desto ähnlicher sind sich die beiden Samples. Abbildung 13 zeigt die Benutzeroberfläche des Snare Drum Navigators.

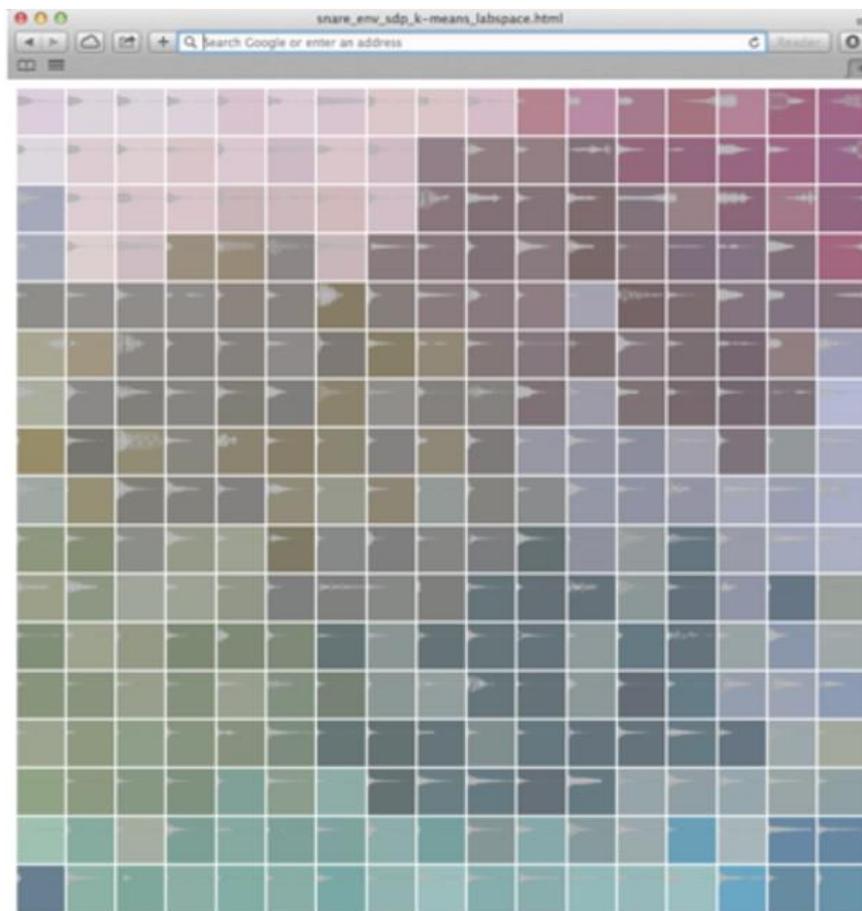


Abbildung 13: Benutzeroberfläche des Snare Drum Navigator

Jedes Sample wird genau einer Gitterposition zugeordnet. Um die Visualisierung der Sample-Ähnlichkeit durch Abstand zu ermöglichen muss bei dieser Zuordnung die paarweise Distanz der Feature Vektoren der paarweisen Distanz der Rechtecke weitestgehend entsprechen. Die Entwickler mussten daher eine Ähnlichkeitserhaltende 1:1 Zuordnung zwischen den Feature-Vektoren der Samples und den Positionen der Rechtecke implementieren. Für diese Ähnlichkeitserhaltende Projektion des mehrdimensionalen Feature Raums auf das Ähnlichkeitsbasierte Gitter verwenden die Entwickler von AudioQuilt „Kernelized sorting“ [34]

Dieses Design bietet zwei Vorteile. Erstens kann der für die Visualisierung der Samples zur Verfügung stehende Bildschirmraum optimal ausgenutzt werden. Zweitens wird eine Überlagerung mehrerer Samples in der Visualisierung verhindert. Wird die Dimensionsreduktion mit einer SOM (Self-organizing Map) bewerkstelligt ist eine Überlagerung mehrerer Samples möglich, da Samples jeweils dem nächstliegenden Knoten der SOM zugeordnet werden. Abbildung 14 zeigt links das Ergebnis der Dimensionsreduktion mittels Kernelized Sorting, rechts wird die Dimensionsreduktion mittels SOM dargestellt. Links ist die Zuordnung mehrerer Samples auf einen Knoten der SOM erkennbar.

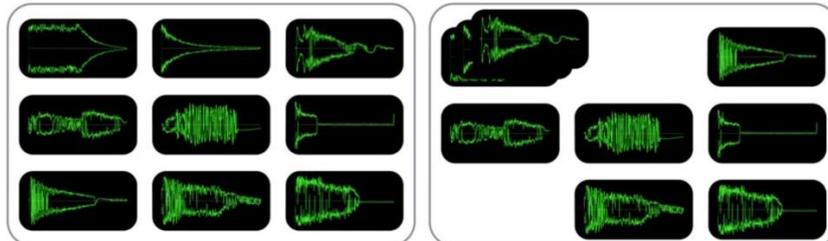


Abbildung 14: Links Dimensionsreduktion mittels Kernelized Sorting, Rechts mittels SOM

„Kernelized sorting“ (2010 in [34] vorgestellt) ermöglicht eine paarweise Zuordnung von Objekten zweier gleichmächtiger Mengen, wenn für jede der beiden Mengen ein Ähnlichkeitsmaß existiert („intra-set distances“ [21]) Ein Ähnlichkeitsmaß das jedem Element der einen Menge Ähnlichkeitswerte für alle Elemente der anderen Menge zuordnet („inter-set distances“ [21]) benötigt der Algorithmus nicht. Die Entwickler von AudioQuilt definieren die Zuordnung mehrerer, hochdimensionaler Feature Vektoren auf bestimmte 2D Positionen als ein solches „set matching problem“ [21] Jedem Element aus der Menge der Feature Vektoren wird ein Element aus der Menge der Gitterkoordinaten zugeordnet. Für jede der beiden Menge existiert ein Ähnlichkeitsmaß. Für die Menge der Feature Vektoren wird die euklidische Distanz der Feature Vektoren verwendet, für die Menge der 2D Gitterkoordinaten der reziproke Wert der euklidischen Distanz.

Der Snare Drum Navigator erstellt pro Sample einen 24-dimensionalen Feature Vektor. Zur Berechnung der Features wird die MIRToolbox [35] verwendet. Vor der Feature Extraktion wird die Hüllkurve des Samples berechnet. Die Dauer der Attackphase der Hüllkurve wird als Feature verwendet. Als spektrale Features werden das temporale Centroid und MFCC Koeffizienten extrahiert. Jeweils 11 MFCC Koeffizienten werden sowohl in der Attackphase als auch in der Releasephase jedes Snaredrum Samples getrennt voneinander extrahiert.

Mehrere Samplebrowser erstellen nach Abschluss der Feature Extraktion eine Ähnlichkeitsmatrix die die paarweise Distanz der Samples im hochdimensionalen Feature Raum anzeigt. Als Distanzmaß wird in vielen Fällen die euklidische Distanz der Feature Vektoren verwendet. Die Ähnlichkeitsmatrix dient als Input für die diversen Dimensionsreduktionsmethoden. Diese Vorgehensweise impliziert jedoch das jedes Feature für das Ähnlichkeitsempfinden der Anwenderin gleichbedeutend ist. Die Entwickler von AudioQuilt gehen hingegen davon aus, daß „... similarity between samples is largely subjective. Audio features that are important to one user may be less important to others.“ [21] Der Snare Drum Navigator bietet daher der Anwenderin die Möglichkeit die Ähnlichkeitsberechnung der Samples ihrem eigenen Ähnlichkeitsempfinden mittels Metric Learning anzupassen.

Der in [36] vorgestellte Metric Learning Algorithmus benötigt als Input der Anwenderin Paare von ähnlichen oder unähnlichen Samples. Mit diesem Input (= Feature Vektoren der als ähnlich oder unähnlich bestimmten Samples) berechnet der Algorithmus ein Distanzmaß, dass die von der Anwenderin bestimmten Ähnlichkeitsbeziehungen berücksichtigt. Die Entwickler des Snare Drum Navigators modifizieren diesen Algorithmus sodass als Input mehrere Gruppen ähnlicher Samples

verwendet werden können. Der modifizierte Algorithmus ermittelt dann ein neues Distanzmaß, das die von der Anwenderin vorgenommenen Gruppierungen berücksichtigt.

Für die Anpassung des Distanzmaßes an das Ähnlichkeitsempfinden der Anwenderin stellt der Snare Drum Navigators eine eigene grafische Benutzeroberfläche zur Verfügung. Abbildung 15 zeigt dieses Interface. Die Anwenderin kann eine beliebige Anzahl an Kreisen erstellen und diesen Samples zu weisen. Anschließend bestimmt sie für jedes Sample eine Gruppenzugehörigkeit. Mit dem Button „Compute Assignment“ wird die Berechnung des Distanzmaßes gestartet.

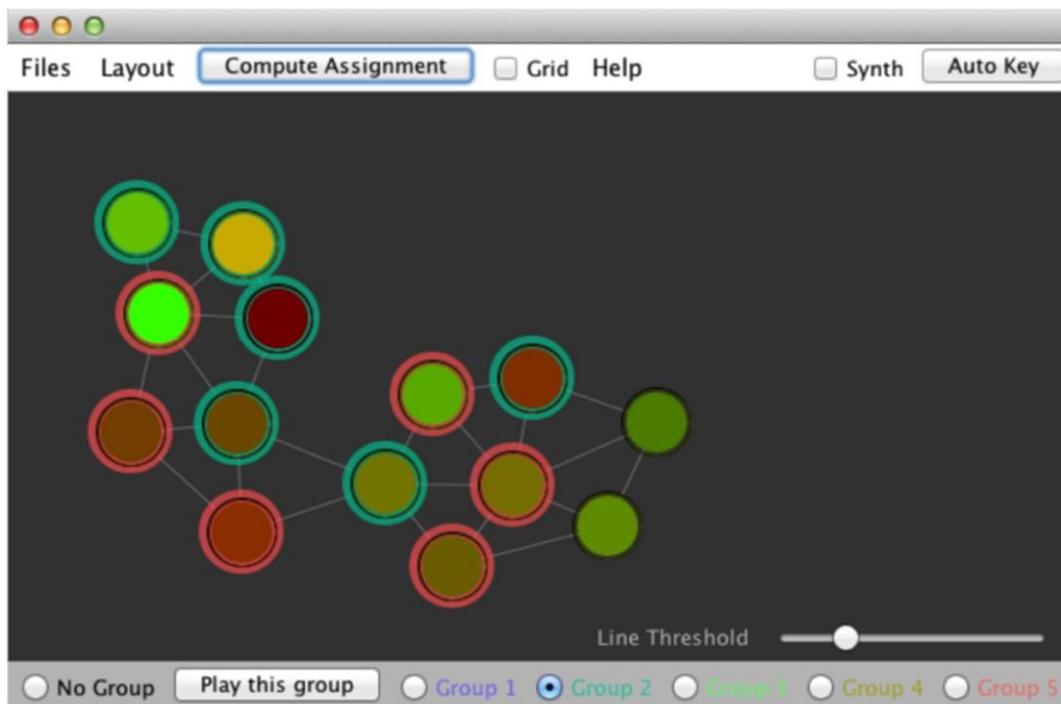


Abbildung 15: Metrik Learning Interface des Snare Drum Navigators

AudioQuilt wurde mittels einer qualitativen Beobachtung und einer quantitativen Online Studie evaluiert. Fünf Musiker, „...familiar with composition using a DAW, and the process of sample selection“ [21] (DAW = Digital Audio Workstation) wurden während der Suche nach einer geeigneten Snare Drum beobachtet. Nach der Auswahl einiger Samples wurde ein semistrukturiertes Interview mit den Teilnehmern durchgeführt. Positiv bewerteten die Musiker den schnellen Zugriff auf die Samples, die das graphische Interface des Snare Drum Navigators bietet. Alle fünf Musiker waren in der Lage Gruppen ähnlicher Samples spontan zu identifizieren und die gemeinsamen klanglichen Eigenschaften dieser Gruppe sprachlich zu beschreiben. In einigen Fällen wiederum deckt sich das Ähnlichkeitsempfinden der Teilnehmer nicht mit der durch Abstand visualisierten, errechneten Ähnlichkeit. Die Teilnehmer entdecken immer wieder Samples die nicht zu ihren Nachbarn passen: „...this snare should be over there“, was a common comment.“ [21]

Eine Anpassung des Distanzmaßes an das Ähnlichkeitsempfinden der Anwenderin wurde während der Beobachtung nicht durchgeführt. Die Entwickler selbst teilten 176 Snare Drum Samples in 12 verschiedene Gruppen auf. Mit diesem Input errechnete der Metrik Learning Algorithmus das für die Beobachtungen verwendete Distanzmaß.

Für die quantitativen Online Tests wurde das grafische Interface in Abbildung 16 verwendet. Die Teilnehmerinnen hören sich ein Beispielsample an (grünes Rechteck links oben) und mussten innerhalb von 60 Sekunden 10 ähnliche Samples in einem Samplegitter mit einem Mausklick markieren. Jedes

Samplegitter besteht aus 100 Samples, 10 davon dem sind dem Bespielsample ähnlich, die anderen 90 Samples unähnlich. Hinweise darauf, dass die Farbe oder der Abstand die Ähnlichkeit der Samples anzeigt enthält das Interface nicht.

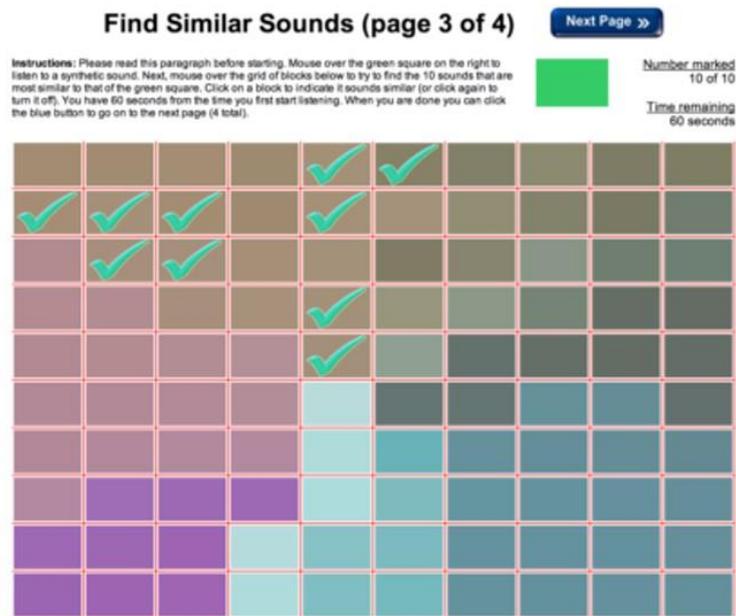


Abbildung 16: Graphisches Userinterface der quantitativen Online Evaluierung

Jede Teilnehmerin führte die Suche in vier verschiedenen Sample-Gittern durch. Die vier Gitter sind die Kombinationen der 2 unabhängigen Testvariablen farbkodierte Ähnlichkeit und abstandskodierte Ähnlichkeit. Die Teilnehmerinnen suchten daher in einem Gitter die Ähnlichkeit durch Farbe und Abstand visualisiert, einem Gitter das weder Farbe noch Abstand verwendet (= farblose, zufällige Anordnung aller Samples), einem Gitter das Ähnlichkeit nur durch Farbe visualisiert und einem Gitter das nur den Abstand zur Ähnlichkeitsvisualisierung verwendet.

Für alle vier Gitter wurde der zeitliche Verlauf der markierten und tatsächlich ähnlichen Samples aufgezeichnet. Die besten Ergebnisse erzielten die Teilnehmerinnen mit dem Gitter indem Ähnlichkeit sowohl durch Farbe als auch durch den Abstand der Samples visualisiert werden. Hier finden die Teilnehmerinnen die ähnlichen Samples am schnellsten.

Für jene zwei Gitter die die Ähnlichkeit durch Abstand visualisieren zeigt der zeitliche Verlauf der richtig markierten Samples eine charakteristische Form. Zu Beginn finden die Teilnehmerinnen nur wenige oder keine ähnlichen Samples, dann steigt die Anzahl der markierten und ähnlichen Samples sprunghaft, innerhalb kürzester Zeit an. Die Autoren vermuten, dass das jener Zeitpunkt ist zu dem die Teilnehmerinnen den Bereich mit den ähnlichen Samples gefunden haben. Im Gegensatz dazu zeigt der Verlauf für die Gitter die keine abstandskodierte Ähnlichkeit verwenden keine Sprünge, sondern einen kontinuierlichen Verlauf.

Obwohl die Teilnehmer keine expliziten Hinweise auf die die Visualisierung von Ähnlichkeit durch Farbe und Abstand bekamen, zeigen die Suchstrategien der Teilnehmerinnen, dass dieses Konzept rasch und intuitiv verstanden wird. Wird den Teilnehmern zuerst das Gitter gezeigt, das keine Farbkodierung und Zufallspositionen der Samples verwendet, so bewegen sie die Maus Zeile für Zeile über das Samplegitter. Wird dann im Anschluss das Gitter gezeigt das Ähnlichkeit sowohl durch Farbe als auch durch die Nähe der Samples visualisiert, so beginnen die Teilnehmerinnen wieder Zeile für

Zeile die Samples abzuspielen, wechseln aber ihre Suchstrategie relativ rasch. Sie durchsuchen einen gleichförmigen Bereich und wechseln dann zum nächsten gleichförmigen Bereich.

### 2.1.9 AudioMetro [2]

Die Entwickler von AudioMetro gehen davon aus, dass große, professionelle Sample-Datenbanken mit einem Stichwortverzeichnis versehen sind. Die Anwenderin kann diese Datenbanken anhand von Tags filtern, erhält aber aufgrund der Größe der Datenbank mehrere hundert Samples als Ergebnis. Die Stichwortsuche ist also nur geeignet den Umfang der in Frage kommenden Samples einzuschränken. Die Autoren formulieren ihre Forschungsfrage daher mit den Worten: „Can contentbased organization be beneficial once a limit is reached when filtering sounds by tag?“ [2]

In [2] werden 4 verschiedene Tests mit mehreren Sample-Browser Prototypen durchgeführt. Als quantitative Testmethodik werden „known-item search tasks“ [2] verwendet. Den Teilnehmerinnen wird ein Sample vorgespielt, das sie anschließend innerhalb einer Minute in einer Samplesammlung suchen müssen. Der Erfolg der Suche und die benötigte Suchzeit werden dabei als Qualitätskriterium des jeweiligen Designs erfasst.

Im ersten Test wird eine Gitteranordnung der Samples mit einer ähnlichkeitsbasierten 2D Karte verglichen. Abbildung 17 zeigt rechts die Gitteranordnung, links die ähnlichkeitsbasierten 2D Karte. Für die ähnlichkeitsbasierte 2D Karte werden als Features MFCCs (zusammengefasst durch mehrere statistische Lagemaße wie z.B. die Standardabweichung), Spectral Flatness und Perceptual Sharpness verwendet. Die Ähnlichkeitserhaltende Projektion auf 2D Positionen wird mittels t-SNE [33] erreicht.

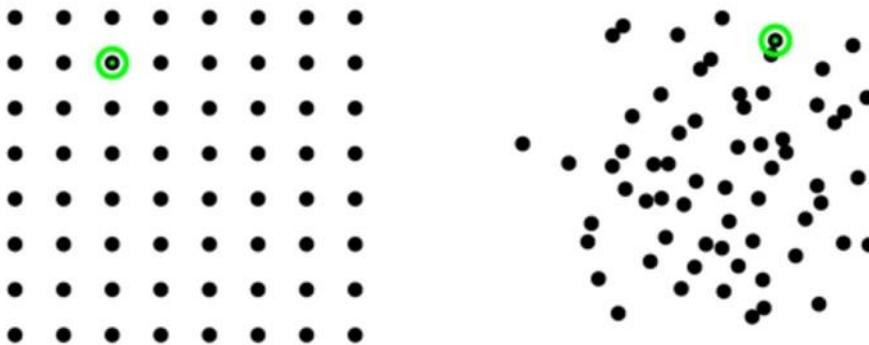


Abbildung 17: Benutzeroberflächen für den ersten Test, Links: Gitteranordnung, Rechts: ähnlichkeitsbasierte 2D Visualisierung

Die Teilnehmerinnen mussten mehrere „Know Item Search Tasks“ durchführen. Pro Task wurde von den Autoren mittels einer Stichwortsuche (z.B. „bell“) 64 Samples aus einer Samplesammlung ausgewählt. In einer kurzen Gewöhnungsphase führten die Teilnehmer 4 Tasks aus. Danach wurden sechs, auf 60 Sekunden limitierte Tasks unmittelbar hintereinander ausgeführt, drei davon mit der Gitteranordnung, die anderen drei mit der ähnlichkeitsbasierten 2D Karte.

Mit der Gitteranordnung wurden 45 von 57 Tasks erfolgreich abgeschlossen, mit der 2D Karte nur 37. Zusätzlich benötigten die Teilnehmerinnen durchschnittlich etwas weniger Zeit für eine Samplesuche, wenn sie die Gitteranordnung verwendeten. Die Autoren folgern daher, dass „... content-based positioning of sounds on a map appeared to be inefficient vs a simple solution“ [2]

Als qualitatives Ergebnis des ersten Testes halten die Autoren fest, dass die Teilnehmerinnen ihre Navigation während der Samplesuche an das jeweilige Interface anpassen. Wird die Gitteranordnung verwendet bewegen die Teilnehmerinnen den Mauscursor zumeist von Zeile zu Zeile. In der 2D Karte wurden zwei verschiedene Navigationsarten beobachtet. Entweder die Teilnehmerinnen umrunden die Punktwolke mit oszillierenden Bewegungen oder sie stellen den Maus Cursor in die Mitte der 2D Karte und bewegen sich dann in ebenfalls oszillierenden Bewegungen nach außen.

Die Teilnehmerinnen des ersten Tests „...are knowledgeable of machine learning methods (some work with dimension reduction techniques such as PCA) and scientific visualization.“ [2] Man kann daher davon ausgehen das ihnen das Konzept der „Visualisierung von Ähnlichkeit durch Abstand“ vertraut ist. Das bedeutet aber nicht das sie in der Lage sind die ähnlichkeitsbasierte 2D Visualisierung effizient zu nützen. Voraussetzung für eine effiziente Benützung ist das Wissen der Anwenderin in welchen Bereich der 2D Karte welche klanglichen Eigenschaften vorherrschen. Dieses Wissen kann sich die Anwenderin nur durch eine längere Interaktion mit der 2D Karte erarbeiten. Sie muss sich mehrere Samples in unterschiedlichen Bereichen der Karte anhören und ein mentales Bild der klanglichen Eigenschaften in den verschiedenen Bereichen formen. Erst dann ist sie in der Lage zielgerichtet in einem bestimmten Bereich der Karte nach einem vorgegebenen Sample zu suchen. Nur ein detailliertes, mentales Bild ermöglicht eine Suche nach dem Prinzip „Das gesuchte Sample hat jene klanglichen Eigenschaften, Samples mit diesen Eigenschaften sind auf der Karte z.B. links oben angeordnet, das gesuchte Sample muss daher in diesem Bereich zu finden sein“

Hatten die Teilnehmer genug Zeit die Semantik der 2D Karten zu erlernen? Die Beschreibung des Testdesigns lässt den Schluss zu, dass die Teilnehmer nur wenig Zeit hatten sich darüber Klarheit zu verschaffen in welchen Bereichen der Karte welche klanglichen Eigenschaften zu finden sind. Auch die Navigationsstrategien der Teilnehmerinnen in der 2D Karte zeigen das sie kein mentales Bild der Karte verwendeten, sondern sich an der Gestalt der Punktwolke orientierten. Ein zielgerichtetes Ansteuern eines bestimmten Bereichs auf der 2D Karte wurde während der Tests nicht beobachtet.

Nachdem die ähnlichkeitsbasierte 2D Karte die Know Item Suche der ersten Tests nicht besser unterstützt als eine Gitteranordnung der Samples führen die Entwickler von AudioMetro eine weitere Studie durch, die klären soll ob eine Visualisierung klanglicher Eigenschaften in der 2D Karte diese Ergebnisse verbessert. Für die zweite Studie greifen die Entwickler auf die Arbeit von Grill und Flexer zurück , die in [29] zeigen, dass die meisten Anwenderinnen ihre Samples unter anderem anhand des Wahrnehmungskonstrukts „high – low“ beurteilen, eine Visualisierung dieser klanglichen Eigenschaft mit der Farbhelligkeit intuitiv verstehen und sich das YAFFE Feature perceptual sharpness (siehe [32]) dazu eignet dieses Konstrukt zu quantifizieren. Die Entwickler von AudioMetro weisen dem Mittelwert der perceptual sharpness Werte die Helligkeit der einzelnen Sampleobjekte in der 2D Karte zu. Die zeitliche Abfolge der perceptual sharpness Werte in den einzelnen Analyseframes wird im Uhrzeigersinn in der Umrandung der einzelnen Sampleobjekte visualisiert.

Für die zweite Studie führten insgesamt 16 Teilnehmerinnen jeweils 5 Know Item Suchen mit der Gitteranordnung und der, um die oben beschriebene Visualisierung erweiterte, ähnlichkeitsbasierte 2D Karte durch. Die Teilnehmerinnen mussten in einer Samplesammlung von 150 industriellen und mechanischen Geräuschen ein bestimmtes Sample finden. Die Ergebnisse zeigen das die Teilnehmerinnen mit der Gitteranordnung statistisch signifikant weniger Zeit benötigten um ein Sample zu finden als mit 2D Karte, in der die Wahrnehmungsdimension „high – low“ durch die Helligkeit und Umrandung der einzelnen Sampleobjekte visualisiert ist. Die Autoren schließen daraus, dass „... these results may be considered negative since a simple baseline solution outperforms a complex system with a layout obtained from a recent dimension reduction technique, carefully chosen feature

extraction, both evaluated algorithmically in previously mentioned references, and glyph representation aiming at supporting audition with vision from perceptual cues.” [2]

In einer dritten Studie wird die gleiche Versuchsanordnung wie in der zweiten Studie verwendet. Um aber auszuschließen, dass die mangelnde Hörerfahrung der Teilnehmerinnen die Ergebnisse beeinflusst werden als Teilnehmerinnen „expert users (in sound auditioning)“ [2] ausgewählt. Die Ergebnisse bleiben die gleichen wie in der zweiten Studie. Mit der ähnlichkeitsbasierten Karte finden die Teilnehmerinnen die gesuchten Samples statistisch signifikant langsamer.

In der zweiten und dritten Studie wurde nur ein „... short 4-task introductory training“ [2] durchgeführt. Für die zweite und dritte Studie ist daher ebenfalls fraglich ob die Teilnehmerinnen genug Zeit hatten ein mentales Bild der klanglichen Eigenschaften in bestimmten Bereichen der Karte zu erstellen. Zusätzlich ist in [2] nicht festgehalten welche Informationen die Teilnehmerinnen der zweiten und dritten Studie vorab über die Visualisierung der Samples erhielten. Mussten sie selbst erkennen, dass die die Helligkeit der klanglichen Eigenschaft „high – low“ entspricht? Nur dieses Vorwissen hätte eine effiziente Samplesuche nach der Logik “Das gesuchte Sample klingt relativ hoch, das gesuchte Sample ist daher ein helles Objekt mit stark gezackter Umrandung“ ermöglicht.

Aus den ersten drei Tests schließen die Entwickler „... that a layout with regular geometry such as the grid directs the search pathway and helps user keep a visual track of their progress in screening collections.” [2] Die Gitteranordnung legt eine Zeile für Zeile Navigation nahe und ermöglicht der Anwenderin leicht festzustellen welche Samples bereits angehört wurden.

Ein qualitatives Ergebnis des dritten Tests sind die Beschwerden der Teilnehmerinnen, dass sich Samples in der 2D Karte überlagern. In der vierten Studie testeten die Entwickler von AudioMetro daher ein ähnlichkeitsbasiertes Gitter ohne Überlagerung der Sampleobjekte. Ziel der Entwickler ist es die Vorteile des Gitters (Navigationshilfe, gute Übersicht, welche Samples bereits angehört wurden) mit den Vorteilen der Visualisierung von Ähnlichkeit mittels Abstand in einem Userinterface zu vereinen. Dazu verwenden die Entwickler den „proximity grid“ [37] Algorithmus, der nach der Dimensionsreduktion mittels t-SNE [33] jeder 2D Position die nächstgelegene Gitterposition zuweist.

Die minimale Seitenlänge des Gitters ist die aufgerundete Quadratwurzel der Sampleanzahl. Mit einer solchen Seitlänge benötigt das Gitter den geringsten Platz, da die Sampleanzahl der Anzahl der Gitterpunkte entspricht. Diese minimale Seitenlänge hat aber eine starke Verzerrung der Ähnlichkeitsverhältnisse zur Folge. Die maximale Seitenlänge des Gitters entspricht der Sampleanzahl. Ein Gitter mit maximaler Seitenlänge weist die geringste Verzerrung der Ähnlichkeitsverhältnisse auf, hat aber den größten Platzverbrauch. Die Entwickler erzielten einen Kompromiss aus minimalen Platzverbrauch und geringer Verzerrung der Ähnlichkeitsverhältnisse indem sie jede Gitterseitenlänge zwischen der minimalen und maximalen Seitenlänge ausprobieren und untersuchen bei welcher Gitterseitenlänge zumindest alle Paare von Feature Vektoren mit geringstem Abstand (im hochdimensionalen Feature Raum) auf benachbarte Gitterpositionen zugewiesen werden. Diese Gitteranordnung erhält den Namen „metro“, da sie wie ein schematischer U- Bahn Plan aussieht. Das „metro“ User Interface verwendet die selbe Visualisierung für die einzelnen Sample-Objekte, die für die zweite Studie verwendet wurde.

In der vierten Studie werdend drei verschiedene User Interfaces getestet: „grid, album,metro“ [2]. „grid“ entspricht der alphabetischen Gitteranordnung, „album“ derselben Gitteranordnung erweitert durch die Visualisierung, die für die zweite Studie entwickelt wurde. Pro Interface führten die 16 Teilnehmerinnen drei Know Item Suchen in drei verschiedenen, durch Schlagwörter gefilterte Samplesammlungen durch. Für die Interfaces „grid“ und „album“ erhielten die Teilnehmerinnen den Hinweis sich die Samples Zeile für Zeile anzuhören. Das „metro“ User Interface wurde den Teilnehmern

mit der Metapher eines U – Bahn Plans erläutert. Benachbarte, ähnlich aussehende Sampleobjekte bilden eine U- Bahn Line. Die Teilnehmerinnen wurden aufgefordert sich die Samples entlang dieser U- Bahn Linien anzuhören. Abbildung 18 zeigt links die „album“ Benutzeroberfläche, rechts die „metro“ Benutzeroberfläche.

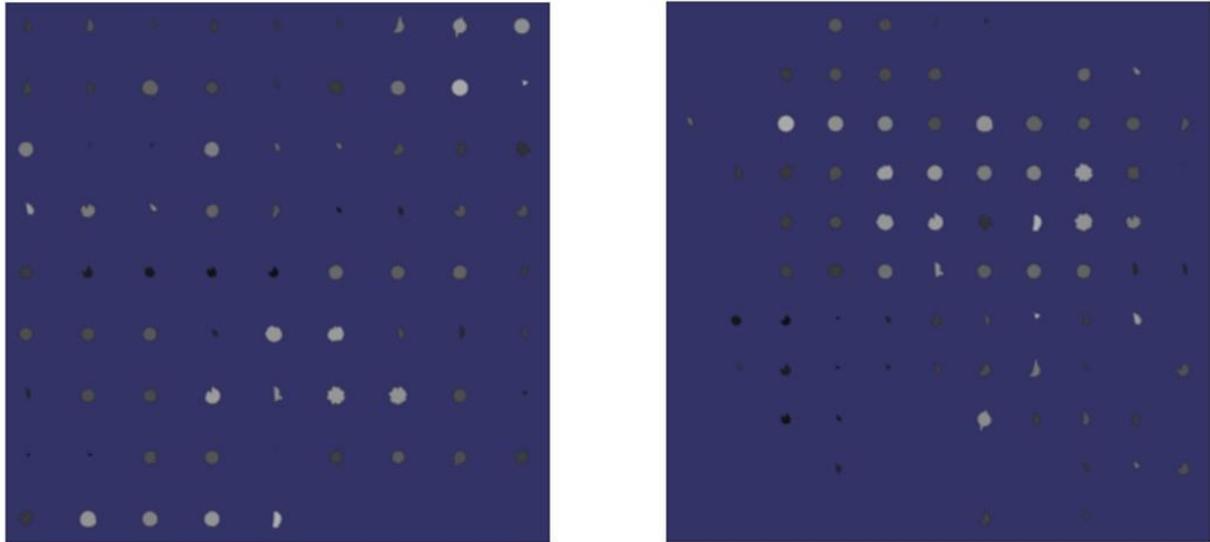


Abbildung 18: Links „album“ Benutzeroberfläche, Rechts „metro“ Benutzeroberfläche

Die Teilnehmerinnen bewerteten „efficiency and pleasurability“ [2] aller drei Interfaces. Beide Bewertungen gewinnt das „metro“ layout mit deutlichem Abstand. Mit statistischer Signifikanz wird mit dem „metro“ Interface die geringste durchschnittliche Suchzeit pro Sample erzielt.

Generell ist fraglich ob eine Know Item Suche eine geeignete Evaluierungsmethode für Samplebrowser ist. Die Autoren selbst halten fest, dass „...sound designers may not necessarily picture target sounds accurately in their head or be able to name further characteristics, leaning closer towards exploratory search.“ [2] Diese Aussage stellt das gesamte Testdesign in Frage. Eine explorative Suche ist dadurch gekennzeichnet, dass die Anwenderin (wie vorhergehenden Zitat beschrieben) keine klare Vorstellung davon hat wie das gesuchte Sample klingt, sie sucht nicht nach genau einem bestimmten Sample. Qualitative Beschreibungen wie die Teilnehmerinnen die verschiedenen getesteten Interfaces für eine explorative Suche einsetzen sind in [2] nicht zu finden.

#### 2.1.10 Drumspace [6]

Musiker setzen Rhythmusmuster aus einzelnen Schlagzeug Samples zusammen. Sie wählen dabei aus Drum-Samplesammlungen die teilweise pro Schlaginstrument 1000 - 10000 Samples umfassen. Die Entwickler von Drumspace konzentrieren sich auf diesen Use Case. Sie entwickeln und testen drei verschiedene Benutzeroberflächen für die Drum-Sample Auswahl während der Erstellung eines Rhythmusmusters: eine listenbasierte Benutzeroberfläche, ein alphabetisch sortiertes 2D Gitter und eine ähnlichkeitsbasierte 2D Karte.

Die listenbasierte Benutzeroberfläche von Drumspace stellt für die Entwickler den in der Praxis vorherrschenden Benützeroberflächentypus für eine Sample-Suche dar: „...the list view resembles as closely as possible the library browsers used in most DAWs: 1 dimension, alphabetical order, with the file names visible.“[6] (DAW = Digital Audio Workstation) Bewegt die Anwenderin die Maus über ein Element der Liste wird das Sample abgespielt. In der 2D Gitter Benutzeroberfläche werden alle Samples mit dem gleichen Anfangsbuchstaben in einer Zeile dargestellt. Für die ähnlichkeitsbasierte Karte werden die Features der TimbreToolbox [38] und der t-SNE [33] Algorithmus für die Dimensionsreduktion verwendet. In den beiden 2D Benutzeroberflächen wird jedes Sample als

einfarbiger Kreis visualisiert. Die Anwenderin kann ein Sample abspielen, indem sie den Mouse Cursor über einem solchen Kreis platziert. Abbildung 19 zeigt links die listenbasierte Benutzeroberfläche, in der Mitte das alphabetisch sortierte 2D Gitter und rechts die ähnlichkeitsbasierte 2D Karte.

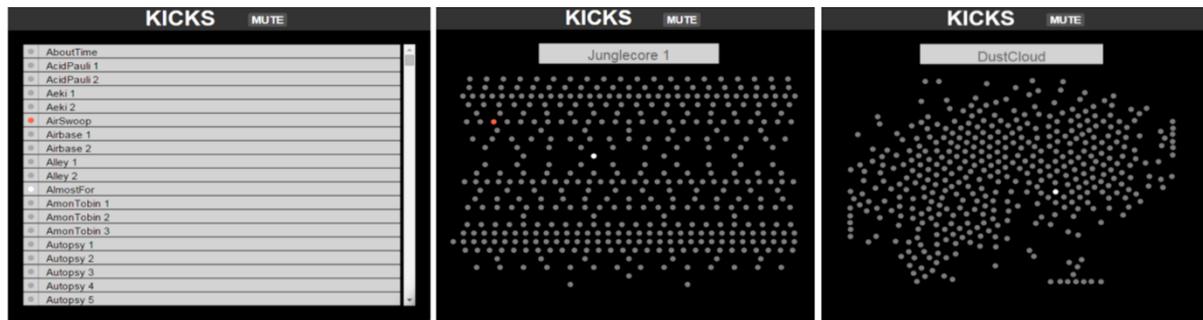


Abbildung 19: Links listenbasierte Benutzeroberfläche, Mitte alphabetisches Gitter, Rechts ähnlichkeitsbasierte 2D Karte

Im Drumspace Prototypen bestehen die Rhythmus muster aus vier verschiedenen Schlagzeuginstrumenten: „Kick, Snare, Open Hi-Hat and Closed Hi-Hat“ [6]. Jede der drei entwickelten Benutzeroberflächen ist in vier Bereiche aufgeteilt, in der die Anwenderin eines dieser 4 verschiedenen Schlagzeuginstrumente auswählen kann. Ein in der Mitte platzierter roter Button startet und stoppt das ausgewählte Rhythmusmuster. Platziert die Anwenderin die Mouse über einem Sample wird es abgespielt, ein Mausklick auf ein Sample und das Sample wird im Rhythmusmuster anstatt des bisher abgespielten Schlaginstrumentes verwendet. Abbildung 20 zeigt die Benutzeroberfläche von Drumspace mit ähnlichkeitsbasierten 2D Karten.

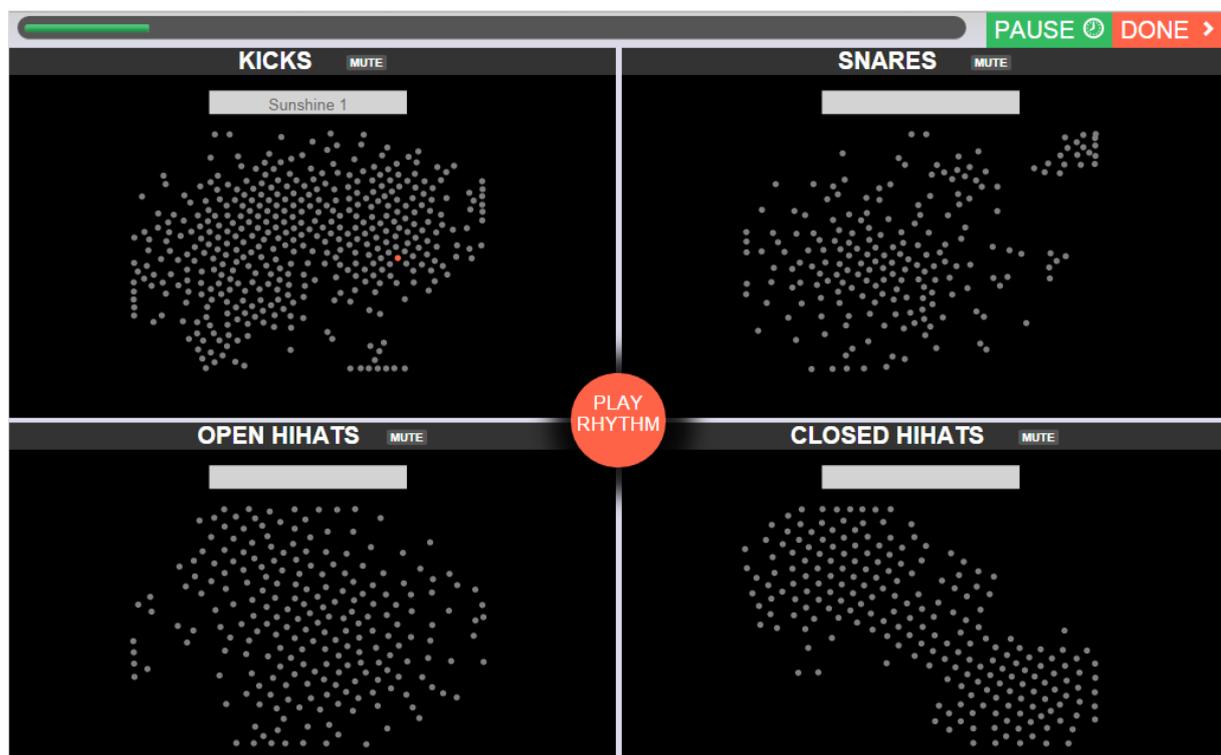


Abbildung 20: Drumspace Benutzeroberfläche mit 4 ähnlichkeitsbasierten 2D Karten für die vier im Rhythmusmuster verwendeten Schlaginstrumente Kicks, Snare, OpenHighhats und Closed Hights

Die Entwickler von Drumspace führten Tests mit 20 Teilnehmerinnen („18 musicians and 2 non musicians“ [6]) durch. Insgesamt mussten die Teilnehmerinnen für 9 Rhythmusmuster passende Schlagzeug Samples finden. Die drei Interfaces wurden dabei abwechselnd verwendet. Die

Teilnehmerinnen hatten pro Rhythmusmuster drei Minuten Zeit für jedes der vier Instrumente des Rhythmusmusters („Kick, Snare, Open Hi-Hat and Closed Hi-Hat“) ein Sample zu finden.

Nach der Erstellung der Rhythmusmuster wurden unstrukturierte Interviews mit den Teilnehmerinnen durchgeführt. Zum Abschluss der ersten Untersuchung wurden die Teilnehmer gefragt welches der drei Interfaces sie bevorzugten. Jeweils 8 Teilnehmer bevorzugten die alphabetisch sortierte 2D Karte bzw. die ähnlichkeitsbasierte 2D Karte. Nur 3 Teilnehmer wählten das listenbasierte Interface. Ein Teilnehmer bekundete keinerlei Präferenz für eins der drei Interfaces.

Mit dem listenbasierten Interface hörten sich die Teilnehmer wesentlich weniger Samples an als mit den beiden 2D Visualisierungen. Für das listenbasierte Interface ergab sich während der Test ein Mittelwert  $M$  an abgespielten Samples von 46,9 pro Rhythmusmuster. Mit der alphabetisch sortierten 2D Karte ( $M = 93.2$ ) und der ähnlichkeitsbasierten 2D Karte ( $M = 101.2$ ) hörten sich die Teilnehmerinnen ungefähr die doppelte Menge an Samples an.

Statistische Analysen zeigten auch, dass die Teilnehmerinnen mit der alphabetisch sortierten Liste vermehrt Samples auswählen deren Anfangsbuchstabe sich im vorderen Teil des Alphabetes befinden. Die Anwenderinnen springen nicht in der List hin und her, sondern hören sich die Samples in der Reihenfolge an, die die Sortierung der List vorgibt. Sie schränken damit ihre Suche unfreiwillig auf bestimmte Samples ein.

Nachdem die Teilnehmerinnen die Samples für ein bestimmtes Rhythmusmuster ausgewählt hatten bewerteten sie auf einer dreistufigen Skala wie zufrieden sie mit der Auswahl der Samples waren. Die statistische Auswertung dieser Daten zeigt, dass die Teilnehmer den entstandenen Rhythmus besser bewerteten, wenn sie die alphabetische 2D Karte verwendeten. Eine statistische Signifikanz konnte für dieses Ergebnis aber nicht nachgewiesen werden.

Aufgrund der oben beschriebenen Testergebnisse halten die Entwickler fest, „...that a 2D visualization provides a better experience than a traditional list interface and seems to improve the exploration. 2D visualizations increase the diversity of samples chosen“ [6] Im Vergleich mit dem listenbasierten Interface bieten die 2D Visualisierungen einen Überblick über die gesamte Samplesammlung und einen wesentlich direkteren Zugriff auf alle Samples. Das Ansteuern aller Samples in einer Liste erfordert wesentlich mehr Mausbewegungen der Anwenderin. Mit den beiden 2D Karten sind die Anwenderinnen in der Lage eine größere Anzahl an Samples in der gleichen Zeitspanne anzuhören. In den qualitativen Interviews beschreiben die Teilnehmer die 2D Karten im Vergleich zu dem listenbasierten Interface mit den Worten „...“easier“, “faster“, “more pleasant“, and “more fun“.“ [6]

Bezüglich des Vergleichs zwischen der ähnlichkeitsbasierten 2D Karte und der alphabetisch sortierten 2D Karte stellen die Autoren aber fest, dass die Ergebnisse „... do not show any significant improvement brought by the arrangement by perceptual similarity“ [6] In den Interviews beschreiben einige Teilnehmerinnen die ähnlichkeitsbasierte Karte als schwer verständlich und ineffizient. „This visualization represents all samples identically and displays no axis or any indicator of the timbre similarity. It lacks transparency and does not allow the user to understand the organization of the sample space. Experimenting with more explicit representations would be necessary in order to design a more intuitive timbre space interface. For instance, the use of colors to indicate some timbral features might improve the understanding of the similarity clustering“ [6]

Warum die Anwenderinnen die ähnlichkeitsbasierten Karten als schwer verständlich und ineffizient empfinden wird in [6] nicht erörtert. Wesentlich diesbezüglich ist die Frage ob den Teilnehmerinnen vor den Tests die Funktionsweise der ähnlichkeitsbasierten 2D Karte erklärt wurde. Zusätzlich ist fraglich ob die Teilnehmerinnen während einer Gewöhnungsphase Zeit hatten die semantische Struktur der Karten zu erlernen. Hatten die Teilnehmerinnen vor den Tests die Chance festzustellen

welche, wie klingenden Samples wo auf der ähnlichkeitsbasierten Karte zu finden sind? In [6] ist weder vermerkt, dass die Funktionsweise der ähnlichkeitsbasierten Karte erklärt wurde noch wird eine Gewöhnungsphase vor den Tests erwähnt. Da es keine kommerziellen Anwendungen gibt die eine Samplesuche durch eine ähnlichkeitsbasierte 2D Karte unterstützen ist anzunehmen, dass die meisten Versuchspersonen zum ersten Mal mit einer solchen Karte eine Samplesuche durchgeführt haben. Die Beschreibung des Testdesigns zeigt das den Usern pro Rhythmusmuster nur drei Minuten Zeit gegeben wurde um vier passende Samples zu finden. Insgesamt führten die User in einem für sie neuartigen User Interface drei Samplesuchen a maximal drei Minuten Dauer durch. Die Teilnehmer hatten also durchschnittlich nur 45 Sekunden Zeit ein bestimmtes Sample für ein Schlaginstrument auszuwählen. Ob Anwenderinnen in der Lage sind die semantische Struktur der Karten im Laufe der Zeit zu erlernen kann damit nicht beantwortet werden.

#### 2.1.11 Konzeptioneller Aufbau von ähnlichkeitsbasierten Sample-Browsern

Nach einer inhaltsbasierten Merkmalsextraktion führen ähnlichkeitsbasierte Sample-Browser eine Dimensionsreduktion durch. Danach werden alle Samples als grafische Objekte visualisiert und der Anwenderin eine Sonifikation dieser Visualisierung zur Verfügung gestellt. Tabelle 1 listet diese 4 gemeinsamen Merkmale (Inhaltsbasierte Features, Dimensionsreduktion, Visualisierung und Sonifikation) für alle, in dieser Arbeit behandelten Sample-Browser auf.

Sample Browser	Inhaltsbasierte Features	Dimensionsreduktion	Visualisierung	Sonifikation
Sonic Browser [12]	Keine	Keine	Samples als grafische Objekte in einem 2 Achsen Streudiagramm mit Zuweisung von Sample-Eigenschaften wie z.B. Dateigröße, Dateityp, Samplingrate und Aufnahmedatum auf die beiden Achsen. Visualisierung derselben Sample-Eigenschaften auch als Größe, Farbe, Form oder Position der Sample-Objekte	Aura – eine bestimmte Anzahl an Samples im Umkreis der Mausposition werden gleichzeitig abgespielt
MARSYAS3D [13]	Marsyas Framework [14] wie z.B. MFCC, Spectral Centroid, Spectral Flux	Principal Component Analysis (PCA)	Samples als Rechtecke in einer ähnlichkeitsbasierten 2D oder 3D Karte. Sample-Rechtecke je nach Gruppenzugehörigkeit (ermittelt mit k-Means Algorithmus) farbcodiert.	Aura
Audio Retrieval Browser [15]	Marsyas Framework [14] wie z.B. MFCC, Spectral Centroid, Spectral Flux	Principal Component Analysis (PCA)	Sample als grafische Objekte in einer ähnlichkeitsbasierten 2D Karte oder Samples als grafische Objekte in einem 2 Achsen Streudiagramm mit Zuweisung von Sample-Eigenschaften auf die beiden Achsen	Aura
SoundTorch [9, 16]	MFCC	Self-organizing Map (SOM)	Samples als Punkte in einer ähnlichkeitsbasierten 2D Karte, Visualisierung der Sample-Waveform im Umriss der Sample -Punkte	Aura (Lichtkegel Metapher)
Sound Navigator [7, 17]	229 verschiedene Features, die im Rahmen des CUIDADO Projekts [26] oder für die inhaltsbasierte Verschlagwortung von Samples [27] implementiert oder entwickelt wurden wie z.B. MFCC	Principal Component Analysis (PCA) oder Multidimensional Scaling (MDS) Variante [7]	Sample-Segmente als Punkte in einer ähnlichkeitsbasierten 2D Karte oder 2 Achsen Streudiagramm mit Zuweisung von Feature-Werten auf die beiden Achsen, Visualisierung einzelner Feature-Werte mittels Farbe der Sample-Segmente	„search-by-interaction paradigm“ [7] Mausbewegungen der Anwenderin triggern mehrere Sample-Segmente in rascher Abfolge
MediaCycle [19], AudioCycle [18]	Harmonie-, Rhythmus- und Klangfarben-Features (z.B. MFCC und mehrere Features, die für das CUIDADO Projekt [26] implementiert oder entwickelt wurden)	Clustering mittels k-Means Algorithmus	Anordnung von Samples in farbkodierten Gruppen rund um ein Referenz-Sample. Der Abstand eines Samples zum Referenz-Samples visualisiert die Ähnlichkeit dieser beiden Samples	Mausklick
Grill und Flexer [20, 29, 30]	Features [30], die 5 Wahrnehmungsqualitäten („personal constructs“ [29]) von Samples quantifizieren	t- SNE [33]	Samples als Punkte in einer ähnlichkeitsbasierten 2D Karte, Visualisierung der 5 Wahrnehmungsqualitäten in einem mit kleinen, grafischen Objekten („tiles“) ausgefüllten Hintergrund. Jede der 5 Wahrnehmungsqualitäten wird durch eine bestimmte grafische Eigenschaft der tiles dargestellt. [20]	Mouse Over
Snare Drum Navigator [21]	MFCC, Log Attack Time, Temporal Centroid,	Kernelized Sorting [34]	Samples als Rechtecke in einem ähnlichkeitsbasierten 2D Gitter, Visualisierung der Ähnlichkeit mittels Farbe der Sample-Rechtecke	Mouse Over
Audio Metro [2]	MFCC, Spectral Flatness	t-SNE [33] + Proximity Grid [37]	Samples als Objekte in einem ähnlichkeitsbasierten 2D Gitter, Visualisierung der Spectral Flatness mittels Helligkeit und der Umrisslinie der Sample-Objekte	Mouse Over
Drumspace [6]	Mehrere Features der Timbre Toolbox [38]	t-SNE [33]	Samples als Punkte in einer ähnlichkeitsbasierten 2D Karte	Mouse Over

Tabelle 1: Inhaltsbasierte Features, Dimensionsreduktion, Visualisierung und Sonifikation der in dieser Arbeit behandelten Sample-Browser

## 2.2 Inhaltsbasierte Merkmalsextraktion

Im Forschungsbereich MIR (Music Information Retrieval) bezeichnet der Begriff „inhaltsbasierte Merkmalsextraktion“ (Contentbased Feature Extraction) die Berechnung von  $n$  Kennzahlen (= Features) ausschließlich aus den digitalen Daten einer Audiodatei. Wie viele andere MIR Anwendungen berechnen ähnlichkeitsbasierte Sample-Browser pro Audiodatei einen  $n$ -dimensionalen Feature-Vektor, der bestimmte akustische und klangliche Eigenschaften der Audiodatei in  $n$  Zahlen abbildet. Diese Feature-Vektoren können die aus psychoakustischer Sicht relevanten Eigenschaften wie Klangfarbe, Lautheit, Schärfe, Tonheit, Rauigkeit, Tonhaltigkeit, Impulshaltigkeit und Schwankungsstärke quantifizieren. In den letzten zwei Jahrzehnten hat die globale MIR Forschungsgemeinde eine große Anzahl an Features entwickelt, die (im Verbund mit diversen Machine Learning Algorithmen) auch musikalische Dimensionen von Samples beschreiben können wie z.B. Tempo, Rhythmus, Tonart, Harmonie und Melodie.

Anhand der (für die jeweilige Anwendung verwendeten) inhaltsbasierten Features lässt sich feststellen welche Sample-Eigenschaft die in dieser Arbeit behandelten Sample-Browser auf ihren ähnlichkeitsbasierten 2D Karten visualisieren. Tabelle 2 zeigt diesen Zusammenhang. Die meisten hier behandelten Sample-Browser setzen inhaltsbasierte Features ein, die verschiedenste MIR Forschungsbereiche zur Modellierung von Klangfarben verwenden wie z.B. MFCCs (Mel Frequency Cepstral Coefficients). Der überwiegende Teil der Sample-Browser ermöglicht daher nur eine ähnlichkeitsbasierte Suche anhand der Klangfarbenunterschiede der einzelnen Samples.

Sample Browser	Sample- Eigenschaft die in der ähnlichkeitsbasierte 2D Karte visualisiert wird
MARSYAS3D [13]	Klangfarbe
Audio Retrieval Browser [15]	Klangfarbe
Sound Torch [9, 16]	Klangfarbe
Sound Navigator [7, 17]	Auswahl der Features durch die Anwenderin bestimmt welche Art der Ähnlichkeitsbeurteilung ermöglicht wird
MediaCycle [19], AudioCycle [18]	Klangfarbe, Harmonie, Rhythmus
Grill und Flexer [20, 29, 30]	5 Wahrnehmungskonstrukte wie z.B. tonal - noisy
Snare Drum Navigator [21]	Klangfarbe
Audio Metro [2]	Klangfarbe
Drumspace [6]	Klangfarbe

Tabelle 2: Sample-Browser und durch inhaltsbasierte Features ermöglichte Ähnlichkeitswahrnehmung

Im Folgenden werden im Abschnitt „*Konzeptioneller Aufbau von inhaltsbasierten Merkmalsextraktoren*“ die algorithmische Verarbeitungskette beschrieben, die bei der Berechnung inhaltsbasierter Features typischerweise verwendet wird. Eine Übersicht über alle inhaltsbasierten Features der behandelten Sample-Browser sprengt den Rahmen dieser Arbeit. Der Abschnitt „*Inhaltsbasierte Features für die Modellierung von Klangfarben*“ behandelt daher nur MFCCs (Mel Frequency Cepstral Coefficients) da diese in einer Vielzahl an MIR Anwendungen erfolgreich zur Modellierung von Klangfarben verwendet werden und einige Features, die durch psychoakustische Studien als Signaleigenschaften identifiziert wurden, die die Klangfarben Ähnlichkeitsbeurteilung wesentlich mitbestimmen.

Die Wahrnehmung von Klangfarben wird in der Psychoakustik seit Jahrzehnten erforscht. Die MIR Forschungsgemeinde beschäftigt sich seit ca. 17 Jahren intensiv mit der Modellierung von Klangfarben, da diesen Modellen in einigen Teilbereichen der MIR Forschung eine zentrale Bedeutung zukommt. Vor allem die erzielten Ergebnisse im MIR Forschungsbereich Instrumentenerkennung lassen Rückschlüsse auf die Benutzerfreundlichkeit der hier behandelten Sample-Browser zu, da in beiden Fällen dieselben Features zur Modellierung von Klangfarben verwendet werden. Diese Ergebnisse und einige Erkenntnisse der psychoakustischen Klangfarben-Forschung werden im Abschnitt „*Klangfarbe – Psychoakustik und MIR Forschung*“ diskutiert.

### 2.2.1 Konzeptioneller Aufbau von inhaltsbasierten Merkmalsextraktoren

Die Features der hier behandelten Sample-Browser werden in den meisten Fällen entweder aus zeit-basierten oder frequenz-basierten Darstellungen eines digitalen Audiosignals berechnet. Abbildung 21 zeigt links eine zeit-basierte Darstellung (Waveform) und rechts eine frequenz-basierte Darstellung (Spektrale Hüllkurve) eines Violinen-Samples und eines Basstrommel-Samples. Die frequenz-basierte Darstellung zeigt harmonische und spektrale Eigenschaften der beiden Samples, die in der Zeit-basierten Darstellung nicht erkannt werden können. In der frequenz-basierten Darstellung des Violinen-Samples sind die harmonischen Obertonschwingungen erkennbar. Die frequenz-basierte Darstellung des Basstrommel-Samples zeigt, dass das Audiosignal im Wesentlichen aus Frequenzanteilen unter 100 Hz besteht.

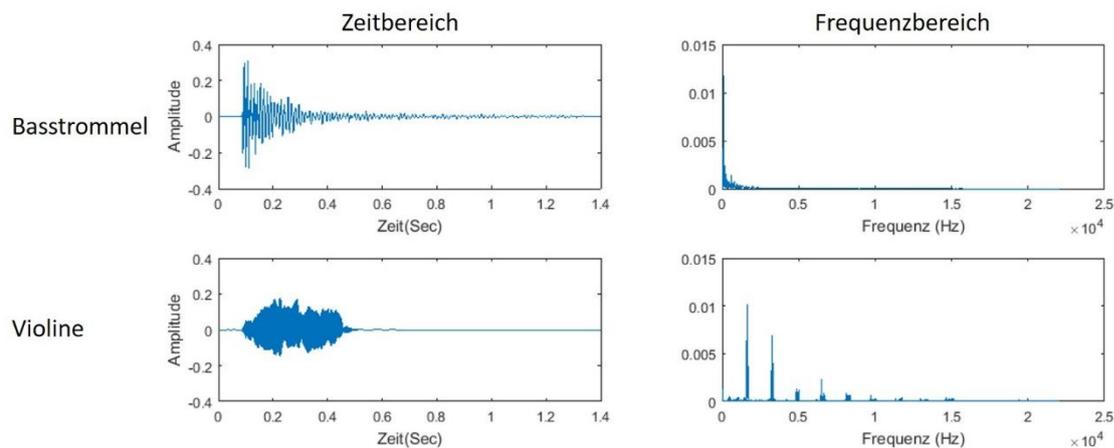


Abbildung 21: Zeit- und Frequenz-basierte Darstellung eines Basstrommel-Samples und eines Violinen-Samples

Beide für Abbildung 21 verwendeten Audiodateien sind mono Audiodateien mit einer Abtastfrequenz von 44,1 kHz. Pro Sekunde bestehen diese beiden Dateien daher aus 44100 Abtastwerten. Für die zeit-basierte Darstellung werden diese Abtastwerte auf einer Zeitachse aufgetragen. Die frequenz-basierte Darstellung wird zumeist mit einer Fourier Transformation aus denselben Abtastwerten berechnet.

Abbildung 22 zeigt konzeptionell den algorithmischen Ablauf bei der Berechnung eines inhaltsbasierten Features (vgl. die umfassendere Darstellung in [39]). Ausgangspunkt ist das mono Violinen-Sample mit 44 100 Abtastwerten pro Sekunden. Erster Verarbeitungsschritt ist das sogenannte „Frameing“ [39]. Aus dem Zahlenstrom der Audiodatei werden kurze aufeinanderfolgende, meist überlappende Segmente (=Frames) entnommen. Für das Framing werden üblicherweise Frames mit einer Dauer von 10 – 50 Millisekunden verwendet. „As the time resolution of the human ear is around 10ms, choosing a frame size that covers a much smaller time range does not make sense.“ [39]. Abbildung 22 (Oben) zeigt einen 0,2 Sekunden langen Ausschnitt aus dem Violinen-Sample mit 17 halbüberlappenden Frames mit einer jeweiligen Dauer von ca. 23 Millisekunden. Die Berechnung von Features aus der zeit-basierten Darstellung dieser Frames (wie z.B. Log Attack Time oder Zero Crossing Rate) benötigt keine weiteren Signaltransformationen. Features, die aus frequenz-basierten Darstellungen der Frames berechnet werden, erfordern als weitere Verarbeitungsschritte die Anwendung einer Fensterfunktion („Windowing“ [39]) und eine anschließende Fourier Transformation. Nach der Fourier Transformation steht pro Frame eine frequenz-basierte Darstellung für die Berechnung verschiedenster Features zur Verfügung. Aus diesen Kurzzeit, frequenz-basierten Signaldarstellungen werden Features, die spektrale und harmonische Eigenschaften von Samples quantifizieren (wie z.B. MFCCs und Spectral Centroid), berechnet. Da Samples eine unterschiedliche Dauer haben werden sie in eine unterschiedliche Anzahl an Frames aufgeteilt. Die nachfolgende Verarbeitung von Feature- Vektoren durch Dimensionsreduktions-

Algorithmen erfordert aber Feature-Vektoren mit gleicher Dimensionalität. Als letzter Verarbeitungsschritt werden daher die für jeden Frame berechneten Feature-Werte (häufig durch statistische Lagemasse wie den Durchschnitt oder die Varianz) aggregiert.

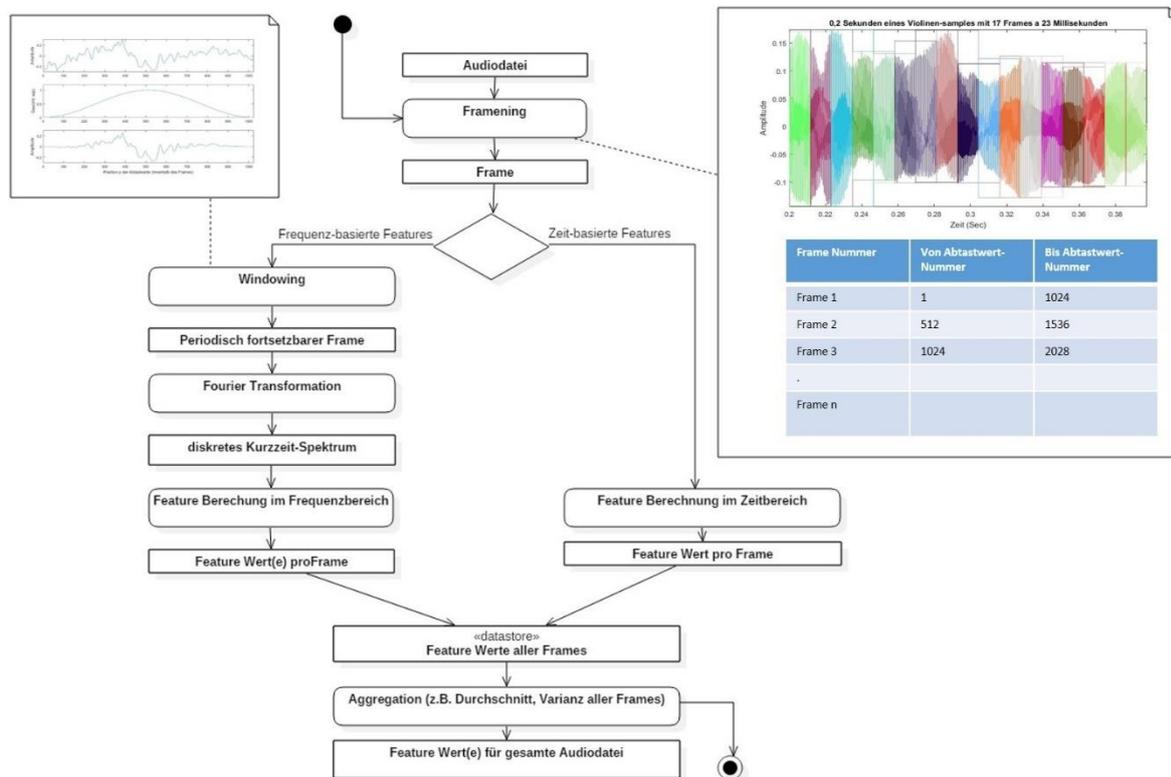


Abbildung 22: Konzeptioneller Aufbau von inhaltsbasierten Merkmalsextraktoren

Da harmonische und spektrale Eigenschaften eines Audiosignals nicht direkt aus zeit-basierten Darstellungen extrahiert werden können, wird für jeden Frame eine frequenz-basierte Darstellung berechnet. Für digitale Audiosignale wird die frequenz-basierte Darstellung der einzelnen Frames meistens mit dem FFT (Fast Fourier Transform) Algorithmus berechnet. Der FFT Algorithmus ist eine rechnerisch effiziente Implementierung der DFT (Diskrete Fourier Transformation) und berechnet pro Frame ein diskretes Kurzzeit-Spektrum. Die Aneinanderreihung dieser Kurzzeit-Spektren entlang einer Zeitachse ermöglicht die Darstellung eines Audiosignals als Spektrogramm.

Ein Spektrogramm zeigt den zeitlichen Verlauf der Frequenzanteile eines Audiosignals und farbkodiert den Schalldruck Pegel (dB) der einzelnen Frequenzanteile. Abbildung 23 zeigt das Spektrogramm für das Violinen- und das Basstrommel-Sample. Für die Berechnung wurden halbüberlappende Frames mit einer Framelänge von 1024 Samples (~ 23ms) und eine Hamming-Fensterfunktion verwendet. Der zeitliche Verlauf des Schalldruckpegels für die jeweiligen Frequenzanteile ist in beiden Spektrogrammen gut erkennbar.

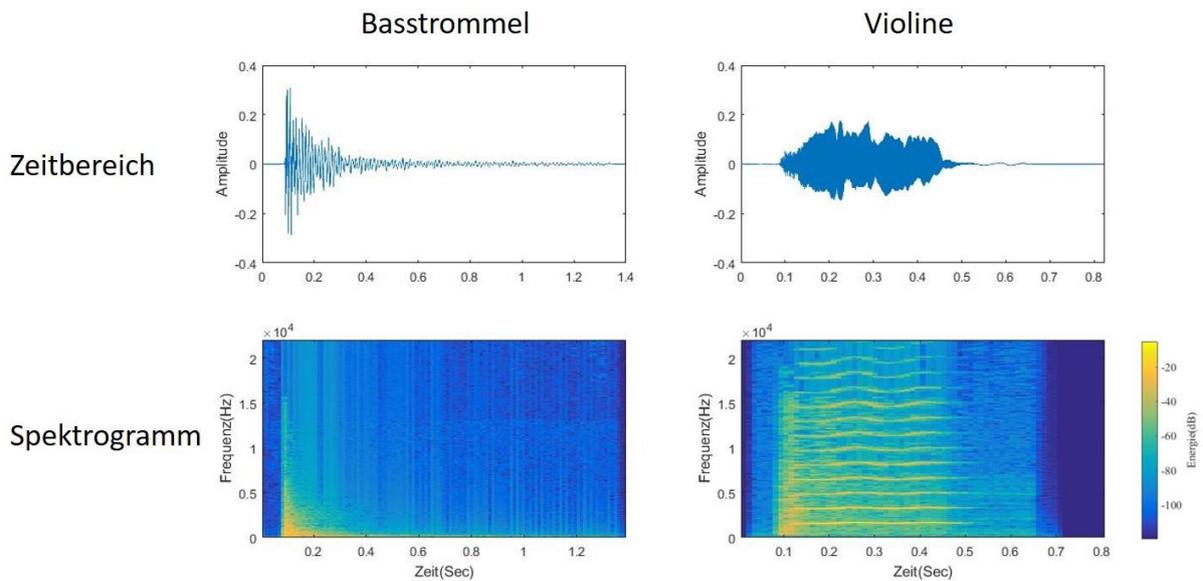


Abbildung 23: Zeit-basierte Darstellung und Spektrogramm eines Basstrommel-Samples und eines Violinen-Samples

Der maximale Frequenzbereich eines (durch eine FFT berechneten) Spektrums entspricht der Hälfte der Abtastfrequenz. Wurde ein Frame mit einer Abtastfrequenz von 44.1 kHz digitalisiert deckt das Amplitudenspektrum des Frames maximal den Frequenzbereich von 0 bis 22.05 Hz ab. Konzeptionell gesehen teilt die FFT diesen Frequenzbereich in gleich große Frequenzbereiche auf und berechnet für jeden dieser Frequenzbereiche einen Energiewert. Um eine effiziente Berechnung der FFT zu gewährleisten wird die Anzahl dieser Frequenzbereiche häufig als nächstgelegene 2er Potenz der Hälfte der Abtastwerte pro Frame festgelegt. Einen Frame mit 1024 Samples teilt die FFT z.B. in 512 gleichgroße Frequenzbereiche. Die Größe dieser Frequenzbereiche entspricht der Abtastfrequenz dividiert durch die Anzahl der Abtastwerte pro Frame. Für einen Frame der aus 1024 Abtastwerten besteht und mit einer Abtastfrequenz von 44.1 kHz digitalisiert wurde ergeben sich damit  $512 \sim 43,1$  Hz große Frequenzbereiche, die im Bereich von 0 bis 22.05 kHz angeordnet sind.

Eine DFT kann nur periodisch fortsetzbare Signale korrekt in die Frequenz-basierte Darstellung transformieren. Wird ein nicht-periodisch fortsetzbares Signal transformiert treten bei der DFT Berechnung Artefakte auf. Die Frequenz-basierte Darstellung eines Frames enthält in diesem Fall einige, im ursprünglichen Signal nicht vorkommende, Frequenzanteile. Vor der Transformation eines Frames in den Frequenzbereich wird daher durch die Anwendung einer Fensterfunktion auf die Abtastwerte eines Frames (= „Windowing“ [39]) das Auftreten dieser Artefakte vermindert. Die Fensterfunktion verwandelt den Frame in ein periodisch fortsetzbares Signal. Häufig werden Hann-, Hamming- oder Blackman- Fensterfunktionen für das Windowing verwendet.

Abbildung 24 zeigt den Windowing Prozess für einen Frame des Violinen-Samples. Abbildung 24 (Oben) zeigt einen Frame des Violinen-Samples, der 1024 Abtastwerte umfasst. Da das Violinen-Sample mit einer Abtastfrequenz von 44.1 kHz digitalisiert wurde entsprechen die 1024 Abtastwerte ungefähr einer Famedauer von 23 Millisekunden. Das Signal des Frames ist nicht periodisch fortsetzbar. Die Multiplikation mit der Fensterfunktion (Hann-Funktion) in Abbildung 24 (Mitte) bewirkt eine Abschwächung des Signals an beiden Enden des Frames, die in Abbildung 24 (Unten) erkennbar ist und verwandelt den Frame in ein periodisch fortsetzbares Signal. Diese Abschwächung führt zu einem Informationsverlust bei der anschließenden Umwandlung in den Frequenzbereich. Um diesen Informationsverlust zu vermeiden werden aufeinander folgende Frames überlappend aus dem Zahlenstrom des Audiosignals entnommen.

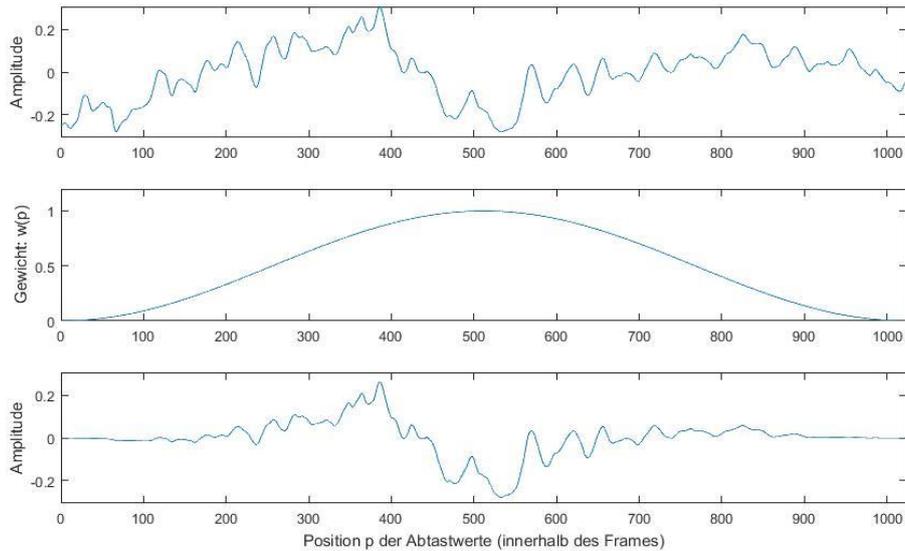


Abbildung 24: Anwendung einer Fensterfunktion auf einen Frame eines Geigen-Samples mit 1024 Samples bzw. ca. 23ms Dauer

Nach der Fourier Transformation werden pro Frame spektrale oder harmonische Features berechnet. Die abschließende Aggregation dieser intraframe Features durch z.B. statistische Lagemasse wie den Durchschnitt oder die Varianz der intraframe Features ermöglicht die Repräsentation jedes Samples als n-dimensionaler Vektor. Die zeitliche Abfolge der Frames wird bei dieser Vorgehensweise ignoriert. Informationen wann welcher Frequenzanteil mit welchem Schalldruck Pegel innerhalb eines Samples auftritt sind in diesen Feature-Vektoren nicht mehr enthalten. Der ideale Anwendungsfall für eine derartige intraframe Aggregation sind daher Samples, deren spektrale Zusammensetzung oder harmonischen Eigenschaften sich während ihrer gesamten Dauer nicht verändern.

### 2.2.2 Inhaltsbasierte Features für die Modellierung von Klangfarben

In den letzten zwei Jahrzehnten sind eine Fülle an inhaltsbasierten Features entwickelt worden. Eine Übersicht über den Forschungsstand bis ins Jahr 2010 bieten Mitrović et al., in [40]. Die Autoren erstellen eine Taxonomie, die einen Vergleich von inhaltsbasierten Features anhand von formalen und semantischen Eigenschaften ermöglicht. Die Darstellung einiger Features der behandelten Sample-Browser orientiert sich an einigen Aspekten dieser Taxonomie. Die Autoren einer aktuelleren Übersicht [41] aus dem Jahr 2016 verwenden eine ähnliche, auf der von Mitrović et al. aufbauende Taxonomie. Diese Übersicht beinhaltet die wesentlichen Neuentwicklungen auf dem Gebiet der inhaltsbasierten Merkmalsextraktion bis ins Jahr 2016.

Aufgrund der unüberschaubaren Anzahl an inhaltsbasierten Features hat es in den letzten 17 Jahren immer wieder Initiativen gegeben inhaltsbasierte Features in Standards zu definieren oder mehrere Features gebündelt als Open Source Software zur Verfügung zu stellen. Einer der ersten Initiativen inhaltsbasierte Features zu standardisieren ist der ISO/IEC 15 938-4: MPEG-7 Audio Standard. Der MPEG 7 Standard definiert 17 Features die Tonhöhe, Rauigkeit, Klangfarbe und harmonische Eigenschaften von digitalen Audiosignalen quantifizieren. (vgl. [42]). Im Rahmen des CUIDADO [26] Projekts erweitern Peeters et al., den MPEG-7Standard und definieren zusätzlich 36 neue inhaltsbasierte Features. Die hier behandelten Sample-Browser verwenden fast ausschließlich Features, die in diesen beiden Standards definiert sind. Die einzigen Ausnahmen stellen der Sound Navigator [7,17] und der Sample-Browser von Grill und Flexer [20] dar. Der Sound Navigator verwendet neben den Cuidado Features einige Features, die für die inhaltsbasierte Verschlagwortung von Samples (siehe [27]) entwickelt wurden. Für den Sample-Browser von Grill und Flexer wurden neuartige Features (siehe

[30]) entwickelt, die bestimmte bipolare Wahrnehmungsdimensionen von Samples quantifizieren. Diese beiden Feature-Gruppen sind nicht im MPEG 7 oder CUIDADO Standard definiert.

Ein der ersten Open Source Software Feature-Sammlungen ist das Marsyas Framework [14] (Jahr 2000), dessen Features für den MARSYAS3D [13] Browser und den Audio Retrieval Browser [15] verwendet werden. Die Features des CUIDADO [26] Projekts werden für den Sound Navigator [7, 17] MediaCycle [19] und AudioCycle [18] Browser verwendet. Die Features der Timbre Toolbox [38] kommen im DrumSpace [6] Browser zum Einsatz. Weitere Beispiele für Open Source Software zur inhaltsbasierten Merkmalsextraktion sind Essentia [43], die MIR Toolbox [35], LibXtract [44] und YAAFE [32]. Eine Evaluierung von Open Source Software für die inhaltsbasierte Merkmalsextraktion bietet [45]. Als Evaluierungskriterien werden die Anzahl der bereitgestellten Features, die Benutzerfreundlichkeit der jeweiligen Software, die angebotenen Dateiformate für die Ausgabe der Features und die Dauer der Feature-Berechnung verwendet. Die Anzahl der Features wird unter anderem danach beurteilt wie viele der im MPEG-7 und Cuidado Standard definierten Features in der jeweiligen Software implementiert sind. Derzeit bietet Essentia [43] die größte Anzahl an Features und eine effiziente Feature-Berechnung.

#### 2.2.2.1 Aus zeit-basierten Signaldarstellungen berechnete Features

Zeitbasierte Features wie ZCR (Zero Crossing Rate), das temporale Centroid, STE (Short-time energy) und Einschwingdauer (z.B. Log Attack Time) werden direkt aus den Abtastwerten eines Frames berechnet. Unter den zeitbasierten Features weist die log attack time den stärksten Zusammenhang mit der Wahrnehmung von Klangfarben auf (siehe Abschnitt „*Psychoakustische Timbre Space Forschung*“). Für die ZCR ist der Zusammenhang mit der Klangfarben Wahrnehmung des Menschen nur vage definiert: „Periodic Sounds tend to have a small value of it, while noisy sounds tend to have a high value of it“ [26] ZCRs „...do not relate to perceptual processing in a straightforward fashion, but they somehow contribute to computational classification accuracy“ [46]

#### 2.2.2.2 Aus frequenz-basierten Signaldarstellungen berechnete Features

Die überwiegende Mehrheit der für die Modellierung von Klangfarben verwendeten Features werden aus frequenz-basierten Signaldarstellungen berechnet.

##### 2.2.2.2.1 Spectral Centroid

Das spektrale Centroid wurde durch mehrere psychoakustischen Studien als wesentliche Bestimmungsgröße der Klangfarben-Ähnlichkeitsbeurteilung identifiziert. (siehe Abschnitt „*Psychoakustische Timbre Space Forschung*“) Die Berechnung des spektralen Centroids läuft nach dem bereits dargestellten Schema in Abbildung 22 ab. Es wird als Schwerpunkt des Spektrums beschrieben und zeigt an in welchem Frequenzbereich des Spektrums die höchsten Energiekonzentrationen auftreten. (vgl. [40])

##### 2.2.2.2.2 MFCC - Mel Frequency Cepstral Coefficients

MFCC sind das mit Abstand meistverwendete Feature in mehreren Teilgebieten des Forschungsbereichs Machine Hearing wie z.B. der Spracherkennung und der MIR Forschung. Eine unüberschaubare Anzahl an MIR Anwendungen verwenden MFCCs zur Modellierung von Klangfarben. 7 von 11 der hier behandelten Sample-Browsern verwenden dieses Feature.

Das Erfolgsgeheimnis von MFCCs ist einerseits eine Modellierung einiger Eigenschaften der menschlichen, auditiven Wahrnehmung. Andererseits stellen sie eine kompakte, niedrig dimensionale Datenrepräsentation der spektralen Hüllkurve eines einzelnen Frames dar, deren einzelnen Feature-Werte keine hohe Korrelation aufweisen und damit die Anforderungen für eine erfolgreiche Verarbeitung durch verschiedenste Machine Learning Algorithmen erfüllt. Die spektrale Hüllkurve gilt als jene physikalische Signaleigenschaft, welche die Wahrnehmung von Klangfarben am stärksten

beeinflusst (siehe Abschnitt „*Psychoakustische Timbre Space Forschung*“). Das Ergebnis der MFCC Berechnung ist ein Vektor aus  $n$  Koeffizienten. In vielen MIR Anwendungen werden Implementierungen des Algorithmus verwendet die  $n = 13 - 20$  Koeffizienten berechnen. Die Koeffizienten kodieren die spektrale Hüllkurve eines Frames ohne wesentliche Informationsverluste. “The first coefficient represents the average power in the spectrum. The second coefficient represents the broad shape of the spectrum - roughly the spectral centroid. Higher-order coefficients represent the finer details of the spectral shape.” [47]

Die menschliche, auditive Wahrnehmung bildet den hörbaren Frequenzbereich von 20 – 20000 HZ nicht linear ab. Wird die physikalische Frequenz eines Tones (Hz) erhöht erfolgt nicht immer eine lineare Steigerung der wahrgenommenen, subjektiven Tonhöhe. Eine Modellierung dieses Zusammenhangs ist die Mel-Frequenz Kurve in Abbildung 25. Die Frequenz-Mel Kurve verläuft im Bereich unter 500 Hz Bereich annähernd linear. In diesem Bereich entspricht z.B. die Verdopplung der physikalischen Frequenz von 200 auf 400 Hz annähernd der Verdoppelung der wahrgenommenen Tonhöhe gemessen in Mel. Im Bereich oberhalb von 500 Hz gilt dieser Zusammenhang nicht mehr. In höheren Frequenzbereichen werden immer größere Frequenzsteigerungen notwendig um z.B. eine Verdopplung der wahrgenommenen Tonhöhe zu bewirken.

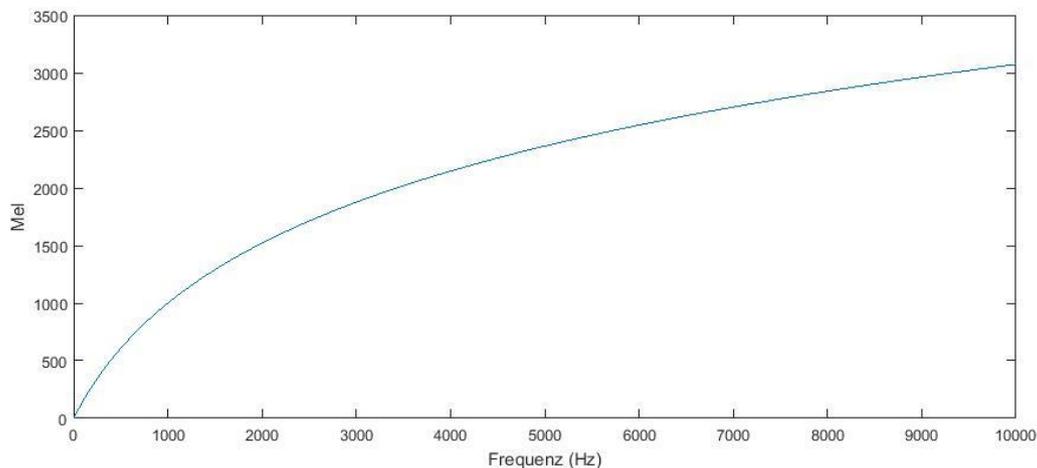


Abbildung 25: Frequenz Mel Kurve - Zusammenhang zwischen physikalischer Frequenz (Hz) und wahrgenommener Tonhöhe in Mel

Auch den physikalischen Schalldruck bildet die menschliche, auditive Wahrnehmung nicht linear ab. Die subjektiv empfundene Lautstärke hängt vom Schalldruckpegel, dem Frequenzspektrum und der zeitlichen Struktur des Schalls ab. Approximativ beschrieben wird der Zusammenhang zwischen dem physikalischen Schalldruck ( $I$ ) und der empfundenen Lautstärke ( $L_n$ ) durch das Weber-Fechnersche Gesetz (Formel 1).

$$L_n \sim \lg(I) \quad (1)$$

Die Berechnung von MFCCs läuft nach dem bereits dargestellten Schema in Abbildung 22 ab. Nach dem Frameing und Windowing wird für jeden der 10 -50 ms langen, überlappenden Frames mit einer DFT ein Spektrum erstellt. Für die Berechnung von MFCCs stehen damit pro Frame z.B. 256 Amplitudenwerte gleich großer Frequenzbereiche zur Verfügung. Diese Frequenzbereiche sind gleichmäßig über den Frequenzbereich von 0 bis Abtastfrequenz/2 Hz verteilt.

Abbildung 26 zeigt konzeptionell den Ablauf der weiteren Berechnungen nach der DFT eines Frames. Die Berechnungen bestehen aus drei Teilschritten (vgl. [48]):

1. Berechnung eines mel-skalierten Spektrums mit überlappenden Dreiecksfiltern
2. Logarithmieren jedes mel-skalierten Frequenzbandes
3. Anwendung der diskreten Cosinus Transformation (DCT) auf die logarithmierten, mel-skalierten Frequenzbänder

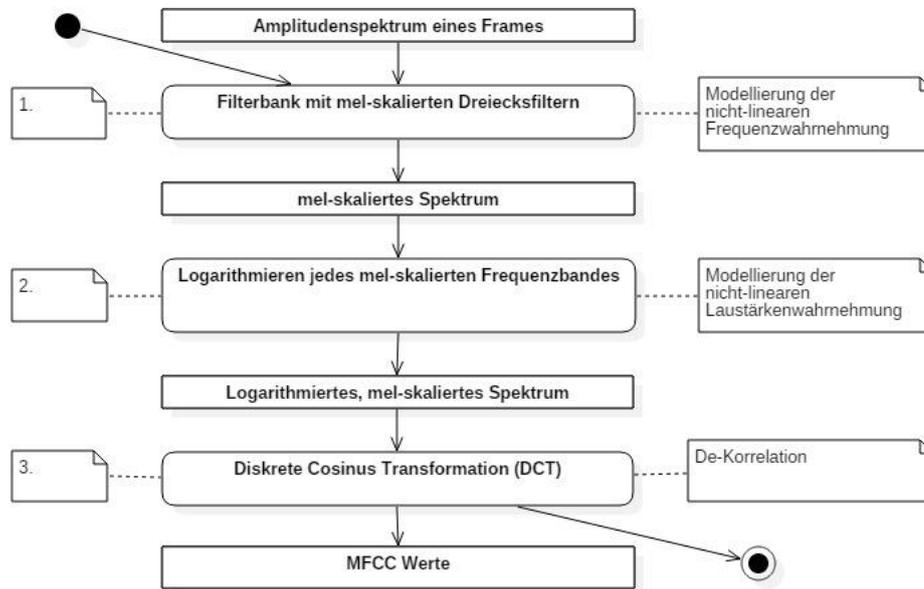


Abbildung 26: MFCC Berechnung nach Erstellung des Amplitudenspektrums mittels FFT

Die z.B. 265 Frequenzbereiche des Spektrums werden durch eine Filterbank bestehend aus Bandbreitenfiltern auf (je nach Implementierung des MFCC Algorithmus) z.B. 20, 24 oder 40 halbüberlappende Frequenzbereiche zusammengefasst. Für die Übertragungsfunktion der Filter werden Dreiecke verwendet. Eine Modellierung der nicht-linearen Frequenzwahrnehmung erfolgt durch die Anpassung der Breite der Filter an die Frequenz-Mel Kurve. Die Breite der Bandfilter in Hz (konzeptionell betrachtet der Frequenzbereich in Hz, den der Filter nicht auf null setzt) wird so gewählt, dass die Filter auf der Mel Skala gleich große Bereiche abdecken. Abbildung 27 verdeutlicht dieses Prinzip vereinfachend mit 3 Filtern im Bereich von 0 bis 10 kHz. Filter deren Bandbreite unterhalb von 500 Hz liegt haben auf der Mel und auf der Frequenzskala ähnliche Werte und eine kleine Bandbreite. Filter deren Bandbreite oberhalb von 500 HZ liegen haben eine mit der Frequenz zunehmende Bandbreite.

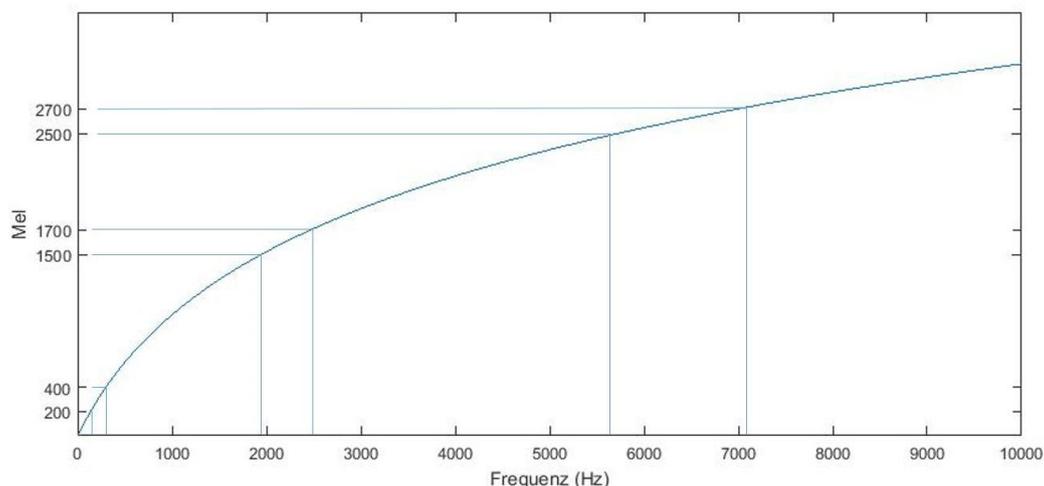


Abbildung 27: Bandbreite von drei mel-skalierten Filtern im Bereich von 0 bis 10 kHz

Nach der Erstellung des Mel Spektrums wird jedes der Mel Frequenzbänder logarithmiert. Mit diesem Schritt wird die nicht-lineare Lautstärkenwahrnehmung modelliert. Die verschiedenen MFCC Implementierungen verwenden für diesen Schritt den natürlichen Logarithmus (ln) oder den dekadischen Logarithmus (lg).

Da sich die Dreiecksfilter für die Mel-Skalierung stark überlappen, weisen die einzelnen Werte der Mel-Frequenzbänder eine hohe Korrelation auf. Die Anwendung der DCT auf die 32 – 40 logarithmierten, mel- skalierten Frequenzbänder ermöglicht die Darstellung dieser Werte in de-korrelierter Form. Die Anwendung der DCT auf eine bestimmte Anzahl an logarithmierten, mel- skalierten Frequenzbändern ergibt dieselbe Anzahl an MFCC Werten. Nicht alle dieser MFCC Werte werden für die Modellierung von Klangfarben verwendet. Der erste MFCC Wert wird für die Modellierung von Klangfarben häufig ignoriert, da er nur die durchschnittliche Energie des gesamten Spektrums anzeigt. Da die wesentlichen Informationen über die spektrale Hüllkurve bereits in den ersten 4 bis 20 MFCC Werten kodiert ist werden MFCC Werte mit höherem Index ebenfalls nicht für die Modellierung von Klangfarben verwendet. (vgl. [48])

### 2.2.3 Klangfarbe – Psychoakustik und MIR Forschung

#### 2.2.3.1 *Psychoakustische Timbre Space Forschung*

In der MIR Forschung werden Klangfarben-Ähnlichkeiten häufig als Distanzen in n-dimensionalen Räumen modelliert. Die Verwendung dieses geometrischen Ähnlichkeitsmodells wurde durch eine Reihe psychoakustischer Studien zum Thema Klangfarbe motiviert, deren Ursprünge bis in die 1970iger Jahre zurückreichen.

Den psychoakustischen Wissensstand bezüglich Klangfarbe beurteilt Licklider im Jahr 1951 folgendermaßen: "...until careful scientific work has been done on the subject, it can hardly be possible to say more about timbre than that it is a 'multidimensional' dimension." (entnommen aus [49]). 1960 wird Klangfarbe im ANSI Standard definiert als: "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus." (entnommen aus [49]). Die ANSI Definition offenbart den multidimensionalen Charakter des psychoakustischen Phänomens Klangfarbe und zeigt, dass die wahrgenommene Klangfarbe eines Samples nicht durch eine einzelne, physikalische Eigenschaft eines Audiosignals bestimmt wird. Aufschlüsse darüber wie die physikalischen Eigenschaften eines Audiosignals und die Beurteilung von Klangfarben-Ähnlichkeiten zusammenhängen liefert die ANSI Definition nicht. Das „Handbuch der Systematischen Musikwissenschaften“ [50] nennt als wesentliche physikalische Bestimmungsgrößen der Klangfarbe von Instrumenten

- den eher tonalen oder eher geräuschhaften Charakter
- die spektrale Hüllkurve
- die zeitliche Hüllkurve
- die Modulation beider Hüllkurven
- den Einschwingvorgang

Den Zusammenhang dieser physikalischen Größen bei der Wahrnehmung von Klangfarben wird aber auch hier nur skizziert: „Die Klangfarbe eines Klages ist ... stark abhängig vom Spektrum, weniger stark abhängig von der zeitlichen Hüllkurve und der Tonhöhe, und nur schwach abhängig von der Lautstärke und der Dauer“ [50]

Eine der ersten wegweisenden Arbeiten um diese Zusammenhänge zu klären ist eine 1977 veröffentlichte Studie von Grey [49]. Er verwendet 16 Instrumenten-Samples mit gleicher Tonhöhe,

Lautheit und Dauer und lässt die Teilnehmerinnen seiner Studie die Klangfarben Ähnlichkeit aller Sample-Paare auf einer Skala von 1 bis 30 bewerten. Grey bestimmt mit einer Ähnlichkeitsstrukturanalyse (MDS-Multidimensionale Skalierung) ein 3-dimensionales Model in dem die Samples als Punkte dargestellt werden und die Distanzen der Sample-Punkte annähernd den Ähnlichkeitsbewertungen der Teilnehmerinnen entsprechen. Abschließend identifiziert Grey drei physikalische Eigenschaften der Samples, welche die räumliche Aufteilung entlang der drei Dimensionen erklärt: „A preliminary psychophysical interpretation of the bases for the similarity judgments found that one dimension related to the spectral energy distribution, while the other two related to various temporal patterns of the tones, namely, the presence of low-amplitude, high-frequency energy in the initial attack segment and the presence of synchronicity in the transients of the higher harmonics along with the closely related level of spectral fluctuation in the tone through time.“ [49]

Die Studie von Grey ist eine von vielen Timbre Space (TS) Studien (z.B. [49] [51], [52], [53]) die versuchen den Zusammenhang zwischen physikalischen Signaleigenschaften und dem Klangfarben-Ähnlichkeitsbeurteilungen der Testteilnehmerinnen zu klären. TS Studien ähneln sich stark in ihrer wissenschaftlichen Methodik. Eine relative geringe Anzahl an synthetischen Instrumentenklangen ( für [49, 51–53] gilt 18 – 34 Instrumenten-Samples) mit gleicher Tonhöhe, Dauer und Lautstärke wird von einer kleinen Anzahl an Versuchsteilnehmern (für [49, 51–53] gilt 18 – 34 Versuchspersonen) paarweise auf einer Ähnlichkeitsskala bewertet. Die Autoren verwenden anschließend eine MDS Variante um einen zumeist 2 oder 3-dimensionalen, euklidischen Raum zu erstellen in dem die Distanz der Sample-Punkte den Ähnlichkeitsbeurteilungen der Teilnehmerinnen entspricht. Die Dimensionalität des Klangfarbenraums wird dabei von den Autoren selbst gewählt. Abschließend ermitteln die Autoren (in den meisten Fällen mit einer Regressionsanalyse) welche inhaltsbasierten Features hohe Korrelationen mit einer der TS Dimensionen aufweisen.

Viele TS Studien stellen übereinstimmend eine hohe Korrelation der Ähnlichkeitsbeurteilungen der verschiedenen Versuchsteilnehmer fest. Die Mehrheit der Versuchsteilnehmerinnen wählt während der Hörtests für ein Sample-Paar nur leicht variierende Skalenwerte. Diese Ergebnisse deuten mehrere Verfasserinnen von TS Studien dahingehend, dass die Ähnlichkeitsbeurteilung, auch wenn sie sicherlich nicht völlig frei von individuellen Eigenschaften der jeweiligen Versuchsteilnehmerin abläuft, stark von physikalischen Signaleigenschaften mitbestimmt wird.

In mehreren TS Studien (z.B. [51, 54, 55]) werden das Spektrale Centroid und die Einschwingdauer (log attack time) übereinstimmend als physikalische Signaleigenschaften identifiziert, die eine hohe Korrelation mit der räumlichen Aufteilung der Sample-Punkte entlang einer der Dimensionen des jeweiligen Klangfarbenraums aufweisen und damit wesentlich die Klangfarben Ähnlichkeitsbeurteilung beeinflussen. In unterschiedlichen Studien werden für das spektrale Spektrale Centroid und die log attack Time eine Korrelation von  $\sim 0.9$  mit einer der TS Dimensionen ermittelt. (vgl. [56]). Zusätzlich zeigen die diversen TS Studien, dass dieser Zusammenhang zwischen den Ähnlichkeitsbeurteilungen und dem Spektralen Centroid bzw. der Log Attack time für verschiedenste Instrumentenkategorien (z.B. Schlagzeug-Samples, Geigen-Samples, Flöten-Samples) gilt. Der Zusammenhang wird sowohl für Sample-Sammlungen, die Samples eines einzigen Instruments beinhalten als auch bei Samples-Sammlungen die mehrere unterschiedliche Instrumente beinhalten bestätigt. Für die dritte Dimension weisen die diversen TS Studien keine Übereinstimmung auf und nennen verschiedenste physikalische Signaleigenschaften, die eine Korrelation mit dieser Dimension aufweisen wie z.B. Spectral deviation, Spectral density, Spectral irregularity, Spectral flux, Decay time, Amplitude envelope, Pitch strength, Attack synchrony, Attack centroid und Noisiness.

Neben den auf Ähnlichkeitsbeurteilungen von Versuchsteilnehmerinnen beruhenden TS wurden mehrere Versuche unternommen Klangfarbenräume mit Hilfe von SOMs (Self-Organizing Map) zu erstellen. Poli et al. (1997) erstellen einen 2-dimensionalen, sogenannten physikalischen TS aus MFCCs. Die Anforderungen an diesen Klangfarbraum definieren die Autoren folgendermaßen: "The concept of acoustic similarity must be coherently mapped to the concept of distance" [57]. Sie stellen fest, dass MFCCs "...is well suited to the representation of all perpetually meaningful sounds besides speech" und die SOM es schafft die 20 verwendeten Instrumenten-Samples in gut unterscheidbare Cluster aufzuteilen. Toiviainen et al. [58] (1998) verwenden 5 synthetisierte, harmonische Klang-Samples mit gleicher Tonhöhe und Dauer aber unterschiedlicher Klangfarbe um in Hörtests Ähnlichkeitsbewertungen zu erheben und anschließend die neuronale Reaktion der Teilnehmerinnen auf die 5 Samples zu messen. Zusätzlich erheben die Autoren die Distanzen der fünf Sample auf einer (zuvor mit 45 anderen Samples trainierten) SOM. "The more dissimilar the subjects rated a given pair of tones, the more dissimilar were the respective neural responses, both as measured directly and as simulated with the SOM" [58]

Die geringe Anzahl an Samples und Teilnehmerinnen werfen die Frage auf ob diese Modelle überhaupt auf eine große Anzahl an heterogenen Samples und Anwenderinnen generalisierbar sind. Die Gültigkeit der Klangfarbenräume von Grey [49], Krumhansl [52] und Peeters et al. [51] wird z.B. durch eine Metastudie aus dem Jahr 2015 in Frage gestellt. Siddiq et al. [59] kommen in ihrer Metastudie zum dem Schluss "...that there's a lack of comparability among and thus a lack of generality to (the compared) TS-studies" [59]. Mit Bezug auf TS Studien werden in [50] die „Vorstellungen und Kenntnisse über die einzelnen Dimensionen von Klangfarbe nur sehr vage, ungesichert und nicht für jeden Fall anwendbar“ bezeichnet. „Lickliders Bemerkung, dass Klangfarbe eine »multidimensional Dimension« ist, bleibt unbestritten, jedoch liegt die Vermutung nahe, dass es besonders bei der Wahrnehmung von musikalischen Klängen verschiedene und je nach Instrument unterschiedliche und unterschiedlich viele Dimensionen zu berücksichtigen gilt.“ Für die dritte Dimension des jeweiligen Klangfarbraums nennen die diversen TS Studien zumeist recht unterschiedliche physikalische Signaleigenschaften, die eine hohe Korrelation mit dieser Dimension aufweisen. Auch in [60] wird dieser Umstand auf spezifische Eigenschaften der verwendeten Sample-Sammlungen zurückgeführt: "...interpretations of the other timbral space dimensions have lacked consensus, probably because the nature of the stimuli used in each experiment determines the perceptual and acoustic features that explain the resulting dimensions and thus cannot be generalized across other sets of stimuli."

Eine der aktuellsten TS Studien [61] wurde 2013 veröffentlicht. Die Autoren konstruieren für die skalierten Ähnlichkeitsbeurteilungen von 33 Musikerinnen für Paare aus 42 Instrumentenklängen (mit gleicher Dauer, Lautstärke und Tonhöhe) mit einer eigens für die Studie entwickelten MDS Variante einen 5-dimensionalen TS. Ausgehend von der Beobachtung, dass bisherige TS Studien eine Korrelation für temporale oder spektrale inhaltsbasierte Features ermitteln, Klangfarbe in der Psychoakustik aber in der Zwischenzeit durch ein Zusammenwirken von temporalen und spektralen Signaleigenschaften erklärt wird, verwenden die Autoren sogenannte „spectrotemporal modulation power spectrum“ [61] (= MPS) und berechnen Korrelationen zwischen 20 Hauptkomponenten der MPS und den einzelnen Dimensionen ihres TS. Die Wahl von MPS für ihre Studie begründen die Autoren unter anderem mit der Ähnlichkeit der MPS Auswertung mit der neuronalen Verarbeitung von Klängen im Gehirn des Menschen. Die Autoren weisen darauf hin, dass die rezeptiven Felder des auditiven Cortex auf bestimmte spektro-temporale Modulationen reagieren und wagen sogar eine Vorhersage welche Art von rezeptiven Feldern die Neurophysiologie im Gehirn des Menschen entdecken wird. Sie gehen davon aus, dass bestimmte rezeptive Felder entdeckt werden, die neuronale Signale wie einzelne Dimensionen ihres TS auswerten: „First, we expect there to be two distinct neural populations tuned to temporal dynamics above or below 6 Hz. Second, we predict that neural computations relevant to

more spectral percepts of timbre will be tightly coupled to pitch computations, whereas those with other relevance will not.” [61] Um einen Vergleich zu den früheren TS Studien zu ermöglichen berechnen die Autoren auch die Korrelation zwischen „traditional acoustic features that separately describe the spectral and temporal envelope“ und den einzelnen Dimensionen ihres 5-dimensionalen TS. Für diesen Features und MPS Features werden ähnliche Korrelationswerte gefunden, was dazu führt, dass Siedenburg et al., [46] (Jahr 2016) diesen 5-dimensionalen TS beurteilen als “...although certainly a valuable way of representing the information content of audio signals, the novel MPS approach did not yield substantial improvements in the fit of acoustic features to MDS dimensions when compared to the classical, audio-descriptor-based approach”

Auch wenn aus den TS Studien im Zeitraum von 1970 - 2000 bis heute kein allgemein anerkanntes Modell von Klangfarben-Ähnlichkeiten hervorgegangen ist wurden im Rahmen der TS Studien aber eine Fülle an inhaltsbasierten Features (z.B. Spectral centroid, Log-attack time, Spectral flux, Attack synchrony, Spectral irregularity, usw.) entwickelt, die in der MIR Forschung in weiterer Folge für eine Vielzahl unterschiedlicher Anwendungsgebiete eingesetzt wurden. Einige dieser Features wurden später in den MPEG 7 [42] Standard oder den CUIDADO [26] Standard aufgenommen. „The results of these studies are not fully consistent due to a number of limitations concerning the sound stimuli and experimental protocol, but it can still be argued that a number of features obtained within these studies do explain perceptual dimensions of timbre and have been successfully used in automatic instrument timbre recognition experiments” [62]

Eine weitere Folge der TS Studien im Zeitraum von 1970 bis 2000 sind erste Ansätze Klangfarben-Ähnlichkeiten als Distanz von Feature-Vektoren in n-dimensionalen, metrischen Räumen zu modellieren. In [56] wird ein Distanzmaß aus „Spectral Centroid, Log Attack-Time, Spectral Irregularity und Spectral Flux“ berechnet. Bereits 1998 nennen die Autoren dieser Studie als Verwendung für dieses Distanzmaß „...the development of a search engine within a database of recorded musical sounds.“ [56]. In [63] entwickeln Peeters et al. 2 euklidische Distanzmaße anhand der inhaltsbasierten Features, die für die Timbre Space Studien in [51–53] verwendet wurden. Auch Peeters et al. nennen als Zweck solcher Distanzmaße die „...facilitation of search-by-similarity operations in databases with musical sounds...“.

### *2.2.3.2 Modellierung von Klangfarben-Ähnlichkeit in der MIR Forschung*

Mit der Jahrtausendwende verwendet die MIR Forschungsgemeinde vermehrt die Methoden der Mustererkennung zur Modellierung von Klangfarben. Konzeptionell gesehen bestehen viele MIR Anwendungen der nächsten 16 Jahre aus einer inhaltsbasierten Merkmalsextraktion und der nachfolgenden Verarbeitung von Feature-Vektoren durch verschiedene Klassifikations-Algorithmen. Nach der inhaltsbasierten Merkmalsextraktion wird durch eine Merkmalsreduktion die Anzahl der für die Klassifikation verwendeten Features verringert (Varianzanalyse, Diskriminanzanalyse). Die Merkmalsreduktion verbessert die rechnerische Effizienz, eliminiert redundante Features und hilft die Überanpassung des Klassifikations-Modells an die verwendete Audiodatenbank zu verringern. Die Klassifikations-Algorithmen werden häufig durch ein Kreuzvalidierungsverfahren trainiert und evaluiert. Für die Kreuzvalidierung wird jedes Element der Audiodatenbank mit einem Klassenlabel annotiert (ground truth) und die gesamte Datenbank in k Teilmengen zerlegt (k-fache Kreuzvalidierung). In k Durchläufen wird jeweils mit k – 1 Teilmengen der Klassifikations-Algorithmus trainiert und mit einer der Teilmengen evaluiert. Der Durchschnitt der Klassifikationsgenauigkeiten (=Accuracy) über alle k-Durchläufe ergibt die Klassifikationsgenauigkeit des Klassifikations-Modells.

Logan [64] (Jahr 2000) ist eine der ersten Autorinnen die MFCCs und einen Klassifikationsalgorithmus für die Unterscheidung von Sprach und Musiksignalen einsetzt. Logan zeigt, dass die Anwendung der Mel-Skala während der Koeffizienten-Berechnung die Klassifikation statisch signifikant verbessert und

dass, die de-korrelierende Wirkung der DCT auch für Koeffizienten gilt, die aus Musiksignalen extrahiert wurden.

2001 verwendet Eronen [65] MFCCs, sowie einige Features, die durch die TS Studien als wesentliche Bestimmungsgrößen für die Klangfarbe identifiziert wurden und einen kNN Algorithmus zur Instrumentenerkennung. Verwendet wird eine Sample-Datenbank aus ca. 5000 Instrumentenklängen. Die Ergebnisse dieser Studie zeigen, dass MFCCs im Vergleich mit dem spektralen Centroid und der Einschwingdauer eine wesentlich bessere Klassifikation ermöglichen und dass die höchsten Erkennungsraten mit Feature-Vektoren erzielt werden, die neben MFCCs weitere Features wie die Einschwingdauer und das spektrale Centroid beinhalten.

Ein Vergleich der diversen Ansätze für die Klassifizierung von Samples einzelner Instrumentenklänge, die bis ins Jahr 2003 entwickelt wurden ist in [66] zu finden. Für die Klassifikation verwenden die vorgestellten Ansätze zu diesem Zeitpunkt bereits z.B. "Hidden Markov Models", "Support Vector Machines" (SVM), "Naive Bayesian Classifiers" oder "Binary trees". Als Features werden in den diversen Ansätzen MFCCs und einige Features des MPEG 7 Standards und CUIDADO Standards verwendet. Die Autoren halten fest, dass der Ansatz aus Feature Extraction + Machine Learning Algorithmus, der ursprünglich mit Samples einzelner Klänge entwickelt wurde auch für Drumloop-Samples eignet, in denen mehrere Instrumentenklänge kombiniert werden.

Die bisher genannten Studien klassifizieren Samples unterschiedlicher Instrumente anhand ihrer Klangfarbe und verwenden dafür dieselben Features wie die hier behandelten Sample-Browser. Die in diesen Studien erzielten Klassifikationsgenauigkeiten können als Indiz betrachtet werden wie geeignet diese Features für die Modellierung von Klangfarben sind. Diese und ähnliche Studien, die ebenfalls Samples unterschiedlicher Instrumente klassifizieren, erzielen Klassifikationsgenauigkeiten zwischen 70 und 95 % (vgl. [67]).

Wie gut sich diese Features für die Modellierung von Klangfarben in ähnlichkeitsbasierten Sample-Browser eigenen lässt sich mit den erzielten Klassifikationsgenauigkeiten dieser Studien aber nicht optimal beurteilen, da der realistische Anwendungsfall für ähnlichkeitsbasierte Sample-Browser nicht die Darstellung unterschiedlicher Instrumente auf einer 2D-Karte ist. Der realistische Anwendungsfall ist die Darstellung und der Zugriff auf unterschiedliche Aufnahmen eines Instruments. (siehe Abschnitt „*Praxisnahes Anwendungsszenario für Sample-Browser*“). Aktuellere Studien, die eine Klassifikation von Samples desselben Instruments durchführen, sind daher für diese Beurteilung besser geeignet. In einer 2015 veröffentlichten Studie von Johnson et al. [68] werden Samples von 17 unterschiedlichen Gitarrenmodellen verwendet. In einer im selben Jahr veröffentlichten Studie [69] werden 1053 Samples von 5 unterschiedlichen Beckenmodellen klassifiziert. Johnson et al. erreichen mit Feature-Vektoren, die unter anderem aus MFCCs, ZCR, Spectral Centroid, und Spectral Flux bestehen, einem kNN Algorithmus und einer 10-fachen Kreuzvalidierung eine Klassifikationsgenauigkeit von 99.94 %. Mit denselben Features und einem SVM Algorithmus wird eine Klassifikationsgenauigkeit von 55.10 % erzielt. „With near 100 % accuracy, the results of cross validation for kNN appear to suffer from high overfit“ [68] In [69] werden mit MFCCs, einem SVM Algorithmus und einer 10-fachen Kreuzvalidierung eine Klassifikationsgenauigkeit von 94.01 % erreicht.

Auch wenn diese Ergebnisse zeigen, dass die Features der hier behandelten Sample-Browser eine hohe Klassifikationsgenauigkeit ermöglichen, sollten sie nicht als Erfolgsgarant bei ähnlichkeitsbasierten 2D Darstellungen betrachtet werden. Die Berechnungen von Klassifikations-Algorithmen unterscheiden sich von den Berechnungen der Dimensionsreduktions-Algorithmen, die ähnlichkeitsbasierte Sample-Browser verwenden. Klassifizierungs-Algorithmen ordnen Samples verschiedenen Klassen anhand der Features mit den höchsten diskriminierenden Eigenschaften zu. Das sind nicht zwangsläufig die Features, welche die Klangfarben Ähnlichkeitswahrnehmung am stärksten beeinflussen.

Psychoakustische Hörtests zeigen, dass die menschliche Klassifizierung von Klängen nur anhand von temporalen Signaleigenschaften nicht möglich ist (vgl. [70]) Trotzdem werden in [69] mit ausschließlich temporalen Features und einem SVM Algorithmus eine Klassifikationsgenauigkeit von 85.85 % erzielt.

### 2.2.3.3 *Das geometrische Ähnlichkeitsmodell: Modellierung des Klangfarben-Ähnlichkeitsempfinden durch Distanzen von Feature-Vektoren*

Alle der hier behandelten Sample-Browser, die eine inhaltsbasierte Merkmalsextraktion einsetzen, verwenden ein geometrisches Ähnlichkeitsmodell. Diese Anwendungen verketteten temporale und spektrale, inhaltsbasierte Features zu Vektoren und berechnen die Distanzen dieser Feature Vektoren. Die Grundannahme dieses Modells ist das sich das Klangfarben-Ähnlichkeitsempfinden von Menschen als Distanz zwischen zwei Punkten in einem mehrdimensionalen, metrischen Raum darstellen lässt und dass metrische Distanzen von Feature-Vektoren sich proportional zu Ähnlichkeitsbeurteilungen von Benutzerinnen verhalten.

#### 2.2.3.3.1 *Distanzen von MFCC Vektoren als Ähnlichkeitsmaß*

Da die Mehrzahl der hier behandelten Sample-Browser MFCCs für die Modellierung von Klangfarben verwenden ist eine Einschätzung notwendig wie geeignet metrische Distanzen (z.B. euklidische Distanzen) von MFCC Werten zur Modellierung von Klangfarben-Ähnlichkeiten sind. Entspricht die Distanz von Punkten in einem 13-dimensionalen Raum aus MFCCs der empfundenen Ähnlichkeit von Testteilnehmerinnen? Terasawa et al., setzen sich in mehreren Publikationen [47, 71–73] mit dieser Fragestellung auseinander.

MFCCs wurden ursprünglich für die Spracherkennung entwickelt, die MIR Forschung hat erst im Laufe der Zeit durch die Entwicklung von Software-Prototypen (z.B. Instrumentenerkennung und Musikgenreerkennung) gezeigt, dass MFCCs auch für Musiksignale eine geeignete Modellierung von Klangfarben darstellen. Terasawa et al. sind wahrscheinlich die ersten Autoren, die den Zusammenhang zwischen MFCC Werten und dem Klangfarben Ähnlichkeitsempfinden von Testteilnehmerinnen durch psychoakustische Hörtests erforschen. (vgl. [72]) Terasawa et al. gehen davon aus, dass die Form der spektralen Hüllkurve die spektrale Dimension des Phänomens Klangfarbe wesentlich bestimmt und MFCCs eine Darstellung dieser Hüllkurve sind.

Der Versuchsaufbau ist in allen vier Veröffentlichungen ähnlich. Die Autoren variieren systematisch 1 oder 2 von 13 MFCC Werten (während alle anderen MFCC Werte konstant gleich gehalten werden), rekonstruieren die spektrale Hüllkurve aus den Koeffizienten und synthetisieren Klänge aus der Hüllkurve. Dauer, Tonhöhe und Lautstärke aller Klänge werden nicht verändert. Die Teilnehmerinnen der Hörtests bewerten die Ähnlichkeit der synthetisierten Klänge auf einer zehnstufigen Skala. Danach postulieren die Autoren, dass die Ähnlichkeitsbeurteilungen und die variierten MFCC Werte in einem linearen, metrischen Zusammenhang stehen. Diese Annahme wird mit einer linearen Regressionsanalyse überprüft.

Abbildung 28 zeigt die Rekonstruktion der spektralen Hüllkurven aus den systematisch variierten MFCCs. Für einen der Tests in [72] werden z.B. die Koeffizienten C3 und C6 eines MFCC-Vektors (bestehen aus C0 bis C12) vierstufig von 0 auf 0,75 erhöht (0, 0.25, 0.5, 0.75). Zeilenweise zeigt die Abbildung die resultierende Hüllkurve, wenn C6 von links (=0) nach rechts (=0.75) stufenweise erhöht wird. Spaltenweise ist die resultierende Hüllkurve bei Erhöhung von C3 von oben (=0) nach unten (=0.75) zu sehen. In [72] werden für Hörtests die Paare [C1,C3], [C3,C4], [C3,C6], [C3,C12] und [C11,C12] wie beschrieben systematisch verändert. Für jedes dieser Paare wird eine Auswahl von 34 (von 256 pro Paar möglichen) MFCC Variationen für die Hörtests verwendet.

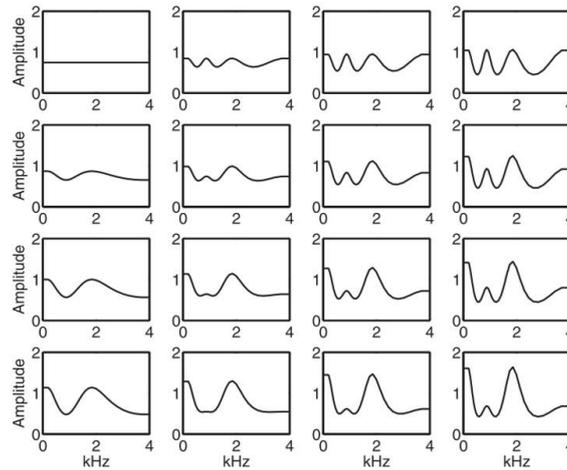


Abbildung 28: Resultierende spektrale Hüllkurven bei vierstufiger Erhöhung von MFCC C3 und C6. Links oben: C3 = 0, C6 = 0. Rechts unten C3 = 0.75, C6 = 0.75

Die Rekonstruktion der Hüllkurve aus nur 13 MFCCs ist verlustbehaftet und zeigt im Vergleich mit der ursprünglichen Hüllkurve eine Glättung. Aus der rekonstruierten Hüllkurve werden der menschlichen Stimme ähnliche Klänge synthetisiert. Danach postulieren die Autoren den in Formel 2 dargestellten, metrischen, euklidischen Zusammenhang zwischen den Ähnlichkeitsbewertungen einer Teilnehmerin und den zwei veränderten MFCC Koeffizienten (z.B. C3 und C6):

$$ax^2 + by^2 = d^2 \quad (2)$$

„This model reflects the idea that the perceptual distance should be described in terms of the Euclidean distance of the spectral-envelope description vectors.” [72].  $d$  entspricht der Angabe der Versuchsteilnehmerin während der Hörtests.  $x$  und  $y$  entsprechen der Distanz zwischen den variierten MFCC Werten (z.B. C3 bzw. C6) der beiden miteinander verglichenen Klänge. ( $x$  stellt in dieser Gleichung immer die Differenz der MFCC Werte mit niedrigerem Index dar,  $y$  bezeichnet die Differenz der MFCC Werte mit höherem Index) Die Schätzung der Regressionskoeffizienten  $a$  und  $b$  erfolgt mit einer multiplen Regressionsanalyse (Methode der kleinsten Quadrate). Der Durchschnitt aller Regressionskoeffizienten  $a$  oder  $b$  über alle Versuchsteilnehmerinnen zeigt, dass  $a$  konstant höher als  $b$  ist. Die Autoren schließen daraus, dass MFCCs mit niedrigem Index die Ähnlichkeitsbeurteilung stärker beeinflussen. „...the degree of contribution of the MFCCs is similar in the low- to mid-order MFCCs with a slightly decreasing trend, and for higher-order MFCCs, the degree of contribution drops more quickly and significantly” [72]

Das Bestimmtheitsmaß  $R^2$  zeigt wie gut die Variation in den Ähnlichkeitsangaben  $d$  durch das postulierte, lineare Regressionsmodell erfasst werden kann.  $R^2$  wird zunächst für jede Teilnehmerin und jedes systematisch veränderte Paar von MFCCs berechnet. Der Durchschnitt der  $R^2$  Werte pro Paar über alle Versuchsteilnehmer (und entsprechendes 95 % Konfidenzintervall) wird in Abbildung 29 dargestellt. „The Euclidean distance in the MFCC explains the spectral envelope perception with an average degree of fit of 74%.” [72] Für eine weitere Versuchsanordnung (siehe [72]) in der nur ein MFCC systematisch variiert wurde ergibt sich ein  $R^2$  Durchschnitt von 85 %. Die Autoren schließen aus diesen Ergebnissen „that the MFCC can be a fair representation of spectral envelope perception, and that spectral envelope perception can be fully described in terms of the Euclidean space constituted by MFCCs.” [72]

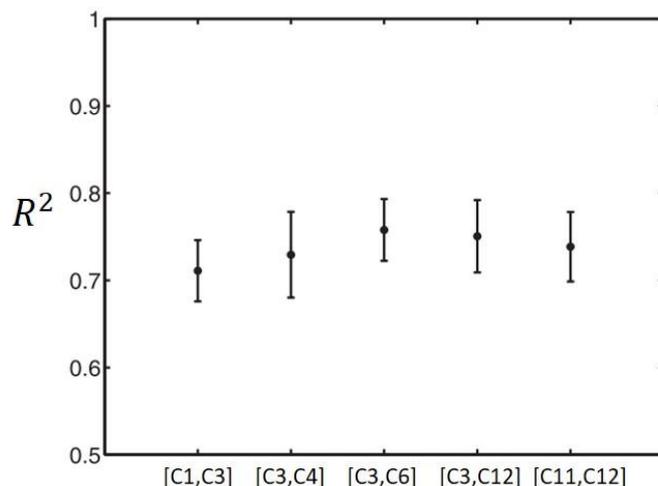


Abbildung 29: Bestimmtheitsmaß  $R^2$  des euklidischen Distanzmodells für Paare von systematisch veränderten MFCC Werten

Die Ergebnisse der Studien von Terasawa et al., [47, 71–73] sind nur ein Indiz dafür, dass die euklidische Distanz von MFCC Vektoren eine adäquate (aber nicht perfekte) Modellierung der spektralen Dimension von Klangfarben-Ähnlichkeiten darstellt. Vorausgesetzt die Ergebnisse sind auf jede Art von Klängen (= beliebig viele der 13 MFCC Werte variieren) und Anwenderinnen übertragbar lässt sich mit der euklidischen Distanz von gewichteten MFCC Vektoren die empfundene, psychoakustische Ähnlichkeit annäherungsweise berechnen. Um in einem praktischen Anwendungsfall ähnliche Ergebnisse wie Terasawa et al. zu erzielen müssten aber alle 13 MFCC Werte mit einem zuvor bestimmten Regressionskoeffizienten ( $a, b, \dots, m$ ) gewichtet werden. Die Werte  $a$  bis  $m$  würden wahrscheinlich wie beschrieben von  $a$  als größtem Wert ausgehend, zunächst ähnliche Werte mit leicht abfallender Tendenz und gegen Ende des MFCC Vektors immer stärker abfallende Werte aufweisen. Die Regressionskoeffizienten für eine bestimmte Sample-Sammlung müssten zusätzlich durch Hörtests und eine lineare Regression ermittelt werden. Die meisten der hier behandelten Sample-Browser berechnen die Distanzen der Feature-Vektoren ohne eine Gewichtung der MFCC Werte. (Nur der Snare Drum Navigator [21] ermöglicht eine Gewichtung der Feature-Werte durch einen Metric Learning Algorithmus). Im günstigsten Fall stellt diese Vorgehensweise ein euklidisches Model von Klangfarben-Ähnlichkeiten mit einem Bestimmtheitsmaß  $R^2$  von knapp unter 74 % dar. Wahrscheinlich weist dieses Modell in der Praxis ein geringeres Bestimmtheitsmaß  $R^2$  auf.

#### 2.2.3.3.2 Gültigkeit des geometrischen Ähnlichkeitsmodells

Die hier behandelten Sample-Browser verketteten in den meisten Fällen MFCCs und einige Features, die im Rahmen der TS Studien entwickelt wurden, zu Feature Vektoren. Auffällig ist dass die meisten der Sample-Browser (und viele MIR Anwendungen) die Distanzen der Feature-Vektoren ohne eine Gewichtung der einzelnen Feature-Werte berechnen. Sowohl die Studien von Terasawa et al. als auch die TS Studien zeigen, dass die diversen Features die Klangfarben-Wahrnehmung unterschiedlich stark beeinflussen. Bei den Studien von Terasawa et al., zeigt der abfallende Trend der Regressionskoeffizienten, dass eine Gewichtung der MFCC Werte notwendig ist um eine optimale Anpassung der berechneten Distanzen an die Ähnlichkeitsbeurteilungen der Teilnehmerinnen zu erzielen. Bei den TS Studien deuten die unterschiedlich hohen Korrelationen der Features mit einzelnen Dimensionen des jeweiligen TS daraufhin, dass eine Gewichtung der Vektoren-Werte für optimale Ergebnisse notwendig wäre. Die ersten Ansätze Klangfarben-Ähnlichkeiten als Distanz von Feature-Vektoren in  $n$ -dimensionalen, metrischen Räumen zu modellieren ([56, 63]) verwenden eine solche Gewichtung. Die Berechnung von Distanzen ohne eine Gewichtung der Feature-Werte impliziert, dass jedes Feature für die Ähnlichkeitsbeurteilung gleichbedeutend ist und steht damit im

Widerspruch zu psychoakustischen Erkenntnissen, die zeigen, dass die gewählten Features für die Klangfarbenwahrnehmung zwar relevante Bestimmungsgrößen sind, diese aber unterschiedlich stark beeinflussen.

Psychologen beurteilen das geometrische Ähnlichkeitsmodell kritisch (vgl. [46]) und verweisen häufig auf das von Tversky in [74] vorgestellte Feature Contrast Model. Das Distanzmaß eines metrischen Raums muss drei Axiome erfüllen: „Minimality“, „Symmetry“ und „The triangle inequality“ [74]. Tversky zeigt anhand mehrerer praktischer Beispiele, dass die menschliche Beurteilung von Ähnlichkeit diese drei Axiome in vielen Fällen nicht erfüllt und stellt damit die Gültigkeit des geometrischen Ähnlichkeitsmodells in Frage. Auch MIR Forscher beurteilen, das geometrische Ähnlichkeitsmodell skeptisch. Rückblickend auf 10 Jahre in denen das geometrische Ähnlichkeitsmodell in einer Vielzahl an MIR Anwendungen verwendet wurde, meinen Mitrović et. al.: „Similarity measures derived from distance metrics are only appropriate to a certain degree, since mathematical metrics usually do not fully match the human perception of similarity.“ [40]

Damit die Modellierung von Klangfarben-Ähnlichkeiten als Distanzen von Feature-Vektoren perfekte Ergebnisse liefert müsste

1. die Klangfarbenwahrnehmung ein Reiz – Reaktionsmechanismus sein, der in allen Menschen gleich abläuft.
2. das menschliche Gehirn, unabhängig davon welche Samples einer bestimmten Sample-Sammlung miteinander verglichen werden, immer dieselben physikalischen Eigenschaften für die Ähnlichkeitswahrnehmung verwenden.

Die Einwände gegen Punkt 1. sind zahlreich. Die Vorstellung impliziert das weder Gedächtnis noch eine musikalische Ausbildung bei der Wahrnehmung und Ähnlichkeitsbeurteilung von Klangfarben eine Rolle spielen. „...timbre has traditionally been treated as a primarily sensory phenomenon that resides “in the moment” and is not subject to long-term familiarisation. Neurophysiological studies on timbre processing have started to provide evidence for the contrary position.“ [75] Das Gehirn von professionellen Musikerinnen zeigt eine verstärkte preattentive Wahrnehmung, wenn sie das von Ihnen gespielte Instrument hören. Die EEG Aufnahmen von Kindern zeigen nach einem Jahr Klavierunterricht verstärkte Gamma-Band (25–100 Hz) Oszillationen. Das selbe, verstärkte Gamma Signal unterscheidet auch die Reaktion von Musikerinnen im Vergleich zu Nicht-Musikerinnen, wenn sie Klänge mit unterschiedlichen Klangfarben hören. Im Vergleich mit Musikern, die nur über geringe Erfahrung mit dem Klavierspiel verfügen, zeigen die Aufnahmen aus dem Gehirnstamm von Pianisten eine höhere Korrelation mit der temporalen Hüllkurve von Klavier-Samples (vgl. [75]). Psychoakustische Hörtests in [75] zeigen das die Klangfarben-Wahrnehmung von der Vertrautheit der Testteilnehmerinnen mit den jeweiligen Klängen, beeinflusst werden.

Punkt 2. impliziert die Vorstellung des Gehirns als neurologisches Messgerät mit einer bestimmten Anzahl an gleichbleibenden Messgrößen. Aufgrund der Ergebnisse mehrerer psychoakustischer Studien weisen die Autoren in [70] daraufhin, dass „...the perceptual system might exploit sensory cues in an opportunistic fashion. Rather than always using the same fixed set of acoustic cues, only the most informative cues are employed with respect to the scenario of a particular perceptual task“ Diese Einschätzung teilen Siedenburg et al., in [46] : „...if confronted with the task of similarity assessment, subjects could rely on the perceptually most salient properties according to which musical tones can be most easily compared with one another.“ Einige neurologische Erkenntnisse „...even questions the very existence of a task-independent, comprehensive (i.e. ‘platonic’) representation of perceptual sound features in auditory cortex“ [46]

2016 betonen Siedenburg et al. mit Bezug sowohl auf die TS Studien und das in der MIR Forschung verwendete geometrische Ähnlichkeitsmodell "...the fact that contemporary models of complex auditory cognition, such as timbre perception, remain coarse approximations of the underlying psychological processes." [46]. Eine im selben Jahr veröffentlichte psychoakustische Studie bezeichnet Klangfarbe als „...one of the “last frontiers” in auditory science“ [70] Die neurologischen Vorgänge im Gehirn des Menschen bei der Wahrnehmung von Klangfarben sind nur ansatzweise erforscht. (vgl. [76]) Da weder Psychoakustiker noch Neurologen davon ausgehen, dass die Klangfarben-Wahrnehmung restlos erforscht ist liegt die Vermutung nahe, dass das geometrische Ähnlichkeitsmodell eine stark vereinfachende Modellierung der Klangfarben-Wahrnehmung ist.

### 2.3 Dimensionsreduktion

Nach der inhaltsbasierten Merkmalsextraktion steht pro Sample ein d-dimensionaler Feature-Vektor zur Verfügung. Pro Sample wurden d Kennzahlen berechnet, die bestimmte akustische Eigenschaften der Samples quantifizieren. Die direkte Präsentation der Ergebnisse der inhaltsbasierten Merkmalsextraktion z.B. in Form einer n x d Tabelle mit n Samples und d Feature-Werten ist wenig sinnvoll. Bestimmte Informationen, die in diesen langen Zahlenketten stecken, sind der menschlichen Wahrnehmung nicht zugänglich. Ein erwachsener Mensch kann nur ca. 4 Dinge gleichzeitig in seinem Kurzzeitgedächtnis verarbeiten (vgl. [77]). Abbildung 30 dient zu Verdeutlichung. Für n=15 Samples wurden jeweils d=13 Feature-Werte berechnet. Selbst bei dieser geringen Anzahl an Samples und Feature-Werten kann eine Anwenderin nicht mehr feststellen, welche Samples für alle Features ähnliche Werte aufweisen oder ob sich ein Sample in allen Feature-Werten wesentlich von allen anderen Samples unterscheidet. Die Anwenderin kann nicht erkennen welche Samples ein Cluster bilden oder ob es sich bei einem bestimmten Sample um einen Ausreißer handelt. Eine Visualisierung mit Hilfe eines Dimensionsreduktion– Algorithmus macht diese Informationen auf einen Blick sichtbar.

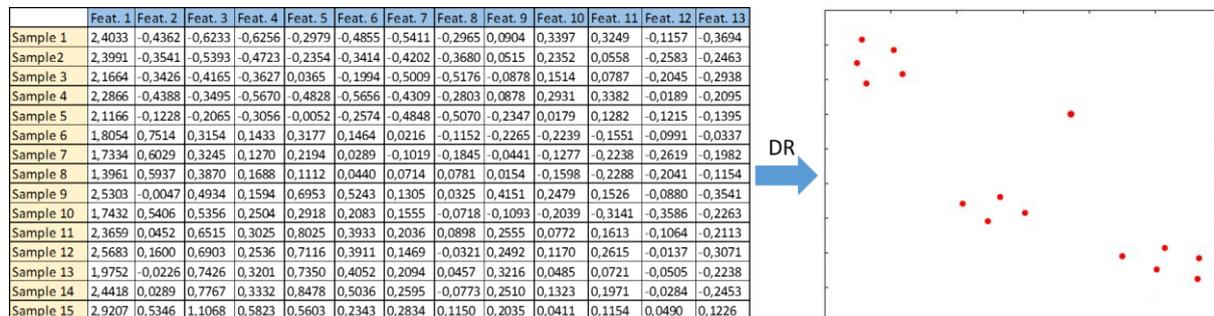


Abbildung 30: Links n x d Tabelle mit n = 15 Samples und d = 13 Feature-Werten. In der Tabelle kann eine Anwenderin aufgrund begrenzter Kapazitäten des Kurzzeitgedächtnisses die Struktur des d-dimensionalen Feature-Raums nicht erkennen. Die Visualisierung mittels DR-Algorithmus (rechts) zeigt auf einen Blick, dass die Struktur aus einem Ausreißer und drei Clustern besteht.

Multivariate Datensätze wie in Abbildung 30 werden für die Dimensionsreduktion als Punkte in einem d-dimensionalen Raum betrachtet. Dimensionsreduktion–Algorithmen bilden die n Punkte des d-dimensionalen Feature-Raums in einen d'-dimensionalen Bild-Raum ab, wobei  $d > d'$  gilt. Ziel aller Dimensionsreduktion–Algorithmen ist es bei dieser Abbildung bestimmte Eigenschaften des d-dimensionalen Raums (z.B. Varianzen, Distanzen, Topologie, lokale Nachbarschaftsverhältnisse der einzelnen Punkte) im d'-dimensionalen Raum so gut als möglich zu erhalten. Mit  $d'=2$  oder  $d'=3$  lassen sich mit DR Algorithmen Visualisierungen multivariater Datensätze wie in Abbildung 30 erstellen. Andere Anwendungsgebiete von DR Algorithmen sind die Filterung und Kompression von Daten, die Auswahl von Features für eine Datenanalyse oder die Erstellung einer kompakteren Datenrepräsentation, die eine nachfolgende Verarbeitung durch Klassifikationsalgorithmen beschleunigen und verbessern kann.

Grundlage für alle genannten Dimensionsreduktion-Anwendungen ist die Beobachtung, dass sich d-dimensionale Daten häufig ohne wesentliche Informationsverluste in eine Darstellung mit niedrigerer Dimensionalität überführen lassen. Für die niedrigste Dimensionalität mit der sich ein d-dimensionaler Datensatz ohne oder mit geringen Informationsverlusten darstellen lässt, wurde der Begriff „intrinsische Dimension“ geprägt. Eine aktuellere, mathematische Definition der intrinsischen Dimension lautet wie folgt: Liegen die d-dimensionalen Punkte eines bestimmten Datensatzes zur Gänze in einem  $d'$ -dimensionalen Teilraum des d-dimensionalen Raums, so weist der d-dimensionale Raum die intrinsische Dimension  $d'$  auf (vgl. [78]).

Wissenschaftlicher Konsens ist, dass eine Dimensionsreduktion auf 2 oder 3 Dimensionen (wie sie für eine Visualisierung notwendig ist) im Allgemeinen nicht ohne einen gewissen Informationsverlust bzw. eine gewisse Verzerrung der Struktur des d-dimensionalen Raums (=Distanzen, Topologie, Nachbarschaftsverhältnisse der einzelnen Punkte) durchgeführt werden kann (vgl. z.B. [79], [80], [81], [82]). Multivariate Datensätze weisen in der Regel eine intrinsische Dimension  $> 2$  oder  $3$  auf und können daher nicht völlig strukturgetreu visualisiert werden. Für eine Dimensionsreduktion, die nicht auf eine Visualisierung, sondern auf eine Auswahl oder Reduktion von Features für die weitere maschinelle Verarbeitung abzielt, kann eine Dimensionalität  $d' > 3$  verwendet werden. In diesem Fall kann  $d'$  der tatsächlichen intrinsischen Dimension des jeweiligen Datensatzes angepasst werden und so eine Dimensionsreduktion ohne wesentliche Informationsverluste erzielt werden.

Im Folgenden werden DR Algorithmen ausschließlich unter dem Aspekt der Visualisierung multivariater Daten behandelt. Mit dem Kürzel DRvis wird daher eine Dimensionsreduktion auf 2 oder 3 Dimensionen mit dem Zweck einer Visualisierung bezeichnet.

Die in dieser Arbeit vorgestellten Sample-Browser verwenden für die DRvis (in chronologischer Reihenfolge)

- eine PCA (Principal Component Analysis) (z.B. in MARSYAS3D [13], Audio Retrieval Browser [15])
- eine SOM (Self-organizing Map) (z.B. in Sound Torch [9])
- eine MDS Variante (Multidimensional Scaling) (z.B. in Sound Navigator [17])
- t-SNE (t-Distributed Stochastic Neighbor Embedding) (z.B. in Drumspace [6], AudioMetro [2], Grill und Flexer [20])

Die Auflistung spiegelt den technologischen Fortschritt der DRvis Forschung von linearen, parametrischen DR Methoden (z.B. PCA) zu nichtlinearen, nicht-parametrischen Methoden (z.B. t-SNE). Die Mehrzahl der modernen, „State of the Art“ DRvis Methoden sind nichtlineare, nicht-parametrische Methoden.

Parametrische DRvis Methoden erstellen für die Abbildung des d-dimensionalen Feature-Raums in den 2 oder 3-dimensionalen Visualisierungs-Raum eine Abbildungsfunktion, deren Form und Parameteranzahl a priori fixiert und unabhängig von der Anzahl der abzubildenden, hochdimensionalen Punkte ist. Die Parameterwerte dieser Abbildungsfunktion werden während der DRvis optimiert. Neuere, nicht-parametrische Methoden erstellen keine Projektionsfunktion. Stattdessen wird während der DRvis jede einzelne Abbildung eines hochdimensionalen Punktes auf sein Pendant im Visualisierungs-Raum optimiert. (vgl. [83])

### 2.3.1 PCA - Principal Component Analysis

PCA ist eine lineare, parametrische Dimensionsreduktions-Methode. Lineare DR Methoden verwenden für die Abbildung des Feature-Raums in den dimensionsreduzierten Raum eine Linearkombination der Koordinaten eines Punktes im Feature-Raum. Wird eine DRvis mit einer PCA durchgeführt wird eine

$d' \times d$  große Projektionsmatrix  $L$  erstellt, deren Komponenten so optimiert werden, dass die Zeilen der Projektionsmatrix  $L$  den Hauptkomponenten des Feature-Raums entsprechen. Eine PCA DRvis ist daher der Versuch die Varianzstruktur des Feature-Raums in der Visualisierung möglichst gut zu erhalten. Die Projektion eines Punkts  $x \in \mathbb{R}^n$  im Feature- Raum auf einen Punkt  $y \in \mathbb{R}^d, d \in \{2,3\}$  im Visualisierungs-Raum erfolgt durch Formel 3.

$$y = L \cdot x \quad (3)$$

Eine derartige lineare Projektion kann nur annähernd linear im Feature-Raum verteilte Punkte strukturgetreu abbilden. Nichtlineare Verteilungen der Punkte im Feature-Raum werden durch eine solche Projektion verzerrt dargestellt. Aufgrund dieser Eigenschaft wird die PCA in mehreren DRvis State of the Art Reviews ([79], [80], [84], [85]) als ungeeignet für die Visualisierung nichtlinear verteilter, multivariater Daten beurteilt. „The neighbourhood relationships between points are lost in projected space and do not preserve their nonlinear relationships of original data“ [80] „A linear projection disrupts the local neighborhood structures of the data sets, and no separated class structures or smooth manifold data are apparent. Hence PCA is not suited for settings where nonlinear effects are expected.“ [79]

Eine PCA dürfte damit insbesondere für die Visualisierung in einem Sample-Browser ungeeignet sein. Eine Anwenderin sich kann jede beliebige Sample-Sammlung in der Visualisierung eines Sample-Browsers darstellen lassen. Damit kann auch die Verteilung der Feature-Vektoren jede beliebige Form annehmen. Eine lineare Anordnung der Sample Feature-Vektoren dürfte für dieses Anwendungsszenario eher die Ausnahme darstellen.

### 2.3.2 MDS (Multidimensional Scaling)

Es existieren mehrere nichtlineare, nicht-parametrische MDS Varianten. Das Ziel aller MDS Varianten ist die Erhaltung der paarweisen Distanzen des Feature-Raums im Visualisierungs-Raum. Die erste MDS Variante wurde bereits in den 1960iger Jahren entwickelt und wird häufig als „classical MDS“ oder „metric MDS“ bezeichnet. Ähnlich wie eine PCA kann diese DRvis Methode nur linear im Feature-Raum angeordnete multivariate Daten strukturgetreu abbilden und dürfte daher ebenfalls für die Visualisierung in einem Sample-Browser ungeeignet sein. (vgl. [79])

Später entwickelte MDS Varianten wie curvilinear component analysis (CCA) und Sammon's mapping versuchen vor allem kleine Distanzen des Feature-Raums strukturgetreu abzubilden. Ihr Design orientiert sich an der Erkenntnis, dass die Erhaltung kleiner Distanzen des Feature-Raums für die Qualität einer DRvis Visualisierung entscheidend sind (vgl. [84]). Beide Methoden werden in [79] aber ähnlich beurteilt wie die klassische MDS Variante und eine PCA: „They have the drawback that, similar to PCA, their capability to capture nonlinearities can be rather limited...“

### 2.3.3 SOM (Self-organizing Map)

Der SOM Algorithmus zielt auf eine Erhaltung der Topologie des Feature Raums im Visualisierungs-Raum ab. Der ursprüngliche SOM Algorithmus wurde bereits 1982 von Teuvo Kohonen veröffentlicht. In den folgenden 35 Jahren wurden eine Vielzahl an SOM Varianten und auf SOM Berechnungen basierende Visualisierungen (z.B. [86], [87], [88]) entwickelt. Als DRvis Methode repräsentiert der Algorithmus neben t-SNE den derzeitigen State of the Art.

### 2.3.4 t-SNE - t-Distributed Stochastic Neighbor Embedding

t-SNE [33] ist eine nichtlineare, nicht-parametrische DRvis Methode, die auf die Erhaltung der lokalen, kleinräumigen Nachbarschaftsverhältnisse des Feature-Raums abzielt. Der Algorithmus wurde als DRvis Methode konzipiert und ist für eine Dimensionsreduktion  $d' > 3$  (wie sie im Bereich Machine

Learning häufig als Verarbeitungsschritt eingesetzt wird) nicht geeignet. (vgl. [33]) Der Algorithmus adaptiert seine Berechnungen an die Struktur der multivariaten Daten, für unterschiedliche Regionen des Feature-Raums werden unterschiedliche Abbildungen in den Visualisierungs-Raum berechnet. (vgl. [89])

Erster Berechnungsschritt von t-SNE ist die Umwandlung aller paarweisen, euklidischen Distanzen der Punkte im Feature-Raum in bedingte Wahrscheinlichkeiten (Formel 4).

$$p_{j|i} = \frac{e^{-\frac{|x_i - x_j|^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{|x_i - x_k|^2}{2\sigma_i^2}}} \quad (4)$$

$p_{j|i}$  ist die bedingte Wahrscheinlichkeit, dass ein gegebener Punkt  $x_i$  einen anderen Punkt  $x_j$  als Nachbar wählt, wenn Nachbarpunkte anhand der Wahrscheinlichkeitsdichte einer, über dem Punkt  $x_i$  zentrierten, Normalverteilung gewählt werden.  $p_{j|i}$  verhält sich proportional zu den Distanzen des Feature -Raums. Liegen die Punkte  $x_i$  und  $x_j$  weit auseinander wird mit Formel 3 (bei geeigneter Wahl von  $\sigma_i^2$ ) eine extrem geringe, gegen null gehende Wahrscheinlichkeit berechnet. Liegen die Punkte eng beieinander ergibt sich eine hohe Wahrscheinlichkeit. Da nur paarweise Wahrscheinlichkeiten berechnet werden und kein Punkt sein eigener Nachbar ist wird  $p_{i|i}$  auf null gesetzt. Die Division mit dem Term  $\sum_{k \neq i} e^{-\frac{|x_i - x_k|^2}{2\sigma_i^2}}$  in Formel 3 dient der Normalisierung der  $p_{j|i}$  Werte.

Die Varianz  $\sigma_i^2$  der über dem Punkt  $x_i$  zentrierten Normalverteilung wird für jeden Punkt  $x_i$  bestimmt und an die Verteilung der Punkte im Feature-Raum angepasst. Ausgehend von der Annahme, dass die Punkte nicht gleichmäßig im Feature- Raum verteilt sind, sondern sich dichtere Regionen (= viele Punkten in einem bestimmten Volumen des Feature-Raums) mit dünnbesetzten Regionen (= wenige Punkte in einem bestimmten Volumen) abwechseln, wird die Varianz an diese Verteilung angepasst. Für dichtere Regionen wird eine geringere Varianz berechnet, für dünnbesetzte Regionen eine höhere Varianz.

Jeder bestimmte Wert von  $\sigma_i^2$  erzeugt eine andere Wahrscheinlichkeitsverteilung  $P_i$  (mit  $p_{j|i}$  Werten für einen bestimmten Punkt  $x_i$ ). T-SNE verwendet eine binäre Suche um eine Varianz  $\sigma_i^2$  zu bestimmen deren erzeugte Wahrscheinlichkeitsverteilung  $P_i$  eine Perplexity (Formel 5) aufweist, die einer von der Anwenderin vorgegebenen Perplexity entspricht. Die von der Anwenderin vorgegebene Perplexity skaliert damit die Varianz  $\sigma_i^2$  der über dem Punkt  $x_i$  zentrierten Normalverteilung so dass jedem Punkt  $x_i$  eine bestimmte, ungefähr gleich große Anzahl an Punkten mit relativ hohen bedingten Wahrscheinlichkeiten  $p_{j|i}$  zugeordnet wird. Die Perplexity kann daher als „smooth measure of the effective number of neighbors“ [33] interpretiert werden.

$$Perp(P_i) = 2^{H(P_i)} \quad (5)$$

Der Term  $H(P_i)$  ist die in bits gemessene Shannon Entropie (Formel 6)

$$H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i} \quad (6)$$

Im nächsten Verarbeitungsschritt wird aus der nicht symmetrischen  $n \times n$  großen Matrix mit allen bedingten Wahrscheinlichkeiten  $p_{j|i}$  mit Formel 7 eine symmetrische  $n \times n$  große Matrix mit Verbundwahrscheinlichkeiten  $p_{ij}$  erstellt. Diese symmetrische Ähnlichkeitsmatrix ermöglicht (im Vergleich zur asymmetrischen Ähnlichkeitsmatrix mit  $p_{j|i}$  Werten) eine effizientere Optimierung des

Visualisierungs-Raums und eine verbesserte Positionierung von Ausreißern im Visualisierungs-Raum. (vgl. [33])

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \quad (7)$$

Ähnlich wie für den Feature-Raum kann auch für den Visualisierungs-Raum eine Ähnlichkeitsmatrix definiert werden. Die Distanz eines Punktes  $y_i$  im Visualisierungs-Raum zu einem anderen Punkt  $y_j$  wird ebenfalls durch eine Verbundwahrscheinlichkeit  $q_{ij}$  dargestellt. Anstatt der Dichtefunktion einer Normalverteilung wird für den Visualisierungs-Raum die Dichtefunktion einer t-Verteilung mit einem Freiheitsgrad verwendet. Die Verbundwahrscheinlichkeit  $q_{ij}$  wird daher wie in Formel 8 dargestellt berechnet. (Im Feature Raum wird die Form der Dichtefunktion für jeden Punkt  $x_i$  einzeln bestimmt. Im Visualisierungs-Raum wird für alle Punkte  $y_i$  dieselbe Dichtefunktion verwendet.  $q_{ii}$  wird wie im Feature-Raum auf null gesetzt)

$$q_{ij} = \frac{(1 + |y_i - y_j|^2)^{-1}}{\sum_{k \neq l} (1 + |y_k - y_l|^2)^{-1}} \quad (8)$$

Die Menge aller  $p_{ij}$  Werte ist eine diskrete Wahrscheinlichkeitsverteilung mit einer Wahrscheinlichkeitsfunktion  $P$ . Ebenso stellt die Menge aller  $q_{ij}$  eine diskrete Wahrscheinlichkeitsverteilung mit Wahrscheinlichkeitsfunktion  $Q$  dar. Sind  $P$  und  $Q$  gleich wird die Struktur des Feature-Raums im Visualisierungs-Raum perfekt abgebildet. Je kleiner der Unterschied zwischen den beiden Wahrscheinlichkeitsverteilungen ist desto besser gelingt die Erhaltung der Struktur des Feature-Raums im Visualisierungs-Raum. t-SNE versucht daher eine Anordnung der Punkte im Visualisierung-Raum zu ermitteln, deren Wahrscheinlichkeitsverteilung sich möglichst geringfügig von der bereits bekannten Wahrscheinlichkeitsverteilung des Feature-Raums unterscheidet.

Die Kullback-Leibler-Divergenz (KL-Divergenz) ist ein Maß für die Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen. Die KL – Divergenz ist eine nicht-negative Zahl und nimmt den Wert null nur dann an, wenn beide Wahrscheinlichkeitsverteilungen gleich sind. Wird in die Definition der KL-Divergenz die beiden Wahrscheinlichkeitsfunktionen  $P$  und  $Q$  eingesetzt ergibt sich Formel 9.

$$KL(Q||P) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (9)$$

t-SNE ermittelt im abschließenden Verarbeitungsschritt mit Hilfe eines Gradienten Verfahrens jene Anordnung von  $y_i$  Punkten im Visualisierung-Raum deren  $KL(Q||P)$  minimal ist. Ausgehend von zufälligen Koordinaten der  $y_i$  Punkten wird die Anordnung der Punkte in mehreren Iterationen optimiert.

### 2.3.5 Quantitative und qualitative Evaluierung von DRvis Methoden

Falls für die hochdimensionalen Punkte Klassenlabels vorhanden sind können die Qualität der Visualisierung sowohl qualitativ als auch quantitativ beurteilt werden. Qualitativ lässt sich optisch feststellen ob die Visualisierung die, durch die Labels gebildete Klasseneinteilung abbildet. Mit der „k-nearest neighbor (kNN) technique“ [79] lässt sich in diesem Fall ein quantitatives Maß für die Qualität der DRvis berechnen. Das Label jedes Punktes wird sowohl im Feature-Raum als auch im Visualisierungs-Raum mit dem Mehrheitslabel seiner k Nachbarn verglichen. Erhöht sich die Anzahl der

falsch klassifizierten Punkte durch die DRvis, so ist das ein Indiz für die schlechte Qualität der Visualisierung. Da Klassenlabels typischerweise für hochdimensionale Daten nicht verfügbar sind wurden mehrere Qualitätsmetriken entwickelt, für deren Berechnung keine Klassenlabels notwendig sind.

Eine Reihe von Publikationen beschäftigen sich mit der Frage mit welchen Qualitätsmetriken die DRvis Ergebnisse quantitativ beurteilt werden können. In [90] wird z.B. das „Trustworthiness / Continuity“ Konzept vorgestellt, in [91] das „mean relative rank errors“ Konzept und in [92] das „co- ranking matrix“ Konzept. Alle in diesen Publikationen vorgestellten Qualitätsmetriken messen unter Anderem wie strukturgetreu k-große Nachbarschaften des Feature-Raums in der Visualisierung abgebildet werden. Das Design dieser Qualitätsmetriken entsprechen damit dem wissenschaftlichen Konsens der DRvis Forschungsgemeinde: Gute Visualisierungen müssen vor allem lokale Nachbarschaftsverhältnisse des Feature- Raums strukturgetreu in der Visualisierung abbilden. Die korrekte Darstellung lokaler, kleinräumiger Nachbarschaftsverhältnisse des Feature-Raums ist für Qualität der Visualisierung wichtiger als die exakte Abbildung aller Nachbarschaftsverhältnisse (vgl. [93]).

Anschaulich lässt sich dieser Konsens anhand zwei Clustern des Feature-Raums erklären. In der Visualisierung sollten sich die zwei Cluster erkennen lassen und die Strukturen innerhalb der Cluster korrekt abgebildet werden. Die Distanz der beiden Cluster zueinander und das exakte Distanzverhältnisse eines Punktes des einen Clusters zu einem Punkt des anderen Clusters ist für eine gute Visualisierung weniger wichtig. Befinden sich in einer Sample-Datenbank 50 Bassdrum-Samples und 50 Violinen-Samples so bilden diese zwei Sample-Gruppen aufgrund ihrer spektralen Eigenschaften zwei Cluster im Feature -Raum. In der Visualisierung sollte die Anwenderin die beiden Gruppen ausmachen können und die Ähnlichkeitsstruktur innerhalb der beiden Gruppen sollte korrekt visualisiert werden. Die Distanz der beiden Cluster zueinander oder die Ähnlichkeit eines beliebigen Violinen-Samples zu einem beliebigen Bassdrum-Sample sind für eine gute Visualisierung weniger ausschlaggebend.

Das Trustworthiness / Continuity Konzept (vgl. [82], [94], [90]) geht davon aus, dass bei einer DRvis zwei Arten von Fehler auftreten können:

1. Ein Punkt der im hochdimensionalen Feature-Raum nicht in der Nachbarschaft eines anderen Punkt liegt, wird im Visualisierung-Raum in der Nachbarschaft des andern Punktes abgebildet. Ein derartiger Fehler verringert die Trustworthiness der Visualisierung, da die Visualisierung Nachbarschaftsverhältnisse zeigt die im Feature-Raum nicht existieren.
2. Ein Punkt, der im Feature-Raum in der Nachbarschaft eines anderen Punktes liegt wird im Visualisierungs-Raum nicht in der Nachbarschaft des Punktes abgebildet. Derartige Fehler verringern die Continuity einer Visualisierung, da nicht alle Nachbarschaftsverhältnisse des Feature-Raums in der Visualisierung abgebildet werden.

Beide Maßzahlen nehmen Werte zwischen 1 und 0 an und werden für unterschiedlich k-große Nachbarschaften berechnet. Zwischen den beiden Maßzahlen besteht ein Zielkonflikt. Analog zu dem aus dem Bereich „Information retrieval“ bekannten Zielkonflikt zwischen precision und recall führt z.B. die Verbesserung der Trustworthiness einer Visualisierung in vielen Fällen zu einer Verschlechterung der Continuity.

In [82] argumentieren die Autoren, dass die Trustworthiness einer Visualisierung für die wahrgenommene Qualität einer DRvis Visualisierung ausschlaggebender ist als die Continuity der Visualisierung. Aus Sicht der Anwenderin sind die Nähe der Punkte zueinander und welche Punkte ein optisches Cluster bilden die hervorstechendsten Eigenschaften von DRvis Visualisierungen. „The

proximities that are visible on the display are salient, and if they are not trustworthy the whole display is misleading. In contrast, if all similar samples cannot be placed proximate, the consequence is only that potentially useful discoveries may be overlooked.“ [82] Die Argumentation bezieht sich auf das Anwendungsszenario „data exploration“ [82], trifft aber wahrscheinlich auch für Sample-Browser zu. Einige Samples bilden ein optisches Cluster im Userinterface des Browsers. Befinden sich in diesem Cluster einander unähnliche Samples, beeinträchtigt das die wahrgenommene Qualität der Visualisierung mehr als ein Sample, das sich nicht in dem Cluster befindet, aber den Samples des Clusters ähnlich ist.

In [90] betonen die Autoren, dass die Qualität einer DRvis Visualisierung wesentlich davon bestimmt wird „...that the small neighborhoods, having a small  $k$ , are thrustworthy.“ In der Studie werden zwei SOM Varianten, zwei MDS Varianten und eine PCA anhand von drei multivariaten Datensätzen und dem Trustworthiness / Continuity Konzept evaluiert. Für alle drei Datensätze weisen die beiden SOM Varianten für Nachbarschaften deren Größe  $k$  5-10 % der Anzahl der Punkte im Feature-Raum nicht übersteigt wesentlich höhere Trustworthiness Werte auf.

In [79] werden einige parametrische und nicht-parametrische DRvis Methoden quantitativ anhand der „co-ranking matrix“ [92] Qualitätsmetrik verglichen. Für den Vergleich wird der MNIST Datensatz verwendet. Der Datensatz besteht aus 60000 28 x 28 pixelgroßen Bildern handgeschriebener Ziffern von 0 – 9. Jedes Bild stellt einen 784-dimensionalen Feature-Vektor dar, der aus den Helligkeitswerten jedes Pixels besteht. Die Ergebnisse zeigen, dass t-SNE im Vergleich sowohl mit einer PCA als auch einigen anderen nichtlinearen, nicht-parametrischen DRvis Methoden (z.B. Manifold charting, Sammon's mapping, Isomap, LLE) die Erhaltung lokaler Nachbarschaften des Feature-Raums wesentlich besser gelingt. Mit der Ausnahme von t-SNE zeigen alle getesteten DRvis Methoden für Nachbarschaftsgrößen  $k < 100$  „...low quality for small neighborhood ranges. Hence these techniques seem inappropriate to inspect high-dimensional real data sets visually.“ [79]

Mit der Zielsetzung DRvis Methoden auf ihre Eignung für Sample-Datenbanken zu prüfen werden in [95] PCA, Isomap und t-SNE miteinander verglichen. Verwendet werden zwei professionelle Sample Datenbanken für die Musikproduktion bzw. für die Film- und Videovertonung. Die Studie behandelt damit die zwei wesentlichen Anwendungsszenarien für Sample Browser mit einem praxisnahen Versuchsaufbau. Als inhaltsbasierte Features werden mehrere statistische Lagemasse von MFCCs und in vier unterschiedlichen Frequenzbändern gemessene Spectral Flatness für die Samples der beiden Datenbanken berechnet. Um die drei getesteten DRvis Methoden quantitativ zu beurteilen wird das Trustworthiness / Continuity Konzept verwendet. Die quantitativen Evaluierungsergebnisse zeigen die wesentlich bessere Qualität der t-SNE Visualisierung. Insbesondere ergeben sich für beide verwendeten Sample-Datenbanken für Nachbarschaftsgrößen unter  $k = 100$  wesentlich höhere Trustworthiness Werte. Man kann daher davon ausgehen, dass t-SNE lokale, kleinräumige Nachbarschaftsverhältnisse, die für die Qualität der Visualisierung ausschlaggebend sind, auch für aus Audio-Samples extrahierte Datensätze wesentlich besser abbildet als Isomap (= eine MDS Variante) oder eine PCA. Im Rahmen der Studie wurden von Testusern die PCA und t-SNE Visualisierungen in einem Sample-Browser auch qualitativ beurteilt. „...end-users acknowledged that t-SNE-based visualizations were more useful and interesting than PCA-based ones. Despite subgroups of some sound categories (like specific timbres of brass instruments, of some rhythmic guitar playing phrases) got some- times farther away from other instances of their category, the visualization created less surprising and annoying phenomena, such as sounds from different categories appearing intermingled in the visualization“ [95]

## 2.4 Zusammenfassung

Kapitel 3 bietet eine Übersicht über einige Sample-Browser Forschungsprototypen, die in den letzten 15 Jahren entwickelt wurden. Die meisten der hier behandelten Sample-Browser führen nach einer inhaltsbasierten Merkmalsextraktion eine Dimensionsreduktion durch. In den Abschnitten „Inhaltsbasierte Merkmalsextraktion“ und „Dimensionsreduktion“ wurde gezeigt, dass beide Verarbeitungsschritte, die Usability dieser Sample-Browser beeinflussen.

Die behandelten Sample-Browser stellen die Klangfarbe von Samples mit einigen Kennzahlen (= Features) dar. Es wurde gezeigt, dass dieses Vorgehen eine stark vereinfachende Modellierung eines komplexen Phänomens ist, dass von einigen Psychoakustikern als „nicht restlos erforscht“ beurteilt wird. Basierend auf diesen Kennzahlen verwenden die meisten der Sample-Browser ein geometrisches Ähnlichkeitsmodell, das von Psychologen stark bezweifelt wird. MIR Forscher weisen darauf hin, dass mit diesem Ähnlichkeitsmodell nur in begrenztem Umfang sinnvolle Retrieval Resultate erzielt wurden.

Die zu 100% strukturgetreue Abbildung des Feature-Raums auf einen 2 oder 3-dimensionalen Visualisierungs-Raum ist mit den derzeitigen State of the Art DRvis Methoden nicht möglich. Viele DRvis Forscher gehen davon aus, dass eine derartige Abbildung für nichtlinear verteilte, multivariate Daten generell nicht möglich ist. Selbst wenn das geometrische Ähnlichkeitsmodell eine perfekte Modellierung der Sample-Klangfarbe darstellen würde, hat die Dimensionsreduktion eine Verzerrung der ursprünglichen Ähnlichkeitsverhältnisse des Feature-Raums zur Folge.

### 3 Analyse der ähnlichkeitsbasierten Sample-Browser und resultierende Forschungsfragen

In der Praxis werden ähnlichkeitsbasierte 2D Karten als Sample-Browsing Werkzeuge kaum verwendet. Die meistverwendeten DAWs (Digital Audio Workstation) bieten für die Sample-Suche nach wie vor nur listenbasierte Interfaces und eine Stichwortsuche. Zu den DAWs, die als Industrie Standard gelten, zählen „Ableton“, „ProTools“, „Logic Pro“, „FL Studio“, „Cubase“, „Reason“, „Sonar“, „Studio One“ und „Bitwig Studio“. Eine informelle Recherche welche Sample-Browsing Interfaces in diesen DAWs implementiert sind zeigt, dass keine der erwähnten DAWs eine ähnlichkeitsbasierte 2D Karte bietet.

Bereits 2001 halten die Entwickler eines Samplebrowsers fest: „Most audio editing tools offer limited capabilities for browsing and editing large collections of audio files (typically only the filename and some form of ID tagging is used).“ [13] 2008 stellen die Entwickler von SoundTorch fest das sich an dieser Situation nichts geändert hat: „For at least the last two decades, the processes for managing, browsing and auditioning sound samples on a computer have remained the same. Auditioning mostly happens file by file, which consumes time and kills creativity“ [9] Diese Analysen sind mit Hinblick auf die erwähnten DAWs auch im Jahr 2016 noch gültig.

Es gibt für die Hersteller von DAWs sicherlich keine technologischen Barrieren, die eine Implementierung von ähnlichkeitsbasierten 2D Karten unmöglich machen. Feature Extraction- und Dimensionsreduktions-Algorithmen sind in der Zwischenzeit relativ leicht zu verwendende und zumeist frei verfügbare Standardtechnologie geworden. Die internationale MIR (Music Information Retrieval) Forschungsgemeinde ist Open Source freundlich und stellt Feature Extraction Toolboxes mit hunderten von Features zur Verfügung. Für die meisten Programmiersprachen existieren eigene Implementierungen der Dimensionsreduktions-Algorithmen.

Nach 15 Jahren universitärer Forschung und einer Vielzahl von vorgestellten Prototypen bleibt der Befund das der Ansatz aus inhaltsbasierter Merkmalsextraktion, Dimensionsreduktion und Sonifikation einer ähnlichkeitsbasierten 2D Karte in der Praxis keine Verwendung gefunden hat. Das dieser Ansatz bisher in keiner kommerziellen DAW implementiert wurde zwingt zu einer kritischen Auseinandersetzung. Bietet dieser Ansatz der Anwenderin keinen wesentlichen Mehrwert zu den herkömmlichen listenbasierten Interfaces? Was sind die Schwächen dieses Ansatzes?

#### 3.1 Usability von ähnlichkeitsbasierten 2D Karten als Sample-Browsing Werkzeuge

Die meisten ähnlichkeitsbasierten Sample-Browser visualisieren einzelne Samples als grafische Objekte auf einer 2D Karte. Der Abstand der einzelnen grafischen Objekte vermittelt der Anwenderin wie ähnlich sich bestimmte Samples sind. Nahe beieinanderliegende Samples weisen eine große Ähnlichkeit auf. Weit entfernt liegende Samples weisen eine geringe Ähnlichkeit auf.

Für diesen Ansatz sprechen Erkenntnisse der Gestaltpsychologie, die später durch die Neurophysiologie bestätigt wurden. Forschungsergebnisse auf beiden Gebieten zeigen, dass der Abstand zwischen grafischen Objekten eine der besten Ähnlichkeitsvisualisierungen ist. Nahe beieinanderliegende graphische Elemente werden intuitiv als zusammengehörig und ähnlich wahrgenommen. Einzelne graphische Elemente ohne direkte Nachbarn werden als „Ausreiser“ erkannt. (vgl. [96]) Abbildung 31 dient zur Verdeutlichung. Eine Betrachterin nimmt in der Abbildung drei Gruppen und ein einzelnes Element als „Ausreiser“ wahr.



Abbildung 31: Ähnlichkeit visualisiert durch Abstand. Hier werden mit einem Blick drei zusammengehörige Gruppen und ein Ausreißer wahrgenommen.

Die ersten Autoren, die im Jahr 2011 einen der wesentlichen Nachteile des Ansatzes beschreiben sind Grill und Flexer: „A drawback of this method is that the projection axes of the low-dimensional space are not pre-defined and/or not interpretable. When used as a visualization as part of a user interface, the systematics of organization consequently have to be learned by the user by means of exploration“ [29] Ähnlich äußern sich im Jahr 2016 Gomez et al. : „This visualization represents all samples identically and displays no axis or any indicator of the timbre similarity. It lacks transparency and does not allow the user to understand the organization of the sample space.“ [6]

Beide Zitate beziehen sich auf ähnlichkeitsbasierte 2D Karten in denen alle Samples mit demselben grafischen Objekt (meistens ein einfarbiger Kreis) dargestellt werden und keine zusätzliche Visualisierung vorhanden ist, die klangliche oder akustische Eigenschaften der Samples darstellt. Die einzige Information die eine solche Darstellung bietet ist, dass sich nah beieinanderliegende Samples ähnlich sind, also bestimmte akustische und klangliche Gemeinsamkeiten haben. Welche akustischen und klanglichen Eigenschaften das sind kann die Anwenderin nicht erkennen ohne sich die Samples anzuhören.

Die Achsen der 2D Karte sind keiner bestimmten klanglichen oder akustischen Eigenschaft zugeordnet, die Karte ist daher nicht sofort durch die Anwenderin interpretierbar. Die Semantik der Karte muss erst durch die Anwenderin mittels Anhören einer Vielzahl an Samples ermittelt werden. Erst dadurch entsteht ein mentales Bild welche Samples mit welchen Eigenschaften wo auf der Karte platziert sind.

Um eine zielgerichtete Suche nach Samples mit bestimmten akustischen oder klanglichen Eigenschaften durchzuführen muss die Anwenderin wissen in welchem Bereich der Karte Samples liegen die diese Eigenschaften aufweisen. Dieses Wissen kann sich die Anwenderin nur durch eine längere Interaktion mit der 2D Karte erarbeiten. Sie muss sich mehrere Samples in unterschiedlichen Bereichen der Karte anhören und ein mentales Bild der klanglichen Eigenschaften in den verschiedenen Bereichen formen. Erst dann ist sie in der Lage bei weiteren Zugriffen auf die Karte zielgerichtet in einem bestimmten Bereich der Karte nach Samples mit bestimmten Eigenschaften zu suchen.

Das Formen dieses mentalen Bildes (= „learning the systematics of organization“ [29], „understanding the organization of the sample space“ [6]) kann bei hunderten oder tausenden Samples einige Zeit in Anspruch nehmen. Abbildung 32 zeigt eine ähnlichkeitsbasierte Karte mit 321 Snaredrum-Samples. Wie lange braucht eine Anwenderin um herauszufinden in welchem Bereich der Karte Snares mit bestimmten akustischen oder klanglichen Eigenschaften liegen? Wo auf der Karte liegen welche, wie klingenden Samples? Die Karte in Abbildung 32 wurde mit denselben Features, die für den Snare Drum Navigator [21] verwendet wurden (log attack time, temporal centroid und MFCC Koeffizienten, die getrennt sowohl für die Attack- und Releasephase berechnet wurden) und dem t-SNE Dimensionsreduktionalgorithmus erstellt.

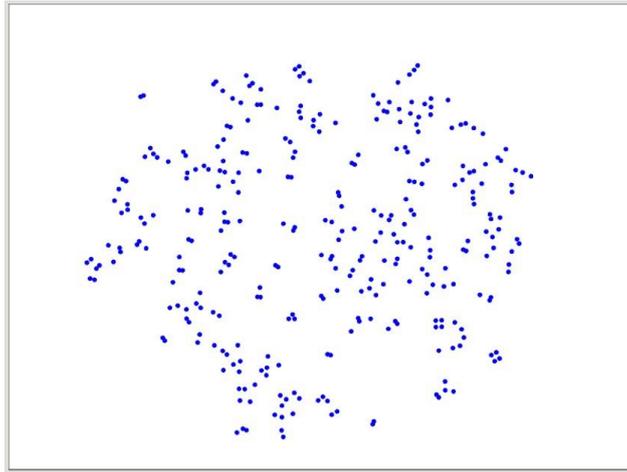


Abbildung 32: Ähnlichkeitsbasierte Karte von 321 Snaredrum-Samples. Wie lange braucht eine Anwenderin um herauszufinden wo auf der Karte welche, wie klingenden Snare-Samples liegen? Eine längere Interaktion mit der 2D Karte ist notwendig um die semantische Struktur der Karte zu erlernen. Für die Karte wurde als Features log attack time, temporal centroid und MFCC Koeffizienten (getrennt für die Attack- und Releasephase berechnet) und der t-SNE Dimensionsreduktionalgorithmus verwendet.

Hinweise, dass Anwenderinnen ähnlichkeitsbasierte 2D Karten wie in Abbildung 23 nicht als wesentliche Verbesserung empfinden, bevor sie die semantische Struktur der Karte durch eine längere Interaktion ermittelt haben sind in [6] und [2] zu finden. In beiden Fällen wurden Usertests durchgeführt ohne den Testteilnehmerinnen die Chance zu geben die semantische Struktur der Karte zu erlernen. “In the light of participants’ feedback, it appears that the similarity-based arrangement was unexpectedly difficult to understand and to use efficiently.” [6] In [2] vergleichen die Autoren eine einfache Gitteranordnung mit einer ähnlichkeitsbasierten 2D Karte: “content-based positioning of sounds on a map appeared to be inefficient vs a simple solution” [2] (content-based positioning= ähnlichkeitsbasierte 2D Karte, simple solution = Gitteranordnung)

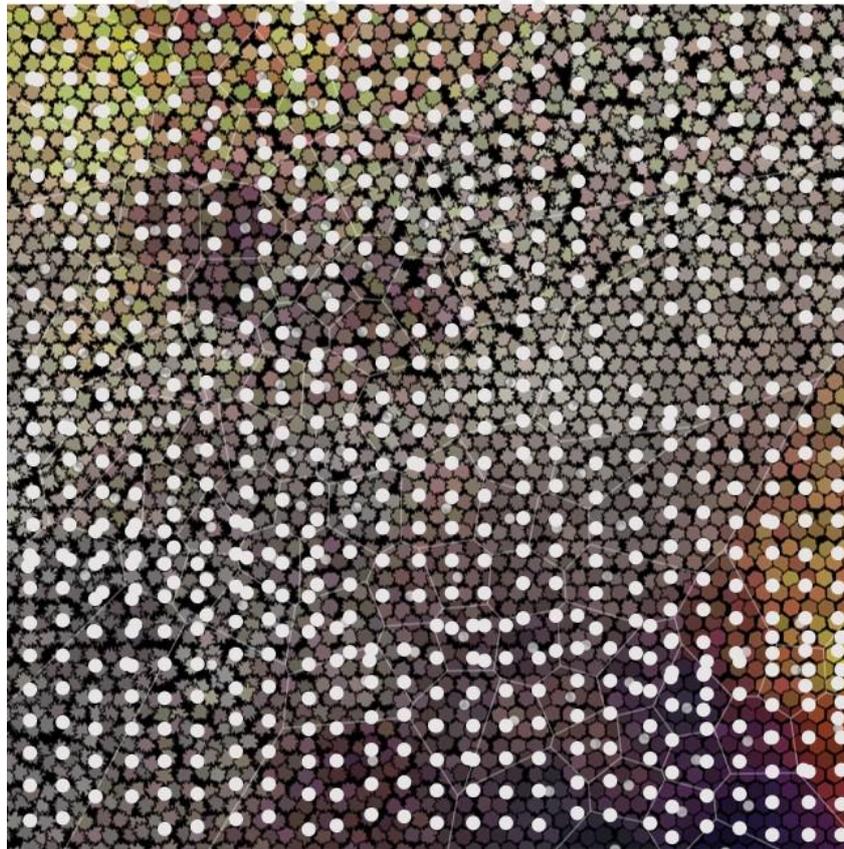
Unklar ist ob Anwenderinnen dieses mentale Bild überhaupt erstellen können. Sind sie in der Lage Bereiche mit ähnlich klingenden Samples zu identifizieren? Sind sie in der Lage die Karte kognitiv in unterschiedliche Bereiche mit gemeinsamen akustischen und klanglichen Eigenschaften aufzuteilen? Vorausgesetzt der Anwenderin gelingt es die Karte kognitiv in unterschiedliche Bereiche aufzuteilen, muss sie sich diese Aufteilung bis zum nächsten Zugriff auf die Karte merken um eine zielgerichtete Suche zu ermöglichen. Verstärkt wird diese Problematik durch die Tatsache, dass die 2D Karten für jede verwendete Sample-Sammlungen unterschiedlich aussehen. Die Anwenderin muss sich für jede Sample-Sammlung merken in welchen Bereichen der Karte welche Samples zu finden sind.

Die bisherige Argumentation in diesem Kapitel soll zeigen, dass die Usability von ähnlichkeitsbasierten 2D Karten ohne eine zusätzliche Visualisierung von akustischen und klanglichen Eigenschaften fragwürdig ist. Diese Benutzeroberflächen (wie sie in Drumspace [6], Soundtorch [9], AudioRetrievalBrowser [15], Marsyas3D [13], AudioMetro [2] verwendet wurden) sind nicht intuitiv verständlich. Grill und Flexer sind die ersten Autoren, die diesen Nachteil erkennen. Ihre gesamte Forschungsarbeit in [20, 29, 30] zielt auf eine intuitive Verständlichkeit der Benutzeroberfläche ab. Die Anwenderin soll durch ihre neuartige Visualisierung auf einen Blick, intuitiv erkennen können wo auf der Karte welche, wie klingenden Samples liegen.

Der dreistufige Designprozess von Grill und Flexer ist als Leitfaden für die Entwicklung von ähnlichkeitsbasierten Sample-Browsern hervorragend geeignet. Als erstes stellt sich die Frage nach welchen klanglichen und akustischen Eigenschaften Anwenderinnen eine bestimmte Art von Samples beurteilen. Der zweite Schritt ist dann die Entwicklung von Features, die diese Eigenschaften statistisch

unabhängig voneinander quantifizieren. Nach der Dimensionsreduktion erfolgt als dritter Schritt die Visualisierung dieser Eigenschaften um der Anwenderin eine intuitive Vorstellung davon zu vermitteln wo auf der Karte Samples mit bestimmten Eigenschaften zu finden sind.

Der als „proof of concept“ [20] im Internet veröffentlichte Prototyp von Grill und Flexer hat aber zwei Schwächen. Der Ansatz ist nur begrenzt skalierbar. Abbildung 33 zeigt den Prototypen von Grill und Flexer mit 400 Samples. Bei hunderten oder tausenden Samples überdecken die grafischen Objekte, die die Samples darstellen (= weiße Punkte) die Visualisierung. Die Wahrnehmung und Dekodierung der Visualisierung durch die Anwenderin wird mit steigender Sampleanzahl immer schwieriger.



*Abbildung 33: Prototyp von Grill und Flexer mit 600 Samples.*

Zweiter Kritikpunkt ist die gleichzeitige Darstellung von fünf verschiedenen klanglichen und akustischen Eigenschaften. Welche und wie viele klangliche oder akustische Eigenschaften visualisiert werden, sollte die Anwenderin je nach Anwendungsfall selbst bestimmen können.

Einige nach der Veröffentlichung von Grill und Flexer vorgestellte Sample-Browser Prototypen verwenden ebenfalls Designkonzepte, die die akustischen und klanglichen Eigenschaften der Samples auf der 2D Karte visualisieren. Der Snare Drum Navigator [21] verwendet farbkodierte Rechtecke, deren Farben Feature-Werte visualisieren. Die Designer von AudioMetro [2] visualisieren den Durchschnitt des perceptual Sharpness Features als Farbwert der Sample-Objekte und den zeitlichen Verlauf des perceptual Sharpness Features in der Umrisslinie der Sample-Objekte.

Neben der Visualisierung beeinflusst auch die Gestalt der gesamten Sample—Objekte auf der 2D Karte die Benutzerfreundlichkeit von ähnlichkeitsbasierten 2D Karten als Sample-Browsing Werkzeuge. Die beiden ähnlichkeitsbasierten 2D Karten in Abbildung 34 wurden für die Evaluierung in [6] und [2] verwendet und sind das Ergebnis einer contentbased Feature Extraction und Dimensionsreduktion. In

beiden Karten sind keine eindeutig getrennten Cluster erkennbar. Die Anwenderin hat eine einzige, diffuse Punktwolke vor sich.

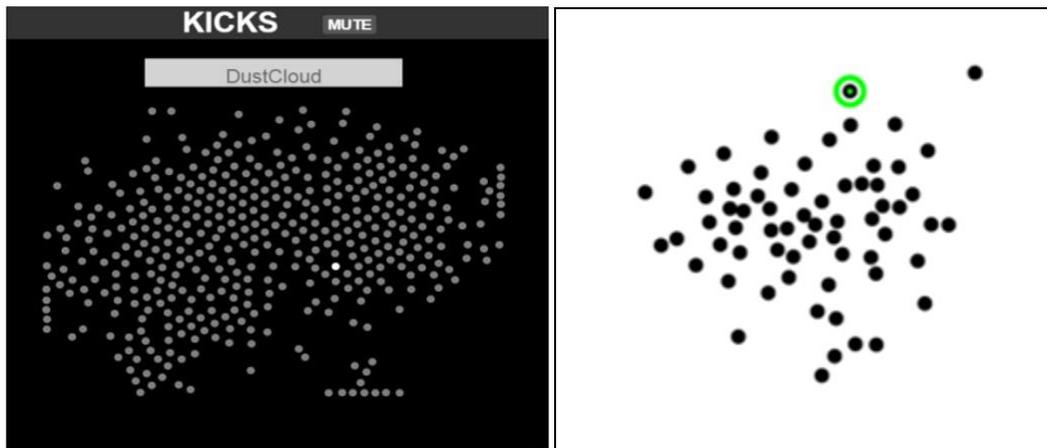


Abbildung 34: Für die Evaluierung von Sample-Browsern verwendete ähnlichkeitsbasierte 2D Karten. Beide Karten sind diffuse Punktwolken ohne erkennbare Cluster.

Abbildung 35 stellt die in [2] verwendete ähnlichkeitsbasierte 2D Karte einer 2D Karte mit klar erkennbaren Clustern gegenüber. In der Karte rechts erkennt die Anwenderin mit einem Blick Gruppen ähnlicher Samples, in der Karte links muss sie die Gruppierung durch eine längere Interaktion mit der Karte selbst herausfinden. Um beim nächsten Zugriff die Karte eine zielgerichtete Suche durchzuführen muss sich die Anwenderin merken wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften liegen. Auch diesen Vorgang unterstützt die Karte rechts wesentlich besser. In dieser Karte kann sich die Anwenderin merken das sich Samples mit bestimmten Eigenschaften z.B. im links untenliegenden Cluster befinden. In der Karte links muss sie sich den Verlauf der Umrisslinie der Gruppe einprägen.

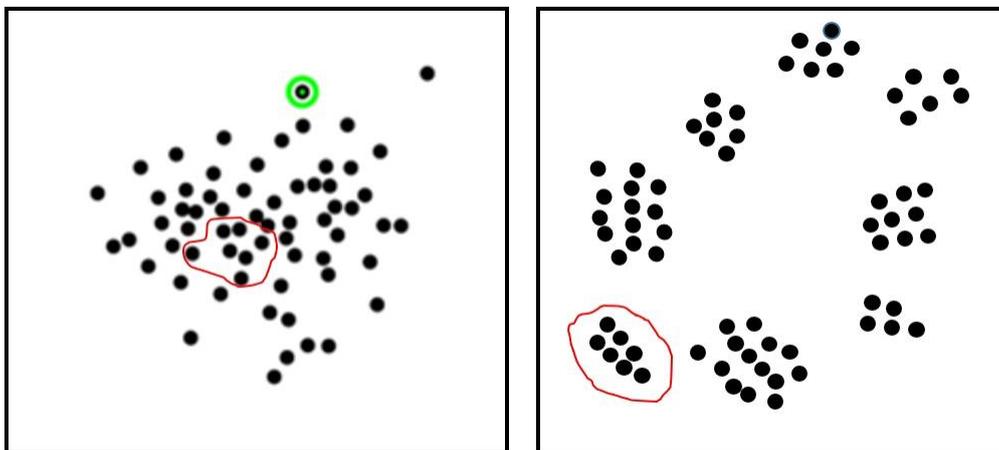


Abbildung 35: Rechts diffuse Punktwolke, links klar erkennbare Cluster.

Zusammenhängende Punktwolken mit relativ gleichmäßigen Abständen zwischen den einzelnen Sample-Objekten erschweren sowohl das Erstellen des mentalen Bildes klanglicher und akustischer Eigenschaften als auch seine Memorisierung für den nächsten Zugriff. Zusätzlich erleichtert eine 2D Karte mit klar erkennbaren Clustern das systematische Ansteuern aller Samples durch die Anwenderin. Die Usability von ähnlichkeitsbasierten 2D Karten hängt daher wesentlich davon ab ob sich mit der Kombination aus contentbased Feature Extraction und Dimensionsreduktion eindeutig getrennte Cluster erzielen lassen oder sich nur zusammenhängende Punktwolken ergeben.

Die Analyse der diversen Forschungsprototypen zeigt, dass ähnlichkeitsbasierte 2D Karten zwei Vorgänge möglichst optimal unterstützen sollten. Erstens sollte die Anwenderin mit einem Blick Gruppen von ähnlichen Samples erkennen können. Die 2D Karte sollten daher immer eindeutig (durch Abstand) getrennte Punktgruppen (= Cluster) anzeigen. Zweitens sollte die 2D Karte das systematische Ansteuern aller Samples durch die Anwenderin unterstützen. Enthält die 2D Karte klar erkennbare Cluster sind beide Vorgänge kein wesentliches Problem für die Anwenderin. Sie erkennt sofort Gruppen ähnlicher Samples und navigiert in dem sie einen Cluster nach dem Anderen ansteuert. Eindeutig getrennte Cluster stellen eine optische Hilfe für die Navigation dar.

Zeigt die Visualisierung nur eine einzige, diffuse Punktwolke ohne erkennbare Cluster werden die Ähnlichkeitswahrnehmung und die Navigation wesentlich schlechter unterstützt. In diesem Fall ist die 2D Karte ein optisches Rätsel für die Anwenderin. Wo liegen die Cluster Grenzen? Wo liegen Gruppen ähnlicher Samples? Eine große, diffuse Punktwolke bietet kaum Anhaltspunkte für eine systematische Navigation. Welcher Weg soll bei der Navigation gewählt werden?

### 3.1.1 Forschungsfragen - Usability von ähnlichkeitsbasierten 2D Karten

Die übergeordnete Forschungsfrage ob sich ähnlichkeitsbasierte 2D Karten als Sample-Browsing Werkzeuge eignen lässt sich durch folgende Detailfragen spezifizieren:

- Ist das Ergebnis der Verarbeitungskette aus contentbased Feature Extraction und Dimensionsreduktion ständig eine diffuse Punktwolke oder lassen sich genau getrennte Cluster erzielen?
- Welche Features und Feature-Kombinationen zeigen im Verbund mit welcher Reduktionsmethode die beste optische Clusterung für eine bestimmte Art von Samples?
- Eine Musikerin, die in hunderten von Snare-Drum Samples die für ihren Song passende Snare-Drum finden möchte, ein Gamesound-Designer der ein Schussgeräusch aus hunderten von Schussgeräuschen auswählt und ein Filmsound-Designer der versucht in hunderten von Wellengeräuschen ein passendes Wellengeräusch zu finden sind typische Anwendungsszenarien für Sample-Browser. Der gemeinsame Nenner dieser drei Anwendungsszenarien ist, dass die Anwenderin ein Sample aus hunderten oder tausenden Samples derselben Sample-Kategorie auswählen möchte. Lassen sich auch in diesem Fall mit der Verarbeitungskette aus Feature Extraktion und Dimensionsreduktion optisch sauber getrennte Cluster herstellen?
- Lässt sich eine schlechte optische Clusterung der 2D Karte durch eine visuelle Darstellung anderer Ähnlichkeitsberechnungen verbessern? Parallel zur Dimensionsreduktion könnte ein Clustering Algorithmus eine Klassenzugehörigkeit pro Sample berechnen. Die Sample-Objekte auf der 2D Karte könnten je nach Clusterzugehörigkeit farbcodiert werden um die Ähnlichkeitswahrnehmung und Navigation zu unterstützen.
- Wie beurteilen Anwenderinnen die Usability von ähnlichkeitsbasierten 2D Karten nachdem sie die semantische Struktur der Karte erlernt haben?
- Sind Anwenderinnen überhaupt in der Lage eine Karte mit hunderten von Samples kognitiv in Bereiche mit unterschiedlichen klanglichen und akustischen Eigenschaften aufzuteilen?
- Können Anwenderinnen sich diese Aufteilung bis zum nächsten Zugriff auf die Karte merken?
- Wissen Anwenderinnen nachdem sie versucht haben die Semantik der Karte zu erlernen beim nächsten Zugriff in welchen Bereichen der Karte sie Samples mit bestimmten klanglichen und akustischen Eigenschaften finden?

### 3.2 Algorithmische Ähnlichkeit und subjektives Ähnlichkeitsempfinden

Die Benutzeroberfläche von MediaCycle in Abbildung 36 stellt aus gestaltpsychologischer Sicht den optischen Idealfall dar. Zusammengehörigkeit und Ähnlichkeit der Samples sind durch Abstand und

Farbe ideal visualisiert. Ähnlichkeitswahrnehmung und Navigation werden dadurch gut unterstützt. Diese optische Visualisierung von Ähnlichkeit muss sich aber mit dem akustischen Ähnlichkeitsempfinden der jeweiligen Anwenderin decken. Die Aufteilung und Anordnung muss für die Anwenderin beim Anhören der einzelnen Gruppen nachvollziehbar sein. Wenn man eine Anwenderin vor die Aufgabe stellt die Samples manuell zu gruppieren hätte sie die selbe Gruppierung gewählt wie der k-Means Algorithmus?

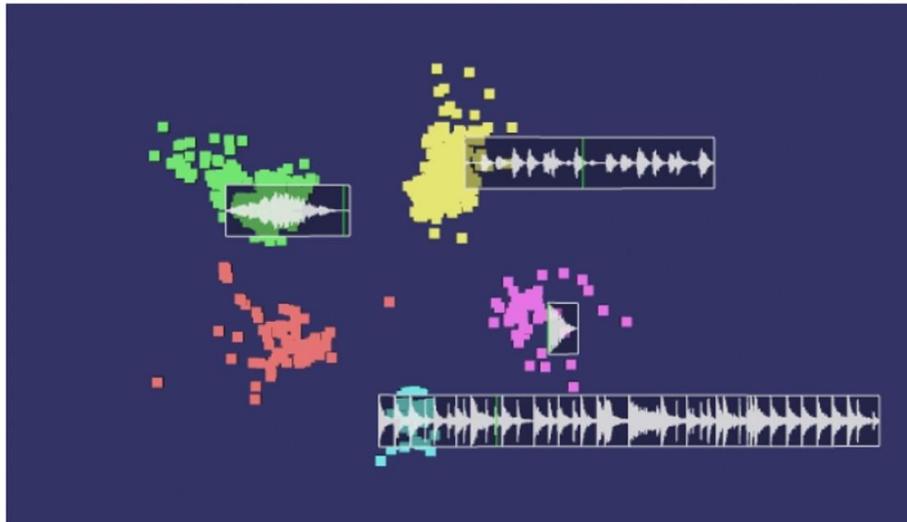


Abbildung 36: Der optische Idealfall aus gestaltpsychologischer Sicht. Abstand und Farbe visualisieren Zusammengehörigkeit und Ähnlichkeit. Aber deckt sich das akustische Ähnlichkeitsempfinden der Anwenderin mit der optischen Darstellung?

Grundlage für die Beurteilung eines ähnlichkeitsbasierten Sample-Browsers sind die individuellen, in einigen Fällen sehr unterschiedlichen, akustischen Ähnlichkeitsvorstellungen der Anwenderinnen. Die Frage in wie weit die algorithmisch ermittelte Ähnlichkeit der Samples mit dem Ähnlichkeitsempfinden der Testteilnehmerinnen übereinstimmt wird in keiner Veröffentlichung ausreichend genug erörtert. Fast alle Autoren begnügen sich damit festzustellen, dass die 2D Karten die Ähnlichkeit der Samples anzeigen. „The TimbreSpace reveals similarity...of sounds.“ [22] „Audio files are automatically arranged in 2D by psychoacoustic similarity“ [9] „The spatial pattern is based on an acoustic analysis that predicts how similar two audio files sound.“ [16] „...yielding clusters of resembling perceptual qualities.“ [20] „...using perceptual similarity algorithms for organizing this 2D visualization, so that two samples that sound similar are placed close to each other“ [6]

Fried et al., [21] sind die einzigen Autoren die eine kurze qualitative Beschreibung zu dieser Fragestellung liefern. Die Testteilnehmer „...were sometimes surprised by certain snares“ (=SnareDrumSamples) „that seemed not to fit their neighbors at all, “this snare should be over there”, was a common comment.“ [21] Das Zitat beschreibt das sich die Ähnlichkeitsempfinden der Teilnehmer in einigen Fällen nicht mit der durch Distanz visualisierten und algorithmisch errechneten Ähnlichkeit deckt.

Kann ein Algorithmus berechnen wie ähnlich eine Anwenderin zwei Samples empfindet? Die Entwickler der hier vorgestellten Sample-Browser beantworten diese Frage offensichtlich mit Ja. Psychologen bezweifeln die Gültigkeit des geometrischen Ähnlichkeitsmodells. Die Klangfarbenwahrnehmung wird grob vereinfachend als ein Reiz-Reaktionsmechanismus modelliert. Diese Modellierung steht im Widerspruch zu einigen aktuellen psychoakustischen und neurologischen Erkenntnissen (siehe Abschnitt „Gültigkeit des geometrischen Ähnlichkeitsmodells“) Selbst wenn die Distanz von Feature-Vektoren exakt dem Ähnlichkeitsempfinden einer Anwenderin

entspricht führt die DR höchstwahrscheinlich zu einer Verzerrung der Distanzverhältnisse des n-dimensionalen Raums. Die Ähnlichkeitsverhältnisse des n-dimensionalen Raums werden in der 2-D Darstellung leicht verzerrt dargestellt. (siehe Kapitel „Dimensionsreduktion“)

### 3.2.1 Forschungsfragen - Algorithmisch berechnete Ähnlichkeit und subjektives Ähnlichkeitsempfinden

Aus der Analyse im vorhergehenden Abschnitt ergeben sich folgende Forschungsfragen:

- Deckt sich das subjektive Ähnlichkeitsempfinden von Anwenderinnen mit der algorithmisch errechneten Ähnlichkeit?
- Die Annahme, dass Algorithmen das subjektive Ähnlichkeitsempfinden von Menschen exakt modellieren, kann als naiv bezeichnet werden. Die praktisch erzielten Ergebnisse der MIR Forschung (z.B. Instrumentenerkennung) lassen die Vermutung zu, dass Algorithmen teilweise in der Lage sind das Ähnlichkeitsempfinden von Anwenderinnen nachzubilden. Sind die Berechnungen der in dieser Arbeit vorgestellten, ähnlichkeitsbasierten Sample-Browser für Anwenderinnen soweit nachvollziehbar das sich daraus ein praktischer Nutzen für das Sample-Browsing ergibt?

### 3.3 Beschleunigtes Sample-Browsing durch selektive, auditive Aufmerksamkeit

In einigen Veröffentlichungen ([9], [16], [12], [13], [15]) wird argumentiert, das sich Sample-Browsing durch das gleichzeitige Anhören mehrerer Samples beschleunigen lässt. Die Autoren dieser Veröffentlichungen verweisen dabei auf den Cocktail Party Effekt, der folgendes Anwendungsszenario ermöglichen soll. Die Anwenderin hört sich einen Mix aus mehreren Samples an, lenkt ihre Aufmerksamkeit gezielt von einem Sample zum nächsten Sample in diesem Mix und kann dadurch eine größere Anzahl an Samples innerhalb einer kürzeren Zeitspanne auf die Eignung für einen bestimmten Einsatzzweck prüfen.

Der Begriff „Cocktail Party Effekt“ wurde 1953 von Cherry in [97] geprägt. „How do we recognize what one person is saying when others are speaking at the same time (the „cocktail party problem“)?“ [97] Wie schafft es das menschliche Gehirn sich selektiv auf eine Schallquelle zu konzentrieren und andere, gleichzeitig zu hörende Schallquellen auszublenden? Cherry nimmt an, dass sich der Cocktailparty-Effekt hauptsächlich durch unterschiedliche Positionen der Schallquellen erklären lässt. Seine Testergebnisse lassen den Schluss zu, dass sobald zwei Sprachsignale aus unterschiedlichen Richtungen kommen, die Versuchsteilnehmer ihre Aufmerksamkeit mühelos auf eines der beiden Sprachsignale lenken können und es schaffen das andere Signal auszublenden.

Um Samples aus unterschiedlichen Richtungen abzuspielen verwendet MARSYAS3D [13] sechzehn Lautsprecher und einen PC der mit zwei 8 Kanal Soundkarten betrieben wird. SoundTorch [9],[16] verwendet ebenfalls mehrere Lautsprecher. In der Praxis haben Musiker, Gamesound-Designer und Toningenieure, die sich mit der Vertonung von Videos und Filmen beschäftigen häufig nur ein Paar Stereolautsprecher oder Kopfhörer zur Verfügung. Der Anwenderin den Eindruck zu vermitteln, dass Samples aus unterschiedlichen Schallquellen mit unterschiedlicher Position stammen ist aber auch auf Kopfhörern und Stereolautsprechern durch eine Anordnung der Samples im Stereopanorama oder mittels Head-Related Transfer Function möglich. (vgl. [98]) Das in der Einleitung skizzierte Anwendungsszenario lässt sich also grundsätzlich auch ohne eine Vielzahl an Lautsprechern verwirklichen.

Stewart [99] bietet eine gute Übersicht über Interfaces die eine Sonifikation für Musik oder Samples anbieten. Viele der dort vorgestellten Interfaces können mehrere Audiosamples gleichzeitig abspielen. Eine während der Evaluierung dieser Interfaces häufig geäußerte Bemerkung der User ist „...that listening to more than one sound source can be confusing and that they can be quickly overwhelmed

with too much information“. [99] Die Autorin schließt daraus: „ There are limits on the number of sound sources that a user should be presented with.“ [99] Die menschliche Fähigkeit mehrere gleichzeitig abgespielte Klänge oder Geräusche zu unterscheiden hat Grenzen. Ab einer gewissen Anzahl von gleichzeitig abgespielten Samples kann die Anwenderin keine einzelnen Sounds mehr identifizieren oder sie nach ihrer Eignung für einen bestimmten Einsatzzweck beurteilen.

Wie viele gleichzeitig zu hörende Samples kann eine Anwenderin unterscheiden und erkennen? Die Entwickler von MARSYAS3D [13] meinen „...that at most 8 simultaneous audio streams are useful.“ Die Designer von SoundTorch stellen in [9] implizit die die Behauptung auf, das professionelle Anwenderinnen in der Lage sind 12 gleichzeitig abgespielte Samples zu unterscheiden. Die Ergebnisse der Hörtests in [100] deuten daraufhin das die Fähigkeit mehrere gleichzeitig abgespielte Samples zu unterscheiden und zu erkennen bereits bei 3 gleichzeitig abgespielten Samples stark beeinträchtigt ist.

In der von Brazil et al. veröffentlichten Studie [100] werden den Versuchsteilnehmern 3, 6 oder 9 Samples gleichzeitig vorgespielt. Die Samples stammen aus Sample-Sammlungen die für die Vertonung von Filmen und Videos eingesetzt werden. Die Studie wertet ein Sample als „erkannt“ sobald eine Teilnehmerin in der Lage war eine korrekte, sprachliche Beschreibung eines Samples zu formulieren. Die Erkennungsrate sinkt dabei von 70,5% (3 gleichzeitig zu hörende Samples) auf 52,6% (5 gleichzeitig zu hörende Samples) und 39,3% (9 gleichzeitig zu hörende Samples). Die qualitativen Interviews der Versuchsteilnehmer nach den Tests zeigen, dass bei den Versuchsteilnehmern eine zunehmende kognitive Überforderung eintritt sobald sie die Tests mit 6 oder 9 gleichzeitig zu hörenden Samples absolvieren. „Several felt that “I know there are way more sounds, I’m just not getting them!”“ [100] Die Ergebnisse in [100] lassen sich nicht direkt auf das Aura Sonifikations-Konzept übertragen da alle Samples als Mono-Samples (d.h. ohne Richtungsinformation) abgespielt wurden, stellen aber die Usability des Aura Sonifikations-Konzepts in Frage.

Die Usability des Aura Sonifikations-Konzepts ist auch durch Maskierungseffekte in Frage gestellt, die bei mehreren, gleichzeitig abgespielten Samples auftreten. Überlagert sich ein bestimmter Frequenzbereich zweier Samples so kann die Anwenderin den Frequenzanteil mit geringer Lautstärke nur mehr abgeschwächt oder gar nicht mehr wahrnehmen. (vgl. [101]) Fraglich ist daher ob eine Anwenderin ein Sample auf Eignung für einen bestimmten Einsatzzweck prüfen kann wenn sie bestimmte spektrale Anteile des Signals nicht hört.

### 3.3.1 Forschungsfragen - Sonifikation

Aus der Analyse im vorhergehenden Abschnitt ergeben sich folgende Forschungsfragen:

- Hat das Aura Sonifikations-Konzept trotz sinkender Erkennungsraten und Maskierungseffekten, die beim Abspielen mehrerer Samples auftreten, einen praktischen Nutzen?
- Gib es eine bestimmte (höchstwahrscheinlich relativ geringe) Anzahl an gleichzeitig abgespielten Samples bei denen das Aura Sonifikations-Konzept noch funktioniert? Ab wie vielen gleichzeitig abgespielten Samples funktioniert das Aura Sonifikations -Konzept aufgrund von sinkenden Erkennungsraten und Maskierungseffekten nicht mehr?
- Schaffen es Anwenderinnen tatsächlich sich selektiv auf ein einzelnes Sample in einem Mix aus mehreren Samples zu konzentrieren und dieses Sample auf seine Eignung für einen bestimmten Einsatzzweck zu beurteilen?
- Wie verwenden Anwenderinnen eine solche Sonifikation?
- Lässt sich das Sample-Browsing durch das Aura Sonifikations-Konzept tatsächlich beschleunigen?
- Gibt es alternative Sonifikations-Konzepte mit größerem praktischen Nutzen?

### 3.4 Zusammenfassung

In diesem Kapitel wurden drei Problemfelder identifiziert, welche die Benutzerfreundlichkeit von ähnlichkeitsbasierten Sample-Browsern beeinträchtigen könnten:

1. Die Visualisierungen in der Benutzeroberfläche eines Sample-Browsers müssen die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind, unterstützen.
2. Der Nutzen von ähnlichkeitsbasierten Sample-Browsern ist stark davon abhängig, ob sich das subjektive, akustische Ähnlichkeitsempfinden von Anwenderinnen mit der algorithmisch errechneten, visualisierten Ähnlichkeit deckt.
3. Die Benutzerfreundlichkeit des Aura-Sonifikationskonzepts ist durch Maskierungseffekte und eine relativ rasch abfallende Erkennungsrate beim gleichzeitigen Abspielen von mehreren Samples in Frage gestellt.

Zu jedem dieser drei Problemfelder wurden in den jeweiligen Abschnitten Forschungsfragen formuliert, die durch einen partizipativen Designprozess geklärt werden sollen. Die Darstellung dieses Designprozesses und der Erkenntnisse, die aus diesem Prozess gewonnen wurden, erfolgt im anschließenden Kapitel.

## 4 Implementierung und Evaluierung des MM-Browsers

Die in dieser Arbeit vorgestellten Sample-Browser wurden häufig durch relative kurze, pro Anwenderin einmalige User-Tests evaluiert. Ähnlichkeitsbasierte Sample-Browser unterscheiden sich in vielen Aspekten deutlich von listenbasierten Interfaces und erfordern eine Lernphase um das volle Potential dieser Interfaces auszuschöpfen. In dieser Arbeit wurde die Idee verfolgt, zwei Anwender durch einen partizipativen Designprozess über einen längeren Zeitraum an einen ähnlichkeits-basierten Sample-Browser zu gewöhnen und erst danach zu beurteilen, wie die beiden Anwender den praktischen Nutzen dieses Design-Konzept einschätzen.

Anwender P. und T. waren am partizipativen Designprozess beteiligt. P. (40 Jahre alt, Beruf: Videocutter) verfügt über 15 Jahre Erfahrung in der Produktion elektronischer Musik. Er verwendet für seine Produktionen Programme wie Fruity Loops, Cubase, Protools und Ableton und setzt seine Stücke fast ausschließlich aus Samples zusammen. P. bezieht seine Samples online. (z.B. [www.bigfishaudio.com](http://www.bigfishaudio.com)) Die rasch wechselnden Sample-Sammlungen durchsucht P. bisher mit listenbasierten Interfaces der genannten Programme. Er schätzt an diesen Interfaces die Möglichkeit eine größere Anzahl an Samples automatisch hintereinander abzuspielen. Er durchsucht die Sample-Sammlungen anhand der Ordnerstruktur Ordner für Ordner oder schränkt die Menge der in Frage kommenden Samples durch eine Stichwortsuche ein falls die Sammlung herstellerseitig mit einem Stichwortverzeichnis versehen wurde.

Anwender T. (43 Jahre alt, Beruf: DJ, Musikproduzent, Videocutter) verfügt über 20 Jahre Erfahrung in der Produktion elektronischer Musik. Er verwendet für seine Produktionen Programme wie Ableton, Cubase, Nuendo und Reason. T. verwendet für seine Produktionen am liebsten Loop-Samples, die er online von vielen unterschiedlichen Websites bezieht. T. durchsucht seine Sample-Sammlungen ähnlich wie Anwender P. anhand der Ordnerstruktur oder verwendet Tags falls die Sammlung herstellerseitig mit einem Stichwortverzeichnis versehen wurde. Keiner der beiden Anwender hat jemals ein eigenes Stichwortverzeichnis angelegt oder ein bestehendes Verzeichnis erweitert.

Entwickelt wurde ein ähnlichkeitsbasierter Sample-Browser, der

- 3 verschiedene Sonifikationskonzepte,
- mehrere Markierungsmöglichkeiten für ausgewählte Samples,
- pro Sample-Sammlung eine Auswahl an unterschiedlichen 2D Karten,
- und zwei ähnlichkeitsbasierte Farbkodierungen der Sample-Punkte

bietet. Abbildung 37 zeigt den während des partizipativen Designprozesses entstandenen Multiple Maps-Browsers (= MM-Browser). Der MM-Browser wurde in der MAX 7 Entwicklungsumgebung implementiert. Für die Implementierung der verschiedenen Programmkomponenten wurden als Programmiersprachen Java, Java Script, Open GL und die Matlab Skriptsprache verwendet.

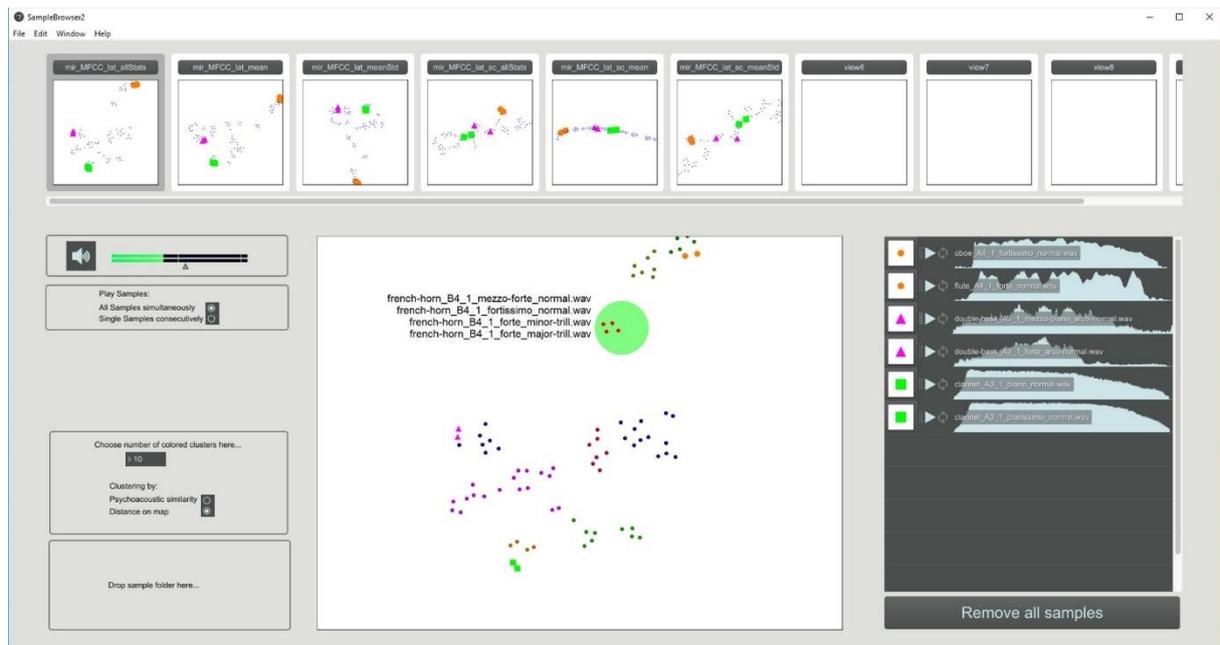


Abbildung 37: Während des partizipativen Designprozesses entwickelter MM-Browser (Multiple Maps Browser)

## 4.1 Usability der ähnlichkeitsbasierten 2D Karten des MM-Browsers

### 4.1.1 Feature-Kombinationen und optische Clusterung

t-SNE stellt den derzeitigen State of the Art für eine DRvis dar. Auf Grund der Ergebnisse in [33], [79] und [93] kann angenommen werden, das sich für multivariate Datensätze, deren Feature-Vektoren im hoch-dimensionalen Feature-Raum in Clustern angeordnet sind, derzeit mit t-SNE die beste optische Clusterung im Visualisierungs-Raum erzielen lässt. Die ursprüngliche Forschungsfrage „*welche Features und Feature-Kombinationen im Verbund mit welcher Reduktionsmethode die beste optische Clusterung für eine bestimmte Art von Samples zeigen*“ kann daher auf die Fragestellung „*welche Feature Kombinationen im Verbund mit t-SNE die beste optische Clusterung liefern*“ reduziert werden.

Die Anzahl der zur Verfügung stehenden Features macht eine qualitative, systematische Evaluierung aller theoretisch verwendbaren Feature-Kombinationen unmöglich. Getestet wurden Feature-Kombinationen von Features der Timbre Toolbox [38], der MIR Toolbox [35] und des Essentia [43] Frameworks. Um die Vergleichbarkeit der Kombinationen zu gewährleisten wurde für alle drei Extraktoren halbüberlappende Frames mit einer Länge von 20 ms verwendet. Unter diesen drei inhaltsbasierten Merkmalsextraktoren bietet die Timbre Toolbox die geringste Anzahl an Features und umfasst (ohne die verschiedenen statistischen Lagemasse, die frame-basierte Features aggregieren) 32 verschiedene Features. Ausgehend von der Überlegung das Samples mit Feature-Kombinationen beschrieben werden, die 10 bis 32 Features enthalten, sind allein mit der Timbre Toolbox theoretisch 4251885323 verschiedene Feature-Kombinationen möglich. Damit ist eine qualitative Evaluierung aller Feature-Kombinationen eines einzelnen Merkmalsextraktors selbst für eine einzelne Sample-Sammlung aus Zeitgründen nicht möglich. Eine Entwicklerin eines Sample-Browsers muss zwangsläufig eine Heuristik anwenden um jene Feature-Kombinationen zur ermitteln, die das menschliche Ähnlichkeitsempfinden am besten modellieren.

Untersucht wurde wie gut das Klangfarben Ähnlichkeitsempfinden durch verschiedene Feature-Kombinationen modelliert wird. Der Mittelwert von MFCC Koeffizienten ist der MIR Standardansatz für eine derartige Modellierung und diente als Referenz für sämtliche getesteten Feature-Kombinationen. Für diese Feature-Kombination wurden die Samples ebenfalls in 20 ms lange,

halbüberlappende Frames aufgeteilt und pro Frame 13 MFCC Koeffizienten berechnet. Der Mittelwert pro Koeffizient wurde über alle Frames eines Samples berechnet.

#### 4.1.2 Evaluierungsmethodik: Gruppen ähnlicher Samples

Für die Tests wurden mehrere Sample-Sammlungen zusammengestellt die Gruppen ähnlicher Samples beinhalten. Für diese Sammlungen wird im weiteren Text das Kürzel „GäS-Sammlung“ verwendet. Die Samples einer Gruppe wurden so ausgesucht, dass sie innerhalb der Gruppe eine möglichst hohe Klangfarben-Ähnlichkeit aufweisen. Zusätzlich wurde darauf geachtet, dass eine Gruppe von Samples eine möglichst große Unähnlichkeit zu den anderen Gruppen der GäS-Sammlung aufweist. Die Gruppeneinteilung ist für eine menschliche Anwenderin durch das Anhören der Samples unmittelbar erkennbar da die GäS-Gruppen jeweils aus Variationen der selben Klangquelle zusammengesetzt wurden (z.B. besteht eine Gruppe ähnlicher Snaredrums aus mehreren Aufnahmen desselben Snaremodells, die durch Effekt- oder EQ-Einstellungen, Spielarten oder Mikrofonierung eine leicht unterschiedliche Klangfarbe aufweisen). Ein ähnlichkeitsbasierter Sample-Browser sollte GäS-Sammlungen als optische, durch Abstand klar getrennte Cluster darstellen. Mit dieser Evaluierungsmethodik kann qualitativ ermittelt werden

1. welche Feature-Kombinationen die beste optische Clusterung liefern und
2. welche Feature-Kombinationen eine Darstellung der Samples ermöglichen, in der sich das Ähnlichkeitsempfinden von Anwenderinnen mit der algorithmisch errechneten und visualisierten Ähnlichkeit deckt.

Tabelle 3 bietet einen Überblick über die verwendeten GäS – Sammlungen (Vollständige Listen mit Quellenangaben und Dateinamen befinden sich im Anhang dieser Arbeit). Für die Anwendungsbereiche Musikproduktion, Film- und Videovertonung sowie Game-Sound Design wurden mehrere GäS-Sammlungen zusammengestellt, sodass sie dem in der Praxis der Anwenderinnen vorkommenden Sample-Browsing Szenarien möglichst nahekommen. Die folgenden Beispiele dienen zur Verdeutlichung. Für den Bereich Musikproduktion wurde z.B. Sample-Sammlungen mit Gruppen ähnlicher Bassdrums, Sample-Sammlungen mit Gruppen ähnlicher Snaredrums und Sample-Sammlungen mit Gruppen ähnlicher Synthesizer-Samples oder Instrumenten-Samples verwendet. Die einzelnen GäS-Sammlungen bestehen aus 13, 14, 20 oder 25 Gruppen ähnlicher Samples, die einzelnen Gruppen umfassen 4, 5 oder 8 Samples. Insgesamt wurden 10 GäS-Sammlungen zusammengestellt und während des partizipativen Designprozesses verwendet.

Name	Anwendungsfall	Beschreibung der Samples	Sampleanzahl pro Gruppe
BD 1	Bassdrum Auswahl	Einzelne Basstrommelschläge mit einer maximalen Dauer von 8 Sekunden. Geringe Klangfarbenunterschiede innerhalb der einzelnen Gruppen.	Insgesamt 100 Samples, aufgeteilt in 25 Gruppen a 4 Samples
BD 2	Bassdrum Auswahl	Einzelne Basstrommelschläge mit einer maximalen Dauer von 3 Sekunden. Größere Klangfarbenunterschiede innerhalb der einzelnen Gruppen als in BD 1.	Insgesamt 100 Samples, aufgeteilt in 25 Gruppen a 4 Samples
SD 1	Snaredrum Auswahl	Einzelne Snareschläge mit einer maximalen Dauer von 2 Sekunden. Geringe Klangfarbenunterschiede innerhalb der einzelnen Gruppen.	Insgesamt 100 Samples, aufgeteilt in 20 Gruppen a 5 Samples
SD 2	Snaredrum Auswahl	Einzelne Snareschläge mit einer maximalen Dauer von 2 Sekunden. Größere Klangfarbenunterschiede innerhalb der einzelnen Gruppen als in SD 1.	Insgesamt 100 Samples, aufgeteilt in 20 Gruppen a 5 Samples
HH 1	Highhat Auswahl	Einzelne Highhat Schläge mit einer maximalen Dauer von 1 Sekunden	Insgesamt 100 Samples, aufgeteilt in 25 Gruppen a 4 Samples
P 1	Pad Sound Auswahl	Flächige, melodiestützende Hintergrundsounds mit einer maximalen Dauer von 18 Sekunden	Insgesamt 100 Samples, aufgeteilt in 25 Gruppen a 4 Samples
M 1	Auswahl eines melodietragenden Sounds	Mikrofonaufnahmen realer Orchesterinstrumente (2 Streich- und 11 Blasinstrumente) mit einer maximalen Dauer von 5 Sekunden	Insgesamt 104 Samples, aufgeteilt in 13 Gruppen a 8 Samples

M 2	Auswahl eines melodietragenden Sounds	Synthesizer Samples, jeweils eine Oktave eines bestimmten Synthesizers, maximale Dauer von 19 Sekunden	Insgesamt 112 Samples, aufgeteilt in 14 Gruppen a 8 Samples
FSD 1	Filmsound Design	Hintergrundgeräusche mit einer maximalen Dauer von 25 Sekunden	Insgesamt 100 Samples, aufgeteilt in 20 Gruppen a 5 Samples
GSD 1	Gamesound Design	Synthetisierte FX Samples mit einer maximalen Dauer von 8 Sekunden	Insgesamt 100 Samples, aufgeteilt in 20 Gruppen a 5 Samples

Tabelle 3: Gäs - Sammlungen

Um festzustellen welche Feature-Kombination die beste optische Clustering ermöglicht wurde pro Gäs-Sammlung ermittelt wieviel Gruppen der Sammlung korrekt als optische Cluster dargestellt werden. Ein optischer Cluster wurde als korrekt dargestellt bewertet, wenn er in der 2D Karte des MM-Browsers durch einen der beiden Anwender als eine durch Abstand von anderen Clustern getrennte, optische Einheit erkannt wurde und alle, sowie nur Samples der jeweiligen Sample-Gruppe enthält. Abbildung 38 zeigt dieses Evaluierungsprinzip anhand der BD 1 Gäs-Sammlung, die aus 25 Gruppen mit jeweils 4 Snaredrum-Samples besteht. Rechts sind zwei optische Cluster mit A und B markiert. In der Mitte ist erkennbar das Cluster A eine korrekte Darstellung der Sample-Gruppe ist. Alle 4 Samples der Gruppe befinden sich innerhalb des optischen Clusters. Rechts ist erkennbar, das Cluster B zwar eine optische Einheit bildet, aber Samples aus unterschiedlichen Gruppen enthält. Cluster B wird daher nicht als korrekt dargestellt bewertet.

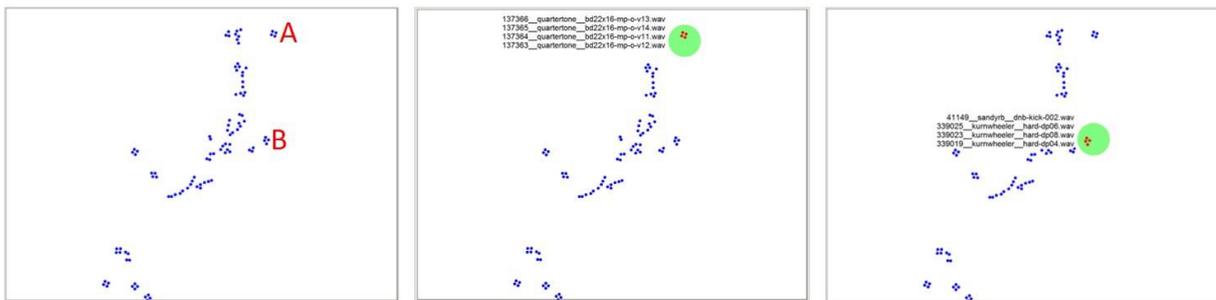


Abbildung 38: Korrekt dargestellt optischer Cluster (A) und nicht korrekt dargestellter optischer Cluster (B) der BD 1 Gäs-Sammlung (25 Gruppen mit jeweils 4 Bassdrum-Samples). Die Karte wurde mit dem Mittelwert von MFCCs und t-SNE erstellt.

Abbildung 39 zeigt das ideale Ergebnis einer DRvis für eine Gäs-Sammlung, die aus 10 Gruppen mit jeweils 5 Snaredrum-Samples besteht. Die Anwenderin kann die 10 Gruppen unmittelbar durch die optischen Cluster erkennen. Die Abbildung ist eine Photoshop-Retusche. Eine derartig gute optische Clustering wurde für keine der Gäs-Sammlungen erreicht obwohl diese so zusammengestellt wurden, dass sie eine optimale optische Clustering ermöglichen (Gleichgroße Gruppen mit minimalen, intragruppen Klangfarben-Unterschieden und maximalen intergruppen Unterschieden). Für einige Gäs-Sammlungen konnte daher eine teilweise korrekte optische Clustering erzielt werden. Dennoch konnte für keine der Gäs-Sammlungen mit einer der getesteten Feature-Kombinationen eine vollständig korrekte Visualisierung erreicht werden in der alle Gruppen einer bestimmten Gäs-Sammlung auch als Cluster abgebildet werden.

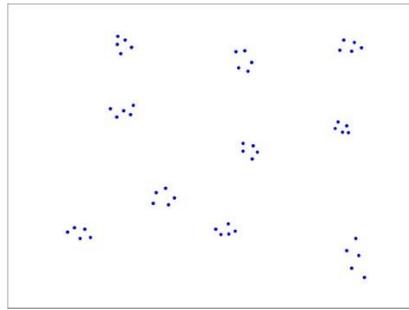


Abbildung 39: Ideales Ergebnis einer DRvis für eine Gäs-Sammlung mit 10 Gruppen à 5 Snaredrum-Samples

#### 4.1.3 Feature-Kombinationen der Timbre Toolbox

Es folgt eine Darstellung der Heuristik, die für die Auswahl der Feature-Kombination der Timbre Toolbox verwendet wurde. Für zwei der verwendeten Feature-Extraktoren wurden mehrere Feature-Kombinationen mit einer möglichst großen Anzahl an Features erstellt. Von allen Features, die eine der beiden Extraktoren zur Verfügung stellt, wurden sämtliche Features verwendet, deren Berechnung für alle Gäs-Sammlungen möglich ist. Im Falle der Timbre Toolbox wurde eine Feature-Kombination erstellt, die sämtliche (-für die Samples aller Gäs-Sammlungen berechenbaren) zeitbasierten und spektralen, frequenzbasierten Features umfasst und deren Features durch unterschiedliche statistische Lagemasse aggregiert wurden. Diese Feature-Kombination (im weiteren Text mit „TT\_max“ bezeichnet, siehe Tabelle 4) wurde durch den Mittelwert, den Mittelwert und den Interquartilsabstand, sowie die beiden vorhergenannten Lagemasse und den Median aggregiert.

Nr.	Feature-Bezeichnung	Dimensionalität
Zeitbasierte Features		
1	Attack	1
2	Decay	1
3	Release	1
4	Log Attack Time	1
5	Attack Slope	1
6	Temporal Centroid	1
7	Effective Duration	1
8	Frequency of Energy Modulation	1
9	Amplitude of Energy Modulation	1
10	Zero Crossing Rate	1
Frequenzbasierte, spektrale Features		
11	Spectral Centroid	1
12	Spectral Spread	1
13	Spectral Skewness	1
14	Spectral Kurtosis	1
15	Spectral Slope	1
16	Spectral Decrease	1
17	Spectral Rolloff	1
18	Spectro-temporal Variation	1
19	Frame Energy	1
20	Spectral Flatness	1
21	Spectral Crest	1

Tabelle 4: Feature-Kombination „TT\_max“ (Für alle Gäs – Sammlungen berechenbare, inhaltsbasierte Features der Timbre Toolbox)

Tabelle 5 zeigt die Ergebnisse für alle Gäs-Sammlungen und die „TT\_max“ Feature-Kombinationen sowie dem Mittelwert von MFCCs. Für 7 der 10 Gäs-Sammlungen lieferte der MFCC Mittelwert eine bessere optische Clusterung als eine der „TT\_max“ Feature-Kombinationen. Für Gäs-Sammlungen mit geräuschhaften Samples (Gäs-Sammlungen BD 1, BD 2, SD 1, HH 1 und FSD 1) konnte mit der MFCC Feature-Kombination eine deutlich höhere Anzahl an korrekt dargestellten Clustern erzielt werden. Keine der „TT\_max“ Feature-Kombinationen zeigte im Vergleich mit dem Mittelwert von MFCCs eine bessere optische Clusterung für Schlagzeug-Samples. Für harmonische Synthesizer-Samples (Gäs-

Sammlung M 2) konnte mit der durch drei Lagemasse aggregierten „TT\_max“ Feature-Kombination eine geringfügig verbesserte optische Clusterung erreicht werden.

Gäs Sammlung	Feature-Kombination	Frame Aggregation	Dimension der Feature-Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	8	7
	TT_max	Mittelwert	21	3	4
	TT_max	Mittelwert, Interquartilsabstand	32	5	5
	TT_max	Mittelwert, Median, Interquartilsabstand	43	5	5
BD 2 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	1	1
	TT_max	Mittelwert	21	0	0
	TT_max	Mittelwert, Interquartilsabstand	32	0	0
	TT_max	Mittelwert, Median, Interquartilsabstand	43	0	0
SD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	4	4
	TT_max	Mittelwert	21	2	2
	TT_max	Mittelwert, Interquartilsabstand	32	2	3
	TT_max	Mittelwert, Median, Interquartilsabstand	43	3	2
SD 2 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	0	0
	TT_max	Mittelwert	21	0	0
	TT_max	Mittelwert, Interquartilsabstand	32	0	0
	TT_max	Mittelwert, Median, Interquartilsabstand	43	0	0
HH 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	2	2
	TT_max	Mittelwert	21	1	1
	TT_max	Mittelwert, Interquartilsabstand	32	1	1
	TT_max	Mittelwert, Median, Interquartilsabstand	43	1	1
P 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	8	7
	TT_max	Mittelwert	21	1	1
	TT_max	Mittelwert, Interquartilsabstand	32	1	1
	TT_max	Mittelwert, Median, Interquartilsabstand	43	1	1
M 1 (13 Gruppen a 8 Samples)	MFCC	Mittelwert	13	1	1
	TT_max	Mittelwert	21	0	0
	TT_max	Mittelwert, Interquartilsabstand	32	0	0
	TT_max	Mittelwert, Median, Interquartilsabstand	43	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC	Mittelwert	13	4	5
	TT_max	Mittelwert	21	3	3
	TT_max	Mittelwert, Interquartilsabstand	32	4	4
	TT_max	Mittelwert, Median, Interquartilsabstand	43	6	5
FSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	6	7
	TT_max	Mittelwert	21	1	1
	TT_max	Mittelwert, Interquartilsabstand	32	3	3
	TT_max	Mittelwert, Median, Interquartilsabstand	43	2	2
GSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	1	0
	TT_max	Mittelwert	21	1	1
	TT_max	Mittelwert, Interquartilsabstand	32	1	1
	TT_max	Mittelwert, Median, Interquartilsabstand	43	2	2

Tabelle 5: Anzahl der korrekt dargestellten optischen Cluster pro Feature-Kombination

Stellvertretend für alle Gäs-Sammlungen mit Schlagzeug-Samples verdeutlicht Abbildung 40 die bessere optische Clusterung durch die MFCC Feature-Kombination anhand der BD 1 Gäs-Sammlung mit 25 Gruppen a 4 Bassdrum-Samples.

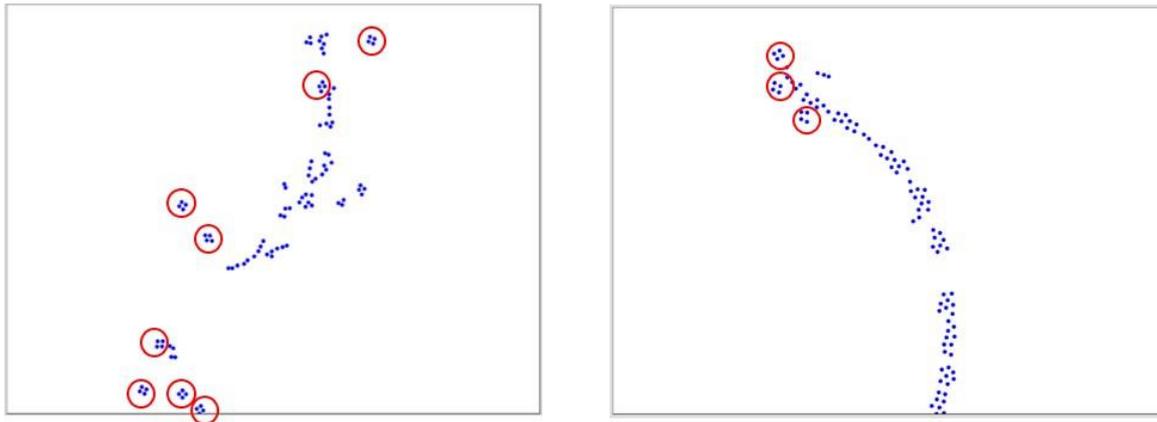


Abbildung 40: Links: 8 korrekte optische Cluster der (mit dem Mittelwert aggregierten) MFCC Feature-Kombination. Rechts: 3 korrekte optische Cluster der (mit dem Mittelwert aggregierten) „TT\_max“ Feature-Kombination.

Die Features der Timbre Toolbox weisen eine hohe Korrelation untereinander auf. (vgl. [38]). Wie viele andere Machine Learning Algorithmen liefert t-SNE für Datensätze deren Features eine geringe Korrelation aufweisen bessere Ergebnisse. (vgl. [33]) Eine Dimensionsreduktion der multivariaten Daten mittels PCA ist daher ein häufiger Verarbeitungsschritt vor dem eigentlichen t-SNE Algorithmus. (Viele t-SNE Implementierungen verwenden den Parameter „inital dimension“ um die Dimensionalität des reduzierten Datensatzes festzulegen.) Die PCA minimiert die Korrelation der Features des reduzierten Datensatzes. Dementsprechend wurde als nächster Schritt die „TT\_max“ Feature Kombinationen mittels PCA auf ein aufgerundetes Drittel ihrer ursprünglichen Dimensionalität reduziert und wiederum mit dem Mittelwert von MFCCs verglichen.

Die Auswertungstabelle für die pca-reduzierten „TT\_max“ Feature Kombinationen befindet sich im Anhang dieser Arbeit (siehe Tabelle „TT\_max\_pca\_vs\_mfcc“). Die pca-reduzierten „TT\_max“ Feature-Kombinationen zeigen im Vergleich mit dem Mittelwert von MFCCs ähnliche Ergebnisse wie die „TT\_max“ Feature-Kombinationen. Für Gäs-Sammlungen mit geräuschhaften Samples (-vor allem Schlagzeug-Samples) konnte mit der MFCC Mittelwert Feature-Kombination eine höhere Anzahl an korrekt dargestellten Clustern erzielt werden. (Nur eine der pca-reduzierten „TT\_max“ Feature-Kombinationen zeigt im Vergleich mit dem Mittelwert von MFCCs eine geringfügig bessere optische Clusterung für die „BD 2“ Gäs-Sammlung) Für die M2 Gäs-Sammlung konnte mit der durch drei Lagemasse aggregierten und pca-reduzierten „TT\_max“ Feature-Kombination eine verbesserte optische Clusterung erreicht werden.

Der Vergleich zwischen den pca-reduzierten TT\_max Feature-Kombinationen und den TT\_max Feature-Kombinationen ist in Tabelle „Vergleich TT\_max vs TT\_max\_pca“ im Anhang dieser Arbeit dargestellt. Von insgesamt 30 Fällen (10 Gäs-Sammlung x 3 unterschiedliche Frame-Aggregationen) bleibt die Anzahl der durch beide Anwender erkannten und korrekten optischen Cluster mit oder ohne pca-Reduzierung in 14 Fällen gleich. In 9 Fällen hat die pca-Reduzierung eine Verbesserung der optischen Clusterung zur Folge. In 7 Fällen führt sie zu einer Verschlechterung der optischen Clusterung.

Das Spektrale Centroid und die Einschwingdauer (log attack time) wurden von der Psychoakustik als zwei wesentliche Bestimmungsgrößen des Klangfarben-Ähnlichkeitsempfinden identifiziert. Dementsprechend wurden mehrere Feature-Kombination getestet, die diese beiden Features enthalten. Von diesen Feature-Kombinationen stellt die Kombination aus der log attack time und dem Mittelwert, dem Median und dem Interquartilsabstand des Spektralen Centroids jene Kombination mit der geringsten Dimensionalität dar (Diese Feature Kombination wird im weiteren Text mit „TT\_lat\_sc“ bezeichnet) Im Vergleich mit dem MFCC Mittelwert zeigen die „TT\_lat\_sc“ Feature-Kombinationen für

8 Gäs-Sammlungen eine wesentlich schlechtere optische Clusterung. Für die BD 2 und die SD 2 Gäs\_Sammlungen ergab sich die selbe Anzahl an korrekten Clustern. Die Auswertungstabelle befindet sich im Anhang dieser Arbeit (siehe Tabelle „TT\_lat\_sc\_vs\_mfcc“).

In [38] zeigen die Entwickler der Timbre Toolbox, dass sich die gesamten Features der Timbre Toolbox in 10 Gruppen einteilen lassen, deren einzelne Features zu den Features der anderen Gruppen eine geringe Korrelation aufweisen und liefern ein „set of guidelines for selecting among the numerous descriptors“ [38]. Die Autoren empfehlen jeweils ein Feature aus einer Gruppe auszuwählen um Feature-Vektoren zusammenzustellen, deren einzelne Features eine geringe Korrelation aufweisen. Die Log attack time und das spektrale Centroid gehören unterschiedlichen Gruppen mit geringer Korrelation an und wurden in jeder Feature-Kombination mit geringer Feature-Korrelation verwendet. Die anderen 8 Features der 10-dimensionalen Feature-Kombinationen mit geringer Feature-Korrelation stammen jeweils aus anderen Gruppen. Insgesamt wurden 4 derartige Feature-Kombinationen getestet. Für diese Kombinationen werden im weiteren Text die Bezeichnungen „TT\_independent\_1“, „TT\_independent\_2“, „TT\_independent\_3“ und „TT\_independent\_4“ verwendet. Tabelle 6 listet die Features dieser vier Kombinationen auf. Die Auswertungstabelle (siehe Tabelle „TT\_independent\_vs\_mfcc“) befindet sich im Anhang dieser Arbeit. Im Vergleich mit dem Mittelwert von MFCCs zeigen die „TT\_independent\_1-4“ Feature-Kombination eine wesentlich schlechtere optische Clusterung für 9 der 10 Gäs-Sammlungen. Für die BD 2 Gäs-Sammlung wurde mit der „TT\_independent\_1“ Feature-Kombination ein korrekter Cluster mehr als mit der MFCC Mittelwert Feature-Kombination erzielt.

	TT_independent_1	TT_independent_2	TT_independent_3	TT_independent_4
1	Spectral Centroid (Med)	Spectral Centroid (Med)	Spectral Centroid (Med)	Spectral Centroid (Med)
2	Inharmonicity (Iqa)	Inharmonicity (Med)	Noisiness (Iqa)	OddEvenRatio (Iqa)
3	Spectral Crest (Med)	Spectral Flatness (Med)	Spectral Crest (Med)	Spectral Flatness (Med)
4	Amplitude Modulation	Spectral Crest (Iqa)	Amplitude Modulation	Spectral Crest (Iqa)
5	Log Attack Time	Log Attack Time	Log Attack Time	Log Attack Time
6	Spectral Centroid (Iqa)	Spectral Slope (Iqa)	Spectral Spread (Iqa)	Spectral Skewness (Iqa)
7	Spectral Flatness (Iqa)	Spectral Flatness (Iqa)	Spectral Flatness (Iqa)	Spectral Flatness (Iqa)
8	Spectral Variation (Iqa)	Spectral Variation (Med)	Spectral Variation (Iqa)	Spectral Variation (Med)
9	Frame Energy (Iqa)	RMS Envelope (Iqa)	Frame Energy (Iqa)	RMS Envelope (Iqa)
10	Frequency Modulation	Frequency Modulation	Frequency Modulation	Frequency Modulation

Tabelle 6: Feature-Kombinationen der Timbre Toolbox mit geringer Feature-Korrelation. Die Kürzel „Med“ (=Median) und „Iqa“ (=Interquartilsabstand) bezeichnen bei framebasierten Features die jeweilige Frame-Aggregation.

#### 4.1.4 Feature-Kombinationen der MIR Toolbox

Es folgt eine Darstellung der Heuristik für die Auswahl der Feature-Kombination der MIR Toolbox [35]. Wiederum wurde eine Feature-Kombination getestet, die alle (-für sämtliche Gäs-Sammlungen berechenbaren) Features der MIR Toolbox enthält. Für diese Feature-Kombination wird im weiteren Text die Bezeichnung „MIR\_max“ verwendet (siehe Tabelle 7). Die „MIR\_max“ Feature-Kombinationen beinhalten MFCCs und 7 (von insgesamt 17) low level Features des MPEG 7 Standards und wurde durch den Mittelwert sowie den Mittelwert und die Standardabweichung aller Frames aggregiert. Zusätzlich wurden die unterschiedlich aggregierten „MIR\_max“ Feature-Kombinationen wie bereits beschrieben mit einer PCA auf ein aufgerundetes Drittel ihrer ursprünglichen Dimensionalität reduziert (Bezeichnung „MIR\_max\_pca“).

Nr.	Feature-Bezeichnung	Dimensionalität
Zeitbasierte Features		
1	Root – Mean – Square Energy	1
2	Zero Crossing Rate	1
3	Low Energy	1
Frequenzbasierte, spektrale Features		
4	Spectral Centroid	1
5	Spectral Brightness	1
6	Spectral Spread	1
7	Spectral Skewnes	1

8	Spectral Kurtosis	1
9	Spectral Rolloff95	1
10	Spectral Rolloff85	1
11	Spectral Entropy	1
12	Spectral Flatness	1
13	Spectral Roughness	1
14	Spectral Irregularity	1
15	Spectral Flux	1
16	MFCC	13
17	Delta MFCC	13
18	Delta delta MFCC	13

Tabelle 7: Feature-Kombination MIR\_max (Für alle Gäs – Sammlungen berechnbare, inhaltsbasierte Features der MIR Toolbox)

Tabelle 8 zeigt die verschiedenen „MIR\_max“ Feature-Kombinationen im Vergleich mit dem Mittelwert von MFCCs. Obwohl die „MIR\_max“ Feature-Kombinationen MFCCs beinhalten verbessern sie die optische Clusterung für nur 4 der 10 Gäs-Sammlungen. Noch deutlicher fällt der Vergleich zwischen dem MFCC Mittelwert und den MIR\_max Varianten, deren framebasierte Features nur mit dem Mittelwert aggregiert wurden aus. In diesem Fall liefert der MFCC Mittelwert für 8 der 10 Gäs-Sammlungen eine bessere optische Clusterung.

Gäs Sammlung	Feature Kombination	Frame Aggregation	Dimension der Feature Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	8	8
	MIR_max	Mittelwert	54	4	4
	MIR_max	Mittelwert, Standardabweichung	92	4	4
	MIR_max_pca	Mittelwert	18	5	4
	MIR_max_pca	Mittelwert, Standardabweichung	32	4	4
BD 2 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	1	1
	MIR_max	Mittelwert	54	1	1
	MIR_max	Mittelwert, Standardabweichung	92	2	2
	MIR_max_pca	Mittelwert	18	2	2
	MIR_max_pca	Mittelwert, Standardabweichung	32	3	3
SD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	4	4
	MIR_max	Mittelwert	54	3	3
	MIR_max	Mittelwert, Standardabweichung	92	6	6
	MIR_max_pca	Mittelwert	18	3	3
	MIR_max_pca	Mittelwert, Standardabweichung	32	6	6
SD 2 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	1	0
	MIR_max	Mittelwert	54	1	1
	MIR_max	Mittelwert, Standardabweichung	92	3	3
	MIR_max_pca	Mittelwert	18	1	1
	MIR_max_pca	Mittelwert, Standardabweichung	32	2	2
HH 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	4	4
	MIR_max	Mittelwert	54	2	2
	MIR_max	Mittelwert, Standardabweichung	92	2	2
	MIR_max_pca	Mittelwert	18	2	2
	MIR_max_pca	Mittelwert, Standardabweichung	32	3	3
P 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	7	8
	MIR_max	Mittelwert	54	3	3
	MIR_max	Mittelwert, Standardabweichung	92	7	7
	MIR_max_pca	Mittelwert	18	3	3
	MIR_max_pca	Mittelwert, Standardabweichung	32	9	9
M 1	MFCC	Mittelwert	13	1	1

(13 Gruppen a 8 Samples)	MIR_max	Mittelwert	54	0	0
	MIR_max	Mittelwert, Standardabweichung	92	0	0
	MIR_max_pca	Mittelwert	18	0	0
	MIR_max_pca	Mittelwert, Standardabweichung	32	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC	Mittelwert	13	5	6
	MIR_max	Mittelwert	54	2	2
	MIR_max	Mittelwert, Standardabweichung	92	5	5
	MIR_max_pca	Mittelwert	18	2	2
FSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	7	7
	MIR_max	Mittelwert	54	2	2
	MIR_max	Mittelwert, Standardabweichung	92	4	4
	MIR_max_pca	Mittelwert	18	2	2
GSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	2	3
	MIR_max	Mittelwert	54	1	1
	MIR_max	Mittelwert, Standardabweichung	92	2	2
	MIR_max_pca	Mittelwert	18	1	1
	MIR_max_pca	Mittelwert, Standardabweichung	32	2	2

Tabelle 8: Feature-Kombination „MIR\_max“ im Vergleich mit dem MFCC Mittelwert

Der Vergleich zwischen den pca-reduzierten „MIR\_max“ Feature-Kombinationen und den „MIR\_max“ Feature-Kombinationen ist in Tabelle „Vergleich MIR\_max vs MIR\_max\_pca“ im Anhang dieser Arbeit dargestellt. Von insgesamt 20 Fällen (10 Gäs-Sammlung x 2 unterschiedliche Frame-Aggregationen) bleibt die Anzahl der durch beide Anwender erkannten und korrekten optischen Cluster mit oder ohne pca-Reduzierung in 15 Fällen gleich. In 5 Fällen hat die pca-Reduzierung eine Verbesserung der optischen Clusterung zur Folge.

Die Ergebnisse für pca-reduzierte Feature-Kombinationen mit maximaler Feature-Anzahl für die Timbre Toolbox und die MIR Toolbox fallen ähnlich aus. In der Mehrzahl der Fälle führt die pca-Reduzierung nicht zu einer Erhöhung der korrekt dargestellten optischen Cluster. Die Standardvorgehensweise bei der Verwendung von t-SNE für eine DRvis (= eine große Anzahl an Features durch eine PCA zu reduzieren) liefert hinsichtlich der korrekten optischen Clusterung der Gäs-Sammlungen für die getesteten Feature-Kombinationen keine wesentliche Verbesserung.

Jene 4 Fälle bei denen eine „MIR\_max“ Feature-Kombination eine bessere optische Clusterung lieferte als der MFCC Mittelwert sind „MIR\_max“ Feature-Kombination deren framebasierte Features mit dem Mittelwert und der Standardabweichung aggregiert wurden. Ob die zusätzliche Frame-Aggregation mit der Standardabweichung oder auf bestimmte Features der „MIR\_max“ Kombinationen zurückzuführen ist konnte mit den Ergebnissen der „MIR\_max“ Tests nicht geklärt werden. Deshalb wurde in einer weiteren Testreihe untersucht ob die delta und DeltaDelta MFCC Features oder die Frame-Aggregation mit zusätzlichen Lagemassen für die Verbesserung der optischen Clusterung ausschlaggebend sind. Tabelle 9 zeigt die Feature-Kombinationen für diese Tests.

Bezeichnung	Features	Frame-Aggregation	Dimensionalität
MFCC_mw	MFCC	Mittelwert	13
MFCC_mw_std	MFCC	Mittelwert, Standardabweichung	26
MFCC_allStats	MFCC	Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	52

MFCC_deltaDelta_mw	MFCC, deltaMFCC, deltaDeltaMFCC	Mittelwert	39
MFCC_deltaDelta_mw_std	MFCC, deltaMFCC, deltaDeltaMFCC	Mittelwert, Standardabweichung	78

Tabelle 9: MFCC Feature-Kombinationen der MIR Toolbox und deren Frame Aggregation

Die Auswertungstabelle für die Feature\_Kombinationen in Tabelle 9 befindet sich im Anhang dieser Arbeit (siehe Tabelle „MIR\_deltaMFCC\_vs\_MFCC“) Für 9 der 10 Gäs-Sammlungen tritt durch die Verwendung zusätzlicher MFCC Lagemasse („MFCC\_mw\_std“ und „MFCC\_allStats“) eine deutliche Verbesserung der optischen Clusterung im Vergleich mit dem MFCC Mittelwert ein. Bei 4 Gäs-Sammlungen lässt sich beobachten, dass die Anzahl der korrekten Cluster mit der Anzahl der verwendeten MFCC Lagemasse ansteigt. Abbildung 41 verdeutlicht diese Ergebnisse stellvertretend anhand der SD 1 Gäs-Sammlung mit 20 Gruppen a 5 Snaredrum-Samples. Die MFCC Mittelwert Feature-Kombination (Rechts) stellt 4 Gruppen korrekt als optischen Cluster dar. Die „MFCC\_mw\_std“ Feature-Kombination (Mitte) stellt 9, die „MFCC\_allStats“ Feature-Kombination (Links) stellt 12 Gruppen korrekt als optischen Cluster dar.

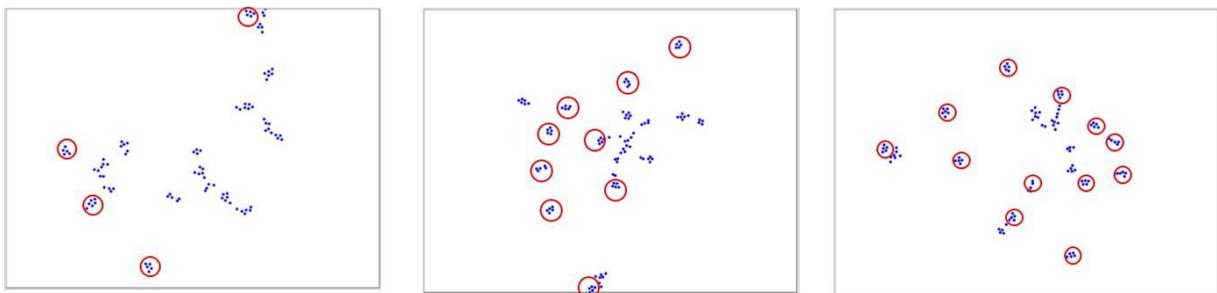


Abbildung 41: Verbesserung der optischen Clusterung durch zusätzliche MFCC Lagemasse

Die delta MFCC und deltaDelta MFCC Features verbessern die optische Clusterung der Gäs-Sammlungen nur geringfügig. Für 6 der 10 Gäs-Sammlungen tritt durch die Verwendung dieses Features keine Verbesserung der optischen Clusterung ein. Für die restlichen 4 Gäs-Sammlungen erhöht die Verwendung des Features die Anzahl der korrekten Cluster nur um einen Cluster (BD 1, HH 1, GSD 1) bzw. um zwei Cluster (P 1).

Weiters wurden mit der MIR Toolbox Feature-Kombination aus MFCC, Log Attack Time und Spektralem Centroid getestet. Das Kalkül für die Wahl dieser drei Features ist die hohe Korrelation mit den Klangfarben-Ähnlichkeitsbeurteilungen von Hörtest-Teilnehmerinnen. Tabelle 10 zeigt diese Feature-Kombinationen.

Bezeichnung	Features	Frame-Aggregation	Dimensionalität
MFCC_allStats	MFCC	Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	52
MFCC_lat_allStats	MFCC, Log Attack Time	Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	53
MFCC_sc_lat_allStats	MFCC, Log Attack Time, Spectral Centroid	Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	57

Tabelle 10: MFCC, Log Attack Time und Spectral Centroid Feature-Kombinationen der MIR Toolbox und deren Frame Aggregation

Die Auswertungstabelle für die Feature-Kombinationen in Tabelle 10 befinden sich im Anhang dieser Arbeit (siehe Tabelle „MIR\_mfcc\_sc\_lat\_allStats“). Im Vergleich mit der „MFCC\_allStats“ Feature-Kombination verbessert die „MFCC\_lat\_allStats“ Feature-Kombination die optische Clusterung für 5 der 10 Gäs-Sammlungen und liefert für die restlichen fünf Gäs-Sammlungen die selbe Anzahl an korrekten, optischen Clustern.

Die „MFCC\_sc\_lat\_allStats“ Feature-Kombinationen hingegen verschlechtert die optische Clusterung für alle Gäs-Sammlungen dramatisch. Die Anzahl der korrekten, optischen Cluster sinkt für jede Gäs-Sammlung unter ein Drittel der optischen Cluster, die mit den beiden anderen Feature-Kombinationen in Tabelle 10 erreicht werden konnte. Ausschlaggebend für diesen Effekt ist die starke Korrelation zwischen MFCC Werten mit niedrigem Index und dem Spektralen Centroid. (vgl. [72]) Beispielhaft verdeutlichen Abbildung 42 und Abbildung 43 dieses Ergebnis anhand der M 2 Gäs- Sammlung, die aus 15 Gruppen a 8 Synthesizer-Samples besteht. Abbildung 43 zeigt, dass die „MFCC\_allStats“ Feature Kombination 7 der 15 Gruppen korrekt als optische Cluster visualisiert (Rechts). Die „MFCC\_sc\_lat\_allStats“ Feature-Kombination (Links) liefert nur zwei optische Cluster.

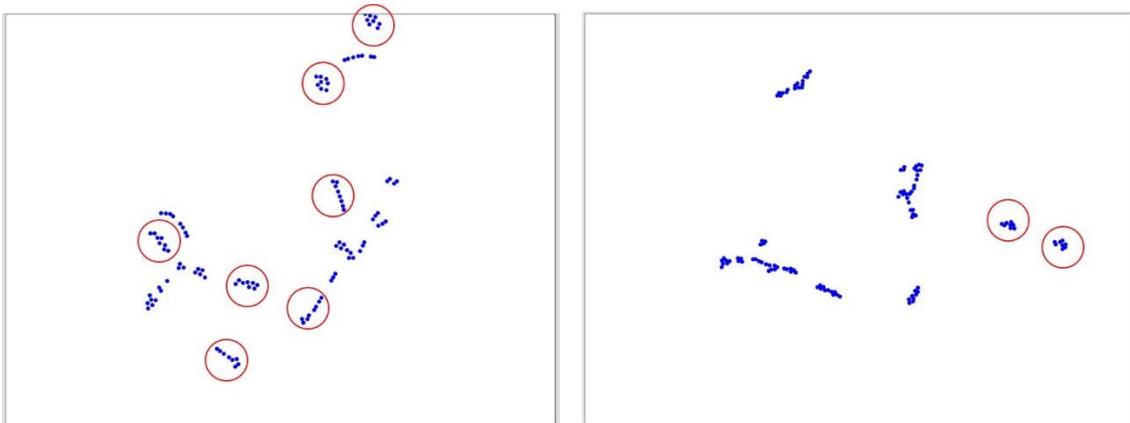
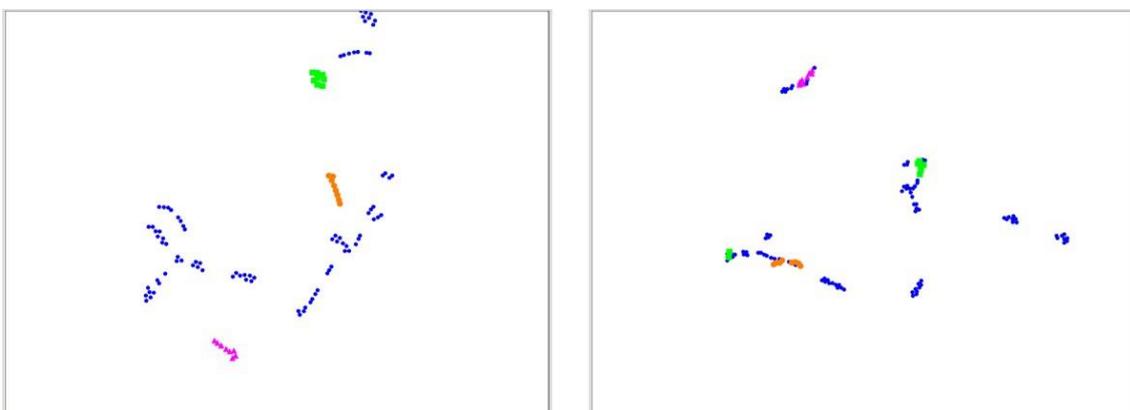


Abbildung 42: Rechts: Die „MFCC\_allStats“ Feature-Kombination liefert 7 optische Cluster. Links: Die „MFCC\_sc\_lat\_allStats“ Feature-Kombination liefert nur zwei optische Cluster.

Abbildung 43 zeigt links 3 optische Cluster (Sample-Gruppe 1 -3) der „MFCC\_allStats“ Feature-Kombination. Die „MFCC\_sc\_lat\_allStats“ Feature-Kombination (Rechts), stellt die drei Cluster nicht korrekt dar.



▲ Sample-Gruppe 1

■ Sample-Gruppe 2

● Sample-Gruppe 2

Abbildung 43: Links: „MFCC\_allStats“, 3 korrekt visualisierte Cluster. Rechts: Darstellung dieser drei Sample-Gruppen durch die „MFCC\_sc\_lat\_allStats“ Feature-Kombination.

Nachdem die „MFCC\_lat\_allStats“ Feature-Kombination für alle Gäs-Sammlungen die beste optische Clusterung lieferte wurde eine Feature-Kombination getestet, die neben (durch vier statistische Lagemasse aggregierte) MFCCs und der Log Attack Time für die Modellierung von Klangfarben häufig verwendete Features umfassen. Tabelle 10 listet die Features dieser Kombination („MIR\_timbre\_allStats“) auf.

Bezeichnung	Features	Frame-Aggregation	Dimensionalität
MIR_timbre_allStats	MFCC, Log Attack Time, Spectral Brightness, Spectral Spread, Spectral Skewness, Spectral Kurtosis, Spectral Rolloff, Spectral Entropy, Spectral Flatness, Zero Crossing Rate, Spectral flux	Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	89

Tabelle 10: „MIR\_mfcc\_timbre“\_allStats Feature-Kombination der MIR Toolbox und deren Frame Aggregation

Die Features der „MIR\_timbre\_allStats“ Kombination sind bereits in den „MIR\_max“ Feature-Kombinationen enthalten. Die „MIR\_max“ Feature-Kombinationen umfassen aber auch das spektrale Centroid, das die optische Clusterung in Verbindung mit MFCCs stark beeinträchtigt. Zusätzlich wurden in den „MIR\_max“ Tests die Klangfarben Features nur durch zwei statistische Lagemasse aggregiert und die „MIR\_max“ Tests deuten darauf hin, dass die Verwendung zusätzlicher Lagemasse die optische Clusterung verbessert. Die Features wurden daher für die „MIR\_timbre\_allStats“ Feature-Kombination durch vier statistische Lagemasse aggregiert.

Die Auswertungstabelle für die „MIR\_timbre\_allStats“ Feature-Kombination befindet sich im Anhang dieser Arbeit (siehe Tabelle „MIR\_timbre\_allstats“). Für 8 der 10 Gäs-Sammlungen liefert die „MIR\_timbre\_allStats“ Feature-Kombination eine deutliche Verschlechterung der optischen Clusterung. Besonders ausgeprägt ist dieser Effekt für Gäs-Sammlungen mit geräuschhaften Samples (BD 1, BD 2, SD 1, SD 2, HH 1 und FSD 1). Für diese Gäs-Sammlungen sinkt die Anzahl der korrekt dargestellten optischen Cluster auf ein Drittel der durch die „MFCC\_lat\_allStats“ korrekt dargestellten Cluster.

#### 4.1.5 Feature-Kombinationen des Essentia Frameworks

Es folgt eine Darstellung der Heuristik für die Auswahl der Feature-Kombination des Essentia [43] Frameworks. Mit dem Essentia Framework wurden Features getestet, die Timbre Toolbox oder die MIR Toolbox nicht beinhalten. Getestet wurden die Features Spectral Contrast und Spectral Valleys, Barkbands, Tristimulus und GFCC, die jeweils mit einer wechselnden Anzahl an statistischen Lagemasse aggregiert wurden. Diese Feature-Kombinationen wurden wiederum mit (durch verschiedene Lagemasse aggregierten) MFCCs verglichen. (Tabelle 11 listet diese Feature-Kombinationen und deren Frame-Aggregation auf). Spectral Contrast und Spectral Valleys wurden getestet da mit diesen Features in [102] bessere Musikgenre-Klassifikationsergebnisse als mit MFCCs erzielt wurden. In [103] zeigen die Autoren das GFCCs im Vergleich mit MFCCs eine bessere Spracherkennung ermöglichen. Das Barkbands Feature misst ähnlich wie MFCCs den Energieanteil in Frequenzbändern, deren Breite und Position der menschlichen, nicht-linearen Tonhöhenwahrnehmung angepasst sind. Das Tristimulus Feature ist eine psychoakustische Modellierung von Klangfarben.

Bezeichnung	Features	Frame-Aggregation	Dimensionalität
ESS_bark_mw	Barkbands	Mittelwert	27
ESS_bark_mw_var	Barkbands	Mittelwert, Varianz	54
ESS_gfcc_mw	GFCC	Mittelwert	13
ESS_gfcc_mw_var	GFCC	Mittelwert, Varianz	26

ESS_mfcc_mw	MFCC	Mittelwert	13
ESS_mfcc_mw_var	MFCC	Mittelwert, Varianz	26
ESS_scsv_mw	Spectral Contrast, Spectral Valleys	Mittelwert	12
ESS_scsv_mw_var	Spectral Contrast, Spectral Valleys	Mittelwert, Varianz	24
ESS_tri_mw	Tristimulus	Mittelwert	3
ESS_tri_mw_var	Tristimulus	Mittelwert, Varianz	6

Tabelle 11: Getestete Feature-Kombinationen des ESSENTIA Frameworks und deren Frame Aggregation

Die Auswertungstabelle für die Feature-Kombinationen in Tabelle 11 befindet sich im Anhang dieser Arbeit (siehe Tabelle „ESS\_feat\_vs\_mfcc“). Unter den in Tabelle 11 aufgelisteten Feature-Kombinationen liefert das Tristimulus Feature die schlechteste optische Clusterung. Die „ESS\_tri\_mw“ Feature-Kombination und die „ESS\_tri\_mw\_var“ Feature-Kombination stellen für 7 der 10 Gäs-Sammlungen keine einzige Sample-Gruppe korrekt als optischen Cluster dar. Für 3 der Gäs-Sammlungen stellen sie jeweils nur eine oder zwei Sample-Gruppen korrekt als optischen Cluster dar. Das Tristimulus-Feature ist daher für ähnlichkeitsbasierte Sample-Browser nicht geeignet.

Das Barkband Feature lieferte im Vergleich mit dem MFCC Mittelwert für 9 von 10 Gäs-Sammlungen eine deutlich niedrigere Anzahl an korrekt dargestellten Clustern. Für diese 9 Gäs-Sammlungen liegt die Anzahl der korrekt dargestellten Sample-Gruppen sowohl für die „ESS\_bark\_mw“ als auch die „ESS\_bark\_mw\_var“ Feature-Kombination unter einem Drittel der korrekt dargestellten Cluster der „ESS\_mfcc\_mw“ Feature-Kombination.

Das Spectral Contrast und Spectral Valleys Feature zeigte im Vergleich mit dem MFCC Mittelwert (unabhängig von der Anzahl der für die Frame-Aggregation verwendeten, statistischen Lagemasse) für 7 der 10 Gäs-Sammlungen eine schlechtere optische Clusterung. Für harmonische Synthesizer-Samples (Gäs-Sammlungen P 1 und M 2) zeigt das Feature aber eine mit dem MFCC Feature vergleichbare optische Clusterung. Die „ESS\_mfcc\_mw“ Feature-Kombination stellt 12 der insgesamt 25 Sample-Gruppen der P 1 Sammlung korrekt als Cluster dar. Die „ESS\_scsv\_mw“ Kombination stellt 9 Sample Gruppen der P1 Sammlung korrekt dar. Für die M 2 Gäs-Sammlung (15 Sample-Gruppen) erzielt die „ESS\_scsv\_mw“ Kombination 9 Cluster, während die „ESS\_mfcc\_mw“ Feature-Kombination nur 7 optische Cluster erzielt. Das Spectral Contrast und Spectral Valleys Feature ist damit genauso wie das MFCC Feature für harmonische Synthesizer-Samples geeignet.

Das GFCC Feature lieferte ähnliche Ergebnisse wie das MFCC Feature. Für 4 der 10 Gäs-Sammlungen (SD 2, P 1, M 1 und M2) lieferten die „ESS\_mfcc\_mw“ und die „ESS\_gfcc\_mw“ Feature-Kombination die gleiche Anzahl an korrekt dargestellten, optischen Clustern. Für 4 Gäs-Sammlungen (BD 2, SD 1, HH 1, FSD 1) konnten mit der „ESS\_gfcc\_mw“ Feature-Kombination eine höhere Anzahl an optischen Clustern erreicht werden. Für 2 Gäs-Sammlungen (BD 1, GSD 1) lieferte die „ESS\_mfcc\_mw“ Feature-Kombination eine höhere Anzahl an Clustern als die „ESS\_gfcc\_mw“ Feature-Kombination. Für 5 der 6 Gäs-Sammlung, bei denen die beiden Feature-Kombinationen eine unterschiedliche Anzahl an Clustern lieferten, gilt aber, dass eine der beiden Feature-Kombinationen nur einen optischen Cluster mehr als die andere Feature-Kombination lieferte. Das GFCC Feature ist damit für die Verwendung in ähnlichkeitsbasierten Sample-Browsern genauso geeignet wie das MFCC Feature.

#### 4.1.6 Auswahl von Schlagzeug-Samples

Eine Musikerin, die in hunderten von Snare-Drum Samples die für ihren Song passende Snare-Drum finden möchte stellt ein typisches Anwendungsszenarium für Sample-Browser dar. Die Anwenderin möchte ein Sample aus hunderten oder tausenden Samples derselben Sample-Kategorie auswählen. Lassen sich auch für dieses Szenario mit der Verarbeitungskette aus Feature Extraktion und Dimensionsreduktion optisch sauber getrennte Cluster erzielen? Für die Beantwortung dieser

Fragestellung wurde die „Total Snare Drums“ Sammlung verwendet, die aus über 10 000 Snaredrum-Samples besteht. Die Sammlung ist thematisch nach Musikgenres in 32 Ordnern aufgeteilt, die jeweils 60 bis 500 Samples beinhalten. Abbildung 44 zeigt die verwendete „Total Snare Drums“ Sammlung. (Download: <https://zero-g.co.uk/products/total-snare-drums-wav-acid-kontakt-exs24-halion-reason-nn-xt-apple-aiff>)

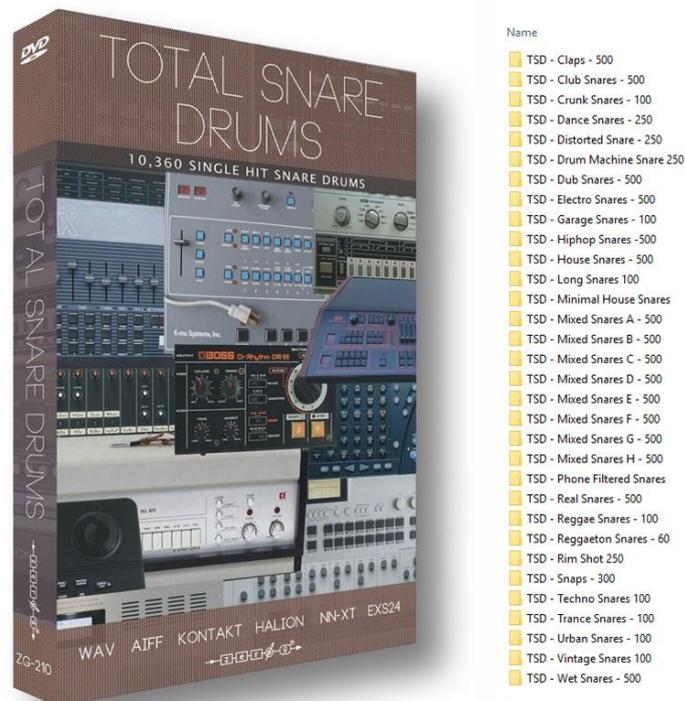


Abbildung 44: „Total Snare Drums Sample“-Sammlung bestehend aus über 10 000 Snaredrum-Samples

Die „Total Snare Drums“ Sammlung besteht aus sogenannten „One Shot“ Samples, aus einzelnen Snare-Drum Schlägen. Alle Samples der Sammlung haben daher den selben charakteristischen, zeitlichen Verlauf, bei dem sich eine Attack- und Releasephase beobachten lassen (siehe Abbildung 45). Die beiden Phasen haben unterschiedliche klangliche und akustische Eigenschaften. Die Attackphase (Stick des Drummers trifft auf das Fell der Snaredrum) ist im Vergleich mit der Releasephase kürzer und durch größere Energiewerte in höheren Frequenzbereichen geprägt. Die Releasephase (Nachschwingen des Fells nach dem Auftreffen des Sticks, der Stick des Drummers berührt das Fell nicht mehr) ist im Vergleich mit der Attackphase länger und durch geringere Energiewerte in höheren Frequenzbereichen geprägt.

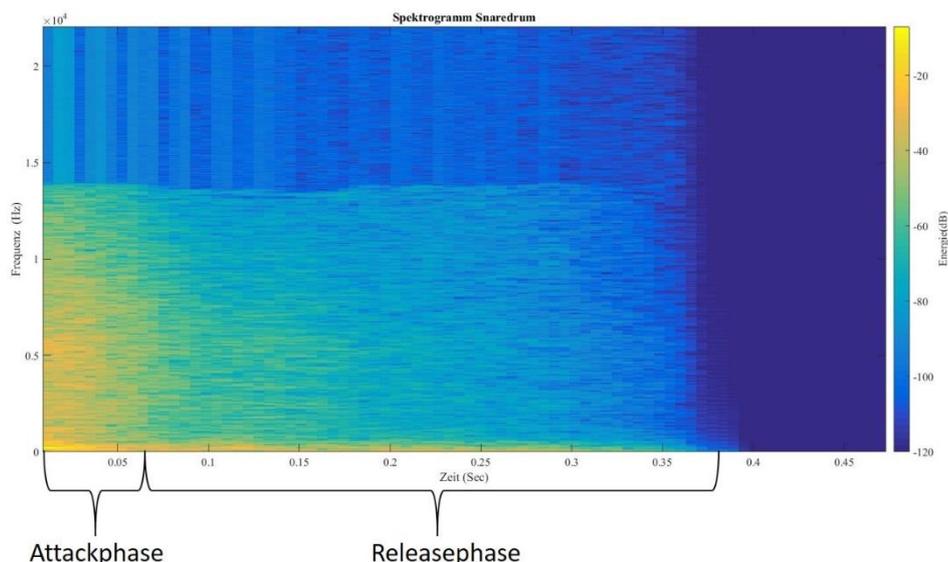


Abbildung 45: Attack- und Releasephase eines Snaredrum-Samples der „Total Snare Drums“ Sammlung

Für die Darstellung der „Total Snare Drums“ Sammlung in der Benutzeroberfläche des MM-Browsers wurden daher 2 Feature-Kombinationen getestet, die auf die Attack/Release Struktur der Samples abzielen. Ähnlich dem für den Snare Drum Navigator (vgl. [21]) verwendeten Ansatz wurden zwei Feature Kombination berechnet, die aus log attack time und getrennt für die Attack- und Releasephase berechneten MFCC Werten bestehen. Die MFCC Werte der Attack- und Releasephase wurden einmal mit dem Mittelwert und für die zweite Feature-Kombination mit dem Mittelwert für die Attackphase sowie dem Mittelwert, der Standardabweichung, der linearen Steigung der Regressionsgeraden und der Shannon Entropy der Autokorrelationsfunktion für die Releasephase aggregiert (siehe Tabelle 12: „MIR\_attRel\_mw“ und „MIR\_attRel\_allStats“).

Getestet wurde auch der MPEG-7 Standard zur Beschreibung von „...percussive sounds (including drum sounds)...“ [104]. Diese Feature-Kombinationen besteht aus der log attack time, dem Temporalen Centroid und dem Spectralen Centroid und wurde mit der Timbre Toolbox berechnet. (siehe Tabelle 12: „TT\_mpeg7\_allStats“ und „TT\_mpeg7\_mw“).

Getestet wurden zusätzlich die Feature-Kombinationen, mit denen sich für die Gäs-Sammlungen mit Schlagzeug Samples die beste optische Clustering erzielen ließen. Diese Feature-Kombinationen sind ebenfalls in Tabelle 12 aufgelistet.

Bezeichnung	Features	Frame-Aggregation	Dimensionalität
ESS_gfcc_mw	GFCC	Mittelwert	13
MIR_attRel_allStats	MFCC, Log Attack Time	Mittelwert für Attackphase, Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion für Releasephase	66
MIR_attRel_mw	MFCC, Log Attack Time	Mittelwert getrennt für Attack- und Releasephase	27
MIR_mfcc_allStats	MFCC	Mittelwert, Standardabweichung, lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	52
MIR_mfcc_lat_allStats	MFCC, Log Attack Time	Mittelwert, Standardabweichung,	53

		lineare Steigung der Regressionsgeraden, Shannon Entropy der Autokorrelationsfunktion	
MIR_mfcc_mw	MFCC	Mittelwert	13
MIR_mfcc_mw_std	MFCC	Mittelwert, Standardabweichung	26
TT_mpeg7_allStats	Log Attack Time, Temporal Centroid, Spectral Centroid	Mittelwert, Median, Interquartilsabstand	7
TT_mpeg7_mw	Log Attack Time, Temporal Centroid, Spectral Centroid	Mittelwert	3

Tabelle 12: Getestete Feature-Kombinationen für die „Total Snare Drums“ Sammlung und deren Frame Aggregation

Die Feature-Kombinationen in Tabelle 12 wurden mit Gäs-Sammlungen mit Schlagzeug-Samples (BD 1, BD 2, SD 1, SD 2, HH 1) getestet (Die Auswertungstabelle für diese Tests befindet sich Anhang dieser Arbeit. Siehe Tabelle „Drums“). Die MPEG-7 Feature-Kombinationen („TT\_mpeg7\_allStats“, „TT\_mpeg7\_mw“) stellen für alle Gäs-Sammlungen mit Schlagzeug-Samples keine oder nur eine Sample-Gruppe korrekt als optischen Cluster dar. Beide MPEG-7 Feature-Kombinationen sind damit für die Darstellung von Schlagzeug-Samples in ähnlichkeitsbasierten Sample-Browsern nicht geeignet.

Bei 4 der 5 Gäs-Sammlungen mit Schlagzeug-Samples lässt sich anhand der Feature-Kombinationen, die ausschließlich durch verschiedene statistische Lagemasse aggregierte MFCCs verwenden („MIR\_mfcc\_allStats“, „MIR\_mfcc\_mw\_std“, „MIR\_mfcc\_mw“) beobachten, dass die Anzahl der korrekt dargestellten optischen Cluster mit der Anzahl der Lagemasse steigt.

Die beste optische Clusterung für Bassdrum-Samples lieferte die „MIR\_mfcc\_lat\_allStats“ Feature-Kombination, die für die BD 1 und BD 2 Sammlung die höchste Anzahl an korrekt dargestellten, optischen Clustern zeigt. Die „MIR\_attRel\_mw“ Feature-Kombination lieferte für die SD 2 und die HH 1 Gäs-Sammmlungen, die höchste Anzahl an korrekt dargestellten Clustern. Beide Feature-Kombinationen beinhalten die Log Attack Time. Damit gilt, dass dieses Feature eine diskriminierende Eigenschaft für die ähnlichkeitsbasierte Darstellung von Drum-Samples besitzt.

Der Ansatz die MFCC Werte für die Attack und-Releasephase getrennt zu berechnen ist für Snaredrum-Samples und Highhat-Samples besonders effektiv. Für die SD 2 und HH 1 Sammlungen lieferte die „MIR\_attRel\_mw“ Feature-Kombination die höchste Anzahl an korrekt dargestellten Clustern. Für 4 der 5 Schlagzeug Gäs-Sammlungen gilt aber das der Ansatz die MFCC Werte über das gesamte Signal zu aggregieren eine gleiche oder höhere Anzahl an optischen Clustern lieferte. Beide Ansätze sind daher gleichermaßen für die Darstellung von Drum-Samples in ähnlichkeitsbasierten Sample-Browsern geeignet.

Die ähnlichkeitsbasierte 2D Karte des Sample-Browsers muss die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind, unterstützen. Alle drei genannten Vorgänge sind wesentlich davon abhängig ob sich mit der Verarbeitungskette aus contentbased Feature Extraction und Dimensionsreduktion eine diffuse Punktwolke oder optisch genau getrennte Cluster erzielen lassen. Abbildung 46 zeigt einen t-SNE Plot aller Samples der „Total Snare Drums“ Sammlung. Erkennbar sind 2 -3 optische Cluster und eine zusammenhängende Punktwolke, welche die drei genannten Vorgänge nicht ausreichend unterstützt.

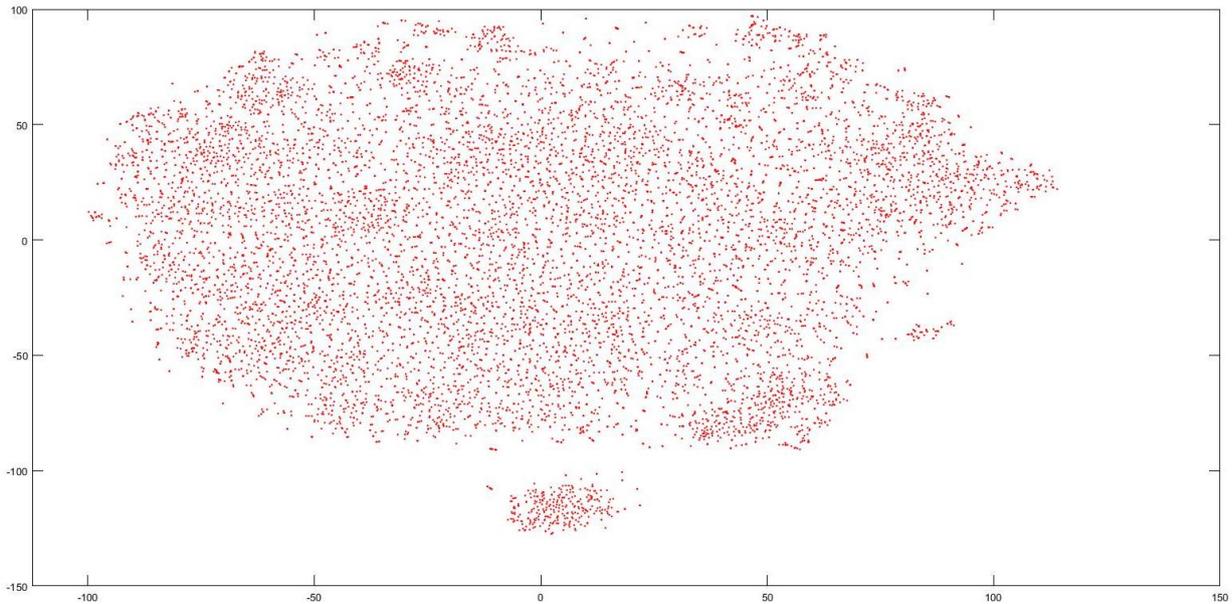


Abbildung 46: t-SNE Plot der MFCC\_allStats Feature-Kombination für die 10 360 Samples der „Total Snare Drums“ Sammlung

Bei der Anwendung der 4 Feature-Kombinationen, mit denen für die Gäs-Sammlungen mit Schlagzeug-Samples die höchste Anzahl an korrekt dargestellten, optischen Clustern erzielt wurden („MIR\_attRel\_allStats“, „MIR\_attRel\_mw“, „MIR\_mfcc\_allStats“ und „MIR\_mfcc\_lat\_allStats“) auf die einzelnen Ordner der „Total Snare Drums“ Sammlung ergaben sich häufig diffuse Punktwolken ohne erkennbare Clustering.

Abbildung 47 zeigt dieses Ergebnis stellvertretend für alle Ordner anhand von zwei Ordnern der „Total Snare Drums“ Sammlung, die 250 bzw. 500 Snaredrum-Samples beinhalten. Die Abbildung zeigt, dass sich für das Anwendungsszenario „Drumsound Selection“ nur diffuse Punktwolken ergeben, welche die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind wesentlich erschweren.

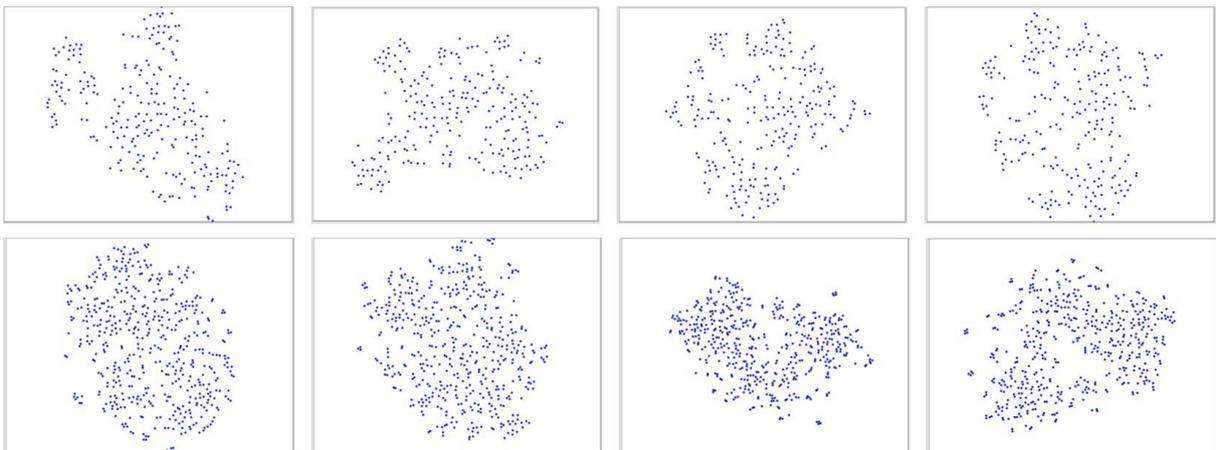


Abbildung 47: t-SNE Plots von zwei Ordnern der „Total Snare Drums“ Sammlung. Pro Zeile von links nach rechts die Plots der „MIR\_attRel\_allStats“, „MIR\_attRel\_mw“, „MIR\_mfcc\_allStats“ und „MIR\_mfcc\_lat\_allStats“ Feature Kombinationen. Erste Zeile: Ordner „TSD - Drum Machine Snare 250“. Zweite Zeile: Ordner „TSD - Real Snare – 500“

#### 4.1.7 Farbkodierung der Sample-Punkte

Das Ergebnis der t-SNE DRvis sind häufig diffuse Punktwolken. In diesen Punktwolken kann eine Anwenderin keine durch Abstand getrennte Gruppen erkennen. Deshalb wurde die Designidee verfolgt die Punktwolken in Farbgruppen aufzuteilen. Durch einen Clustering-Algorithmus wird pro Sample eine Clusterzugehörigkeit berechnet und der jeweilige Sample-Punkt auf der (mittels t-SNE erstellten) 2D-Karte je nach Clusterzugehörigkeit farbkodiert. Durch die Farbkodierung sollte in den Punktwolken erkennbare Farbgruppen entstehen, welche die Ähnlichkeitswahrnehmung, die Navigation und das Memorisieren der Karte besser unterstützen als eine einfarbige, diffuse Punktwolke. Die Anzahl der Cluster in die eine Sample-Sammlung aufgeteilt wird bestimmt die Anwenderin im Userinterface des MM-Browsers.

Um zu ermitteln welcher Clustering Algorithmus das menschliche Klangfarben-Ähnlichkeitsempfinden gut modelliert wurden alle hierarchischen Clustering Algorithmen der Weka [105] Machine Learning Software quantitativ miteinander verglichen. Für diesen Vergleich wurden Feature-Daten aller Samples der 10 Gäs-Sammlungen mit Gäs-Gruppennummern versehen und mit Weka eine „Classes to Cluster“ Evaluierung durchgeführt. Als Feature-Daten für die Clustering wurden die „MFCC\_mw\_std“ und „MFCC\_allStats“ Feature-Kombinationen verwendet. Als Clusteranzahl wurde die Gruppenanzahl der jeweiligen Gäs-Sammlung verwendet. Die Testmethode ermöglicht eine Einschätzung ob die Clustering Algorithmen unter günstigen Umständen für Anwenderinnen nachvollziehbare, sinnvolle Ergebnisse liefern. Stellen die durch die Farbkodierung entstehenden, farblichen Cluster eine Gruppierung dar, die sich mit dem menschlichen Ähnlichkeitsempfinden deckt? Tabelle 13 und 14 listen die Ergebnisse der „Classes to Cluster“ Evaluierung auf. Unter den getesteten Clustering Algorithmen lieferte der Ward Algorithmus die besten Ergebnisse.

	BD 1	BD 2	SD 1	SD 2	HH 1	P 1	M 1	M 2	FSD 1	GSD 1
SINGLE	41	66	61	78	72	72	76	71,42	83	79
COMPLETE	25	34	34	45	40	50	47	50	50	62
AVERAGE	32	41	55	61	45	57	58	64,28	65	67
MEAN	34	41	45	53	34	43	45,19	44,64	54	47
CENTROID	37	55	58	76	69	75	78,84	74,10	82	81
WARD	23	35	16	43	35	36	49,03	19,64	38	39
ADJCOMPLETE	82	75	82	86	85	85	82,69	89,28	86	88
NEIGHBOR_JOINING	96	96	95	95	96	96	92,30	92,85	95	95

Tabelle 13: Prozentsatz der durch die hierarchischen Clustering Algorithmen von Weka falsch geclusterten Samples pro Gäs-Sammlung für die „MFCC\_allStats“ Feature-Kombination. Grüne Felder kennzeichnen den geringsten Prozentsatz pro Gäs-Sammlung

	BD 1	BD 2	SD 1	SD 2	HH 1	P 1	M 1	M 2	FSD 1	GSD 1
SINGLE	47	69	86	83	78	70	83	66,07	81	78
COMPLETE	29	47	56	60	52	46	64	49,10	60	50
AVERAGE	36	50	63	69	70	54	73	59,82	75	58
MEAN	38	48	56	62	57	46	56	45,53	64	55
CENTROID	40	56	70	82	74	68	78	63,39	76	71
WARD	28	43	23	44	48	38	48	20,53	40	44
ADJCOMPLETE	87	83	87	88	87	85	88	89,28	89	83
NEIGHBOR_JOINING	96	96	95	95	96	96	92	92,85	95	95

Tabelle 14: Prozentsatz der durch die hierarchischen Clustering Algorithmen von Weka falsch geclusterten Samples pro Gäs-Sammlung für die „MFCC\_mw\_std“ Feature-Kombination. Grüne Felder kennzeichnen den geringsten Prozentsatz pro Gäs-Sammlung

Das auf der euklidischen Distanz von Feature-Vektoren basierende Ward Clustering hat aber in Kombination mit ähnlichkeitsbasierten t-SNE 2D Karte einen wesentlichen Nachteil. Die durch Distanz visualisierte Ähnlichkeit stimmt in vielen Fällen nicht mit der durch die Farbkodierung visualisierten Ähnlichkeit überein. Abbildung 48 zeigt links das Ergebnis des Ward Clusterings bei dem in Kombination mit der t-SNE DRvis nur wenige, zusammenhängende Farbgruppen entstehen. Deshalb wurde in einem weiteren Designschnitt ein Clustering basierend auf den x, y Koordinaten der Sample-Punkte implementiert (Abbildung 48 rechts). Beide Anwender bestätigen, dass Ihnen mit der auf der

Distanz der Sample-Punkte basierender Farbkodierung besser gelingt in den zusammenhängenden Punktwolken der „Total Snare Drums“ Sammlung Untergruppen zu erkennen. Eine schlechte optische Clusterung (=mittels Distanz) auf der 2D Karte kann also durch eine auf der Distanz der Sample-Punkte beruhende Farbkodierung verbessert werden.

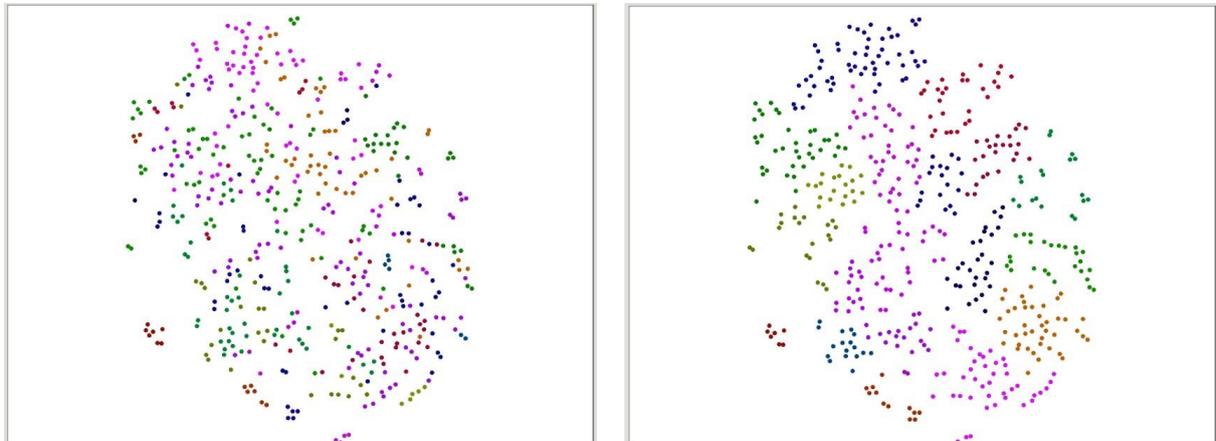


Abbildung 48: Links: Farbkodierung durch Ward Clustering basierend auf den Feature-Daten der „MFCC\_allStats“ Feature-Kombination. Rechts: Farbkodierung durch Ward Clustering basierend auf den x, y Koordinaten der Sample-Punkte

#### 4.1.8 Semantische Struktur von ähnlichkeitsbasierten 2D Karten

Beide Anwender sind in der Lage die Karten der „Total Snare Drums“ Sammlung kognitiv in Bereiche mit unterschiedlichen klanglichen und akustischen Eigenschaften aufzuteilen. Beim Erkunden der Karte sind sie in der Lage die einzelnen Bereiche verbal zu beschreiben: „...oben sind die modernen House und Elektronik Sachen, weiter rechts sind die Live Hall Sachen, links unten sind dann, ich sag jetzt mal „normalen“, trockenen Snares, hier die trockenen Stick Sachen,...“. „...Sub-bassigen sind hier, Kick hier, die haben sehr viel Hall...“ Beide Anwender haben kein Problem damit die semantische Struktur der Karte beim Anhören der Samples zu erkennen.

Beide Anwender haben aber Probleme sich die semantische Struktur der Karte bis zum nächsten Zugriff zu merken. Nachdem sie die semantische Struktur der Karte erlernt haben wissen beide Anwender beim nächsten Zugriff auf die jeweilige Karte meistens nicht mehr in welchen Bereichen der Karte sie Samples mit bestimmten klanglichen und akustischen Eigenschaften finden. Das gilt sowohl für Zugriffe auf eine bestimmte Karte innerhalb derselben Designsession (= Erlernen der semantischen Struktur und erneuter Zugriff innerhalb ca. einer Stunde) als auch für Zugriffe zwischen zwei Designsessions (= Erlernen der semantischen Struktur in einer Designsession und erneuter Zugriff in der nächsten Designsession ca. im Abstand von 2 – 3 Wochen). Eine zielgerichtete Suche nach dem Prinzip „ich suche Samples mit diesen klanglichen oder akustischen Eigenschaften und diese Art von Samples finde ich in der Karte links oben“ gelingt den beiden Anwendern zumeist nicht. Eine erneute Suche in bereits bekannten Karten fühlt sich für beide Anwender eher wie ein zufallsgesteuerter „Try and Error“ Prozess an.

Im Allgemeinen dürften Anwenderinnen, sobald sie ein 2D Punktediagramm vor sich haben, intuitiv davon ausgehen, dass den beiden Achsen eine bestimmte Bedeutung zukommt. Bei der ersten Designsession (-beide Anwender benützten eine ähnlichkeitsbasierte Karte ohne weiter Erläuterung-) gehen beide Anwender automatisch davon aus, dass den beiden Achsen eine bestimmte klangliche oder akustische Eigenschaft zugeordnet ist und die Samples anhand dieser Eigenschaften angeordnet sind. Die Erläuterung das den beiden Achsen keine Bedeutung zukommt nehmen beide Anwender mit Skepsis und leichtem Staunen zur Kenntnis.

Von dieser Vorstellung könne sich beide Anwender im Laufe des gesamten Designprozesses nur schwer lösen. Beide Anwender bringen immer wieder Designvorschläge, bei denen den beiden Achsen eine klare Bedeutung zukommt. Sie thematisieren häufig welche Eigenschaften den beiden Achsen zukommen sollte. Ein Anwender macht mehrmals den Vorschlag den Ecken der 2D Karte bestimmte akustische oder klangliche Eigenschaften zuzuordnen und die Anordnung der Samples an diesen vier Eigenschaften auszurichten. Bei Beendigung des Designprozesses glauben beide Anwender, dass eine genaue Bedeutung der beiden Achsen eine zielgerichtete Suche besser unterstützen würde als eine ähnlichkeitsbasierte 2D Karte.

Der von Grill und Flexer in [29] identifizierte, wesentliche Nachteil von ähnlichkeitsbasierten 2D Karten bleibt für beide Anwender selbst nach einer monatelangen Gewöhnungsphase bestehen. Jede ähnlichkeitsbasierte 2D Karte hat ihre eigene semantische Struktur. Den beiden Achsen ist keine Bedeutung zugeordnet. Damit stehen beide Anwender bei jeder Karte erneut vor der Frage „warum die Samples so und nicht anders angeordnet sind“ „Man fragt sich schon nach welchen Kriterien die Samples dargestellt werden... das Samples, die sich ähnlich sind Gruppen bilden ist schon klar... aber warum ist diese Gruppe von Samples links und nicht da oben?... Weil sie lauter sind, leiser sind, weil sie länger oder kürzer sind?“ Beiden Anwendern fehlt auf den ähnlichkeitsbasierten 2D Karten ein fixes Koordinatensystem, das für alle Sample-Sammlungen gleichbleibt. „Ein bisschen fehlt mir schon zu wissen... wenn ich in diese Richtung fahr mit der Maus... was wird sich damit ändern? Wenn ich das schon wüsste würde ich mir leichter tun.“ „Das sich orientieren würde besser klappen, wenn eine links, rechts oder ABCD Aufteilung (= Ecken der 2D Karte mit bestimmten akustischen oder klanglichen Eigenschaften) hätte. Weil ich schon weiß, na, die da drüben... das wir immer ärger.. immer blehener...das will ich eigentlich nicht“

#### 4.2 Algorithmische Ähnlichkeit und subjektives Ähnlichkeitsempfinden

Zur Beantwortung der Forschungsfrage *ob sich das subjektive Ähnlichkeitsempfinden von Anwenderinnen mit der algorithmisch errechneten Ähnlichkeit deckt* wurden die beiden Anwender aufgefordert die Anordnung der Sample-Punkte auf den 2D Karten aller Gäs-Sammlungen zu analysieren.

Konzeptionell lässt sich eine 100 prozentige Übereinstimmung des subjektiven Ähnlichkeitsempfindens und der errechneten Ähnlichkeit wie folgt definieren. Vor die Aufgabe gestellt die Samples nach Ähnlichkeit auf einer 2D Karte anzuordnen erstellt eine Anwenderin exakt dieselbe Verteilung der Sample-Punkte wie ein ähnlichkeitsbasierter Sample-Browser. Für keine der Gäs-Sammlungen konnte mit einer der getesteten Feature-Kombinationen eine derartige Anordnung erzielt werden. In einem mehrmonatigen Designprozess beantworten die beiden Anwender die Frage ob sich die Anordnung und Abstände der Punkte zu 100 % mit ihrem Ähnlichkeitsempfinden deckt nicht einmal mit „Ja“.

Beide Anwender beurteilen die Anordnungen der Punkte dennoch als „nachvollziehbar“ bzw. als „Da fallen immer ein paar heraus...aber generell machts Sinn“ Sie sehen in den für sie falsch platzierten Sample-Punkten nicht den Hauptgrund für den eingeschränkten praktischen Nutzen von ähnlichkeitsbasierten 2D Karten als Sample-Browsing Werkzeuge. Die für jede Karte andersartige semantische Struktur halten beide Anwender für wesentlich problematischer.

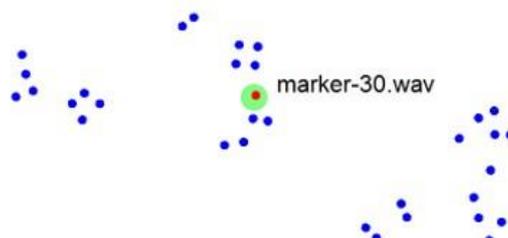
Häufig geäußerte Kritikpunkte der beiden Anwender sind:

- Die Darstellung von Samples aus mehreren, unterschiedlichen Gruppen ähnlicher Samples als ein einziger optischer Cluster
- Die Darstellung von Samples einer Gruppe als einzelne, über die 2D Karte verstreute Sample-Punkte

- Der Abstand mehrerer Sample-Punkte auf der 2D Karte ist ungefähr gleich groß und relativ gering. Beim Anhören dieser Samples wird den Anwendern aber bewusst das es sich nur teilweise um äußerst ähnliche Samples und teilweise doch deutlich unterschiedliche Samples handelt
- Auf der Verbindungsgeraden zwischen zwei Samples aus der selben Gruppe wurde ein Sample aus einer anderen Gruppe plaziert
- Fast alle Samples einer bestimmten Gruppe werden als optischer Cluster dargestellt. 1 bis 2 Samples der Gruppe werden aber ausserhalb des Clusters plaziert
- Alle Samples einer Gruppe werden als Cluster visualisiert. In dem Cluster befindet sich aber zusätzlich noch ein Sample aus einer anderen Gruppe

### 4.3 Sonifikation

Es wurden drei verschiedenen Sonifikations-Konzepte implementiert und mit den Gäs-Sammlungen sowie einigen Sample-Sammlungen der beiden Anwender getestet: Mouse Over, Aura und eine Aura Variante, bei der die Samples innerhalb des Aura-Kreises einzeln, hintereinander abgespielt werden. Die Mouse Over Sonifikation ermöglicht nur ein sequentielles Ansteuern und Abspielen einzelner Samples. Im Vergleich mit der Mouse Over Sonifikation bevorzugen beide Anwender für das Ansteuern und Abspielen einzelner Samples die Aura Sonifikation bei der die Größe des Aura-Kreises etwas größer als ein einzelner Sample-Punkt gewählt wurde (siehe Abbildung 49). Beide Anwender weisen darauf hin, dass Ihnen mit dieser Einstellung das Ansteuern einzelner Sample-Punkt einfacher gelingt als mit der Mouse Over Sonifikation.



*Abbildung 49: Ein Aura-Kreis, dessen Umfang etwas grösser ist als der Umfang der Sample-Punkte eignet sich für das Ansteuern einzelner Samples besser als ein Mouse Cursor*

Trotz der mit steigender Sample-Anzahl rasch sinkenden Erkennungsraten und der beim Abspielen mehrerer Samples auftretenden Maskierungseffekte hat die Aura Sonifikation einen praktischen Nutzen. Beide Anwendern entwickelten unabhängig voneinander folgendes Verwendungsmuster für die Aura Sonifikation. Sie benützen diese Sonifikation um sich einen „akustischen Überblick“ zu verschaffen. Mit einer Aura Größe, die eine geringe Anzahl an Samples gleichzeitig abspielt, gleiten beide Anwender über die Karte um sich klar zu machen welche klanglichen und akustischen Eigenschaften Samples in einer bestimmten Region der Karte aufweisen. Nachdem die Anwender auf diese Art und Weise eine Region mit den gewünschten Eigenschaften ausgemacht haben, verkleinern sie den Umfang des Aura-Kreises auf ein bisschen mehr als der Umfang eines Sample-Punktes. Mit dieser Einstellung hören sich die Anwender die Samples dieser Region einzeln an um eine Auswahl zu treffen.

Die Anzahl der gleichzeitig abgespielten Samples mit der es beiden Anwendern gelingt sich einen „akustischen Überblick“ zu verschaffen (-also für sie wesentliche klangliche und akustische Gemeinsamkeiten der gleichzeitig abgespielten Samples zu erkennen-) ist aufgrund von sinkenden Erkennungsraten und Maskierungseffekten relativ gering und von der Art der Samples abhängig. Für

Schlagzeug- und Percussion-Samples wählen die Anwender eine Aura-Größe, die nicht mehr als 3 oder 4 Samples gleichzeitig abspielt. Je harmonischer und flächiger die Samples der jeweiligen Samplesammlung sind desto mehr Samples spielen die Anwender gleichzeitig ab während sie sich einen „akustischen Überblick“ verschaffen. Für harmonische Keyboard Sounds wählen sie eine Auragröße mit der nicht mehr als 6 Samples gleichzeitig abgespielt werden. Für Samples realer, klassischer Orchesterinstrumente stellen 8- 10 Samples die Obergrenze dar bei der beide Anwender es noch schaffen klangliche und akustische Eigenschaften der Samples zu erkennen. Bei Aura-Größen, die eine höhere Anzahl an Samples abspielen nehmen beide Anwender nur noch eine „Kakophonie“ wahr, in der der gemeinsame Nenner der Samples für sie nicht mehr erkennbar ist.

Dem in [9], [16], [12], [13] und [15] für die Aura Sonifikation beschriebenen Anwendungsszenario (=Auswahl eines Samples direkt in einem Mix aus mehreren Samples mittel selektiver, auditiver Aufmerksamkeit) stehen beide Anwender bei Beendigung des Designprozesses ablehnend gegenüber. Beide Anwender schaffen es auch nach einer längeren Gewöhnungs- und Trainingsphase nicht sich selektiv auf ein einzelnes Sample in einem Mix aus mehreren Samples zu konzentrieren und dieses Sample auf seine Eignung für einen bestimmten Einsatzzweck zu beurteilen. Bei der Auswahl eines Samples für einen bestimmten Einsatzzweck hören sich beide Anwender die in Frage kommenden Samples immer zuerst einzeln an. Zitate der beiden Anwender dazu: „Auch wenn ich mir 4 anhören kann und sagen kann, da ist eins dabei, trotzdem muss ich mir sie dann einzeln anhören“. „Würdest du Dir zutrauen aus diesem Mix (= 6 Synthesizer-Samples) ein Sample auszusuchen? Nein! Aber ich kann feststellen ob das alles Pads (= flächige, harmonische Hintergrund Sounds) sind, oder eine Lead Synth (=Melodietragendes Instrument) dabei ist. Da ist kein Lead Synth dabei, das kannst auch im Vorbeigehen sagen („Vorbeigehen“ meint hier ein mittelschnelles Ziehen des Aura Kreises über eine kleine Anzahl an Samples). Aber dann endet man beim Einzeln anhören, so oder so“

Obwohl das in [9], [16], [12], [13] und [15] für die Aura Sonifikation beschriebene Anwendungsszenario in der Praxis nicht funktioniert ermöglicht die Aura Sonifikation ein beschleunigtes Sample-Browsing. Ausschlaggebend für den Zeitgewinn während des Sample-Browsing ist aber nicht die Auswahl eines Samples direkt in einem Mix aus mehreren Samples mittel selektiver, auditiver Aufmerksamkeit, sondern die Verwendung der Aura Sonifikation um sich einen „akustischen Überblick“ zu verschaffen. Beide Anwender sagen übereinstimmend aus, dass sie durch den „akustischen Überblick“ schneller nicht gewünschte Samples ausschließen können und damit eine größere Anzahl an Samples in der selben Zeit abarbeiten können als mit listenbasierten Interfaces.

Der Vergleich der beiden Aura Sonifikationen fällt zugunsten der Aura Variante aus, bei der die Samples innerhalb des Aura-Kreises einzeln, hintereinander abgespielt werden. Für beide Anwender stellt diese Aura Variante ein alternatives Sonifikations-Konzept mit größerem praktischen Nutzen dar. Abbildung 50 zeigt links die eigentliche Aura Sonifikation und rechts die Aura Variante. Gerade abgespielte Samples werden rot dargestellt.

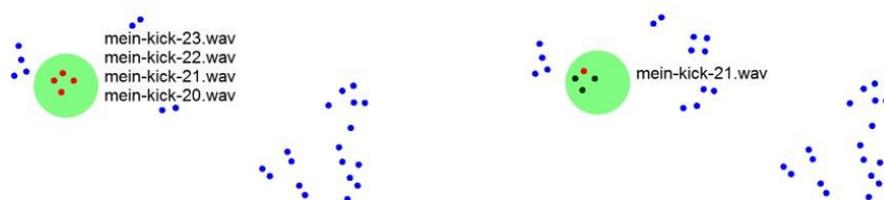


Abbildung 50: Links: ursprüngliche Aura Variante, Rechts: Aura Variante, bei der die Samples innerhalb des Aura-Kreises einzeln, hintereinander abgespielt werden

#### 4.4 Zusammenfassung

In Kapitel 6 wurde gezeigt, dass vor allem Feature-Kombinationen, die statistische Lagemasse von MFCCs und die log attack time enthalten, die beste optische Clusterung für Gäs-Sammlungen mit Schlagzeug-Samples liefern. Die Anwendung dieser Feature-Kombinationen auf eine kommerzielle, professionelle Sample-Sammlung lieferte aber nur diffuse Punktwolken ohne erkennbare optische Cluster. Damit sind für den realen Anwendungsfall drei wesentliche Vorgänge stark beeinträchtigt: die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind.

Um diese drei Vorgänge auch für diffuse Punktwolken zu erleichtern wurden zwei unterschiedliche Farbkodierungen der Sample-Punkte getestet. Eine Clusterung mit dem Ward Algorithmus basierend auf den x, y Koordinaten der Sample-Punkte erleichtert die optische Ähnlichkeitswahrnehmung und die systematische Navigation für Punktwolken ohne erkennbare optische Clusterung mittels Distanz. Die Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind gelingt aber beiden Anwendern auch mit der Farbkodierung nicht.

Die für jede ähnlichkeitsbasierte 2D Karte unterschiedliche semantische Struktur wurde als die wesentliche Schwäche dieses Designkonzepts identifiziert. Beiden Anwendern gelingt es grundsätzlich die semantische Struktur der Karten durch Anhören zu ermitteln. Ein Memorisieren der semantischen Struktur gelingt den beiden Anwendern aber in der Regel nicht. Eine zielgerichtete Suche in Karten, deren semantische Struktur bereits einmal ermittelt wurde, ist daher nicht möglich. Jeder erneute Zugriff auf eine Karte gleicht eher einem zufallsgesteuerten „Try and Error“ Prozess.

Die Aura Sonifikation hat trotz der mit steigender Sample-Anzahl rasch sinkenden Erkennungsraten und der beim Abspielen mehrerer Samples auftretenden Maskierungseffekte einen praktischen Nutzen als „akustische Überblicksfunktion“. Dennoch bevorzugen beide Anwender eine Aura Variante, die Samples einzeln, hintereinander abspielt.

## 5 Zusammenfassung und Ausblick

Für Musiker, Gamesound-Designer und Toningenieure, die sich mit der Vertonung von Filmen und Videos beschäftigen, ist Sample-Browsing ein Bestandteil des täglichen Arbeitsablaufes. Während sich die Methoden und Benutzeroberflächen mit denen die genannten Berufsgruppen ihre Samplesuche durchführen kaum verändert haben, ist der Umfang der zur Verfügung stehenden Samples in den letzten 15 Jahren exponentiell gestiegen. Das kreative Potential derartig großer Sample-Mengen kann weder mit listenbasierten Interfaces noch mit einer Stichwortsuche ausgeschöpft werden. Als Lösung für die „Data Overload“ Problematik im Bereich Sample-Browsing hat die internationale MIR Forschungsgemeinde einige ähnlichkeitsbasierte Sample-Browser entwickelt (siehe Abbildung 51). Nach 15 Jahren universitärer Forschung und einer Vielzahl an vorgestellten Forschungsprototypen bleibt aber der Befund, dass sich ähnlichkeitsbasierter Sample-Browser in der Praxis nicht durchgesetzt haben. Die zentrale Fragestellung dieser Arbeit ist daher die Identifikation jener Faktoren, welche die Usability dieses Ansatzes beeinträchtigen.

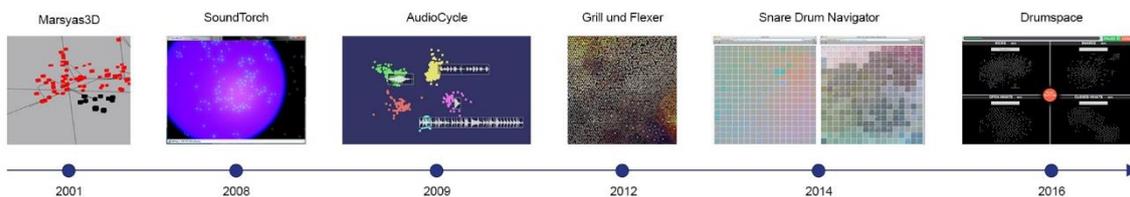


Abbildung 51: Chronologische Darstellung von ähnlichkeitsbasierten Sample-Browsern der letzten 15 Jahre

Ähnlichkeitsbasierte Sample-Browser verwenden die in Abbildung 52 dargestellte algorithmische Verarbeitungskette. Nach einer inhaltsbasierten Merkmalsextraktion wird der n-dimensionale Feature-Raum durch verschiedene DRvis Algorithmen auf einen 2- oder 3-dimensionalen Bildraum reduziert.

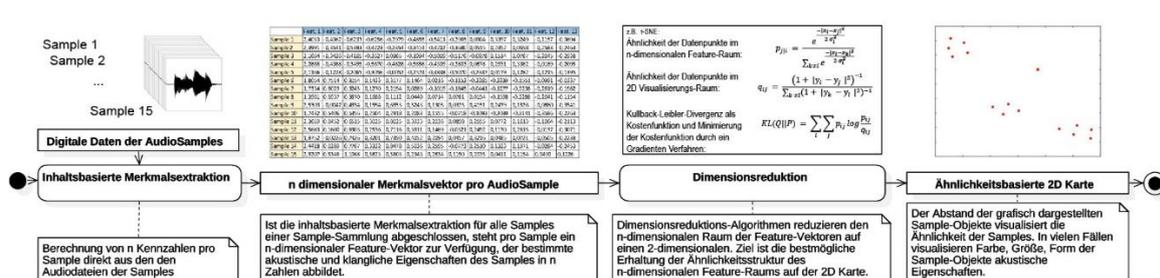


Abbildung 52: Algorithmische Verarbeitungskette von ähnlichkeitsbasierten Sample-Browsern

Sowohl die inhaltsbasierte Merkmalsextraktion als auch die DRvis beeinflussen die Usability von ähnlichkeitsbasierten Sample-Browsern. Die Darstellung der Klangfarbe von Samples durch Feature-Vektoren ist eine stark vereinfachte Modellierung eines komplexen Phänomens, dass von einigen Psychoakustikern als „nicht restlos erforscht“ beurteilt wird. Die Distanz der Feature-Vektoren wird als Maß für die Ähnlichkeit der Samples verwendet. Die Gültigkeit dieses geometrischen Ähnlichkeitsmodells wird von Psychologen stark bezweifelt. MIR Forscher weisen darauf hin, dass mit diesem Ähnlichkeitsmodell nur in begrenztem Umfang sinnvolle Retrieval Resultate erzielt wurden.

Die zu 100% strukturgetreue Abbildung des Feature-Raums auf einen 2- oder 3-dimensionalen Visualisierungs-Raum ist mit den derzeitigen State of the Art DRvis Methoden nicht möglich. Viele

DRvis Forscher gehen davon aus, dass eine derartige Abbildung für nichtlinear verteilte, multivariate Daten generell nicht möglich ist. Selbst wenn das geometrische Ähnlichkeitsmodell eine perfekte Modellierung des Klangfarben-Ähnlichkeitsempfindens darstellen würde, hat die Dimensionsreduktion eine Verzerrung der ursprünglichen Ähnlichkeitsverhältnisse des Feature-Raums zur Folge.

Aufgrund der in den beiden vorhergehenden Absätzen skizzierten Schwächen der algorithmischen Verarbeitungskette ist die Diskrepanz zwischen der algorithmisch ermittelten und durch Abstand visualisierten Ähnlichkeit der Samples und dem subjektiven Ähnlichkeitsempfinden der Anwenderinnen ein Faktor, der die Usability von ähnlichkeitsbasierten Sample-Browsern negativ beeinflusst. In einem mehrmonatigen Designprozess beantworten die Anwender die Frage ob sich die Anordnung und Abstände der Sample-Punkte zu 100 % mit ihrem Ähnlichkeitsempfinden deckt nicht einmal mit „Ja“. Der Normalfall sind umfangreiche Schilderungen wie eine bestimmte Anordnung verändert werden müsste um exakt dem Ähnlichkeitsempfinden der beiden Anwender zu entsprechen.

Die Visualisierungen in der Benutzeroberfläche eines Sample-Browsers müssen die optische Ähnlichkeitswahrnehmung, die systematische Navigation, sowie die Erstellung und Memorisierung eines mentalen Bildes, wo auf der Karte Samples mit bestimmten akustischen und klanglichen Eigenschaften zu finden sind, unterstützen. Alle drei genannten Vorgänge werden durch zusammenhängende Punktwolken nur unzureichend unterstützt. Das typische Anwendungsszenario für Sample-Browser in der Musikproduktion bzw. der Film- und Videovertonung ist die Auswahl aus hunderten oder tausenden, ähnlicher Samples (z.B. die Auswahl einer Snaredrum aus hunderten Snaredrum-Samples). Aufgrund der Gleichartigkeit der Samples weist der Feature-Raum der Samples eine geringe Varianz auf, die wiederum dazu führt, dass auf den 2D Karten von ähnlichkeitsbasierten Sample-Browsern nicht mehrere klar voneinander getrennte Cluster, sondern ein oder zwei zusammenhängende Punktwolken abgebildet werden. Die algorithmische Verarbeitungskette des MM-Browsers lieferte für in der Praxis verwendete Sample-Sammlungen zumeist nur eine einzige, diffuse Punktwolke.

Die für jede ähnlichkeitsbasierte 2D Karte unterschiedliche semantische Struktur wurde als gravierende Schwäche von ähnlichkeitsbasierten Sample-Browsern identifiziert. Anwenderinnen gelingt es grundsätzlich die semantische Struktur der Karten durch Anhören der Samples zu ermitteln. Ein Memorisieren der semantischen Struktur gelingt den Anwenderinnen aber in der Regel nicht. Eine zielgerichtete Suche in Karten, deren semantische Struktur bereits einmal ermittelt wurde, ist daher nicht möglich. Jeder erneute Zugriff auf eine Karte gleicht (-wie bei einem listenbasierten Interface-) eher einem zufallsgesteuerten „Try and Error“ Prozess.

Bei Beendigung des partizipativen Designprozesses betrachten beide Anwender den ähnlichkeitsbasierten MM-Browser skeptisch. Beide Anwender können sich eine ähnlichkeitsbasierte 2D Benutzeroberfläche nur als Alternative zu einem listenbasierten Interface vorstellen. Falls beide Interfacetypen in der Audio-Workstation ihrer Wahl implementiert sind, können sich beide Anwender vorstellen, ein solches Interface gelegentlich zu verwenden. Keiner der beiden Anwender hält den MM-Browser für eine ideale Lösung des Sample-Browsing Problems. Ähnlichkeits-basierte 2D Karten haben im Vergleich mit listenbasierten Interfaces eigene, für sie charakteristische Schwächen und stellen daher wie diese keine ideale Lösung für das Sample-Browsing Problem dar.

Was sind die Designvorschläge der beiden Anwender bei Beendigung des Designprozesses? Anwender P. wünscht sich ein listenbasiertes Interface mit sortierbaren Spalten. Er schlägt vor einzelne Features wie das Spektrale Centroid, die Log Attack Time, Spectral Brightness, Spectral Roughness, Spectral Irregularity, usw. als Sortierkriterium für die einzelnen Spalten zu verwenden. Er fordert weiters die

Entwicklung von Features für bestimmte Eigenschaften einer bestimmten Sample-Kategorie um das Feature dann als Sortierkriterium zu verwenden. Für Drum-Samples wünscht er sich z.B. ein Feature, dass Drum-Samples mit viel Hall („wet“) von Samples mit geringem Hallanteil („dry“) unterscheidet. Die Spalte würde er dann mit „dry/wet“ beschriften.

Anwender T. hält eine Gitteranordnung der Samples mit genau festgelegter Bedeutung der beiden Achsen für zielführender als eine ähnlichkeitsbasierte 2D Karte. Er möchte den beiden Achsen je nach Anwendungsfall eine andere Bedeutung selbst zuordnen können. Mit diesem Konzept möchte vor allem eine für ihn leicht nachvollziehbare und für jede Sample-Sammlung gleichbleibende Navigation ermöglichen.

## 6 Literaturverzeichnis

1. Knees, Peter & Kristina Andersen. 2016. Searching for Audio by Sketching Mental Images of Sound. *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval - ICMR '16*, 95–102. New York, USA: ACM Press. doi:10.1145/2911996.2912021.
2. Frisson, Christian, Stéphane Dupont, Willy Yvart, Nicolas Riche, Xavier Siebert & Thierry Dutoit. 2014. AudioMetro. *Proceedings of the 9th Audio Mostly on A Conference on Interaction With Sound - AM '14*, 1–8. New York, USA: ACM Press. doi:10.1145/2636879.2636880.
3. Brazil, E. & M. Fernstrom. 2003. Audio information browsing with the Sonic Browser. *Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization - CMV 2003*, 26–31. London, UK: IEEE. doi:10.1109/CMV.2003.1215000.
4. Woods, D. D., E. S. Patterson, E. M. Roth & K. Christoffersen. 1999. Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 43*: 174–178. doi:10.1177/154193129904300310.
5. Andersen, Kristina & Florian Grote. 2015. GiantSteps. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, 2:2295–2300. New York, USA: ACM Press. doi:10.1145/2702613.2732868.
6. Turquois, Chloé, Martin Hermant, Daniel Gómez-Marín & Sergi Jordà. 2016. Exploring the Benefits of 2D Visualizations for Drum Samples Retrieval. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval - CHIIR '16*, 329–332. New York, USA: ACM Press. doi:10.1145/2854946.2854996.
7. Schwarz, Diemo, Norbert Schnell & Sebastien Gulluni. 2009. Scalability in Content-Based Navigation of Sound Databases. *Proceedings of the International Computer Music Conference*, 253–258. Montreal, Canada.
8. Cano, P, M Koppenberger, P Herrera, O Celma & V Tarasov. 2004. Sound Effect Taxonomy Management in Production Environments. *Proceedings of 25th International AES Conference*. London, UK.
9. Heise, S, M Hlatky & J Loviscach. 2008. Soundtorch: Quick browsing in large audio collections. *Audio Engineering Society Convention 125*. San Francisco, USA.
10. Gerber, Aurna, Alta Van der Merwe & Ronell Alberts. 2007. Practical Implications of Rapid Development Methodologies. *Proceedings of the 2007 Computer Science and IT Education Conference*, 233–245. Covington, USA.
11. Solomon, Paul. 1995. The think aloud method: A practical guide to modelling cognitive processes. *Information Processing & Management 31*: 906–907. doi:10.1016/0306-4573(95)90031-4.
12. Brazil, Eoin & Mikael Fernstrom. 2001. Sonic browsing: An auditory tool for multimedia asset management. *Proceedings of the 2001 International Conference on Auditory Display*. Espoo, Finland.
13. Tzanetakis, George & Perry Cook. 2001. Marsyas3D: a Prototype Audio Browser-Editor Using a Large Scale Immersive Visual and Audio Display. *Proceedings of the 2001 International Conference on Auditory Display*. Espoo, Finland.
14. Tzanetakis, George & Perry Cook. 2000. MARSYAS: a framework for audio analysis. *Organised Sound 4*. doi:10.1017/S1355771800003071.
15. Brazil, Eoin & Mikael Fernström. 2002. Enhancing sonic browsing using audio information

- retrieval. *Proceedings of the 2002 International Conference on Auditory Display*, 1–6. Kyoto, Japan.
16. Heise, Sebastian, Michael Hlatky & Jörn Loviscach. 2009. Aurally and visually enhanced audio search with soundtorch. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*, 3241. New York, USA: ACM Press. doi:10.1145/1520340.1520465.
  17. Schwarz, Diemo & Norbert Schnell. 2009. Sound Search by Content-Based Navigation in Large Databases. *Proceedings of the SMC 2009 - 6th Sound and Music Computing Conference*, 253–258. Porto, Portugal.
  18. Dupont, Stephane, Thomas Dubuisson, Jerome Urbain, Raphael Sebbe, Nicolas D'Alessandro & Christian Frisson. 2009. AudioCycle: Browsing Musical Loop Libraries. *2009 Seventh International Workshop on Content-Based Multimedia Indexing*, 1:73–80. IEEE. doi:10.1109/CBMI.2009.19.
  19. Dupont, Stephane, Christian Frisson, Xavier Siebert & Damien Tardieu. 2012. Browsing sound and music libraries by similarity. *Audio Engineering Society Convention 128*. London, UK.
  20. Grill, Thomas & Arthur Flexer. 2012. Visualization of perceptual qualities in textural sounds. *International Computer Music Conference*, 589–596. Ljubljana, Slovenien.
  21. Fried, Ohad, Zeyu Jin, Reid Oda & Adam Finkelstein. 2014. AudioQuilt: 2D Arrangements of Audio Samples using Metric Learning and Kernelized Sorting. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 281–286. London, UK.
  22. Tzanetakis, George & Perry Cook. 2000. 3D Graphics Tools For Sound Collections. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, 7–10. Verona, Italien.
  23. Johnson, B. & B. Shneiderman. 1991. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. *Proceeding Visualization '91*, 284–291. IEEE Comput. Soc. Press. doi:10.1109/VISUAL.1991.175815.
  24. Lamping, John, Ramana Rao & Peter Pirolli. 1995. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*, 401–408. New York, USA: ACM Press. doi:10.1145/223904.223956.
  25. Vinet, Hugues, Perfecto Herrera, Francois Pachet, Hugues Vinet, Perfecto Herrera, Francois Pachet & Hugues Vinet. 2002. The CUIDADO Project. *International Conference on Music Information Retrieval*, 197–203. Paris, Frankreich.
  26. Peeters, Geoffroy. 2004. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report 54*: 1–25. doi:10.1234/12345678.
  27. Peeters, Geoffroy G & Emmanuel Deruty. 2008. Automatic morphological description of sounds. *The Journal of the Acoustical Society of America* 123: 3801–3801. doi:10.1121/1.2935492.
  28. Morrison, Alistair & M Chalmers. 2003. Improving hybrid MDS with pivot-based searching. *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*, 85–90. IEEE. doi:10.1109/INFVIS.2003.1249012.
  29. Grill, Thomas, Arthur Flexer & Stuart Cunningham. 2011. Identification of perceptual qualities in textural sounds using the repertory grid method. *Proceedings of the 6th Audio Mostly Conference on A Conference on Interaction with Sound - AM '11*: 67–74.

doi:10.1145/2095667.2095677.

30. Grill, Thomas. 2012. Constructing high-level perceptual audio descriptors for textural sounds. *SMC2012 - 9th Sound and Music Computing Conference 2012*, 235. Kopenhagen, Dänemark.
31. Cunningham, Stuart. 2010. Applying personal construct psychology in sound design using a repertory grid. *Proceedings of the 5th Audio Mostly Conference on A Conference on Interaction with Sound - AM '10*. New York, USA: ACM Press: 1–6. doi:10.1145/1859799.1859807.
32. Mathieu, Benoit, Slim Essid, Thomas Fillon, Jacques Prado & Gaëel Richard. 2010. YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software. *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 441–446. Utrecht, Holland.
33. Van Der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing Data using t-SNE Laurens. *Journal of Machine Learning Research* 9: 187–202.
34. Quadrianto, Novi, Alexander J. Smola, Le Song & Tinne Tuytelaars. 2010. Kernelized sorting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32: 1809–1821. doi:10.1109/TPAMI.2009.184.
35. Lartillot, Olivier, Petri Toivainen & Tuomas Eerola. 2007. A Matlab Toolbox for Music Information Retrieval. *Data Analysis, Machine Learning and Applications*: 261–268. doi:10.1007/978-3-540-78246-9\_31.
36. Xing, Eric P., Andrew Y. Ng, Michael I. Jordan & Stuart Russell. 2003. Distance Metric Learning, With Application To Clustering With Side-Information. *Advances in Neural Information Processing Systems 15*, 505–512. MIT Press.
37. Basalaj, Wojciech. 2000. Proximity visualisation of abstract data. *PhD thesis, University of Cambridge*.
38. Peeters, Geoffroy, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis & Stephen McAdams. 2011. The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America* 130: 2902. doi:10.1121/1.3642604.
39. Knees, Peter & Markus Schedl. 2016. *Music Similarity and Retrieval*. Vol. 36. The Information Retrieval Series. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-662-49722-7.
40. Mitrović, Dalibor, Matthias Zeppelzauer & Christian Breiteneder. 2010. Features for Content-Based Audio Retrieval. *Advances in Computers*, 78:71–150. doi:10.1016/S0065-2458(10)78003-7.
41. Alías, Francesc, Joan Socoró & Xavier Sevillano. 2016. A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Applied Sciences* 6: 143. doi:10.3390/app6050143.
42. Mitrovic, Dalibor, Matthias Zeppelzauer & Horst Eidenberger. 2007. Analysis of the data quality of audio descriptions of environmental sounds. *Journal of Digital Information Management* 5: 48–54.
43. Bogdanov, D, Nicolas Wack, Emilia Gómez, Sankalp Gulati, P Herrera, O Mayor, G Roma, J Salamon, J Zapata & Xavier Serra. 2013. ESSENTIA: an Audio Analysis Library for Music Information Retrieval. *International Society for Music Information Retrieval Conference (ISMIR 2013)*, 493–498. Curitiba, Brasilien.
44. Bullock, Jamie. 2007. LibXtract: A Lightweight Library for Audio Feature Extraction. *Proc.*

*International Computer Music Conference*, 3–6. Göteborg, Schweden.

45. Moffat, David, David Ronan & Joshua D. Reiss. 2015. An Evaluation of Audio Feature Extraction Toolboxes. *Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx-15)*: 1–7. doi:10.13140/RG.2.1.1471.4640.
46. Siedenburg, Kai, Ichiro Fujinaga & Stephen McAdams. 2016. A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology. *Journal of New Music Research* 45. Routledge: 27–41. doi:10.1080/09298215.2015.1132737.
47. Terasawa, Hiroko, Malcolm Slaney & Jonathan Berger. 2005. Perceptual Distance in Timbre Space. *Eleventh Meeting of the International Conference on Auditory Display*, 5:61–68. Limerick, Irland.
48. Lerch, Alexander. 2012. Instantaneous Features. *An Introduction to Audio Content Analysis*, 31–69. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118393550.ch3.
49. Grey, John M. 1977. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America* 61: 1270–1277. doi:10.1121/1.381428.
50. Motte-Haber, Helga de la & Günther Rötter. 2005. *Handbuch der Systematischen Musikwissenschaft*. Laaber-Verlag.
51. McAdams, Stephen, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete & Jochen Krimphoff. 1995. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research* 58: 177–192. doi:10.1007/BF00419633.
52. Krumhansl, Carol L. 1989. Why is musical timbre so hard to understand? *Structure and perception of electroacoustic sound and music*, 9:43–53.
53. Lakatos, Stephen. 2000. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics* 62: 1426–1439. doi:10.3758/BF03212144.
54. Grey, John M & John W. Gordon. 1978. Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America* 63: 1493–1500. doi:10.1121/1.381843.
55. Iverson, P & Carol L. Krumhansl. 1993. Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America* 94: 2595–2603. doi:10.1121/1.407371.
56. Misdariis, Nicolas, Bennett Smith, Daniel Presssmitzer, Patrick Susini & Stephen McAdams. 1998. Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres. *The Journal of the Acoustical Society of America* 103: 3005–3006. doi:10.1121/1.421751.
57. De Poli, Giovanni & Paolo Prandoni. 1997. Sonological models for timbre characterization\*. *Journal of New Music Research* 26: 170–197. doi:10.1080/09298219708570724.
58. Toiviainen, Petri, Mari Tervaniemi, Jukka Louhivuori, Marieke Saher, Minna Huutilainen & Risto Näätänen. 1998. Timbre similarity: Convergence of neural, behavioral, and computational approaches. *Music Perception* 16: 223–241.
59. Saleh Siddiq, Christoph Reuter, Isabella Czedik-Eysenberg, Denis Knauf. 2015. TOWARDS THE COMPARABILITY AND GENERALITY OF TIMBRE SPACE STUDIES. *Proceedings of the Third Vienna Talk on Music Acoustics*, 16–19. Wien, Österreich.
60. Alluri, Vinoo & Petri Toiviainen. 2010. Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre. *Music Perception* 27: 223–242. doi:10.1525/mp.2010.27.3.223.

61. Elliott, Taffeta M, Liberty S Hamilton & Frédéric E Theunissen. 2013. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America* 133: 389–404. doi:10.1121/1.4770244.
62. Richard, Gael, Shiva Sundaram & Shrikanth Narayanan. 2013. An Overview on Perceptually Motivated Audio Indexing and Classification. *Proceedings of the IEEE* 101: 1939–1954. doi:10.1109/JPROC.2013.2251591.
63. Peeters, Geoffroy, Stephen McAdams & Perfecto Herrera. 2000. Instrument Sound Description in the Context of MPEG-7. *Computer* 2000: 1–4.
64. Logan, Beth. 2000. Mel Frequency Cepstral Coefficients for Music Modeling. *International Symposium on Music Information Retrieval*, 28:11p. Massachusetts, USA. doi:10.1.1.11.9216.
65. Eronen, A. 2001. Comparison of features for musical instrument recognition. *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, 19–22. IEEE. doi:10.1109/ASPAA.2001.969532.
66. Herrera-Boyer, Perfecto, Geoffroy Peeters & Shlomo Dubnov. 2003. Automatic Classification of Musical Instrument Sounds. *Journal of New Music Research* 32: 3–21. doi:10.1076/jnmr.32.1.3.16798.
67. Patil, Kailash, Daniel Pressnitzer, Shihab Shamma & Mounya Elhilali. 2012. Music in Our Ears: The Biological Bases of Musical Timbre Perception. *PLoS Computational Biology* 8. doi:10.1371/journal.pcbi.1002759.
68. Johnson, David & George Tzanetakis. 2015. Guitar model recognition from single instrument audio recordings. *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 2015–Novem:370–375. IEEE. doi:10.1109/PACRIM.2015.7334864.
69. Souza, Vinicius M A, Gustavo E A P A Batista & Nilson E. Souza-Filho. 2015. Automatic classification of drum sounds with indefinite pitch. *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015–Septe:1–8. IEEE. doi:10.1109/IJCNN.2015.7280342.
70. Siedenburg, Kai, Kiray Jones-Mollerup & Stephen McAdams. 2016. Acoustic and Categorical Dissimilarity of Musical Timbre: Evidence from Asymmetries Between Acoustic and Chimeric Sounds. *Frontiers in Psychology* 6: 1–17. doi:10.3389/fpsyg.2015.01977.
71. Terasawa, H., M. Slaney & J. Berger. 2005. The thirteen colors of timbre. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, 323–326. IEEE. doi:10.1109/ASPAA.2005.1540234.
72. Terasawa, Hiroko, Jonathan Berger & Shoji Makino. 2012. In search of a perceptual metric for timbre: Dissimilarity judgments among synthetic sounds with MFCC-derived spectral envelopes. *AES: Journal of the Audio Engineering Society* 60: 674–685.
73. Terasawa, Hiroko, Malcolm Slaney & Jonathan Berger. 2006. A statistical model of timbre perception. *ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition (SAPA 2006)*, 18–23.
74. Tversky, Amos. 1977. Features of similarity. *Psychological Review* 84: 327–352. doi:10.1037/0033-295X.84.4.327.
75. Siedenburg, Kai & Stephen McAdams. 2016. The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory* 8211. Taylor & Francis: 1–15. doi:10.1080/09658211.2016.1197945.
76. Town, Stephen M & Jennifer K Bizley. 2013. Neural and behavioral investigations into timbre

- perception. *Frontiers in Systems Neuroscience* 7: 88. doi:10.3389/fnsys.2013.00088.
77. Buschman, Timothy J, Markus Siegel, Jefferson E Roy & Earl K Miller. 2011. Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences* 108: 11252–11255. doi:10.1073/pnas.1104666108.
  78. Campadelli, P, E Casiraghi, C Ceruti & A Rozza. 2015. Intrinsic Dimension Estimation : Relevant Techniques and a Benchmark Framework. *Mathematical Problems in Engineering*.
  79. Gisbrecht, Andrej & Barbara Hammer. 2015. Data visualization by nonlinear dimensionality reduction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5: 51–73. doi:10.1002/widm.1147.
  80. Najim, Safa. 2014. Information visualization by dimensionality reduction: a review. *Journal of Advanced Computer Science & Technology* 3: 101–112. doi:10.14419/jacst.v3i2.2746.
  81. Venna, J., S. Kaski, H. Aidos, K. Nybo & J. Peltonen. 2010. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research* 11: 451–490.
  82. Kaski, Samuel, Janne Nikkilä, Merja Oja, Jarkko Venna, Petri Törönen & Eero Castrén. 2003. Trustworthiness and metrics in visualizing similarity of gene expression. *BMC bioinformatics* 4: 48. doi:10.1186/1471-2105-4-48.
  83. Bunte, Kerstin & John Aldo Lee. 2015. Unsupervised dimensionality reduction : the challenges of big data visualisation. *Esann*: 22–24.
  84. Verleysen, Michel & John A Lee. 2013. Nonlinear Dimensionality Reduction for Visualization. *Neural Information Processing*: 617–622.
  85. Engel, Daniel, Lars Hüttenberger & Bernd Hamann. 2012. A Survey of Dimension Reduction Methods for High-dimensional Data Analysis and Visualization. *Visualization of Large and Unstructured Data Sets: Applications in Geospatial Planning, Modeling and Engineering - Proceedings of IRTG 1131 Workshop 2011* i: 135--149. doi:10.4230/OASlcs.VLUDS.2011.135.
  86. Mayer, Rudolf, Taha Abdel Aziz & Andreas Rauber. 2007. Visualising Class Distribution on Self-organising Maps. *Proceedings of the International Conference on Artificial Neural Networks (ICANN'07)*, 4669:359–368. doi:10.1007/978-3-540-74695-9\_37.
  87. Neumayer, Robert, Rudolf Mayer, Georg Polzlbauer & Andreas Rauber. 2007. The Metro Visualisation of Component Planes for Self-Organising Maps. *2007 International Joint Conference on Neural Networks*, 2788–2793. IEEE. doi:10.1109/IJCNN.2007.4371401.
  88. Pampalk, Elias, Andreas Rauber & Dieter Merkl. 2002. Using Smoothed Data Histograms for Cluster Visualization in Self-Organizing Maps. *International Conference on Artificial Neural Networks (ICANN'02)*, 871–876. doi:10.1007/3-540-46084-5\_141.
  89. Wattenberg, Martin, Fernanda Viégas & Ian Johnson. 2016. How to Use t-SNE Effectively. *Distill* 1: e2. doi:10.23915/distill.00002.
  90. Venna, Jarkko & Samuel Kaski. 2001. Neighborhood Preservation in Nonlinear Projection Methods: An Experimental Study. *Artificial Neural Networks — ICANN 2001*, ed. Georg Dorffner, Horst Bischof, and Kurt Hornik, 2130:485–491. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/3-540-44668-0\_68.
  91. Lee, John A & Michel Verleysen. 2007. *Nonlinear dimensionality reduction*. Springer Science & Business Media.
  92. Lee, John A. & Michel Verleysen. 2009. Quality assessment of dimensionality reduction: Rank-

- based criteria. *Neurocomputing* 72: 1431–1443. doi:10.1016/j.neucom.2008.12.017.
93. Lee, John A. & Michel Verleysen. 2014. Two key properties of dimensionality reduction methods. *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 163–170. IEEE. doi:10.1109/CIDM.2014.7008663.
  94. Venna, Jarkko & Samuel Kaski. 2006. Local multidimensional scaling. *Neural Networks* 19: 889–899. doi:10.1016/j.neunet.2006.05.014.
  95. Dupont, Stephane, Thierry Ravet, Cecile Picard-Limpens & Christian Frisson. 2013. Nonlinear dimensionality reduction approaches applied to music and textural sounds. *2013 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE. doi:10.1109/ICME.2013.6607550.
  96. Ware, Colin. 2008. *Visual Thinking: For Design*. Morgan Kaufmann Publishers Inc.
  97. Cherry, E. Colin. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America* 25: 975. doi:10.1121/1.1907229.
  98. Lorho, Gaëtan, Juha Marila & Jarmo Hiipakka. 2001. Feasibility of Multiple Non-Speech Sounds Presentation Using Headphones. *Proceedings of the 2001 International Conference on Auditory Display*, 32–37. Espoo, Finland.
  99. Stewart, Rebecca. 2010. Spatial Auditory Display for Acoustics and Music Collections. *Doctoral dissertation, School of Electronic Engineering and Computer Science Queen Mary, University of London*.
  100. Brazil, Eoin, Mikael Fernström & John Bowers. 2009. Exploring concurrent auditory icon recognition. *Proceedings of the 15th International Conference on Auditory Display*. Copenhagen, Denmark.
  101. Moore, Brian C. J. 2012. *An introduction to the psychology of hearing*. Brill.
  102. Akkermans, Vincent, Joan Serrà & Perfecto Herrera. 2009. Shape-based spectral contrast descriptor. *Proceedings of the 6th Sound and Music Computing Conference (SMCC'09)*, 143–148. Porto, Portugal.
  103. Shao, Yang, Zhaozhang Jin & Deliang Wang. 2009. AN AUDITORY-BASED FEATURE FOR ROBUST SPEECH RECOGNITION. *Cognitive Science*: 4625–4628.
  104. Pampalk, Elias, Perfecto Herrera & Masataka Goto. 2008. Computational Models of Similarity for Drum Samples. *IEEE Transactions on Audio, Speech, and Language Processing* 16: 408–423. doi:10.1109/TASL.2007.910783.
  105. Frank, Eibe, Mark A Hall & Ian H Witten. 2016. The WEKA Workbench. Online Appendix “Data Mining: Practical Machine Learning Tools and Techniques.” *Morgan Kaufmann, Fourth Edition*.

## 7 Abbildungsverzeichnis

Abbildung 1: kopiert von der Website <http://www.avid.com/pro-tools>

Abbildung 2: kopiert aus [12]

Abbildung 3: kopiert aus [13]

Abbildung 4: kopiert aus [15]

Abbildung 5: kopiert aus [16]

Abbildung 6: kopiert aus [7]

Abbildung 7: kopiert aus [7]

Abbildung 8: kopiert aus [18]

Abbildung 9: kopiert aus [29]

Abbildung 10: kopiert aus [20]

Abbildung 11: kopiert aus [20]

Abbildung 12: kopiert aus [20]

Abbildung 13: kopiert aus [21]

Abbildung 14: kopiert aus [21]

Abbildung 15: kopiert aus [21]

Abbildung 16: kopiert aus [21]

Abbildung 17: kopiert aus [2]

Abbildung 18: kopiert aus [2]

Abbildung 19: kopiert aus [6]

Abbildung 20: kopiert aus [6]

Abbildung 28: kopiert aus [72]

Abbildung 29: kopiert aus [72]

Abbildung 33: kopiert aus [20]

Abbildung 34: kopiert aus [6] und [2]

Abbildung 35: rechte Hälfte des Bildes kopiert aus [2]

Abbildung 36: kopiert aus [18]

Abbildung 45: kopiert von der Website <https://zero-g.co.uk/>

## 8 Anhang

### 8.1 Auswertungstabellen der getesteten Feature-Kombinationen

#### 8.1.1 Tabelle „Drums“

Gäs Sammlung	Feature Kombination	Dimension der Feature Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	ESS_gfcc_mw	13	6	6
	MIR_attRel_allStats	66	8	8
	MIR_attRel_mw	27	7	7
	MIR_mfcc_allStats	52	7	8
	MIR_mfcc_lat_allStats	53	9	8
	MIR_mfcc_mw	13	8	8
	MIR_mfcc_mw_std	26	7	7
	TT_mpeg7_allStats	7	1	0
BD 2 (25 Gruppen a 4 Samples)	ESS_gfcc_mw	13	2	3
	MIR_attRel_allStats	66	3	3
	MIR_attRel_mw	27	3	2
	MIR_mfcc_allStats	52	3	3
	MIR_mfcc_lat_allStats	53	5	5
	MIR_mfcc_mw	13	1	1
	MIR_mfcc_mw_std	26	1	1
	TT_mpeg7_allStats	7	1	1
SD 1 (20 Gruppen a 5 Samples)	ESS_gfcc_mw	13	7	9
	MIR_attRel_allStats	66	10	10
	MIR_attRel_mw	27	7	7
	MIR_mfcc_allStats	52	12	12
	MIR_mfcc_lat_allStats	53	11	11
	MIR_mfcc_mw	13	4	4
	MIR_mfcc_mw_std	26	9	9
	TT_mpeg7_allStats	7	0	0
SD 2 (20 Gruppen a 5 Samples)	ESS_gfcc_mw	13	4	4
	MIR_attRel_allStats	66	4	4
	MIR_attRel_mw	27	7	6
	MIR_mfcc_allStats	52	4	4
	MIR_mfcc_lat_allStats	53	4	4
	MIR_mfcc_mw	13	2	2
	MIR_mfcc_mw_std	26	4	4
	TT_mpeg7_allStats	7	1	1
HH 1 (25 Gruppen a 4 Samples)	ESS_gfcc_mw	13	5	5
	MIR_attRel_allStats	66	7	7
	MIR_attRel_mw	27	6	6
	MIR_mfcc_allStats	52	5	5
	MIR_mfcc_lat_allStats	53	7	7
	MIR_mfcc_mw	13	3	3
	MIR_mfcc_mw_std	26	5	5
	TT_mpeg7_allStats	7	0	0
	TT_mpeg7_mw	3	0	0

### 8.1.2 Tabelle "ESS\_feat\_vs\_mfcc"

Gäs Sammlung	Feature Kombination	Dimension der Feature Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	ESS_bark_mw	27	2	2
	ESS_bark_mw_var	54	2	3
	ESS_gfcc_mw	13	6	5
	ESS_gfcc_mw_var	26	1	1
	ESS_mfcc_mw	13	7	7
	ESS_mfcc_mw_var	26	1	1
	ESS_scsv_mw	12	3	3
	ESS_scsv_mw_var	24	4	4
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0
BD 2 (25 Gruppen a 4 Samples)	ESS_bark_mw	27	0	0
	ESS_bark_mw_var	54	1	1
	ESS_gfcc_mw	13	3	3
	ESS_gfcc_mw_var	26	0	0
	ESS_mfcc_mw	13	2	2
	ESS_mfcc_mw_var	26	2	2
	ESS_scsv_mw	12	0	0
	ESS_scsv_mw_var	24	0	0
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0
SD 1 (20 Gruppen a 5 Samples)	ESS_bark_mw	27	0	0
	ESS_bark_mw_var	54	0	0
	ESS_gfcc_mw	13	10	9
	ESS_gfcc_mw_var	26	1	1
	ESS_mfcc_mw	13	6	6
	ESS_mfcc_mw_var	26	2	2
	ESS_scsv_mw	12	5	5
	ESS_scsv_mw_var	24	5	5
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	1	1
SD 2 (20 Gruppen a 5 Samples)	ESS_bark_mw	27	1	1
	ESS_bark_mw_var	54	0	0
	ESS_gfcc_mw	13	4	4
	ESS_gfcc_mw_var	26	0	0
	ESS_mfcc_mw	13	4	4
	ESS_mfcc_mw_var	26	0	0
	ESS_scsv_mw	12	1	1
	ESS_scsv_mw_var	24	0	0
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0
HH 1 (25 Gruppen a 4 Samples)	ESS_bark_mw	27	3	3
	ESS_bark_mw_var	54	3	3
	ESS_gfcc_mw	13	5	5
	ESS_gfcc_mw_var	26	1	1
	ESS_mfcc_mw	13	3	4
	ESS_mfcc_mw_var	26	1	1
	ESS_scsv_mw	12	4	4
	ESS_scsv_mw_var	24	2	2
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0
P 1 (25 Gruppen a 4 Samples)	ESS_bark_mw	27	4	4
	ESS_bark_mw_var	54	4	4
	ESS_gfcc_mw	13	13	12
	ESS_gfcc_mw_var	26	1	1
	ESS_mfcc_mw	13	12	13
	ESS_mfcc_mw_var	26	1	1
	ESS_scsv_mw	12	9	9
	ESS_scsv_mw_var	24	9	9
	ESS_tri_mw	3	1	1
	ESS_tri_mw_var	6	1	1
M 1	ESS_bark_mw	27	1	1
	ESS_bark_mw_var	54	2	2

(13 Gruppen a 8 Samples)	ESS_gfcc_mw	13	0	0
	ESS_gfcc_mw_var	26	0	0
	ESS_mfcc_mw	13	0	0
	ESS_mfcc_mw_var	26	0	0
	ESS_scsv_mw	12	0	0
	ESS_scsv_mw_var	24	0	0
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0
M 2 (15 Gruppen a 8 Samples)	ESS_bark_mw	27	0	0
	ESS_bark_mw_var	54	0	0
	ESS_gfcc_mw	13	7	7
	ESS_gfcc_mw_var	26	3	3
	ESS_mfcc_mw	13	7	7
	ESS_mfcc_mw_var	26	3	3
	ESS_scsv_mw	12	4	4
	ESS_scsv_mw_var	24	9	9
	ESS_tri_mw	3	1	1
	ESS_tri_mw_var	6	1	1
FSD 1 (20 Gruppen a 5 Samples)	ESS_bark_mw	27	5	5
	ESS_bark_mw_var	54	4	4
	ESS_gfcc_mw	13	9	9
	ESS_gfcc_mw_var	26	0	0
	ESS_mfcc_mw	13	7	8
	ESS_mfcc_mw_var	26	0	0
	ESS_scsv_mw	12	6	6
	ESS_scsv_mw_var	24	2	2
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0
GSD 1 20 Gruppen a 5 Samples)	ESS_bark_mw	27	1	1
	ESS_bark_mw_var	54	1	1
	ESS_gfcc_mw	13	6	6
	ESS_gfcc_mw_var	26	0	0
	ESS_mfcc_mw	13	7	6
	ESS_mfcc_mw_var	26	0	0
	ESS_scsv_mw	12	3	3
	ESS_scsv_mw_var	24	3	3
	ESS_tri_mw	3	0	0
	ESS_tri_mw_var	6	0	0

### 8.1.3 Tabelle „MIR\_deltaMFCC\_vs\_MFCC“

Gäs Sammlung	Feature Kombination	Dimension der Feature Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC_mw	13	7	7
	MFCC_mw_std	26	7	7
	MFCC_allStats	52	8	9
	MFCC_deltaDelta_mw	39	8	8
	MFCC_deltaDelta_mw_std	78	8	8
BD 2 (25 Gruppen a 4 Samples)	MFCC_mw	13	1	1
	MFCC_mw_std	26	1	1
	MFCC_allStats	52	3	4
	MFCC_deltaDelta_mw	39	1	1
	MFCC_deltaDelta_mw_std	78	1	1
SD 1 (20 Gruppen a 5 Samples)	MFCC_mw	13	4	4
	MFCC_mw_std	26	9	9
	MFCC_allStats	52	12	12
	MFCC_deltaDelta_mw	39	4	4
	MFCC_deltaDelta_mw_std	78	9	9
SD 2 (20 Gruppen a 5 Samples)	MFCC_mw	13	2	2
	MFCC_mw_std	26	5	5
	MFCC_allStats	52	5	5
	MFCC_deltaDelta_mw	39	2	2
	MFCC_deltaDelta_mw_std	78	4	4
HH 1 (25 Gruppen a 4 Samples)	MFCC_mw	13	4	4
	MFCC_mw_std	26	6	6
	MFCC_allStats	52	7	6
	MFCC_deltaDelta_mw	39	5	5
	MFCC_deltaDelta_mw_std	78	6	6
P 1 (25 Gruppen a 4 Samples)	MFCC_mw	13	5	5
	MFCC_mw_std	26	5	5
	MFCC_allStats	52	7	7
	MFCC_deltaDelta_mw	39	7	7
	MFCC_deltaDelta_mw_std	78	5	6
M 1 (13 Gruppen a 8 Samples)	MFCC_mw	13	1	1
	MFCC_mw_std	26	1	1
	MFCC_allStats	52	0	0
	MFCC_deltaDelta_mw	39	1	1
	MFCC_deltaDelta_mw_std	78	1	1
M 2 (15 Gruppen a 8 Samples)	MFCC_mw	13	4	4
	MFCC_mw_std	26	7	7
	MFCC_allStats	52	8	7
	MFCC_deltaDelta_mw	39	4	4
	MFCC_deltaDelta_mw_std	78	6	6
FSD 1 (20 Gruppen a 5 Samples)	MFCC_mw	13	7	7
	MFCC_mw_std	26	8	8
	MFCC_allStats	52	10	10
	MFCC_deltaDelta_mw	39	6	6
	MFCC_deltaDelta_mw_std	78	8	8
GSD 1 20 Gruppen a 5 Samples)	MFCC_mw	13	1	1
	MFCC_mw_std	26	2	2
	MFCC_allStats	52	2	2
	MFCC_deltaDelta_mw	39	2	2
	MFCC_deltaDelta_mw_std	78	2	2

#### 8.1.4 Tabelle „MIR\_mfcc\_sc\_lat\_allStats“

Gäs Sammlung	Feature Kombination	Dimension der Feature Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC_allStats	52	9	9
	MFCC_Lat_allStats	53	9	9
	MFCC_Sc_Lat_allStats	57	2	2
BD 2 (25 Gruppen a 4 Samples)	MFCC_allStats	52	4	4
	MFCC_Lat_allStats	53	5	5
	MFCC_Sc_Lat_allStats	57	1	1
SD 1 (20 Gruppen a 5 Samples)	MFCC_allStats	52	11	11
	MFCC_Lat_allStats	53	12	12
	MFCC_Sc_Lat_allStats	57	3	3
SD 2 (20 Gruppen a 5 Samples)	MFCC_allStats	52	5	5
	MFCC_Lat_allStats	53	5	5
	MFCC_Sc_Lat_allStats	57	2	2
HH 1 (25 Gruppen a 4 Samples)	MFCC_allStats	52	6	6
	MFCC_Lat_allStats	53	6	6
	MFCC_Sc_Lat_allStats	57	2	1
P 1 (25 Gruppen a 4 Samples)	MFCC_allStats	52	7	7
	MFCC_Lat_allStats	53	9	9
	MFCC_Sc_Lat_allStats	57	3	3
M 1 (13 Gruppen a 8 Samples)	MFCC_allStats	52	0	0
	MFCC_Lat_allStats	53	1	1
	MFCC_Sc_Lat_allStats	57	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC_allStats	52	7	7
	MFCC_Lat_allStats	53	8	8
	MFCC_Sc_Lat_allStats	57	2	2
FSD 1 (20 Gruppen a 5 Samples)	MFCC_allStats	52	10	9
	MFCC_Lat_allStats	53	11	11
	MFCC_Sc_Lat_allStats	57	0	0
GSD 1 (20 Gruppen a 5 Samples)	MFCC_allStats	52	2	2
	MFCC_Lat_allStats	53	2	2
	MFCC_Sc_Lat_allStats	57	1	1

### 8.1.5 Tabelle „MIR\_timbre\_allstats“

Gäs Sammlung	Feature Kombination	Dimension der Feature Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC_Lat_allStats	53	9	9
	MFCC_Timbre_allStats	89	3	3
BD 2 (25 Gruppen a 4 Samples)	MFCC_Lat_allStats	53	6	6
	MFCC_Timbre_allStats	89	2	1
SD 1 (20 Gruppen a 5 Samples)	MFCC_Lat_allStats	53	11	11
	MFCC_Timbre_allStats	89	4	4
SD 2 (20 Gruppen a 5 Samples)	MFCC_Lat_allStats	53	5	5
	MFCC_Timbre_allStats	89	0	0
HH 1 (25 Gruppen a 4 Samples)	MFCC_Lat_allStats	53	6	6
	MFCC_Timbre_allStats	89	2	1
P 1 (25 Gruppen a 4 Samples)	MFCC_Lat_allStats	53	0	0
	MFCC_Timbre_allStats	89	0	0
M 1 (13 Gruppen a 8 Samples)	MFCC_Lat_allStats	53	0	0
	MFCC_Timbre_allStats	89	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC_Lat_allStats	53	8	8
	MFCC_Timbre_allStats	89	5	5
FSD 1 (20 Gruppen a 5 Samples)	MFCC_Lat_allStats	53	11	11
	MFCC_Timbre_allStats	89	3	3
GSD 1 20 Gruppen a 5 Samples)	MFCC_Lat_allStats	53	2	2
	MFCC_Timbre_allStats	89	1	1

### 8.1.6 Tabelle „TT\_independent\_vs\_mfcc“

Gäs Sammlung	Feature-Kombination	Dimension der Feature-Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC (Mw)	13	7	8
	TT_independent_1	10	3	4
	TT_independent_2	10	2	2
	TT_independent_3	10	4	4
	TT_independent_4	10	1	1
BD 2 (25 Gruppen a 4 Samples)	MFCC (Mw)	13	1	1
	TT_independent_1	10	2	2
	TT_independent_2	10	0	0
	TT_independent_3	10	1	1
	TT_independent_4	10	0	0
SD 1 (20 Gruppen a 5 Samples)	MFCC (Mw)	13	5	4
	TT_independent_1	10	0	0
	TT_independent_2	10	0	0
	TT_independent_3	10	1	1
	TT_independent_4	10	0	0
SD 2 (20 Gruppen a 5 Samples)	MFCC (Mw)	13	1	1
	TT_independent_1	10	0	0
	TT_independent_2	10	0	0
	TT_independent_3	10	0	1
	TT_independent_4	10	0	0
HH 1 (25 Gruppen a 4 Samples)	MFCC (Mw)	13	4	4
	TT_independent_1	10	0	0
	TT_independent_2	10	1	1
	TT_independent_3	10	0	0
	TT_independent_4	10	1	1
P 1 (25 Gruppen a 4 Samples)	MFCC (Mw)	13	8	8
	TT_independent_1	10	2	2
	TT_independent_2	10	0	0
	TT_independent_3	10	0	0
	TT_independent_4	10	0	0
M 1 (13 Gruppen a 8 Samples)	MFCC (Mw)	13	1	1
	TT_independent_1	10	0	0
	TT_independent_2	10	0	0
	TT_independent_3	10	0	0
	TT_independent_4	10	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC (Mw)	13	5	5
	TT_independent_1	10	0	0
	TT_independent_2	10	0	0
	TT_independent_3	10	3	3
	TT_independent_4	10	0	0
FSD 1 (20 Gruppen a 5 Samples)	MFCC (Mw)	13	6	6
	TT_independent_1	10	0	0
	TT_independent_2	10	0	0
	TT_independent_3	10	0	0
	TT_independent_4	10	0	0
GSD 1 (20 Gruppen a 5 Samples)	MFCC (Mw)	13	2	2
	TT_independent_1	10	0	0
	TT_independent_2	10	0	0
	TT_independent_3	10	0	0
	TT_independent_4	10	1	1

### 8.1.7 Tabelle „TT\_lat\_sc\_vs\_mfcc“

Gäs Sammlung	Feature-Kombination	Frame Aggregation	Dimension der Feature-Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	8	8
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	0	0
BD 2 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	1	1
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	1	1
SD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	4	4
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	0	0
SD 2 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	1	1
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	1	1
HH 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	2	3
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	0	0
P 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	5	5
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	2	2
M 1 (13 Gruppen a 8 Samples)	MFCC	Mittelwert	13	1	1
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC	Mittelwert	13	5	4
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	1	1
FSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	7	6
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	1	1
GSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	1	2
	TT_lat_sc	Mittelwert, Median Interquartilsabstand	4	0	0

### 8.1.8 Tabelle „TT\_max\_pca\_vs\_mfcc“

Gäs Sammlung	Feature-Kombination	Frame Aggregation	Dimension der Feature-Kombination	Erkannte optische Cluster Anwender P.	Erkannte optische Cluster Anwender T.
BD 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	7	7
	TT_max_pca	Mittelwert	7	3	3
	TT_max_pca	Mittelwert, Interquartilsabstand	11	4	4
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	4	4
BD 2 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	1	1
	TT_max_pca	Mittelwert	7	2	1
	TT_max_pca	Mittelwert, Interquartilsabstand	11	0	0
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	0	0
SD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	4	4
	TT_max_pca	Mittelwert	7	2	2
	TT_max_pca	Mittelwert, Interquartilsabstand	11	2	2
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	2	3
SD 2 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	1	0
	TT_max_pca	Mittelwert	7	0	0
	TT_max_pca	Mittelwert, Interquartilsabstand	11	0	0
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	0	0
HH 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	3	3
	TT_max_pca	Mittelwert	7	1	1
	TT_max_pca	Mittelwert, Interquartilsabstand	11	1	1
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	2	1
P 1 (25 Gruppen a 4 Samples)	MFCC	Mittelwert	13	1	1
	TT_max_pca	Mittelwert	7	0	0
	TT_max_pca	Mittelwert, Interquartilsabstand	11	0	0
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	0	0
M 1 (13 Gruppen a 8 Samples)	MFCC	Mittelwert	13	1	1
	TT_max_pca	Mittelwert	7	0	0
	TT_max_pca	Mittelwert, Interquartilsabstand	11	0	0
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	0	0
M 2 (15 Gruppen a 8 Samples)	MFCC	Mittelwert	13	4	5
	TT_max_pca	Mittelwert	7	2	2
	TT_max_pca	Mittelwert, Interquartilsabstand	11	6	7
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	7	8
FSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	7	7
	TT_max_pca	Mittelwert	7	0	0
	TT_max_pca	Mittelwert, Interquartilsabstand	11	3	3
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	3	3
GSD 1 (20 Gruppen a 5 Samples)	MFCC	Mittelwert	13	1	2
	TT_max_pca	Mittelwert	7	0	0
	TT_max_pca	Mittelwert, Interquartilsabstand	11	1	1
	TT_max_pca	Mittelwert, Median, Interquartilsabstand	15	1	2

### 8.1.9 Tabelle „Vergleich MIR\_max vs MIR\_max\_pca“

GäS Sammlung	Frame Aggregation	MIR_max ohne pca	MIR_max_pca	Differenz "ohne pca" - "pca-reduziert"
BD 1	Mittelwert	8	9	-1
	Mittelwert, Standartabweichung	8	8	0
BD 2	Mittelwert	2	4	-2
	Mittelwert, Standartabweichung	4	6	-2
SD 1	Mittelwert	6	6	0
	Mittelwert, Standartabweichung	12	12	0
SD 2	Mittelwert	2	2	0
	Mittelwert, Standartabweichung	6	4	2
HH 1	Mittelwert	4	4	0
	Mittelwert, Standartabweichung	4	6	-2
P 1	Mittelwert	6	6	0
	Mittelwert, Standartabweichung	14	18	-4
M 1	Mittelwert	0	0	0
	Mittelwert, Standartabweichung	0	0	0
M 2	Mittelwert	4	4	0
	Mittelwert, Standartabweichung	10	10	0
FSD 1	Mittelwert	4	4	0
	Mittelwert, Standartabweichung	8	8	0
GSD 1	Mittelwert	2	2	0
	Mittelwert, Standartabweichung	4	4	0

### 8.1.10 Tabelle „Vergleich TT\_max vs TT\_max\_pca“

GäS Sammlung	Frame Aggregation	TT_max	TT_max_pca	Differenz "ohne pca" - "pca-reduziert"
BD 1	Mittelwert	7	6	1
	Mittelwert, Interquartilsabstand	10	8	2
	Mittelwert, Median, Interquartilsabstand	10	8	2
BD 2	Mittelwert	0	3	-3
	Mittelwert, Interquartilsabstand	0	0	0
	Mittelwert, Median, Interquartilsabstand	0	0	0
SD 1	Mittelwert	4	4	0
	Mittelwert, Interquartilsabstand	5	4	1
	Mittelwert, Median, Interquartilsabstand	5	5	0
SD 2	Mittelwert	0	0	0
	Mittelwert, Interquartilsabstand	0	0	0
	Mittelwert, Median, Interquartilsabstand	0	0	0
HH 1	Mittelwert	2	2	0
	Mittelwert, Interquartilsabstand	2	2	0
	Mittelwert, Median, Interquartilsabstand	2	3	-1
P 1	Mittelwert	2	5	-3
	Mittelwert, Interquartilsabstand	2	4	-2
	Mittelwert, Median, Interquartilsabstand	2	3	-1
M 1	Mittelwert	0	0	0
	Mittelwert, Interquartilsabstand	0	0	0
	Mittelwert, Median, Interquartilsabstand	0	0	0
M 2	Mittelwert	3	4	-1
	Mittelwert, Interquartilsabstand	8	13	-5
	Mittelwert, Median, Interquartilsabstand	11	15	-4
FSD 1	Mittelwert	2	0	2
	Mittelwert, Interquartilsabstand	6	6	0
	Mittelwert, Median, Interquartilsabstand	4	6	-2
GSD 1	Mittelwert	2	0	2
	Mittelwert, Interquartilsabstand	2	2	0
	Mittelwert, Median, Interquartilsabstand	4	3	1

## 8.2 Gäs -Sammlungen

### 8.2.1 BD 1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="http://freesound.org/people/Samulis/packs/21073/">http://freesound.org/people/Samulis/packs/21073/</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Härte des Anschlags.	375257_samulis_bass-drum-bdrum-f-1 375258_samulis_bass-drum-bdrum-f-2 375259_samulis_bass-drum-bdrum-ff-1 375260_samulis_bass-drum-bdrum-fff-1	wav wav wav wav
2	<a href="http://bedroomproducersblog.com/2013/02/05/500-free-kicks-by-markus-hakala/">http://bedroomproducersblog.com/2013/02/05/500-free-kicks-by-markus-hakala/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch Anpassung der Bassdrum an vier unterschiedliche Tonhöhen	mhak kick 201 G mhak kick 202 F mhak kick 204 F# mhak kick 212 A	wav wav wav wav
3	<a href="http://themetalkickdrum.com/free-stuff/kicks.html">http://themetalkickdrum.com/free-stuff/kicks.html</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	TMKDD4 - dying_kick_01 TMKDD4 - dying_kick_02 TMKDD4 - dying_kick_03 TMKDD4 - dying_kick_04	wav wav wav wav
4	<a href="http://themetalkickdrum.com/free-stuff/kicks.html">http://themetalkickdrum.com/free-stuff/kicks.html</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	TMKDD2 - darkness_kick_01 TMKDD2 - darkness_kick_02 TMKDD2 - darkness_kick_03 TMKDD4 - dying_kick_04	wav wav wav wav
5	<a href="http://themetalkickdrum.com/free-stuff/kicks.html">http://themetalkickdrum.com/free-stuff/kicks.html</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	TMKDD7 - Fate_kick_01 TMKDD7 - Fate_kick_02 TMKDD7 - Fate_kick_03 TMKDD7 - Fate_kick_04	wav wav wav wav
6	<a href="https://freesound.org/people/crispydinner/packs/17071/">https://freesound.org/people/crispydinner/packs/17071/</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen.	276767_crispydinner_crnd-pndlrgrbt-kick-marker-1 276807_crispydinner_crnd-pndlrgrbt-kick-marker-2 276831_crispydinner_crnd-pndlrgrbt-kick-marker-3 276776_crispydinner_crnd-pndlrgrbt-kick-marker-4	wav wav wav wav
7	<a href="https://freesound.org/people/crispydinner/packs/15497/">https://freesound.org/people/crispydinner/packs/15497/</a>	Synthetisierte Bassdrum der TR 909 Drum Machine. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	252731_crispydinner_crnd-upgrkick-011 252730_crispydinner_crnd-upgrkick-012 252732_crispydinner_crnd-upgrkick-013 252733_crispydinner_crnd-upgrkick-014	wav wav wav wav
8	<a href="https://freesound.org/people/crispydinner/packs/8287/">https://freesound.org/people/crispydinner/packs/8287/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen	131799_crispydinner_lyr-kck-mk4-r1-marker-1 131798_crispydinner_lyr-kck-mk4-r1-marker-2 131797_crispydinner_lyr-kck-mk4-r1-marker-3 131796_crispydinner_lyr-kck-mk4-r1-marker-4	wav wav wav wav
9	<a href="http://99sounds.org/kick-drum/">http://99sounds.org/kick-drum/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch Anpassung der Bassdrum an vier unterschiedliche Tonhöhen	sub_kick_01_C sub_kick_28_A sub_kick_31_A# sub_kick_34_B	wav wav wav wav
10	<a href="https://freesound.org/people/sandyrb/packs/2614/">https://freesound.org/people/sandyrb/packs/2614/</a>	Gemischt aus Mikrofonaufnahmen und synthetisierten Bassdrums. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen	41148_sandyrb_dnb-kick-001 41149_sandyrb_dnb-kick-002 41150_sandyrb_dnb-kick-003 41151_sandyrb_dnb-kick-004	wav wav wav wav
11	<a href="https://freesound.org/people/zgump/packs/5467/">https://freesound.org/people/zgump/packs/5467/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Kick Anteile der einzelnen Bassdrums.	83254_zgump_club-kick-0301 83255_zgump_club-kick-0302 83256_zgump_club-kick-0303 83261_zgump_club-kick-0308	wav wav wav wav
12	<a href="https://freesound.org/people/waveplay/packs/12562/">https://freesound.org/people/waveplay/packs/12562/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	237065_waveplay_deadmau5-style-kick-drum-alternate-2 237066_waveplay_deadmau5-style-kick-drum-alternate-3 237067_waveplay_deadmau5-style-kick-drum-alternate-4 237068_waveplay_deadmau5-style-kick-drum-alternate-5	wav wav wav wav
13	<a href="https://freesound.org/people/SBAudio/packs/21201/">https://freesound.org/people/SBAudio/packs/21201/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	377423_sbaudio_kick-01 377424_sbaudio_kick-02 377425_sbaudio_kick-03 377426_sbaudio_kick-04	wav wav wav wav
14	<a href="https://freesound.org/people/zgump/packs/6852/">https://freesound.org/people/zgump/packs/6852/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	107827_zgump_kick-pack-0801 107829_zgump_kick-pack-0802 107831_zgump_kick-pack-0803 107832_zgump_kick-pack-0804	wav wav wav wav
15	<a href="https://freesound.org/people/sandyrb/packs/2471/">https://freesound.org/people/sandyrb/packs/2471/</a>	Gemischt aus Mikrofonaufnahmen und synthetisierten Bassdrums. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen	38901_sandyrb_mdvp-kick-001 38902_sandyrb_mdvp-kick-002 38903_sandyrb_mdvp-kick-003 38904_sandyrb_mdvp-kick-004	wav wav wav wav
16	<a href="https://freesound.org/people/Stomp/packs/1250/">https://freesound.org/people/Stomp/packs/1250/</a>	Synthetisierte Bassdrum des Clavia Nordrack 2. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen	20391_stomp_nr2-bd03-01 20392_stomp_nr2-bd04-03 20393_stomp_nr2-bd05-02 20394_stomp_nr2-bd05-04	aif aif aif aif
17	<a href="https://freesound.org/people/jesterdyne/packs/5303/">https://freesound.org/people/jesterdyne/packs/5303/</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Härte des Anschlags.	81228_jesterdyne_bd1 81229_jesterdyne_bd2 81230_jesterdyne_bd3 81231_jesterdyne_bd4	wav wav wav wav
18	<a href="https://freesound.org/people/freakrush/packs/2735/">https://freesound.org/people/freakrush/packs/2735/</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Härte des Anschlags.	43366_freakrush_bassdrum3 43367_freakrush_bassdrum5 43368_freakrush_bassdrum6 43369_freakrush_bassdrum7	wav wav wav wav
19	<a href="https://freesound.org/people/kurnwheeler/">https://freesound.org/people/kurnwheeler/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen	339019_kurnwheeler_hard-dp04 339025_kurnwheeler_hard-dp06 339024_kurnwheeler_hard-dp07 339023_kurnwheeler_hard-dp08	wav wav wav wav
20	<a href="https://freesound.org/people/k3rn3lp4n1k/packs/8926/">https://freesound.org/people/k3rn3lp4n1k/packs/8926/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	144881_k3rn3lp4n1k_kick05 144880_k3rn3lp4n1k_kick06 144879_k3rn3lp4n1k_kick07 144878_k3rn3lp4n1k_kick08	wav wav wav wav
21	<a href="https://freesound.org/people/altemark/packs/1643/">https://freesound.org/people/altemark/packs/1643/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche EQ Einstellungen.	26486_altemark_bd01 26487_altemark_bd02 26488_altemark_bd03 26489_altemark_bd04	wav wav wav wav
22	<a href="https://freesound.org/people/jobro/packs/4932/">https://freesound.org/people/jobro/packs/4932/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen.	76152_jobro_klubkick1 76155_jobro_klubkick2 76156_jobro_klubkick3 76157_jobro_klubkick4	wav wav wav wav
23	<a href="https://freesound.org/people/loofa/packs/5762/">https://freesound.org/people/loofa/packs/5762/</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Härte des Anschlags.	88720_loofa_kik-005 88721_loofa_kik-006 88722_loofa_kik-007 88723_loofa_kik-008	aif aif aif aif
24	<a href="https://freesound.org/people/quartertone/packs/8586/">https://freesound.org/people/quartertone/packs/8586/</a>	Mikrofonaufnahme der selben Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Härte des Anschlags.	137363_quartertone_bd22x16-mp-o-v12 137364_quartertone_bd22x16-mp-o-v11 137365_quartertone_bd22x16-mp-o-v14 137366_quartertone_bd22x16-mp-o-v13	wav wav wav wav
25	<a href="https://freesound.org/people/Pder/packs/6748/">https://freesound.org/people/Pder/packs/6748/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch unterschiedliche Effekteinstellungen	106290_pder_mein-kick-1 106327_pder_mein-kick-2 106337_pder_mein-kick-3 106338_pder_mein-kick-4	wav wav wav wav

## 8.2.2 BD 2

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="https://freesound.org/people/zgump/packs/5495/">https://freesound.org/people/zgump/packs/5495/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	102173_zgump_club-kick-0701 102174_zgump_club-kick-0702 102175_zgump_club-kick-0703 102176_zgump_club-kick-0704	wav wav wav wav
2	<a href="https://freesound.org/people/Pder/packs/6748/">https://freesound.org/people/Pder/packs/6748/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	106290_pder_mein-kick-1 106327_pder_mein-kick-2 106337_pder_mein-kick-3 106338_pder_mein-kick-4	wav wav wav wav
3	<a href="https://freesound.org/people/1%20Death%20God/packs/7509/">https://freesound.org/people/1%20Death%20God/packs/7509/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	120136_i-death-god_h-kick02 120137_i-death-god_h-kick03 120138_i-death-god_h-kick04 120139_i-death-god_h-kick05	wav wav wav wav
4	<a href="https://freesound.org/people/crispydinner/packs/8287/">https://freesound.org/people/crispydinner/packs/8287/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	131796_crispydinner_lyr-kck-mk4-r1-marker-4 131797_crispydinner_lyr-kck-mk4-r1-marker-3 131798_crispydinner_lyr-kck-mk4-r1-marker-2 131799_crispydinner_lyr-kck-mk4-r1-marker-1	wav wav wav wav
5	<a href="https://freesound.org/people/opm/packs/121/">https://freesound.org/people/opm/packs/121/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	2085_opm_kk-set1 2086_opm_kk-set2 2087_opm_kk-set3 2089_opm_kk-set5	wav wav wav wav
6	<a href="https://freesound.org/people/Stomp/packs/1250/">https://freesound.org/people/Stomp/packs/1250/</a>	Synthetisierte Bassdrum des Clavia Nordrack 2. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	20909_stomp_nr2-bd22-05 20910_stomp_nr2-bd23-01 20914_stomp_nr2-bd24-03 20917_stomp_nr2-bd25-03	aif aif aif aif
7	<a href="https://freesound.org/people/waveplay/packs/12562/">https://freesound.org/people/waveplay/packs/12562/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	218038_waveplay_deep-hard-bass-kick-g4 218039_waveplay_deep-hard-bass-kick-a4 218041_waveplay_deep-hard-bass-kick-c5 218042_waveplay_deep-hard-bass-kick-f4	ogg ogg ogg ogg
8	<a href="https://freesound.org/people/waveplay/packs/12562/">https://freesound.org/people/waveplay/packs/12562/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	237065_waveplay_deadmau5-style-kick-drum-alternate-2 237068_waveplay_deadmau5-style-kick-drum-alternate-5 237072_waveplay_deadmau5-style-kick-drum-alternate-9 237075_waveplay_deadmau5-style-kick-drum-alternate-12	wav wav wav wav
9	<a href="https://freesound.org/people/crispydinner/packs/15497/">https://freesound.org/people/crispydinner/packs/15497/</a>	Synthetisierte Bassdrum des TR 909. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	252719_crispydinner_crdn-upgrkick-004 252720_crispydinner_crdn-upgrkick-003 252721_crispydinner_crdn-upgrkick-002 252722_crispydinner_crdn-upgrkick-001	wav wav wav wav
10	<a href="https://freesound.org/people/Walter_Odington/packs/1579/">https://freesound.org/people/Walter_Odington/packs/1579/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche EQ und Kompressoreinstellungen	25603_walter-odington_amen-kick-1-a 25604_walter-odington_amen-kick-1-b 25605_walter-odington_amen-kick-1-c 25606_walter-odington_amen-kick-2-a	aif aif aif aif
11	<a href="https://freesound.org/people/Hard3eat/packs/10701/">https://freesound.org/people/Hard3eat/packs/10701/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	266777_hard3eat_synth-kick-4 266778_hard3eat_synth-kick-5 266781_hard3eat_synth-kick-6 266783_hard3eat_synth-kick-8	wav wav wav wav
12	<a href="https://freesound.org/people/crispydinner/packs/17071/">https://freesound.org/people/crispydinner/packs/17071/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	276707_crispydinner_crdn-pndlrgrt-kick-marker-12 276730_crispydinner_crdn-pndlrgrt-kick-marker-130 276735_crispydinner_crdn-pndlrgrt-kick-marker-62 276746_crispydinner_crdn-pndlrgrt-kick-marker-66	wav wav wav wav
13	<a href="https://freesound.org/people/Harha/packs/18234/">https://freesound.org/people/Harha/packs/18234/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	324806_harha_kick-01-f2 324809_harha_kick-06-g-2 324813_harha_kick-08-f2 324815_harha_kick-10-g-2	wav wav wav wav
14	<a href="https://freesound.org/people/AndreOdate/packs/18537/">https://freesound.org/people/AndreOdate/packs/18537/</a>	Aus mehreren Samples gelayerte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	330041_andreonate_dre-kick-10 330042_andreonate_dre-kick-09 330043_andreonate_dre-kick-08 330045_andreonate_dre-kick-11	wav wav wav wav
15	<a href="https://freesound.org/people/kurnwheeler/sounds/339018/">https://freesound.org/people/kurnwheeler/sounds/339018/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	339019_kurnwheeler_hard-dp04 339020_kurnwheeler_hard-dp03 339021_kurnwheeler_hard-dp02 339025_kurnwheeler_hard-dp06	wav wav wav wav
16	<a href="https://freesound.org/people/sandyrb/packs/2295/">https://freesound.org/people/sandyrb/packs/2295/</a>	Hybrid aus Mikrofonaufnahme und synthetisierter Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	35989_sandyrb_ac10-kick-001 35990_sandyrb_ac10-kick-002 35991_sandyrb_ac10-kick-003 35992_sandyrb_ac10-kick-004	wav wav wav wav
17	<a href="https://freesound.org/people/Samulis/packs/21073/">https://freesound.org/people/Samulis/packs/21073/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	375257_samulis_bass-drum-bdrum-f-1 375258_samulis_bass-drum-bdrum-f-2 375259_samulis_bass-drum-bdrum-ff-1 375260_samulis_bass-drum-bdrum-fff-1	wav wav wav wav
18	<a href="https://freesound.org/people/SBAudio/packs/21201/">https://freesound.org/people/SBAudio/packs/21201/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	377402_sbaudio_reverb-kick-09 377404_sbaudio_reverb-kick-07 377405_sbaudio_reverb-kick-08 377406_sbaudio_reverb-kick-05	wav wav wav wav
19	<a href="https://freesound.org/people/sandyrb/packs/2471/">https://freesound.org/people/sandyrb/packs/2471/</a>	Hybrid aus Mikrofonaufnahme und synthetisierter Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	38901_sandyrb_mdvp-kick-001 38902_sandyrb_mdvp-kick-002 38903_sandyrb_mdvp-kick-003 38904_sandyrb_mdvp-kick-004	wav wav wav wav
20	<a href="https://freesound.org/people/pestilenza/packs/5106/">https://freesound.org/people/pestilenza/packs/5106/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	40612_pestilenza_kick-03 40613_pestilenza_kick-04 40614_pestilenza_kick-05 40615_pestilenza_kick-06	wav wav wav wav
21	<a href="https://freesound.org/people/sandyrb/packs/2614/">https://freesound.org/people/sandyrb/packs/2614/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	41152_sandyrb_dnb-kick-005 41153_sandyrb_dnb-kick-006 41154_sandyrb_dnb-kick-007 41155_sandyrb_dnb-kick-008	wav wav wav wav
22	<a href="https://freesound.org/people/zgump/packs/5495/">https://freesound.org/people/zgump/packs/5495/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	83762_zgump_kick-pack-0701 83763_zgump_kick-pack-0702 83764_zgump_kick-pack-0703 83765_zgump_kick-pack-0704	wav wav wav wav
23	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-2294">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-2294</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	CYcdh_AcouKick-01 CYcdh_AcouKick-02 CYcdh_AcouKick-03 CYcdh_AcouKick-04 CYcdh_KesKick-01	wav wav wav wav wav
24	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-2294">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-2294</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	CYcdh_KesKick-04 CYcdh_KesKick-06 CYcdh_KesKick-08	wav wav wav
25	<a href="http://bedroomproducersblog.com/2013/02/05/500-free-kicks-by-markus-hakala/">http://bedroomproducersblog.com/2013/02/05/500-free-kicks-by-markus-hakala/</a>	Synthetisierte Bassdrum. Minimale Klangfarbenvariation durch Anpassung der Bassdrum an vier unterschiedliche Tonhöhen	mhak_kick_201_G mhak_kick_220_F mhak_kick_240_A# mhak_kick_260_G	wav wav wav wav

## 8.2.3 FSD 1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="https://freesound.org/people/Benboncan/packs/6541/">https://freesound.org/people/Benboncan/packs/6541/</a>	"Hammer & Amboss" Minimale Klangfarbenunterschiede durch unterschiedliche Schläge	103628_benboncan_large-anvil-steel-hammer	wav
			103629_benboncan_large-anvil-steel-hammer-1	wav
			103630_benboncan_large-anvil-steel-hammer-2	wav
			103631_benboncan_large-anvil-steel-hammer-3	wav
			103632_benboncan_large-anvil-steel-hammer-4	wav
2	<a href="http://freesound.org/people/samuelburt/packs/8357/">http://freesound.org/people/samuelburt/packs/8357/</a>	"Affe Daxophon" Minimale Klangfarbenunterschiede durch unterschiedliche Spielweisen	133143_samuelburt_ape-rubba-32-talky	aiff
			133144_samuelburt_ape-rubba-33-talky	aiff
			133145_samuelburt_ape-rubba-34-talky	aiff
			133146_samuelburt_ape-rubba-35-talky	aiff
			133147_samuelburt_ape-rubba-36-talky	aiff
3	<a href="http://freesound.org/people/samuelburt/packs/8357/">http://freesound.org/people/samuelburt/packs/8357/</a>	"Bestie Daxophon" Minimale Klangfarbenunterschiede durch unterschiedliche Spielweisen	133258_samuelburt_beast-sigh-01	aiff
			133259_samuelburt_beast-sigh-02	aiff
			133260_samuelburt_beast-sigh-03	aiff
			133261_samuelburt_beast-sigh-04	aiff
			133262_samuelburt_beast-sigh-05	aiff
4	<a href="https://freesound.org/people/RHumphries/packs/111/">https://freesound.org/people/RHumphries/packs/111/</a>	"Applause" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	1921_rumphries_rbh-applause-01-big	wav
			1922_rumphries_rbh-applause-02-big	wav
			1923_rumphries_rbh-applause-03-big	wav
			1924_rumphries_rbh-applause-04-big	wav
			1925_rumphries_rbh-applause-05-big	wav
5	<a href="https://freesound.org/people/Cratic/packs/13033/">https://freesound.org/people/Cratic/packs/13033/</a>	"Zerbrechendes Glas" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	204120_cratic_glass01-proc	flac
			204122_cratic_glass02-proc	flac
			204124_cratic_glass04-proc	flac
			204143_cratic_glass05-proc	flac
			213168_arnaud-coutancier_cork-bouchon-04	wav
6	<a href="https://freesound.org/people/arnaud%20coutancier/packs/2937/">https://freesound.org/people/arnaud%20coutancier/packs/2937/</a>	"Weinflasche Öffnen" Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	213169_arnaud-coutancier_cork-bouchon-03	wav
			213170_arnaud-coutancier_cork-bouchon-02	wav
			213171_arnaud-coutancier_cork-bouchon-01	wav
			213172_arnaud-coutancier_cork-bouchon-05	wav
			216937_funkymuskrat_moon-fauna-64	wav
7	<a href="https://freesound.org/people/funkymuskrat/packs/13747/">https://freesound.org/people/funkymuskrat/packs/13747/</a>	"Greifvogel Schrei" Minimale Klangfarbenunterschiede durch unterschiedliche Layeringsounds	216938_funkymuskrat_moon-fauna-63	wav
			216940_funkymuskrat_moon-fauna-65	wav
			216943_funkymuskrat_moon-fauna-62	wav
			216944_funkymuskrat_moon-fauna-61	wav
			236278_delphidebrain_jazz-the-dog-howl-bark-055	wav
8	<a href="https://freesound.org/people/delphidebrain/packs/14500/">https://freesound.org/people/delphidebrain/packs/14500/</a>	"Bellender Hund" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	236279_delphidebrain_jazz-the-dog-howl-bark-054	wav
			236280_delphidebrain_jazz-the-dog-howl-bark-053	wav
			236281_delphidebrain_jazz-the-dog-howl-bark-052	wav
			236282_delphidebrain_jazz-the-dog-howl-bark-051	wav
			264690_michael-klier_bundeswehr-bag-heavy-medium-movement-processed-24bit96khz-004	wav
9	<a href="https://freesound.org/people/Michael%20Klier/packs/16309/">https://freesound.org/people/Michael%20Klier/packs/16309/</a>	"Rucksack" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	264691_michael-klier_bundeswehr-bag-heavy-medium-movement-processed-24bit96khz-003	wav
			264693_michael-klier_bundeswehr-bag-heavy-medium-movement-processed-24bit96khz-005	wav
			264696_michael-klier_bundeswehr-bag-heavy-medium-movement-processed-24bit96khz-002	wav
			264697_michael-klier_bundeswehr-bag-heavy-medium-movement-processed-24bit96khz-001	wav
			270505_littlerobotsoundfactory_breath-scared-03	wav
10	<a href="https://freesound.org/people/LittleRobotSoundfactory/packs/16688/">https://freesound.org/people/LittleRobotSoundfactory/packs/16688/</a>	"Ängstliches Atmen" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	270506_littlerobotsoundfactory_breath-scared-02	wav
			270507_littlerobotsoundfactory_breath-scared-01	wav
			270511_littlerobotsoundfactory_breath-scared-05	wav
			270512_littlerobotsoundfactory_breath-scared-04	wav
			273322_arnaud-coutancier_fall-4-on-floor-cushion	wav
11	<a href="https://freesound.org/people/arnaud%20coutancier/packs/2937/">https://freesound.org/people/arnaud%20coutancier/packs/2937/</a>	"Hinfallen" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	273323_arnaud-coutancier_fall-2-on-floor-cushion	wav
			273324_arnaud-coutancier_fall-1-on-floor-cushion	wav
			273325_arnaud-coutancier_fall-6-on-floor-cushion	wav
			273326_arnaud-coutancier_fall-5-on-floor-cushion	wav
			377723_danarobinsondesigns@gmail.com_b30	ogg
12	<a href="https://freesound.org/people/danarobinsondesigns@gmail.com/packs/21217/">https://freesound.org/people/danarobinsondesigns@gmail.com/packs/21217/</a>	"Kleine Glocke" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen und Spielarten	377725_danarobinsondesigns@gmail.com_b29	ogg
			377727_danarobinsondesigns@gmail.com_b28	ogg
			377729_danarobinsondesigns@gmail.com_b27	ogg
			377731_danarobinsondesigns@gmail.com_b26	ogg
			39_anton_glass-e1-ff	wav
14	<a href="https://freesound.org/people/Anton/packs/4/">https://freesound.org/people/Anton/packs/4/</a>	"Klingendes Glas" Minimale Klangfarbenunterschiede durch unterschiedliche Tonhöhen	42_anton_glass-f-ff	wav
			45_anton_glass-f-ff	wav
			48_anton_glass-g-ff	wav
			51_anton_glass-g-ff	wav
			59990_qubodup_swosh-03.flac	flac
15	<a href="https://freesound.org/people/qubodup/packs/12143/">https://freesound.org/people/qubodup/packs/12143/</a>	"Bambus Hieb" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	59992_qubodup_swosh-sword-swing.flac	flac
			59994_qubodup_swosh-07.flac	flac
			59996_qubodup_swoosh-wind-09.flac	flac
			59997_qubodup_swosh-10.flac	flac
			68765_aboe_canpnc01	wav
16	<a href="https://freesound.org/people/aboe/packs/4424/">https://freesound.org/people/aboe/packs/4424/</a>	"Dose Zerdrücken" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	68766_aboe_canpnc02	wav
			68767_aboe_canpnc03	wav
			68768_aboe_canpnc04	wav
			68769_aboe_canpnc05	wav
			68868_aboe_canslice1	wav
17	<a href="https://freesound.org/people/aboe/packs/4424/">https://freesound.org/people/aboe/packs/4424/</a>	"Dose Zerreißen" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	68869_aboe_canslice2	wav
			68870_aboe_canslice3	wav
			68871_aboe_canslice4	wav
			68872_aboe_canslice5	wav
			82_plagasul_uhraa	wav
19	<a href="https://freesound.org/people/plagasul/packs/3/">https://freesound.org/people/plagasul/packs/3/</a>	"Indio Kampflaute" Minimale Klangfarbenunterschiede durch unterschiedliche Läufe	83_plagasul_uoh	wav
			84_plagasul_uueh	wav
			85_plagasul_jeeh	wav
			86_plagasul_oh-h	wav
			zapsplat_foley_raincoat_polyester_movements_001	mp3
18	<a href="https://www.zapsplat.com/sound-effect-packs/raincoat-foley/">https://www.zapsplat.com/sound-effect-packs/raincoat-foley/</a>	"Polyester" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	zapsplat_foley_raincoat_polyester_movements_002	mp3
			zapsplat_foley_raincoat_polyester_movements_003	mp3
			zapsplat_foley_raincoat_polyester_movements_004	mp3
			zapsplat_foley_raincoat_polyester_movements_005	mp3
			zapsplat_human_eat_popcorn_001_11394	mp3
19	<a href="https://www.zapsplat.com/sound-effect-packs/popcorn/">https://www.zapsplat.com/sound-effect-packs/popcorn/</a>	"Kaugeräusch Popcorn" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	zapsplat_human_eat_popcorn_002_11395	mp3
			zapsplat_human_eat_popcorn_003_11396	mp3
			zapsplat_human_eat_popcorn_004_11397	mp3
			zapsplat_human_eat_popcorn_005_11398	mp3
			zapsplat_impact_metal_oil_drum_empty_001_10589	mp3
20	<a href="https://www.zapsplat.com/sound-effect-packs/metal-oil-drum-impacts/">https://www.zapsplat.com/sound-effect-packs/metal-oil-drum-impacts/</a>	"Ölfass" Minimale Klangfarbenunterschiede durch unterschiedliche Aufnahmen	zapsplat_impact_metal_oil_drum_empty_002_10590	mp3
			zapsplat_impact_metal_oil_drum_empty_003_10591	mp3
			zapsplat_impact_metal_oil_drum_empty_004_10592	mp3
			zapsplat_impact_metal_oil_drum_empty_005_10593	mp3

## 8.2.4 GSD 1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="https://www.zapsplat.com/sound-effect-packs/cartoon-boings/">https://www.zapsplat.com/sound-effect-packs/cartoon-boings/</a>	"Boing", "Hüpfgeräusch". Aufnahme einer Maultrommel. Minimale Klangfarbenunterschiede durch unterschiedliche Spielweise und Effekteinstellungen	zapsplat_cartoon_boings_jaw_harp_001	mp3
			zapsplat_cartoon_boings_jaw_harp_002	mp3
			zapsplat_cartoon_boings_jaw_harp_003	mp3
			zapsplat_cartoon_boings_jaw_harp_004	mp3
			zapsplat_cartoon_boings_jaw_harp_005	mp3
2	<a href="https://freesound.org/people/B_Lamerichs/packs/16234/">https://freesound.org/people/B_Lamerichs/packs/16234/</a>	"Sucess" Tonfolge. Minimale Klangfarbenunterschiede durch unterschiedliche Tonhöhen	263686_b-lamerichs_ableton-game-sound-effects-09-02-2015-1	mp3
			263699_b-lamerichs_ableton-game-sound-effects-09-02-2015-2	mp3
			263701_b-lamerichs_ableton-game-sound-effects-09-02-2015-3	mp3
			263705_b-lamerichs_ableton-game-sound-effects-09-02-2015-5	mp3
			263706_b-lamerichs_ableton-game-sound-effects-09-02-2015-4	mp3
3	<a href="https://freesound.org/people/pan14/packs/16196/">https://freesound.org/people/pan14/packs/16196/</a>	"Beep" Minimale Klangfarbenunterschiede durch verschiedene Effekteinstellungen	263126_pan14_tone-beep-lower-slower	wav
			263127_pan14_tone-beep-large-verb	wav
			263128_pan14_tone-beep-amb-verb	wav
			263130_pan14_tone-beep-slower-lower-large-verb	wav
			263131_pan14_tone-beep-slower-lower-amb-verb	wav
4	<a href="http://sweetsoundeffects.com/lightsaber-sounds/">http://sweetsoundeffects.com/lightsaber-sounds/</a>	"Laserschwert" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds.	Spin 1	wav
			Spin 2	wav
			Spin 3	wav
			Spin 4	wav
			Spin 5	wav
5	<a href="http://freesound.org/people/Jovica/packs/294/">http://freesound.org/people/Jovica/packs/294/</a>	"Elektronischer Hintergrund" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds	5471_jovica_attack-zound-94	wav
			5472_jovica_attack-zound-95	wav
			5473_jovica_attack-zound-96	wav
			5474_jovica_attack-zound-97	wav
			5475_jovica_attack-zound-98	wav
6	<a href="https://freesound.org/people/LittleRobotSoundFactory/packs/16689/">https://freesound.org/people/LittleRobotSoundFactory/packs/16689/</a>	"Laserpistole" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds	270546_littlerobotsoundfactory_laser-06	wav
			270547_littlerobotsoundfactory_laser-05	wav
			270548_littlerobotsoundfactory_laser-04	wav
			270549_littlerobotsoundfactory_laser-03	wav
			270551_littlerobotsoundfactory_laser-07	wav
7	<a href="http://freesound.org/people/LittleRobotSoundFactory/packs/17311/">http://freesound.org/people/LittleRobotSoundFactory/packs/17311/</a>	"Taktile Sound bei Buttonberührung" Minimale Klangfarbenunterschiede durch verschiedene Aufnahmen des selben Geräuschs	290415_littlerobotsoundfactory_mouth-02	wav
			290416_littlerobotsoundfactory_mouth-01	wav
			290419_littlerobotsoundfactory_mouth-06	wav
			290421_littlerobotsoundfactory_mouth-04	wav
			290422_littlerobotsoundfactory_mouth-03	wav
8	<a href="https://freesound.org/people/waveplay/packs/12565/">https://freesound.org/people/waveplay/packs/12565/</a>	"Failure Sound" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds	221767_waveplay_game-fx-2	wav
			221768_waveplay_game-fx-3	wav
			221769_waveplay_game-fx-4	wav
			221770_waveplay_game-fx-5	wav
			221771_waveplay_game-fx-6	wav
9	<a href="https://opengameart.org/content/512-sound-effects-8-bit-style">https://opengameart.org/content/512-sound-effects-8-bit-style</a>	"Menu Select Sound" Minimale Klangfarbenunterschiede durch verschiedene Tonhöhen	sfx_menu_move1	wav
			sfx_menu_move2	wav
			sfx_menu_move3	wav
			sfx_menu_move4	wav
			sfx_menu_move5	wav
10	<a href="https://opengameart.org/content/512-sound-effects-8-bit-style">https://opengameart.org/content/512-sound-effects-8-bit-style</a>	"Alien Todesschrei" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds	sfx_deathscream_alien2	wav
			sfx_deathscream_alien3	wav
			sfx_deathscream_alien4	wav
			sfx_deathscream_alien5	wav
			sfx_deathscream_alien6	wav
11	<a href="https://awolfe.home.xs4all.nl/samples.html">https://awolfe.home.xs4all.nl/samples.html</a>	"Bleep 2" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds.	FX-FiltergameleanBlewoop	wav
			FX-Filtergameleanfreaky	wav
			FX-FiltergameleanOne	wav
			FX-FiltergameleanShort	wav
			FX-FiltergameleanWooBleep	wav
12	<a href="https://freesound.org/people/ionicmusic/packs/12529/">https://freesound.org/people/ionicmusic/packs/12529/</a>	"Robo Voice" Minimale Klangfarbenunterschiede durch Aussprache unterschiedlicher Wörter durch den selben Vocoder	196850_ionicmusic_race-robot-danger	wav
			196852_ionicmusic_race-robot-third-place	wav
			196863_ionicmusic_race-robot-full-speed	wav
			196868_ionicmusic_race-robot-game-over	wav
			196879_ionicmusic_race-robot-shut-down	wav
13	<a href="https://awolfe.home.xs4all.nl/samples.html">https://awolfe.home.xs4all.nl/samples.html</a>	"Power Up Sound" Minimale Klangfarbenunterschiede durch unterschiedlich lange Attack- und Releaseanteile	FX-ER-Kraftwerkcity	wav
			FX-ER-Kraftwerkcity2	wav
			FX-ER-Kraftwerkcity3	wav
			FX-ER-Kraftwerkcity5	wav
			FX-ER-Kraftwerkcity6almost	wav
14	<a href="https://www.zapsplat.com/sound-effect-packs/magic-wand/">https://www.zapsplat.com/sound-effect-packs/magic-wand/</a>	"Whoosh" Minimale Klangfarbenunterschiede durch verschiedene Layeringsounds.	zapsplat_magic_wand_whoosh_burst_001_12546	mp3
			zapsplat_magic_wand_whoosh_burst_002_12547	mp3
			zapsplat_magic_wand_whoosh_burst_003_12548	mp3
			zapsplat_magic_wand_whoosh_burst_004_12549	mp3
			zapsplat_magic_wand_whoosh_burst_005_12550	mp3
15	<a href="http://freesound.org/people/LittleRobotSoundFactory/packs/">http://freesound.org/people/LittleRobotSoundFactory/packs/</a>	"Dämon" Minimale Klangfarbenunterschiede durch Aussprache unterschiedlicher Wörter durch den selben Vocoder	316364_littlerobotsoundfactory_demon-03	wav
			316365_littlerobotsoundfactory_demon-02	wav
			316366_littlerobotsoundfactory_demon-01	wav
			316367_littlerobotsoundfactory_demon-00	wav
			316370_littlerobotsoundfactory_demon-05	wav
16	<a href="http://freesound.org/people/taren1965/packs/8339/">http://freesound.org/people/taren1965/packs/8339/</a>	"Alien2" Minimale Klangfarbenunterschiede durch unterschiedliche Tonhöhen	132475_taren1965_caught-alien-3	aif
			132476_taren1965_caught-alien-2	aif
			132477_taren1965_caught-alien-1	aif
			132481_taren1965_caught-alien-5	aif
			132482_taren1965_caught-alien-4	aif
17	<a href="http://freesound.org/people/jalastram/packs/17815/">http://freesound.org/people/jalastram/packs/17815/</a>	"Retro Game Sound" Minimale Klangfarbenunterschiede durch unterschiedliche Akkorde	318210_jalastram_retro-game-sounds-sfx-23	wav
			318213_jalastram_retro-game-sounds-sfx-26	wav
			318214_jalastram_retro-game-sounds-sfx-27	wav
			318215_jalastram_retro-game-sounds-sfx-24	wav
			318216_jalastram_retro-game-sounds-sfx-25	wav
18	<a href="http://freesound.org/people/Wdfourtee/packs/12179/">http://freesound.org/people/Wdfourtee/packs/12179/</a>	"Waffengeräusch" Minimale Klangfarbenunterschiede durch verschiedene Aufnahmen des selben Geräuschs	192050_wdfourtee_mx1993-yarstick-swoosh-4	wav
			192051_wdfourtee_mx1993-yarstick-swoosh-3	wav
			192053_wdfourtee_mx1993-yarstick-swoosh-1	wav
			192054_wdfourtee_mx1993-yarstick-swoosh-6	wav
			192055_wdfourtee_mx1993-yarstick-swoosh-5	wav
19	<a href="http://freesound.org/people/B_Lamerichs/packs/12248/">http://freesound.org/people/B_Lamerichs/packs/12248/</a>	"Event Sound" Minimale Klangfarbenunterschiede durch unterschiedliche Tonhöhen und Layeringsounds	193253_b-lamerichs_eventsounds4-l	mp3
			193254_b-lamerichs_eventsounds4-k2	mp3
			193257_b-lamerichs_eventsounds4-l2	mp3
			193258_b-lamerichs_eventsounds4-m2	mp3
			193259_b-lamerichs_eventsounds4-m	mp3
20	<a href="http://freesound.org/people/LittleRobotSoundFactory/packs/16687/">http://freesound.org/people/LittleRobotSoundFactory/packs/16687/</a>	"Water Footstep" Minimale Klangfarbenunterschiede durch verschiedene Aufnahmen des selben Geräuschs	270422_littlerobotsoundfactory_footstep-water-01	wav
			270423_littlerobotsoundfactory_footstep-water-00	wav
			270425_littlerobotsoundfactory_footstep-water-02	wav
			270427_littlerobotsoundfactory_footstep-water-04	wav
			270429_littlerobotsoundfactory_footstep-water-03	wav

## 8.2.5 HH 1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	KHats Clsd-01 KHats Clsd-02 KHats Clsd-03 KHats Clsd-04	wav wav wav wav
2	<a href="https://freesound.org/people/minorr/sounds/">https://freesound.org/people/minorr/sounds/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte und offen/geschlossen	104221_minorr_hhat-paiste-302-14-f 104223_minorr_hhat-paiste-302-14-p 104229_minorr_hhat-paiste-302-14-semi-open-f 104231_minorr_hhat-paiste-302-14-semi-open-p	wav wav wav wav
3	<a href="https://freesound.org/people/quartertone/packs/8588/">https://freesound.org/people/quartertone/packs/8588/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	137426_quartertone_hh13inkz-eso-v02 137427_quartertone_hh13inkz-eso-v01 137428_quartertone_hh13inkz-eso-v04 137429_quartertone_hh13inkz-eso-v03	wav wav wav wav
4	<a href="https://freesound.org/people/quartertone/packs/8656/">https://freesound.org/people/quartertone/packs/8656/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	138686_quartertone_chh18x20-1-bwsc-v04 138687_quartertone_chh18x20-1-bwsc-v03 138688_quartertone_chh18x20-1-bwsc-v06 139689_quartertone_chh18x20-1-bwsc-v05	wav wav wav wav
5	<a href="https://freesound.org/people/tlbbits/packs/889/">https://freesound.org/people/tlbbits/packs/889/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte und offen/geschlossen	16272_tlbbits_hh-a-custom-full-open-hard 16273_tlbbits_hh-a-custom-half-open-hard 16274_tlbbits_hh-a-custom-half-open-soft 16275_tlbbits_hh-a-custom-mostly-open-hard	wav wav wav wav
6	<a href="https://freesound.org/people/AlienXXX/packs/11587/">https://freesound.org/people/AlienXXX/packs/11587/</a>	Beatboxing sounds. Minimale Klangfarbenunterschiede durch unterschiedliche Läufe	183675_alienxxx_openhat-004 183676_alienxxx_openhat-003 183677_alienxxx_openhat-002 183678_alienxxx_openhat-001	wav wav wav wav
7	<a href="https://freesound.org/people/concierge_weetr/packs/119/">https://freesound.org/people/concierge_weetr/packs/119/</a>	Eko rhythmaker (mid-70's drum computer) Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen und Spielarten	2028_concierge-weetr_hat2 2029_concierge-weetr_hat3 2030_concierge-weetr_openhat1 2031_concierge-weetr_openhat2	wav wav wav wav
8	<a href="https://freesound.org/people/TwistedLemon/packs/120/">https://freesound.org/people/TwistedLemon/packs/120/</a>	Aufnahme eines analogen Drum Computers. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	2045_twistedlemon_mattel-hh1 2046_twistedlemon_mattel-hh2 2047_twistedlemon_mattel-hh3 2048_twistedlemon_mattel-hh4	wav wav wav wav
9	<a href="https://freesound.org/people/Walter_Odington/packs/1583/">https://freesound.org/people/Walter_Odington/packs/1583/</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	25680_walter-odington_blake-hall-hat-2 25681_walter-odington_blake-hall-hat-3 25682_walter-odington_blake-hall-hat-4 25683_walter-odington_blake-hall-hat-5	aiff aiff aiff aiff
10	<a href="https://freesound.org/people/Walter_Odington/packs/1583/">https://freesound.org/people/Walter_Odington/packs/1583/</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	25693_walter-odington_hackney-hat-1 25694_walter-odington_hackney-hat-2 25695_walter-odington_hackney-hat-3 25696_walter-odington_hackney-hat-4	aiff aiff aiff aiff
11	<a href="https://freesound.org/people/stomachache/packs/1813/">https://freesound.org/people/stomachache/packs/1813/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	29779_stomachache_hat1 29780_stomachache_hat2 29784_stomachache_hattigt1 29785_stomachache_hattigt2	wav wav wav wav
12	<a href="https://freesound.org/people/shpira/packs/17760/">https://freesound.org/people/shpira/packs/17760/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	317088_shpira_hat-cl2 317089_shpira_hat-cl1 317092_shpira_hat-cl4 317093_shpira_hat-cl3	wav wav wav wav
13	<a href="https://freesound.org/people/cabled_mess/packs/19219/">https://freesound.org/people/cabled_mess/packs/19219/</a>	Aufnahme eines analogen Drum Computers. Minimale Klangfarbenunterschiede durch Gate und EQ	339281_cabled-mess_hihat-closed-raw-07-gate-eq 339282_cabled-mess_hihat-closed-raw-06-gate-eq 339283_cabled-mess_hihat-closed-raw-05-gate-eq 339284_cabled-mess_hihat-closed-raw-04-gate-eq	wav wav wav wav
14	<a href="https://freesound.org/people/sandyrb/packs/2265/">https://freesound.org/people/sandyrb/packs/2265/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	35633_sandyrb_real-closed-hat-001 35634_sandyrb_real-closed-hat-002 35635_sandyrb_real-closed-hat-003 35636_sandyrb_real-closed-hat-004	wav wav wav wav
15	<a href="http://freesound.org/people/SBAudio/packs/21201/">http://freesound.org/people/SBAudio/packs/21201/</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	377354_sbaudio_hh-open-04 377356_sbaudio_hh-open-06 377357_sbaudio_hh-open-05 377359_sbaudio_hh-open-07	wav wav wav wav
16	<a href="https://freesound.org/people/Robinhood76/packs/3940/">https://freesound.org/people/Robinhood76/packs/3940/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	61555_robinhood76_aa084-sabian-aax-hihat-14-closed-weak 61556_robinhood76_aa085-sabian-aax-hihat-14-closed-normal 61557_robinhood76_aa086-sabian-aax-hihat-14-closed-strong 61558_robinhood76_aa087-sabian-aax-hihat-14-closed-powerful	wav wav wav wav
17	<a href="https://freesound.org/people/CBeeching/packs/4857/">https://freesound.org/people/CBeeching/packs/4857/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	75034_cbeeching_hat-04-41 75035_cbeeching_hat-04-43 75036_cbeeching_hat-04-5 75037_cbeeching_hat-04	wav wav wav wav
18	<a href="http://freedrumkits.net/download/freedrumkits-net-808-hi-hats-zip/">http://freedrumkits.net/download/freedrumkits-net-808-hi-hats-zip/</a>	808 Drum Machine. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	8080HH3_TapeSat 8080HH3_X 8080HH3_X2 8080HH3_X3	wav wav wav wav
19	<a href="http://www.stayonbeat.com/2013/03/05/free-hi-hat-sounds/">http://www.stayonbeat.com/2013/03/05/free-hi-hat-sounds/</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	909_long_open_hat 909_new_hat 909_open_hat 909_open_hat_short_verb	wav wav wav wav
20	<a href="https://archive.org/details/OpenPathMusic44V1">https://archive.org/details/OpenPathMusic44V1</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte und offen/geschlossen	cymbal-hihat-open-stick-1 cymbal-hihat-open-stick-2 cymbal-hihat-openclose-stick cymbal-hihat-stick	wav wav wav wav
21	<a href="http://www.producerspot.com/download-free-edm-drums-drum-samples-kit-by-producerspot">http://www.producerspot.com/download-free-edm-drums-drum-samples-kit-by-producerspot</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	edm-hihat-16 edm-hihat-17 edm-hihat-18 edm-hihat-20	wav wav wav wav
22	<a href="https://soundpacks.com/free-sound-packs/271-hi-hat-samples/">https://soundpacks.com/free-sound-packs/271-hi-hat-samples/</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	hihat_013a hihat_013B hihat_014a hihat_014b	wav wav wav wav
23	<a href="https://www.kvraudio.com/forum/viewtopic.php?t=384248">https://www.kvraudio.com/forum/viewtopic.php?t=384248</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	open short 01 open short 02 open short 03 open short 04	wav wav wav wav
24	<a href="http://www.musicradar.com/news/tech/sampleradar-477-free-ultimate-hi-hat-samples-639419">http://www.musicradar.com/news/tech/sampleradar-477-free-ultimate-hi-hat-samples-639419</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch Schlaghärte	UHH_AcoustiHatB_Hit-01 UHH_AcoustiHatB_Hit-02 UHH_AcoustiHatB_Hit-03 UHH_AcoustiHatB_Hit-04	wav wav wav wav
25	<a href="http://www.musicradar.com/news/tech/sampleradar-477-free-ultimate-hi-hat-samples-639419">http://www.musicradar.com/news/tech/sampleradar-477-free-ultimate-hi-hat-samples-639419</a>	Synthetisierte High-Hat. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	UHH_ElectroHatC_Hit-01 UHH_ElectroHatC_Hit-02 UHH_ElectroHatC_Hit-03 UHH_ElectroHatC_Hit-04	wav wav wav wav

## 8.2.6 M1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"bass-clarinet" A3 und C3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	bass-clarinet_A3_1_forte_normal.wav	mp3
			bass-clarinet_A3_1_fortissimo_normal.wav	mp3
			bass-clarinet_A3_1_pianissimo_normal.wav	mp3
			bass-clarinet_A3_1_piano_normal.wav	mp3
			bass-clarinet_C3_1_forte_normal.wav	mp3
			bass-clarinet_C3_1_mezzo-forte_normal.wav	mp3
			bass-clarinet_C3_1_pianissimo_normal.wav	mp3
			bass-clarinet_C3_1_piano_normal.wav	mp3
2	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"bassoon" A3 und C3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	bassoon_A3_1_forte_normal.wav	mp3
			bassoon_A3_1_fortissimo_normal.wav	mp3
			bassoon_A3_1_mezzo-forte_normal.wav	mp3
			bassoon_A3_1_mezzo-piano_normal.wav	mp3
			bassoon_C3_1_forte_normal.wav	mp3
			bassoon_C3_1_fortissimo_normal.wav	mp3
			bassoon_C3_1_mezzo-piano_normal.wav	mp3
			bassoon_C3_1_piano_normal.wav	mp3
3	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"cello" A3 und C3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	cello_A3_15_forte_arco-normal.wav	mp3
			cello_A3_1_forte_arco-normal.wav	mp3
			cello_A3_1_fortissimo_arco-normal.wav	mp3
			cello_A3_1_mezzo-piano_arco-normal.wav	mp3
			cello_C3_15_piano_arco-normal.wav	mp3
			cello_C3_1_forte_arco-normal.wav	mp3
			cello_C3_1_fortissimo_arco-normal.wav	mp3
			cello_C3_1_mezzo-piano_arco-normal.wav	mp3
4	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"clarinet" A3 und E3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	clarinet_A3_1_forte_normal.wav	mp3
			clarinet_A3_1_fortissimo_normal.wav	mp3
			clarinet_A3_1_pianissimo_normal.wav	mp3
			clarinet_A3_1_piano_normal.wav	mp3
			clarinet_E3_1_forte_normal.wav	mp3
			clarinet_E3_1_fortissimo_normal.wav	mp3
			clarinet_E3_1_pianissimo_normal.wav	mp3
			clarinet_E3_1_piano_normal.wav	mp3
5	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"contrabassoon" A3 und C3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	contrabassoon_A3_1_forte_normal.wav	mp3
			contrabassoon_A3_1_fortissimo_normal.wav	mp3
			contrabassoon_A3_1_mezzo-forte_normal.wav	mp3
			contrabassoon_A3_1_piano_normal.wav	mp3
			contrabassoon_C3_1_forte_normal.wav	mp3
			contrabassoon_C3_1_fortissimo_normal.wav	mp3
			contrabassoon_C3_1_mezzo-forte_normal.wav	mp3
			contrabassoon_C3_1_piano_normal.wav	mp3
6	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"cor-anglais" A3 und B3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	cor-anglais_A3_1_forte_normal.wav	mp3
			cor-anglais_A3_1_fortissimo_normal.wav	mp3
			cor-anglais_A3_1_mezzo-forte_normal.wav	mp3
			cor-anglais_A3_1_mezzo-piano_normal.wav	mp3
			cor-anglais_B3_1_forte_normal.wav	mp3
			cor-anglais_B3_1_fortissimo_normal.wav	mp3
			cor-anglais_B3_1_mezzo-forte_normal.wav	mp3
			cor-anglais_B3_1_mezzo-piano_normal.wav	mp3
7	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"double-bass" A3 und C3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	double-bass_A3_1_forte_arco-normal.wav	mp3
			double-bass_A3_1_fortissimo_arco-normal.wav	mp3
			double-bass_A3_1_mezzo-forte_arco-normal.wav	mp3
			double-bass_A3_1_mezzo-piano_arco-normal.wav	mp3
			double-bass_C3_1_forte_arco-normal.wav	mp3
			double-bass_C3_1_fortissimo_arco-normal.wav	mp3
			double-bass_C3_1_mezzo-forte_arco-normal.wav	mp3
			double-bass_C3_1_mezzo-piano_arco-normal.wav	mp3
8	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"flute" A4 und E4. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	flute_A4_1_forte_normal.wav	mp3
			flute_A4_1_mezzo-forte_normal.wav	mp3
			flute_A4_1_mezzo-piano_normal.wav	mp3
			flute_A4_1_pianissimo_normal.wav	mp3
			flute_E4_1_forte_normal.wav	mp3
			flute_E4_1_mezzo-forte_normal.wav	mp3
			flute_E4_1_mezzo-piano_normal.wav	mp3
			flute_E4_1_pianissimo_normal.wav	mp3
9	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"french-horn" A4 und B4. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	french-horn_A4_15_mezzo-forte_normal.wav	mp3
			french-horn_A4_15_piano_normal.wav	mp3
			french-horn_A4_1_mezzo-forte_normal.wav	mp3
			french-horn_A4_1_piano_normal.wav	mp3
			french-horn_B4_15_fortissimo_normal.wav	mp3
			french-horn_B4_15_mezzo-forte_normal.wav	mp3
			french-horn_B4_1_fortissimo_normal.wav	mp3
			french-horn_B4_1_mezzo-forte_normal.wav	mp3
10	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"oboe" A4 und C4. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	oboe_A4_1_forte_normal.wav	mp3
			oboe_A4_1_fortissimo_normal.wav	mp3
			oboe_A4_1_mezzo-forte_normal.wav	mp3
			oboe_A4_1_piano_normal.wav	mp3
			oboe_C4_1_forte_normal.wav	mp3
			oboe_C4_1_fortissimo_normal.wav	mp3
			oboe_C4_1_mezzo-forte_normal.wav	mp3
			oboe_C4_1_piano_normal.wav	mp3
11	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"saxophone" A3 und C4. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	saxophone_A3_1_forte_normal.wav	mp3
			saxophone_A3_1_fortissimo_normal.wav	mp3
			saxophone_A3_1_pianissimo_normal.wav	mp3
			saxophone_A3_1_piano_normal.wav	mp3
			saxophone_C4_1_forte_normal.wav	mp3
			saxophone_C4_1_fortissimo_normal.wav	mp3
			saxophone_C4_1_mezzo-forte_normal.wav	mp3
			saxophone_C4_1_pianissimo_normal.wav	mp3
12	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"trombone" A4 und C4. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	trombone_A4_1_forte_normal.wav	mp3
			trombone_A4_1_fortissimo_normal.wav	mp3
			trombone_A4_1_mezzo-forte_normal.wav	mp3
			trombone_A4_1_pianissimo_normal.wav	mp3
			trombone_C4_1_forte_normal.wav	mp3
			trombone_C4_1_fortissimo_normal.wav	mp3
			trombone_C4_1_mezzo-forte_normal.wav	mp3
			trombone_C4_1_pianissimo_normal.wav	mp3
13	<a href="https://www.philharmonia.co.uk/explore/sound_samples">https://www.philharmonia.co.uk/explore/sound_samples</a>	"tuba" A3 und C3. Minimale Klangfarbenunterschiede durch unterschiedliche Spielarten	tuba_A3_1_forte_normal.wav	mp3
			tuba_A3_1_fortissimo_normal.wav	mp3
			tuba_A3_1_mezzo-forte_normal.wav	mp3
			tuba_A3_1_mezzo-piano_normal.wav	mp3
			tuba_C3_1_forte_normal.wav	mp3
			tuba_C3_1_fortissimo_normal.wav	mp3
			tuba_C3_1_mezzo-forte_normal.wav	mp3
			tuba_C3_1_mezzo-piano_mute.wav	mp3

## 8.2.7 M2

GruppennQuelle	Beschreibung	Dateinamen	Dateiform	
1	<a href="https://freesound.org/people/modularsamples/packs/17184/">https://freesound.org/people/modularsamples/packs/17184/</a>	Akai ax 80. Oktave c2 - c3	278500_modularsamples_akai-ax80-diplodocus-c2-36-ewp6	aiff
			278506_modularsamples_akai-ax80-diplodocus-d2-38-ezud	aiff
			278512_modularsamples_akai-ax80-diplodocus-e2-40-nyla	aiff
			278518_modularsamples_akai-ax80-diplodocus-f2-42-ioef	aiff
			278524_modularsamples_akai-ax80-diplodocus-g2-44-0osp	aiff
			278527_modularsamples_akai-ax80-diplodocus-a2-45-kyhw	aiff
			278533_modularsamples_akai-ax80-diplodocus-b2-47-1t6y	aiff
			278536_modularsamples_akai-ax80-diplodocus-c3-48-mqgh	aiff
			279418_modularsamples_akai-ax80-layered-c2-36-1z9h	aiff
			279426_modularsamples_akai-ax80-layered-d2-38-3g4m	aiff
			279434_modularsamples_akai-ax80-layered-e2-40-frfd	aiff
			279442_modularsamples_akai-ax80-layered-f2-42-wh7u	aiff
			279450_modularsamples_akai-ax80-layered-g2-44-r8ll	aiff
			279455_modularsamples_akai-ax80-layered-a2-45-c0a	aiff
279462_modularsamples_akai-ax80-layered-b2-47-j8d6	aiff			
279466_modularsamples_akai-ax80-layered-c3-48-v3qr	aiff			
3	<a href="https://freesound.org/people/modularsamples/packs/17241/">https://freesound.org/people/modularsamples/packs/17241/</a>	Arturia microbrute. Oktave c1 - c2	281325_modularsamples_arturia-microbrute-dynamic-square-c1-24c0-yneu	aiff
			281333_modularsamples_arturia-microbrute-dynamic-square-d1-26d0-dffm	aiff
			281341_modularsamples_arturia-microbrute-dynamic-square-e1-28e0-4ys0	aiff
			281349_modularsamples_arturia-microbrute-dynamic-square-f1-30f-0-cq7a	aiff
			281357_modularsamples_arturia-microbrute-dynamic-square-g1-32g-0-316s	aiff
			281361_modularsamples_arturia-microbrute-dynamic-square-a1-33a0-zb9	aiff
			281369_modularsamples_arturia-microbrute-dynamic-square-b1-35b0-jlwh	aiff
			281373_modularsamples_arturia-microbrute-dynamic-square-c2-36c1-r0to	aiff
			283318_modularsamples_elektron-analog-four-elaborate-75-c3-c2elaborate75-48-64	aiff
			283326_modularsamples_elektron-analog-four-elaborate-75-d3-d2elaborate75-50-64	aiff
			283334_modularsamples_elektron-analog-four-elaborate-75-e3-e2elaborate75-52-64	aiff
			283342_modularsamples_elektron-analog-four-elaborate-75-f3-f2elaborate75-54-64	aiff
			283350_modularsamples_elektron-analog-four-elaborate-75-g3-g2elaborate75-56-64	aiff
			283354_modularsamples_elektron-analog-four-elaborate-75-a3-a2elaborate75-57-64	aiff
283362_modularsamples_elektron-analog-four-elaborate-75-b3-b2elaborate75-59-64	aiff			
283366_modularsamples_elektron-analog-four-elaborate-75-c4-c3elaborate75-60-64	aiff			
5	<a href="https://freesound.org/people/modularsamples/packs/17294/">https://freesound.org/people/modularsamples/packs/17294/</a>	Fender chroma polaris. Oktave c3 - c4	287757_modularsamples_fender-chroma-polaris-nervous-energy-c3-48c2-gjhx	wav
			287773_modularsamples_fender-chroma-polaris-nervous-energy-d3-50d2-41h0	wav
			287789_modularsamples_fender-chroma-polaris-nervous-energy-e3-52e2-laie	wav
			287805_modularsamples_fender-chroma-polaris-nervous-energy-f3-54f-2-10n7	wav
			287821_modularsamples_fender-chroma-polaris-nervous-energy-g3-56g-2-pujh	wav
			287829_modularsamples_fender-chroma-polaris-nervous-energy-a3-57a2-g0ei	wav
			287845_modularsamples_fender-chroma-polaris-nervous-energy-b3-59b2-5lqp	wav
			287853_modularsamples_fender-chroma-polaris-nervous-energy-c4-60c3-rzrb	wav
			293130_modularsamples_korg-mono-poly-details-bass-c3-details-bass-48-32	aiff
			293138_modularsamples_korg-mono-poly-details-bass-d3-details-bass-50-32	aiff
			293146_modularsamples_korg-mono-poly-details-bass-e3-details-bass-52-32	aiff
			293154_modularsamples_korg-mono-poly-details-bass-f3-details-bass-54-32	aiff
			293162_modularsamples_korg-mono-poly-details-bass-g3-details-bass-56-32	aiff
			293166_modularsamples_korg-mono-poly-details-bass-a3-details-bass-57-32	aiff
293174_modularsamples_korg-mono-poly-details-bass-b3-details-bass-59-32	aiff			
293178_modularsamples_korg-mono-poly-details-bass-c4-details-bass-60-32	aiff			
7	<a href="https://freesound.org/people/modularsamples/packs/17462/">https://freesound.org/people/modularsamples/packs/17462/</a>	Oberheim xpander. Otave c4 - c5	300862_modularsamples_oberheim-xpander-sin-pad-c4-sin-pad-60-32	aiff
			300868_modularsamples_oberheim-xpander-sin-pad-d4-sin-pad-62-32	aiff
			300874_modularsamples_oberheim-xpander-sin-pad-e4-sin-pad-64-32	aiff
			300880_modularsamples_oberheim-xpander-sin-pad-f4-sin-pad-66-32	aiff
			300886_modularsamples_oberheim-xpander-sin-pad-g4-sin-pad-68-32	aiff
			300889_modularsamples_oberheim-xpander-sin-pad-a4-sin-pad-69-32	aiff
			300895_modularsamples_oberheim-xpander-sin-pad-b4-sin-pad-71-32	aiff
			300898_modularsamples_oberheim-xpander-sin-pad-c5-sin-pad-72-32	aiff
			303155_modularsamples_roland-jupiter-4-4-lead-c4-jp-organ-60-127-pi17	aiff
			303157_modularsamples_roland-jupiter-4-4-lead-d4-jp-organ-62-127-owz9	aiff
			303159_modularsamples_roland-jupiter-4-4-lead-e4-jp-organ-64-127-kdqq	aiff
			303161_modularsamples_roland-jupiter-4-4-lead-f4-jp-organ-66-127-ynay	aiff
			303163_modularsamples_roland-jupiter-4-4-lead-g4-jp-organ-68-127-lymz	aiff
			303164_modularsamples_roland-jupiter-4-4-lead-a4-jp-organ-69-127-kv96	aiff
303166_modularsamples_roland-jupiter-4-4-lead-b4-jp-organ-71-127-zcvr	aiff			
303167_modularsamples_roland-jupiter-4-4-lead-c5-jp-organ-72-127-4h9z	aiff			
9	<a href="https://freesound.org/people/modularsamples/packs/17507/">https://freesound.org/people/modularsamples/packs/17507/</a>	Roland jupiter. Oktave c2 - c3	304375_modularsamples_roland-jupiter-4-organic-hell-c2-organic-hell-36-127-zeae	aiff
			304377_modularsamples_roland-jupiter-4-organic-hell-d2-organic-hell-38-127-z0yj	aiff
			304379_modularsamples_roland-jupiter-4-organic-hell-e2-organic-hell-40-127-l4ve	aiff
			304381_modularsamples_roland-jupiter-4-organic-hell-f2-organic-hell-42-127-mltz	aiff
			304383_modularsamples_roland-jupiter-4-organic-hell-g2-organic-hell-44-127-47zc	aiff
			304385_modularsamples_roland-jupiter-4-organic-hell-a-2-organic-hell-46-127-e1gg	aiff
			304386_modularsamples_roland-jupiter-4-organic-hell-b2-organic-hell-47-127-d4j2	aiff
			304387_modularsamples_roland-jupiter-4-organic-hell-c3-organic-hell-48-127-85iz	aiff
			306488_modularsamples_roland-jx-3p-inspace-c-4-61c-3-55xd	aiff
			306490_modularsamples_roland-jx-3p-inspace-d-4-63d-3-u8j	aiff
			306491_modularsamples_roland-jx-3p-inspace-e-4-64e3-4xfy	aiff
			306493_modularsamples_roland-jx-3p-inspace-f4-66f-3-23zr	aiff
			306495_modularsamples_roland-jx-3p-inspace-g4-68g-3-6k4l	aiff
			306496_modularsamples_roland-jx-3p-inspace-a-4-69a3-clx0	aiff
306498_modularsamples_roland-jx-3p-inspace-b4-71b3-kanw	aiff			
306499_modularsamples_roland-jx-3p-inspace-c5-72c4-oucx	aiff			
11	<a href="https://freesound.org/people/modularsamples/packs/17548/">https://freesound.org/people/modularsamples/packs/17548/</a>	Roland jx 3p. Oktave c2 - c3	306585_modularsamples_roland-jx-3p-ivor-evan-biggurun-c2-0-1kdf	aiff
			306587_modularsamples_roland-jx-3p-ivor-evan-biggurun-d2-0-1sog	aiff
			306589_modularsamples_roland-jx-3p-ivor-evan-biggurun-e2-0-uoql	aiff
			306591_modularsamples_roland-jx-3p-ivor-evan-biggurun-f2-0-0lww	aiff
			306593_modularsamples_roland-jx-3p-ivor-evan-biggurun-g2-0-8yfn	aiff
			306594_modularsamples_roland-jx-3p-ivor-evan-biggurun-a2-0-8yfn	aiff
			306596_modularsamples_roland-jx-3p-ivor-evan-biggurun-b2-0-8rwp	aiff
			306597_modularsamples_roland-jx-3p-ivor-evan-biggurun-c3-0-664l	aiff
			307352_modularsamples_roland-jx-3p-slowone-c2-36c1-7ykk	aiff
			307354_modularsamples_roland-jx-3p-slowone-d2-38d1-n9g	aiff
			307356_modularsamples_roland-jx-3p-slowone-e2-40e1-pncg	aiff
			307358_modularsamples_roland-jx-3p-slowone-f2-42f-1-oktt	aiff
			307360_modularsamples_roland-jx-3p-slowone-g2-44g-1-ks8y	aiff
			307361_modularsamples_roland-jx-3p-slowone-a2-45a1-tws3	aiff
307362_modularsamples_roland-jx-3p-slowone-a-2-46a-1-166m	aiff			
307364_modularsamples_roland-jx-3p-slowone-c3-48c2-ctp9	aiff			
13	<a href="https://freesound.org/people/modularsamples/packs/17609/">https://freesound.org/people/modularsamples/packs/17609/</a>	Roland sh2. Oktave c5 -c6	311314_modularsamples_roland-sh-2-wow-bass-c5-72c4-xxcu	aiff
			311316_modularsamples_roland-sh-2-wow-bass-d5-74d4-v3r6	aiff
			311318_modularsamples_roland-sh-2-wow-bass-e5-76e4-1k0j	aiff
			311320_modularsamples_roland-sh-2-wow-bass-f5-78f-4-711s	aiff
			311322_modularsamples_roland-sh-2-wow-bass-g5-80g-4-pbry	aiff
			311323_modularsamples_roland-sh-2-wow-bass-a5-81a4-e8ah	aiff
			311325_modularsamples_roland-sh-2-wow-bass-b5-83b4-qxkm	aiff
			311326_modularsamples_roland-sh-2-wow-bass-c6-84c5-krd7	aiff
			315120_modularsamples_yamaha-cs-30l-velocity-bass-c3-velocity-bass-48-32	aiff
			315128_modularsamples_yamaha-cs-30l-velocity-bass-d3-velocity-bass-50-32	aiff
			315136_modularsamples_yamaha-cs-30l-velocity-bass-e3-velocity-bass-52-32	aiff
			315144_modularsamples_yamaha-cs-30l-velocity-bass-f3-velocity-bass-54-32	aiff
			315152_modularsamples_yamaha-cs-30l-velocity-bass-g3-velocity-bass-56-32	aiff
			315156_modularsamples_yamaha-cs-30l-velocity-bass-a3-velocity-bass-57-32	aiff
315164_modularsamples_yamaha-cs-30l-velocity-bass-b3-velocity-bass-59-32	aiff			
315168_modularsamples_yamaha-cs-30l-velocity-bass-c4-velocity-bass-61-32	aiff			
14	<a href="https://freesound.org/people/modularsamples/packs/17680/">https://freesound.org/people/modularsamples/packs/17680/</a>	Yamaha cs 30l. Oktave c3 - c4	315120_modularsamples_yamaha-cs-30l-velocity-bass-c3-velocity-bass-48-32	aiff
			315128_modularsamples_yamaha-cs-30l-velocity-bass-d3-velocity-bass-50-32	aiff
			315136_modularsamples_yamaha-cs-30l-velocity-bass-e3-velocity-bass-52-32	aiff
			315144_modularsamples_yamaha-cs-30l-velocity-bass-f3-velocity-bass-54-32	aiff
			315152_modularsamples_yamaha-cs-30l-velocity-bass-g3-velocity-bass-56-32	aiff
			315156_modularsamples_yamaha-cs-30l-velocity-bass-a3-velocity-bass-57-32	aiff
			315164_modularsamples_yamaha-cs-30l-velocity-bass-b3-velocity-bass-59-32	aiff
			315168_modularsamples_yamaha-cs-30l-velocity-bass-c4-velocity-bass-61-32	aiff

## 8.2.8 P 1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="https://freesound.org/people/Keider/packs/6680/">https://freesound.org/people/Keider/packs/6680/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	104895_keider_beyond-c5 104901_keider_beyond-e5 104906_keider_beyond-g-4 104907_keider_beyond-g-5	wav wav wav wav
2	<a href="https://freesound.org/people/Iovica/packs/7547/">https://freesound.org/people/Iovica/packs/7547/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	120720_jovica_the-real-virus-08-bandpass-vowelpad-c2 120726_jovica_the-real-virus-08-bandpass-vowelpad-e1 120727_jovica_the-real-virus-08-bandpass-vowelpad-e2 120733_jovica_the-real-virus-08-bandpass-vowelpad-g-2	flac flac flac flac
3	<a href="https://freesound.org/people/Iovica/packs/811/">https://freesound.org/people/Iovica/packs/811/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	15128_jovica_airy-pad-c5 15133_jovica_airy-pad-e4 15134_jovica_airy-pad-e5 15139_jovica_airy-pad-g-5	wav wav wav wav
4	<a href="https://freesound.org/people/Corsica_S/packs/1455/">https://freesound.org/people/Corsica_S/packs/1455/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	23357_corsica-s_creatures-c-1 23363_corsica-s_creatures-d-1 23369_corsica-s_creatures-e-1 23375_corsica-s_creatures-f-1	flac flac flac flac
5	<a href="https://freesound.org/people/Corsica_S/packs/1459/">https://freesound.org/people/Corsica_S/packs/1459/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	23458_corsica-s_argonsky-a4 23464_corsica-s_argonsky-b4 23470_corsica-s_argonsky-c4 23476_corsica-s_argonsky-d4	flac flac flac flac
6	<a href="https://freesound.org/people/Tom_McLaughlin/packs/1537/">https://freesound.org/people/Tom_McLaughlin/packs/1537/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	24587_tom-mclaughlin_01-flute-pad-tmc 24588_tom-mclaughlin_02-flute-pad-tmc 24589_tom-mclaughlin_03-flute-pad-tmc 24590_tom-mclaughlin_04-flute-pad-tmc	aif aif aif aif
7	<a href="https://freesound.org/people/gis_sweden/sounds/249176/">https://freesound.org/people/gis_sweden/sounds/249176/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	249173_gis-sweden_120bpmacantholabrus6s-a 249174_gis-sweden_120bpmacantholabrus4s-g 249175_gis-sweden_120bpmacantholabrus6s-g 249176_gis-sweden_120bpmacantholabrus6s-d	wav wav wav wav
8	<a href="https://freesound.org/people/LS/sounds/25008/">https://freesound.org/people/LS/sounds/25008/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	25008_ls_soft-pad-2 25009_ls_soft-pad-3 25010_ls_soft-pad-4 25011_ls_soft-pad	wav wav wav wav
9	<a href="https://freesound.org/people/Jagadamba/packs/15626/">https://freesound.org/people/Jagadamba/packs/15626/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	254686_jagadamba_basspad011-sterео 254688_jagadamba_basspad010-sterео 254692_jagadamba_basspad012-sterео 254702_jagadamba_basspad03-sterео	wav wav wav wav
10	<a href="https://freesound.org/people/Garzul/packs/16452/">https://freesound.org/people/Garzul/packs/16452/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	266801_garzul_minimoog-vibrating-c-4 266806_garzul_minimoog-vibrating-d-4 266810_garzul_minimoog-vibrating-e-4 266812_garzul_minimoog-vibrating-f-4	wav wav wav wav
11	<a href="https://freesound.org/people/modularsamples/packs/17269/">https://freesound.org/people/modularsamples/packs/17269/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	284143_modularsamples_elektron-analog-four-grit-pad-c2-36c1-x5ry 284144_modularsamples_elektron-analog-four-grit-pad-c2-37c-1-862h 284145_modularsamples_elektron-analog-four-grit-pad-d2-38d1-99wf 284146_modularsamples_elektron-analog-four-grit-pad-d-2-39d-1-ipp8	wav wav wav wav
12	<a href="https://freesound.org/people/malmen/packs/18282/">https://freesound.org/people/malmen/packs/18282/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	325574_malmen_pad2-d1 325575_malmen_pad2-c1 325579_malmen_pad2-f1 325580_malmen_pad2-e1	wav wav wav wav
13	<a href="https://freesound.org/people/deleted_user_2195044/packs/17948/">https://freesound.org/people/deleted_user_2195044/packs/17948/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	330147_deleted-user-2195044_subatra-08 330148_deleted-user-2195044_subatra-07 330151_deleted-user-2195044_subatra-10 330152_deleted-user-2195044_subatra-09	wav wav wav wav
14	<a href="https://freesound.org/people/Hoerspielwerkstatt_HEF/packs/21352/">https://freesound.org/people/Hoerspielwerkstatt_HEF/packs/21352/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	380043_hoerspielwerkstatt-hef_ch3-100bpm 380044_hoerspielwerkstatt-hef_ch2-100bpm 380050_hoerspielwerkstatt-hef_ch4-100bpm 380051_hoerspielwerkstatt-hef_ch9-90bpm	wav wav wav wav
15	<a href="https://freesound.org/people/gunbladez/sounds/390454/">https://freesound.org/people/gunbladez/sounds/390454/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	390451_gunbladez_live-space-pad-04 390452_gunbladez_live-space-pad-03 390453_gunbladez_live-space-pad-02 390454_gunbladez_live-space-pad-01	wav wav wav wav
16	<a href="https://freesound.org/people/Hoerspielwerkstatt_HEF/packs/17953/">https://freesound.org/people/Hoerspielwerkstatt_HEF/packs/17953/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	400361_hoerspielwerkstatt-hef_strings-20 400363_hoerspielwerkstatt-hef_strings-17 400367_hoerspielwerkstatt-hef_strings-19 400368_hoerspielwerkstatt-hef_strings-18	wav wav wav wav
17	<a href="https://freesound.org/people/deep%20into%20perspectives/packs/3063/">https://freesound.org/people/deep%20into%20perspectives/packs/3063/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	47945_deep-into-perspectives_layered6 47946_deep-into-perspectives_layered7 47947_deep-into-perspectives_layered8 47948_deep-into-perspectives_layered9	wav wav wav wav
18	<a href="https://freesound.org/people/Iovica/packs/3385/">https://freesound.org/people/Iovica/packs/3385/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	52515_jovica_pad-009-bending-hypnosis-c2 52516_jovica_pad-009-bending-hypnosis-c3 52520_jovica_pad-009-bending-hypnosis-f2 52522_jovica_pad-009-bending-hypnosis-f3	flac flac flac flac
19	<a href="https://freesound.org/people/humphreyswill/packs/3920/">https://freesound.org/people/humphreyswill/packs/3920/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	61122_humphreyswill_texas-ranger-001 61123_humphreyswill_texas-ranger-002 61124_humphreyswill_texas-ranger-003 61125_humphreyswill_texas-ranger-004	aif aif aif aif
20	<a href="https://freesound.org/people/Iovica/packs/346/">https://freesound.org/people/Iovica/packs/346/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	6697_jovica_steampipe-9-saturated-pad-c4 6698_jovica_steampipe-9-saturated-pad-c5 6703_jovica_steampipe-9-saturated-pad-e4 6708_jovica_steampipe-9-saturated-pad-g-4	wav wav wav wav
21	<a href="https://freesound.org/people/Iovica/packs/5785/">https://freesound.org/people/Iovica/packs/5785/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	89662_jovica_layers-13-fryday-30 89663_jovica_layers-13-fryday-31 89664_jovica_layers-13-fryday-32 89665_jovica_layers-13-fryday-33	flac flac flac flac
22	<a href="http://www.musicradar.com/news/tech/sampleradar-382-free-ambient-samples-385417">http://www.musicradar.com/news/tech/sampleradar-382-free-ambient-samples-385417</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	AM_GuitVerber120A-01 AM_GuitVerber120A-02 AM_GuitVerber120A-03 AM_GuitVerber120A-04	wav wav wav wav
23	<a href="http://www.musicradar.com/news/tech/sampleradar-382-free-ambient-samples-385417">http://www.musicradar.com/news/tech/sampleradar-382-free-ambient-samples-385417</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	AM_MellonPump80A-01 AM_MellonPump80A-02 AM_MellonPump80A-03 AM_MellonPump80A-04	wav wav wav wav
24	<a href="http://www.musicradar.com/news/tech/sampleradar-382-free-ambient-samples-385417">http://www.musicradar.com/news/tech/sampleradar-382-free-ambient-samples-385417</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	AM_RhodesVerb120A-01 AM_RhodesVerb120A-02 AM_RhodesVerb120A-03 AM_RhodesVerb120A-04	wav wav wav wav
25	<a href="http://99sounds.org/simply-space/">http://99sounds.org/simply-space/</a>	Synthetisierter Pad Sound. Minimale Klangfarbenunterschiede durch Tonhöhe oder Frequenzanteile	Starrynight_C_1 Starrynight_C_2 Starrynight_C_3 Starrynight_C_4	wav wav wav wav

## 8.2.9 SD 1

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="http://freesound.org/people/altemark/packs/1643/">http://freesound.org/people/altemark/packs/1643/</a>	Auf einer 909 Drum Machine synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	26706_altemark_sn33.wav 26707_altemark_sn34.wav 26708_altemark_sn35.wav 26709_altemark_sn36.wav 26710_altemark_sn37.wav	wav wav wav wav wav
2	<a href="https://freesound.org/people/crispydinner/packs/13248/">https://freesound.org/people/crispydinner/packs/13248/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	208309_crispydinner_ppu-snare-marker-2.wav 208337_crispydinner_ppu-snare-marker-3.wav 208351_crispydinner_ppu-snare-marker-1.wav 208394_crispydinner_ppu-snare-marker-5.wav 208415_crispydinner_ppu-snare-marker-4.wav	wav wav wav wav wav
3	<a href="https://freesound.org/people/duskstep/packs/9877/">https://freesound.org/people/duskstep/packs/9877/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	159683_duskstep_heavy-snare-1.wav 160837_duskstep_heavy-snare-2.wav 160844_duskstep_heavy-snare-4.wav 160862_duskstep_heavy-snare-3.wav 160863_duskstep_heavy-snare-5.wav	wav wav wav wav wav
4	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	CYcDh_SonUpsd-01 CYcDh_SonUpsd-02 CYcDh_SonUpsd-03 CYcDh_SonUpsd-04 CYcDh_SonUpsd-05	wav wav wav wav wav
5	<a href="https://freesound.org/people/sandyrb/packs/5541/">https://freesound.org/people/sandyrb/packs/5541/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	84076_sandyrb_kbsd-special-velocity001 84081_sandyrb_kbsd-special-velocity005 84086_sandyrb_kbsd-special-velocity010 84091_sandyrb_kbsd-special-velocity015 84095_sandyrb_kbsd-special-velocity019	wav wav wav wav wav
6	<a href="https://freesound.org/people/sandyrb/packs/5202/">https://freesound.org/people/sandyrb/packs/5202/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	79752_sandyrb_ino-snare-001-noh 79754_sandyrb_ino-snare-002-noh 79756_sandyrb_ino-snare-003-noh 79758_sandyrb_ino-snare-004-noh 79760_sandyrb_ino-snare-005-noh	wav wav wav wav wav
7	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	CYcDh_LudFlamA-01 CYcDh_LudFlamA-02 CYcDh_LudFlamA-03 CYcDh_LudFlamA-04 CYcDh_LudFlamA-05	wav wav wav wav wav
8	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460</a>	Mikrofonaufnahme eines Rimshoots. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	PearlPiccolo Side SnrOff-01 PearlPiccolo Side SnrOff-02 PearlPiccolo Side SnrOff-03 PearlPiccolo Side SnrOff-04 PearlPiccolo Side SnrOff-05	wav wav wav wav wav
9	<a href="http://freesound.org/people/quartertone/packs/8839/">http://freesound.org/people/quartertone/packs/8839/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	142376_quartertone_sd14x08tama-hp-0sn-hdc-v14 142377_quartertone_sd14x08tama-hp-0sn-hdc-v13 142378_quartertone_sd14x08tama-hp-0sn-hdc-v14 142379_quartertone_sd14x08tama-hp-0sn-hdc-v15 142380_quartertone_sd14x08tama-hp-0sn-hdc-v16	wav wav wav wav wav
10	<a href="http://freesound.org/people/sandyrb/packs/5982/">http://freesound.org/people/sandyrb/packs/5982/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	92245_sandyrb_mpm14x512-d6-velo1 92246_sandyrb_mpm14x512-d6-velo2 92246_sandyrb_mpm14x512-d6-velo2 92248_sandyrb_mpm14x512-d6-velo4 92249_sandyrb_mpm14x512-d6-velo5	wav wav wav wav wav
11	<a href="http://freesound.org/people/quartertone/packs/9411/">http://freesound.org/people/quartertone/packs/9411/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	152110_quartertone_sd13x03-pearl-mp-0sn-hdc-v02 152115_quartertone_sd13x03-pearl-mp-0sn-hdc-v01 152116_quartertone_sd13x03-pearl-mp-0sn-hdc-v04 152117_quartertone_sd13x03-pearl-mp-0sn-hdc-v03 152119_quartertone_sd13x03-pearl-mp-0sn-hdc-v05	wav wav wav wav wav
12	<a href="http://freesound.org/people/quartertone/packs/8839/">http://freesound.org/people/quartertone/packs/8839/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	142528_quartertone_sd14x08tama-hp-tsn-hdc-v04 142529_quartertone_sd14x08tama-hp-tsn-hdc-v03 142530_quartertone_sd14x08tama-hp-tsn-hdc-v02 142531_quartertone_sd14x08tama-hp-tsn-hdc-v01 142535_quartertone_sd14x08tama-hp-tsn-hdc-v05	wav wav wav wav wav
13	<a href="https://freesound.org/people/Suicidity/packs/5326/">https://freesound.org/people/Suicidity/packs/5326/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	81683_suicidity_snare-of-god-wet-030 81684_suicidity_snare-of-god-wet-031 81685_suicidity_snare-of-god-wet-032 81686_suicidity_snare-of-god-wet-033 81687_suicidity_snare-of-god-wet-034	wav wav wav wav wav
14	<a href="http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460">http://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229460</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	CYcDh_TamRim-01 CYcDh_TamRim-02 CYcDh_TamRim-03 CYcDh_TamRim-04 CYcDh_TamRim-05	wav wav wav wav wav
15	<a href="http://freesound.org/people/Tr4ck3r/packs/18998/">http://freesound.org/people/Tr4ck3r/packs/18998/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	335921_tr4ck3r_xtrullor-snare-04-e 335922_tr4ck3r_xtrullor-snare-03-c 335923_tr4ck3r_xtrullor-snare-02-f 335924_tr4ck3r_xtrullor-snare-01-d 335928_tr4ck3r_xtrullor-snare-05-g	wav wav wav wav wav
16	<a href="http://freesound.org/people/Veiler/packs/13345/">http://freesound.org/people/Veiler/packs/13345/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Samples aus denen die Snare zusammengesetzt wurden	209880_veiler_hi-snare-small 209881_veiler_ez-bfd-smaller 209882_veiler_ez-bfd-small 209883_veiler_ez-bfd-mix-snare 209884_veiler_snare-2-smaller	wav wav wav wav wav
17	<a href="http://freesound.org/people/Walter_Odington/packs/1580/">http://freesound.org/people/Walter_Odington/packs/1580/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	25656_walter-odington_blake-hall-snare-1 25657_walter-odington_blake-hall-snare-2 25658_walter-odington_blake-hall-snare-3 25659_walter-odington_blake-hall-snare-4 25660_walter-odington_blake-hall-snare-5	aif aif aif aif aif
18	<a href="http://freesound.org/people/robberisurp/packs/189/">http://freesound.org/people/robberisurp/packs/189/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	3144_robberisurp_dm-snare-bb5-5-close 3145_robberisurp_dm-snare-bb5-5-close 3146_robberisurp_dm-snare-tama-steel6 3147_robberisurp_dm-snare-ringger 3148_robberisurp_dm-snare-sizzle	wav wav wav wav wav
19	<a href="https://freesound.org/people/ErrorCell/packs/2342/">https://freesound.org/people/ErrorCell/packs/2342/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	36761_errorcell_sky-snare001 36762_errorcell_sky-snare002 36764_errorcell_sky-snare004 36765_errorcell_sky-snare005 36766_errorcell_sky-snare006	aif aif aif aif aif
20	<a href="https://freesound.org/people/Suicidity/packs/5493/">https://freesound.org/people/Suicidity/packs/5493/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	83573_suicidity_fat-snare-of-god-001 83574_suicidity_fat-snare-of-god-002 83575_suicidity_fat-snare-of-god-003 83576_suicidity_fat-snare-of-god-004 83577_suicidity_fat-snare-of-god-005	wav wav wav wav wav

## 8.2.10 SD 2

Gruppennummer	Quelle	Beschreibung	Dateinamen	Dateiformat (Download)
1	<a href="https://freesound.org/people/crispydinner/packs/7430/">https://freesound.org/people/crispydinner/packs/7430/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	118303_crispydinner_lr-sn-0020	wav
			118304_crispydinner_lr-sn-0021	wav
			118305_crispydinner_lr-sn-0022	wav
			118306_crispydinner_lr-sn-0023	wav
			118307_crispydinner_lr-sn-0024	wav
2	<a href="https://freesound.org/people/quartertone/packs/8858/">https://freesound.org/people/quartertone/packs/8858/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Mikrofonierung und Dämpfung der Snare	143542_quartertone_sd14x08tama-mlp-tsn-rs-v01	wav
			143543_quartertone_sd14x08tama-mlp-tsn-rs-v08	wav
			143549_quartertone_sd14x08tama-mlp-tsn-rs-v12	wav
			143553_quartertone_sd14x08tama-mlp-tsn-rs-v16	wav
			152087_quartertone_sd13x03-pearl-mp-0sn-hdc-v09	wav
			152110_quartertone_sd13x03-pearl-mp-0sn-hdc-v02	wav
			152117_quartertone_sd13x03-pearl-mp-0sn-hdc-v03	wav
			152161_quartertone_sd13x03-pearl-mp-0sn-hdc-v01	wav
			152197_quartertone_sd13x03-pearl-mp-0sn-rs-v13	wav
			160863_duskstep_heavy-snare-5	wav
3	<a href="https://freesound.org/people/quartertone/packs/9411/">https://freesound.org/people/quartertone/packs/9411/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Mikrofonierung und Dämpfung der Snare	160865_duskstep_heavy-snare-7	wav
			160866_duskstep_heavy-snare-6	wav
			160867_duskstep_heavy-snare-9	wav
			160868_duskstep_heavy-snare-8	wav
			208340_crispydinner_ppu-snare-marker-101	wav
4	<a href="https://freesound.org/people/duskstep/packs/9877/">https://freesound.org/people/duskstep/packs/9877/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	208341_crispydinner_ppu-snare-marker-100	wav
			208343_crispydinner_ppu-snare-marker-106	wav
			208344_crispydinner_ppu-snare-marker-105	wav
			208345_crispydinner_ppu-snare-marker-104	wav
			209880_veiler_hi-snare-small	wav
5	<a href="https://freesound.org/people/crispydinner/packs/13248/">https://freesound.org/people/crispydinner/packs/13248/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	209883_veiler_ez-bfd-mix-snare	wav
			209886_veiler_snare-2-fat	wav
			209888_veiler_snare-3-trr	wav
			209889_veiler_snare-2	wav
			212205_alexthegr81_tapesnare	wav
6	<a href="http://freesound.org/people/Veiler/packs/13345/">http://freesound.org/people/Veiler/packs/13345/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Samples aus denen die Snare zusammengesetzt wurden	212208_alexthegr81_tapesnare-15	wav
			212211_alexthegr81_tapesnare-18	wav
			212230_alexthegr81_tapesnare-20	wav
			264784_johnthewizar_snare-10	wav
			264785_johnthewizar_snare-1	wav
7	<a href="https://freesound.org/people/alexthegr81/packs/13483/">https://freesound.org/people/alexthegr81/packs/13483/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	264787_johnthewizar_snare-15	wav
			264798_johnthewizar_snare-5	wav
			264801_johnthewizar_snare-22	wav
			271271_soneproject_snare-1	wav
			271275_soneproject_snare-3	wav
8	<a href="https://freesound.org/people/johnthewizar/packs/16315/">https://freesound.org/people/johnthewizar/packs/16315/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Mikrofonierung und Dämpfung der Snare	271276_soneproject_snare-2	wav
			271277_soneproject_snare-5	wav
			271278_soneproject_snare-4	wav
			290601_modularsamples_fender-chroma-polaris-storm-drum-e-5-77f4-dqf1	wav
			290633_modularsamples_fender-chroma-polaris-storm-drum-f-5-79g4-p6aw	wav
9	<a href="https://freesound.org/people/soneproject/packs/16734/">https://freesound.org/people/soneproject/packs/16734/</a>	Aus mehreren Samps zusammengesetzte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Samples aus denen die Snare zusammengesetzt wurden	290649_modularsamples_fender-chroma-polaris-storm-drum-g5-80g-4-pg08	wav
			290694_modularsamples_fender-chroma-polaris-storm-drum-b5-82b4-vvqx	wav
			290714_modularsamples_fender-chroma-polaris-storm-drum-c6-84c5-akm0	wav
			3144_robbsiesurp_dm-snare-bb5-5-close	wav
			3145_robbsiesurp_dm-snare-bb6-5-close	wav
10	<a href="https://freesound.org/people/modularsamples/packs/17318/">https://freesound.org/people/modularsamples/packs/17318/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	3146_robbsiesurp_dm-snare-tama-steel6	wav
			3147_robbsiesurp_dm-snare-ringger	wav
			3148_robbsiesurp_dm-snare-sizzle	wav
			328508_modularsamples_modular-hits-06-35	aiff
			328509_modularsamples_modular-hits-06-33	aiff
11	<a href="https://freesound.org/people/robbsiesurp/packs/189/">https://freesound.org/people/robbsiesurp/packs/189/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	328510_modularsamples_modular-hits-06-32	aiff
			328511_modularsamples_modular-hits-06-31	aiff
			328512_modularsamples_modular-hits-06-43	aiff
			335921_tr4ck3r_xtrullor-snare-04-e	wav
			335922_tr4ck3r_xtrullor-snare-03-c	wav
12	<a href="https://freesound.org/people/alexthegr81/packs/18417/">https://freesound.org/people/alexthegr81/packs/18417/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	335923_tr4ck3r_xtrullor-snare-02-f	wav
			335924_tr4ck3r_xtrullor-snare-01-d	wav
			335928_tr4ck3r_xtrullor-snare-05-g	wav
			36761_errorcell_sky-snare001	aif
			36765_errorcell_sky-snare005	aif
13	<a href="http://freesound.org/people/Tr4ck3r/packs/18998/">http://freesound.org/people/Tr4ck3r/packs/18998/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	36770_errorcell_sky-snare010	aif
			36775_errorcell_sky-snare015	aif
			36780_errorcell_sky-snare020	aif
			79736_sandyrb_brz-snare-001-noh	wav
			79738_sandyrb_brz-snare-002-noh	wav
14	<a href="https://freesound.org/people/ErrorCell/packs/2342/">https://freesound.org/people/ErrorCell/packs/2342/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Syntheseinstellungen	79740_sandyrb_brz-snare-003-noh	wav
			79742_sandyrb_brz-snare-004-noh	wav
			79744_sandyrb_brz-snare-005-noh	wav
			81654_suicidity_snare-of-god-wet-001	wav
			81663_suicidity_snare-of-god-wet-010	wav
15	<a href="https://freesound.org/people/sandyrb/packs/5202/">https://freesound.org/people/sandyrb/packs/5202/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte und Mikrofonierung	81673_suicidity_snare-of-god-wet-020	wav
			81683_suicidity_snare-of-god-wet-030	wav
			81688_suicidity_snare-of-god-wet-035	wav
			82725_sandyrb_12x8-snare-06-01	wav
			82728_sandyrb_12x8-snare-e22-01	wav
16	<a href="https://freesound.org/people/Suicidity/packs/5326/">https://freesound.org/people/Suicidity/packs/5326/</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Effekteinstellungen	82731_sandyrb_12x8-snare-m201-01	wav
			82734_sandyrb_12x8-snare-sm57-01	wav
			82736_sandyrb_12x8-snare-sm57-03	wav
			92488_sandyrb_tfu213x3-atm250-velo02	wav
			92529_sandyrb_tfu213x3-e22-velo03	wav
17	<a href="https://freesound.org/people/sandyrb/packs/5427/">https://freesound.org/people/sandyrb/packs/5427/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte und Mikrofonierung	92554_sandyrb_tfu213x3-m203-velo08	wav
			92562_sandyrb_tfu213x3-nd868-velo06	wav
			92596_sandyrb_tfu213x3-sm7b-velo10	wav
			CYcDh_LudSnrOffc-01	wav
			CYcDh_LudSnrOffc-03	wav
18	<a href="https://freesound.org/people/sandyrb/packs/5988/">https://freesound.org/people/sandyrb/packs/5988/</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	CYcDh_LudSnrOffc-05	wav
			CYcDh_LudSnrOffc-06	wav
			CYcDh_LudSnrOffc-08	wav
			D_SNARE_01	wav
			E_SNARE_01	wav
19	<a href="p://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229">p://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229</a>	Mikrofonaufnahme. Minimale Klangfarbenunterschiede durch unterschiedliche Schlaghärte	F_SNARE_1	wav
			G_SNARE_1	wav
			H_SNARE_1	wav
			I_SNARE_1	wav
			J_SNARE_1	wav
20	<a href="p://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229">p://www.musicradar.com/news/drums/sampleradar-1000-free-drum-samples-229</a>	Synthetisierte Snaredrum. Minimale Klangfarbenunterschiede durch unterschiedliche Grundfrequenzen	K_SNARE_1	wav
			L_SNARE_1	wav
			M_SNARE_1	wav
			N_SNARE_1	wav
			O_SNARE_1	wav

