
Bachelorarbeit

Herr
Johannes Otto Fritsch

KI zur Verstärkeremulation:
Vergleich von Gitarren-
Röhrenverstärkern mit KI-
generierten digitalen Abbildern
von hochverzerrten Sounds

2023

Fakultät **Medien**

Bachelorarbeit

KI zur Verstärkeremulation: Vergleich von Gitarren- Röhrenverstärkern mit KI- generierten digitalen Abbildern von hochverzerrten Sounds

Autor:

Johannes Otto Fritsch

Studiengang:

Media and Acoustical Engineering

Seminargruppe:

MG18wA-B

Erstprüfer:

Prof. Dr.-Ing. Michael Hösel

Zweitprüfer:

Prof. Dr.-Ing. Christian Roschke

Mittweida, Dezember 2023

Bibliographische Angaben

Fritsch, Otto:

KI zur Verstärkeremulation: Vergleich von Gitarren-Röhrenverstärkern mit KI-generierten digitalen Abbildern von hochverzerrten Sounds

AI for amplifier emulation: Comparison between guitar tube amplifiers and AI-generated digital replicas of highly distorted sounds

59 Seiten, Hochschule Mittweida, University of Applied Sciences,
Fakultät Medien, Bachelorarbeit, 2023

Abstract

In dieser Arbeit wird mittels eines qualitativen Vergleichs untersucht, ob KI-basierte Modeling-Gitarrenverstärker in der Lage sind, einen Röhrenverstärker klanglich exakt zu emulieren. Dazu werden Emulationen von zwei verschiedenen Röhrenverstärkern erstellt, um diese, als auch die Röhrenverstärker mit denselben Stimuli eines Gitarrensings zu speisen. Aus den Signalen soll ein Differenzsignal gebildet werden, welches lediglich Signalanteile beinhaltet, in welchem sich Röhrenverstärker und Emulation unterscheiden. Die Signale werden hinsichtlich der Parameter Dynamik, Spektralverhalten und Timbre auf Unterschiede untersucht. Die Resultate werden analysiert und es wird gezeigt, dass minimale messbare Unterschiede existieren und worin diese liegen.

I. Inhaltsverzeichnis

I.	Inhaltsverzeichnis	I
II.	Abkürzungsverzeichnis	III
III.	Abbildungsverzeichnis	IV
1	Einleitung.....	1
2	Gitarrenverstärkung.....	3
2.1	Grundlagen der Elektrotechnik	3
2.1.1	Audiosignale	3
2.1.2	Bauelemente	4
2.2	Grundlagen von Verzerrung	6
2.2.1	Lineare Verzerrungen.....	7
2.2.2	Nichtlineare Verzerrungen.....	7
2.2.3	Verzerrung von Gitarrensignalen.....	9
2.3	Gitarrenverstärker.....	10
2.3.1	Röhrenverstärker	11
2.3.2	Transistorverstärker	16
2.3.3	Der „Röhrensound“	17
2.3.4	Modeling-Verstärker	18
2.4	Lautsprecherkabinette und Mikrofonierung.....	19
3	Künstliche Intelligenz und deren Anwendung zur Verstärker-Emulation	21
3.1	Grundlagen maschinellen Lernens	22
3.1.1	Entwurf eines maschinellen Lernsystems.....	22
3.1.2	Unterscheidung maschineller Lernsysteme	24
3.2	Deep Learning	25
3.3	Künstliche Neuronale Netze	25
3.3.1	Aufbau und Funktionsweise	26
3.3.2	Netztopologien	27
3.4	Emulation von Röhrenverstärkern mittels künstlicher neuronaler Netze ...	28
3.4.1	Relevante Modelle	29
3.4.2	Konzepte und Training	32
3.4.3	Evaluation der Ergebnisse	35
3.4.4	Kommerzielle Angebote	37
4	Parameter der musikalischen Audioanalyse.....	39
4.1	Dynamik	39
4.1.1	Lautstärke	39

4.1.2	Schwingungsphasen	39
4.2	Spektralverhalten.....	40
4.2.1	Tonhöhe und Frequenz	41
4.2.2	Fourier-Transformation.....	41
4.3	Timbre	42
4.3.1	Mel-Spektrum.....	42
4.3.2	Mel-Frequenz-Cepstral-Koeffizienten (MFCC).....	42
5	Methodik	44
5.1	Versuchsaufbau	44
5.1.1	Wahl der Stimuli	45
5.1.2	Versuchsdurchführung	45
5.1.3	Nachbereitung.....	46
5.2	Parameter und deren Darstellung.....	47
6	Ergebnisse.....	48
6.1	Analyse der Messergebnisse.....	48
6.1.1	Stimulus 1: Ausklingender Akkord	48
6.1.2	Stimulus 2: Akkordwechsel.....	50
6.1.3	Stimulus 3: Arpeggio	51
6.1.4	Stimulus 4: Kurzes Riff	53
6.1.5	Stimulus 5: Kurzes Solo	53
6.2	Auswertung der Ergebnisse.....	55
6.2.1	Zusammenfassung der Analyse	55
6.2.2	Fazit	57
6.3	Ausblick auf weitere Forschungen	58
IV.	Literaturverzeichnis	VI
V.	Anhang.....	IX
	Anhangsverzeichnis	IX
VI.	Eigenständigkeitserklärung	XXXV

II. Abkürzungsverzeichnis

CNN	<i>Convolutional neural network</i>
dBFS	<i>Decibels relative to full scale</i>
DSP	<i>digitaler Signalprozessor</i>
DSV	<i>digitale Signalverarbeitungssysteme</i>
ESR	<i>Error-to-Signal Ratio</i>
HiFi	<i>High Fidelity</i>
LMS	<i>Last-Mean-Square</i>
LSTM	<i>Long short-term memory</i>
MAE	<i>Mean Absolute Error</i>
MFCC	<i>Mel-Frequenz-Cepstral-Koeffizienten</i>
MSE	<i>Mean Squared Error</i>
MUSHRA	<i>Multiple stimuli with hidden reference and anchor</i>
NARX	<i>Nonlinear autoregressive exogenous model</i>
PA	<i>Public Adress</i>
RMSE	<i>Root Mean Square Error</i>

III. Abbildungsverzeichnis

Abbildung 1: Summen- und Differentöne zweier Sinustöne bei Intermodulationsverzerrung (Dickreiter: Handbuch der Tonstudioteknik, S. 622).....	9
Abbildung 2: Eine simple Vorverstärkerstufe (Lähdevaara, J.: The science of electric guitars and guitar electronics, S. 489).....	12
Abbildung 3: Filterschaltung AA763 (Zollner, M.: Physik der Elektrogitarre, S. 819)	13
Abbildung 4: Typische Kathodyn-Schaltung mit erzeugten Ausgangsspannungen (Dailey, D. J.: Electronics for Guitarists, S. 212)	14
Abbildung 5: Endstufe der Klasse A (Lähdevaara, J.: The science of electric guitars and guitar electronics, S. 504).....	15
Abbildung 6: Gegentakt-Endstufe (Lähdevaara, J.: The science of electric guitars and guitar electronics, S. 505).....	16
Abbildung 7: Ordnung der KI-betreffenden Begriffe nach Ralf T. Kreuzer und Marie Sirrenberg (Kreuzer, R; Sirrenberg, M.: Künstliche Intelligenz verstehen, S. 4).....	21
Abbildung 8: Schematische Darstellung einer Einheit. (Ernst, H. et al: Grundkurs Informatik, S. 801)	26
Abbildung 9: Einige Aktivierungsfunktionen (Ernst, H. et al: Grundkurs Informatik, S. 801)	27
Abbildung 10: Beispiel einer Faltung. B sind numerische Eingabedaten eines Bildes, M der Faltungsfilter und B*M die resultierende sogenannte Aktivierungskarte (Ernst, H. et al.: Grundkurs Informatik, S. 818)	30
Abbildung 11: LSTM-Block mit einer Zelle. Die schwarz gefüllten Kreise stellen Gewichtungen dar. In dieser Form wird der Block zwischen die Einheiten zweier Schichten geschaltet (Graves, A.: Supervised sequence labelling with recurrent neural networks, S. 39).....	32
Abbildung 12: Struktur des verwendeten NARX-Netzwerks (Covert, J.; Livingston, D. L.: A vacuum-tube guitar amplifier model using a recurrent neural network, S. 3).....	33
Abbildung 13: Topologie der LSTM-Methode (Schmitz, T.; Embrechts, J. J.: Nonlinear Real-Time Emulation of a Tube Amplifier with a Long Short Term Memory Neural-Network)...	33
Abbildung 14: Modell des Faltungsnetzes (Damskäg et al.: Deep Learning for Tube Amplifier Emulation, S. 2).....	34
Abbildung 15: Die Amplitude eines Signals in Abhängigkeit von der Zeit unterteilt in Attack, Decay, Sustain und Release (Devi, S et al.: A Study: Analysis of Music Features for Musical Instrument Recognition and Music Similarity Search In: International Journal of Computer Science and Informatics, S. 182).....	40
Abbildung 16: Versuchsaufbau.....	44

Abbildung 17: Versuchsaufbau mit JCM900	46
Abbildung 18: Obertonverlust während Releasephase beim JCM900 (Oben Röhrenverstärker, unten ToneX)	50
Abbildung 19: Differenzsignal des Revv Amp. Hohe Amplituden bei den Akkordanschlägen sind zu erkennen.....	51
Abbildung 20: Verflachte Kurve des Vergleichssignals im Kontrast zum Referenzsignal .	52
Abbildung 21: Zusätzliche Obertöne im Vergleichssignal (unten) beim JCM900.	54

1 Einleitung

Ein Gitarrenverstärker macht die Gitarre längst nicht nur laut, sondern ist, genauso wie die Lautsprecherbox – das Kabinett - welches den Klang abstrahlt, so maßgeblich für den Klang, dass man sie als zum Instrument zugehörig betrachten kann.¹ Klassischerweise kommen dabei Röhrenverstärker zum Einsatz. Häufig Modelle, die 50 Jahre oder älter sind. Dem Klang eines guten Röhrenverstärkers ist eben nichts hinzuzufügen und grundlegend hat sich an ihnen über die letzten Jahrzehnte auch nichts geändert, was Jim Marshall, Gründer von *Marshall Amplification*, einem der bis heute führenden Hersteller von Gitarrenverstärkern, meint, wenn er sagt:

*„Since 1962, we've basically made the same amplifier. There's hardly any difference. What we do is about getting the Marshall sound“.*²

Doch dass der Klang grundlegend keiner Verbesserung bedarf, bedeutet nicht, dass die Entwicklung stillsteht. Während moderne Röhrenverstärker sich von ihren Vorgängern meist lediglich durch höhere Leistung und größeren Funktionsumfang unterscheiden,³ schreitet die Entwicklung besonders im Digitalen voran. Bereits seit Beginn der 90er Jahre versuchen sogenannte Modeling-Verstärker den Klang von Röhrenverstärkern mittels digitaler Signalverarbeitung nachzubilden. Der Vorteil ist klar: Röhrenverstärker und Kabinette, welche zu jedem Auftritt und jeder Studio-Session transportiert und vor Ort aufgebaut werden müssen, sind groß und schwer. Zudem sind sie viel zu leistungsstark für den Heimgebrauch. Dazu gesellt sich noch der Wartungsaufwand, die Wartungskosten und die Fragilität der in den Verstärkern verbauten Röhren. All dies entfällt bei Modeling-Verstärkern, welche zur Signalverarbeitung auf Digitaltechnik setzen und somit deutlich leichter zu handhaben sind. Doch dass eben jene Modeling-Verstärker - wurden sie auch über die Jahre immer besser - ihren Vorbildern gegenüber dennoch klanglich das Nachsehen hatten, stand bisher für viele Gitarristen stets außer Frage.⁴

Ein großer Schritt nach vorn gelang 2011 der deutschen Firma *Kemper GmbH* mit dem *Kemper Profiling Amp*, der damit wirbt, den Klang eines analogen Verstärkers vollständig nachzubilden.⁵ Dieser setzt auf das sogenannte Blackbox-Modeling, einer Methode, bei welcher mittels Eingangs- und Ausgangsmessungen an analogen Verstärkern, deren Verhalten erfasst und der Klang emuliert wird (Siehe Kapitel 2.3.4).

Die Performance des Kemper Profiling Amps wurde bereits mehreren Untersuchungen unterzogen, unter anderem auch an der Hochschule Mittweida. Herr Dau verglich in seiner

¹ Vgl. Zollner, M.: Physik der Elektrogitarre, Regensburg 2014, S. 765.

² Maloof, R.: Jim Marshall - The Father of Loud. The Story of the Man Behind the World's Most Famous Guitar Amplifiers, Milwaukee 2004, S. 74.

³ Vgl. Jan-Peter Herbst, Christoph Reuter, Isabella Czedik-Eysenberg: Guitar profiling technology in metal music production: public reception, capability, consequences and perspectives. In: Metal Music Studies (2018), S. 481–506, hier S. 482.

⁴ Vgl. ebd., S. 483.

⁵ Kemper GmbH: Profiler Overview, <https://www.kemper-amps.com/profiler/overview>.

Arbeit⁶ die Emulationen des Kempers mit Referenzverstärkern und kam zum Ergebnis, dass er zwar eine gute Klangqualität liefert, die Emulationen aber immer noch von ihren Referenzen unterscheidbar sind. Ein Fazit, welches auch von anderen Studien^{7, 8} bestätigt wird.

An diesem Punkt setzen nun Produkte wie *das Quad Cortex* von *Neural DSP* oder *ToneX* von *IK Multimedia* an. Anders als beim Kemper Profiling Amp ist künstliche Intelligenz (KI) ihr Ansatz zum Blackbox-Modeling. Beim sogenannten *Capturing* werden Testsignale in das Original-Setup gespeist und am Ende der Signalkette wieder abgenommen. Ein künstliches neuronales Netz lernt dabei das Verhalten des Setups kennen und im Ergebnis entsteht ein digitales Abbild desselbigen, welches „ununterscheidbar vom Original“⁹ sein soll. Auf diese Weise können Setups wie einzelne Effekte, Verstärker oder ganze Signalketten digital nachgebildet werden.

Dieser Prozess kann von Laien ausgeführt werden, beansprucht meist nur wenige Minuten und auch besondere Hardware wird nicht benötigt. Die Emulationen (*Captures* oder *Tone Models* genannt) können gespeichert, jederzeit aufgerufen und es kann schnell zwischen ihnen hin- und hergeschaltet werden. Kombiniert mit den zur Verfügung stehenden Austauschplattformen für die *Captures*, kann man die Gitarre an den PC anschließen und auf sämtliche Setups zugreifen, die man sich vorstellen kann, schnell zwischen ihnen wechseln und sie verwenden, wo man will. Überall kann man jenen perfekten Röhrenverstärkerklang haben, von dem eingangs gesprochen wurde. Zumindest falls das Versprechen wahr ist und jene Emulationen tatsächlich so bahnbrechend sind, wie behauptet wird.

Ob dies der Fall ist, und die *Captures* sich tatsächlich nicht mehr vom Original unterscheiden, ob die endlosen Versuche, einen Röhrenverstärker endlich perfekt digital zu imitieren, also endlich zu einem Ende kommen, soll in dieser Arbeit durch einen qualitativen Vergleich von Referenzsetups mit ihren digitalen Abbildern geklärt werden. Bezogen werden soll sich dabei besonders auf die komplexen hochverzerrten Sounds, welche vordergründig im Rock- und Metalbereich zum Einsatz kommen.

Um dies zu untersuchen, sollen zuerst Grundlagen, einerseits über Gitarrenverstärkung und andererseits über die Grundlagen künstlicher Intelligenz, bereitgestellt werden, um schließlich Studien über Ansätze zur KI-Emulation von Gitarrenverstärkern auszuwerten. Nach Erläuterungen zu einigen Parametern der qualitativen Audioanalyse, soll die Methodik vorgestellt, die Ergebnisse dargestellt und ausgewertet werden, um schließlich eine Antwort auf die Frage zu geben, ob die KI-Technik das Modeling endlich perfektioniert.

⁶ Terence Dau: Qualitativer Vergleich der Performance verschiedener Gitarrenamps mit den entsprechenden Emulationen des Kemper Profiling Amps. Bachelorarbeit, Mittweida 2022 2022.

⁷ Jan-Peter Herbst, Christoph Reuter, Isabella Czedik-Eysenberg: Guitar profiling technology in metal music production: public reception, capability, consequences and perspectives.

⁸ Düvel, N.; Kopiez, R.; Wolf, A.; Weihe, P.: Confusingly Similar: Discerning between Hardware Guitar Amplifier Sounds and Simulations with the Kemper Profiling Amp. In: *Music & Science* (2020), 1-16.

⁹ ToneX, https://www.ikmultimedia.com/products/tonex/#nl_form_area (1. Juni 2023).

2 Gitarrenverstärkung

2.1 Grundlagen der Elektrotechnik

Will man die Funktionsweise und den Aufbau traditioneller sowie moderner Technik zur Signalverarbeitung bei der Gitarrenverstärkung verstehen, sind grundlegende elektronische Kenntnisse vonnöten. Während ein Verständnis von elektrotechnischen und akustischen Grundbegriffen an dieser Stelle vorausgesetzt wird, soll auf Audiosignale und die relevantesten elektronischen Bauelemente eingegangen werden.

2.1.1 Audiosignale

Als Schall werden hörbare mechanische Schwingungen, Wellen und Felder bezeichnet, welche sich in einem elastischen, das meint aus schwingungsfähigen Teilchen bestehenden, Medium – wie Luft, Wasser oder Festkörpern, ausbreiten. Regt eine Schallquelle in ihrer unmittelbaren Umgebung die Teilchen (z.B. Luftteilchen) zum Schwingen an, wird der emittierte Schall durch die Weitergabe der Schwingung auf benachbarte Teilchen übertragen.¹⁰

Durch elektro-akustische Wandler, bspw. Mikrofone oder die Tonabnehmer einer E-Gitarre, können diese schallenergetischen Schwingungen entnommen, in mechanische Schwingung und in daraus resultierende elektrische Energie, umgesetzt werden. Bei einem Mikrofon regt der Luftschall eine Membran zum Schwingen an und die mechanische Schwingung wird in eine proportionale Wechselspannung umgewandelt.¹¹ Auf diesem zugrundeliegenden Mechanismus basierend, gibt es unterschiedliche Wandlerprinzipien, welche an dieser Stelle nicht weiter vertieft werden sollen.

Bei einem solchen, zu einem Schallereignis proportionalen Spannungsverlauf, spricht man von einem Signal, in diesem Fall einem Audiosignal. Also eine durch mathematische Funktionen oder Zahlenfolgen beschriebene, sich verändernde Größe, welche Informationen repräsentiert, hier den Spannungsverlauf eines Mikrofons. Beim beschriebenen Signal handelt es sich um ein analoges Signal, was sich darin auszeichnet, dass es zeit- und wertkontinuierlich ist. Derartige Signale können als Funktion der Zeit oder der Frequenz beschrieben werden. Bei ersterem ergibt sich eine Aneinanderreihung von Zuständen zu einzelnen Zeitpunkten $x(t)$, während die Funktion der Frequenz $X(\omega)$, auch Spektrum genannt (Siehe Kapitel 4.2), als Gewichtungsfunktion verstanden werden kann, mit der harmonische Verläufe, also reine sinusförmige Schwingungen und deren Anteil am Gesamtsignal beschrieben werden. Ihre Summe ergibt wiederum das Zeitsignal.¹²

Da Computer und digitale Signalprozessoren nur digitale Signale verarbeiten können, ist es hierfür erforderlich, das kontinuierliche Signal in eine Folge regelmäßig aufeinander

¹⁰ Vgl. Dickreiter, M.: Handbuch der Tonstudioteknik, 7. Aufl., München 2008, S. 9–10.

¹¹ Vgl. ebd., S. 136–139.

¹² Vgl. Weinzierl, S. (Hrsg.): Handbuch der Audiotechnik, Berlin u. a. 2008, S. 5–6.

folgende Abtastwerte, sogenannte Samples, zu zerlegen. Hierbei spricht man von einem diskreten Signal. Dieses wird gewonnen, indem das kontinuierliche Signal durch den Einsatz einer Sample-and-Hold-Schaltung abgetastet wird. Bei dieser Schaltung wird in den zeitlichen Abständen der Abtastperiode das kontinuierliche Signal durch eine Anordnung von Halbleiterschaltern geschaltet, um dahinterliegende, als Speicher dienende Kondensatoren aufzuladen. Die Abtastfrequenz f_s muss hierbei größer als die doppelte höchste vorkommende Frequenz f_{\max} des Nutzsignals sein, besagt das von Claude Shannon formulierte Abtasttheorem: $f_s > 2f_{\max}$. Der auf die Sample-and-Hold-Schaltung folgende Quantisierer vergleicht die Abtastwerte mit einer gestuften Skala und ordnet dem Signal jeweils Werte zu, welche der Ordnungsnummer der einzelnen Stufen entsprechen. Im Ergebnis erhält man ein zeit- und wertdiskretes Signal. Der beschriebene Gesamtvorgang wird Analog-/Digital-Wandlung (A/D-Wandlung) genannt.¹³

2.1.2 Bauelemente

Für die Verstärkertechnik relevant sind insbesondere die Elektronenröhre, der Transistor und digitale Signalverarbeitungssysteme (DSV). Die beiden Erstgenannten können als Weiterentwicklung des Prinzips einer Diode gesehen werden. Eine Diode ist ein unsymmetrisch aufgebauter Zweipol¹⁴ mit zwei Anschlüssen, welche als Anode (A) und Kathode (K) bezeichnet werden. Sie lässt den Strom nur in eine Richtung (Flussrichtung) durch. In die andere Richtung, die Sperrichtung, ist hingegen kein Stromfluss möglich. Der Stromfluss erfolgt generell nur, wenn eine gewisse Spannung, welche in Flussrichtung der Diode anliegt, überschritten wird. Im Gegensatz zu bspw. Kondensatoren oder Ohm'schen Widerständen handelt es sich bei der Diode um ein nichtlineares Bauelement.¹⁵

Transistoren

Eine Halbleiterdiode besteht aus zwei unterschiedlich dotierten Halbleiter-Schichten: der p- und n-leitenden Schicht. In der n-leitenden Schicht befinden sich überschüssige Elektronen, welche in die p-leitende Schicht diffundieren können, in welcher ein Elektronenmangel herrscht. Im stromlosen Zustand entsteht dadurch eine positive Raumladung am Rand der Sperrschicht im n-Leiter und eine Negative im p-Leiter. Diese Raumladungen werden verstärkt, wenn eine positive Spannung zwischen n- und p-Leiter angelegt wird. Die Ladungsträger werden aus der Sperrschicht verdrängt und der Stromfluss unterbrochen. Wird die Spannung dagegen zwischen p- und n-Leiter angelegt, verstärkt das die Diffusion der Elektronen und der Strom steigt exponentiell mit der angelegten Spannung an.¹⁶

Ein Transistor kann als Diode mit einem zusätzlichen p-n-Übergang verstanden werden. Es gibt pnp- und npn-Transistoren, womit die Reihenfolge der unterschiedlich dotierten Gebiete beschrieben wird.¹⁷ Beispielhaft soll der npn-Transistor beschrieben werden,

¹³ Vgl. Dickreiter: Handbuch der Tonstudioteknik, S. 663–668.

¹⁴ Hering, E.; Bressler, K.; Gutekunst, Jürgen (Hrsg.): Elektronik für Ingenieure und Naturwissenschaftler, 7. Aufl., Berlin u. a. 2017, S. 146.

¹⁵ Vgl. Stiny, L.: Aktive elektronische Bauelemente. Aufbau, Struktur, Wirkungsweise, Eigenschaften und praktischer Einsatz diskreter und integrierter Halbleiter-Bauteile, 2. Aufl., Wiesbaden 2015, S. 71–72.

¹⁶ Vgl. Hering/ Bressler/ Gutekunst: Elektronik für Ingenieure und Naturwissenschaftler, S. 147.

¹⁷ Vgl. Hahn, W.: Elektronik-Praktikum für Informatiker, Berlin 1971, S. 17.

welcher aus drei Elektroden besteht: Dem negativ dotierten Emitter, der positiv dotierten Basis und dem negativ dotierten Kollektor. Wird die Basis-Emitter-Diode an einen kleinen Gleichstrom angeschlossen und der Transistor in eine elektrische Schaltung eingebaut, bringt dieser Basisstrom Ladungsträger in die Basis-Kollektor-Diode, wodurch diese leitfähig wird. Der Basisstrom erzeugt dabei einen deutlich größeren Kollektorstrom, welcher nur wenig von der Kollektor-Emitter-Spannung abhängig ist.¹⁸

Man unterscheidet generell zwischen Bipolartransistoren und Feldeffekttransistoren. Das beschriebene Wirkungsprinzip ist den stromgesteuerten Bipolartransistoren zuzuordnen. Feldeffekttransistoren werden im Gegensatz dazu durch Spannung gesteuert. Beide können eingesetzt werden, um elektrische Signale zu verstärken oder zu schalten.¹⁹

Elektronenröhren

Eine Diode kann auch in Form einer Elektronenröhre gebaut werden. Das auch als Vakuumdiode bezeichnete Bauelement besteht ebenfalls aus einer Kathode, welche durch einen Glühfaden beheizt wird, und einer Anode. Beide sind verbaut in einem luftleeren Glasgefäß. Die beheizte Kathode sendet Elektronen aus, ein Vorgang der als Glühemission bezeichnet wird, wodurch sich vor der Kathode eine Elektronenwolke als negative Raumladung bildet. Besitzt die Anode ein positives Potential gegenüber der Kathode, saugt diese die Elektronen an. Mit steigender Anodenspannung wächst auch der Anodenstrom. Wird eine negative Spannung an die Anode angelegt, fließt bis auf einen geringen Anlaufstrom kein Strom.

Fügt man eine zusätzliche Elektrode zwischen Kathode und Anode, das sogenannte „Gitter“ hinzu, wird die Diode zur Triode. Über die Regelung der negativen Gitterspannung lässt sich der Anodenstrom steuern, da mit zunehmender negativer Gitterspannung die von der Kathode ausgesendeten Elektronen abgestoßen werden. Für den Gebrauch im Hochleistungs- und Hochfrequenzbereich können weitere Gitter ergänzt werden, dann spricht man z.B. von Tetroden und Pentoden, je nach Anzahl der Elektroden.²⁰

Digitale Signalverarbeitungssysteme

Bei DSV handelt es sich um einen Verbund von mehreren Bauteilen, welche die Verarbeitung analoger Signale mittels digitaler Systeme ermöglichen. Das Konzept der digitalen Signalprozessierung hat die Musikbranche und die Gesellschaft im Allgemeinen stark beeinflusst. Digitale Aufzeichnungen, ob von Musik, Film oder Bildern wären ohne DSV nicht möglich, genauso wenig wie die digitale Bearbeitung selbiger.²¹

Umweltsignale, so wie jene, die wir durch unsere Sinnesorgane aufnehmen, wie Licht oder Schall, werden mittels Sensoren erfasst und in elektrische Spannungen umgewandelt. Ein Eingangsverstärker verstärkt die meist sehr schwachen Spannungen und begrenzt ihre Bandbreite auf den Bereich, der von Interesse ist. Das vorbereitete analoge Signal wird von

¹⁸ Vgl. Hering/ Bressler/ Gutekunst: Elektronik für Ingenieure und Naturwissenschaftler, S. 167–168.

¹⁹ Vgl. ebd., S. 167.

²⁰ Vgl. Böhmer, E.; Ehrhardt, D.; Oberschelp, W.: Elemente der angewandten Elektronik. Kompendium für Ausbildung und Beruf : mit umfangreichem Bauteilekatalog, 17. Aufl., Wiesbaden u. a. 2018, S. 96–97.

²¹ Vgl. Tan, L.: Digital Signal Processing. Fundamentals and Applications, s.l. 2007, S. 1.

einer Sample-and-Hold-Schaltung periodisch abgetastet und zeitlich quantisiert, um anschließend auch pegelmäßig quantisiert zu werden (A/D-Wandlung).²² Das dabei entstehende digitale Signal wird von einem digitalen Signalprozessor (DSP) behandelt, entsprechend der Programmierung des DSP. Bei einem DSP kann es sich um einen Mikroprozessor, einen normalen PC oder einen fortgeschrittenen Mikrocontroller handeln. Generell kann der DSP durch Software gesteuert werden. Durch den DSP wird ein bearbeitetes Ausgangssignal generiert, welches durch einen D/A-Wandler und Ausgangsverstärker zurück in ein analoges Signal zurückgewandelt werden kann.²³

Die digitale Signalverarbeitung hat den Vorteil der Langzeitstabilität, da die digitalen Daten nicht altern, und leicht reproduzierbar sind. Durch sie sind viele technische Errungenschaften erst möglich geworden, wie die moderne Bild- und Audiodbearbeitung, Mustererkennung oder moderne Kommunikationstechnik.²⁴

2.2 Grundlagen von Verzerrung

Verzerrungen, also allgemeine Veränderungen eines Signals im zeitlichen Verlauf,²⁵ sind nicht nur grundsätzliche Störungen eines Signals, sondern in der Musik längst eine gewünschte und gewollt herbeigeführte Klangqualität. Bspw. verzerrte Gitarrensignale sind zentrales Element von Rockmusik.²⁶ Dietmar Elflein beschreibt im Fazit zu seinen Untersuchungen über die klanglichen Eigenschaften des Heavy Metal *Schwermetallanalysen*²⁷ die Wirkung verzerrter Signale auf die Musikrichtung wie folgt:

„[...] die Klangfarbe [muss] jedoch bestimmte allgemeine Charakteristika erfüllen, damit ein Tonträger als Traditionsstrom des Heavy Metal gehörig wahrgenommen wird. Das wichtigste Element ist dabei die deutliche Dominanz der verzerrten Gitarrenklangfarbe. [...] Zudem lässt sich auf dem Weg vom Hard Rock über den Classic Metal zum Extreme Metal eine Zunahme des Verzerrungsgrades konstatieren, die neben den Saiteninstrumenten auch den Stimmklang einschließt.“²⁸

Wie eingangs beschrieben, ist Verzerrung prinzipiell eine Abweichung des Signals im zeitlichen Verlauf. Sie liegt also vor, wenn die Wellenform am Ausgang eines Systems nicht identisch zur Wellenform am Eingang ist. Allerdings nur prinzipiell, denn es existieren einige Signaländerungen, die man nicht als Verzerrung versteht. Dazu gehört eine Abweichung der Verstärkung, wenn das Gesamtsignal mit einer konstanten multipliziert wird, da diese die Wellenform nicht verändert und durch Anhebung oder Absenkung des Ausgangssignal

²² Vgl. Bähring, H.: Anwendungsorientierte Mikroprozessoren. Mikrocontroller und Digitale Signalprozessoren, 4. Aufl., Berlin 2010, S. 47–49.

²³ Vgl. Tan: Digital Signal Processing, S. 2.

²⁴ Vgl. Bähring: Anwendungsorientierte Mikroprozessoren, S. 47.

²⁵ Dickreiter: Handbuch der Tonstudiotechnik, S. 617.

²⁶ Vgl. Herbst, J.-P.: Distortion and Rock Guitar Harmony. In: Music Perception 36 (2019), S. 335–352, hier S. 337.

²⁷ Elflein, D.: Schwermetallanalysen. Die musikalische Sprache des Heavy Metal, Bielefeld 2010.

²⁸ Ebd., S. 305.

leicht zu kompensieren ist. Ebenso ist eine Invertierung des Signals keine Verzerrung, da der Frequenzgehalt der gleiche bleibt.²⁹

Verzerrungen entstehen durch Glieder der Signalübertragung, deren Eigenschaften von der Frequenz oder Amplitude abhängig sind. Man unterscheidet zwischen linearen und nichtlinearen Verzerrungen.³⁰

Dieses Kapitel dient dazu, den Einfluss von Verzerrung auf Gitarrensignale und deren Wirkung zu beschreiben. Zuvor sollen aber einige Grundlagen zum Thema dargeboten werden.

2.2.1 Lineare Verzerrungen

Wird ein Signal frequenzabhängig verstärkt oder gedämpft, spricht man von linearer Verzerrung. Dementsprechend können lineare Verzerrungen durch zuvor genannte Kriterien lediglich bei Signalen auftreten, welche komplexer sind als ein Sinussignal. Lineare Verzerrungen können das Klangbild einer Aufnahme beeinflussen, indem sie sie bspw. heller, spitzer, dumpfer, dicker oder präsenter machen. All diese Ausdrücke beziehen sich auf die Veränderung des Höreindrucks eines Signals durch die frequenzabhängige Verstärkung oder Dämpfung.³¹

Werden die Amplituden bestimmter Frequenzen des Signals durch ein Systemglied verstärkt oder gedämpft, spricht man von Amplitudenverzerrung. Die Amplitudenverzerrung, die ein System oder Bauteil an einem Signal bewirkt, bezeichnet man als dessen Frequenzgang.

Ebenfalls von linearer Verzerrung spricht man, wenn ein Signal dispergiert wird, also wenn unterschiedliche Frequenzen mit unterschiedlicher Verzögerung übertragen werden. In diesem Fall handelt es sich um Phasenverzerrung.³²

2.2.2 Nichtlineare Verzerrungen

Wird ein Signal durch amplitudenabhängige Verstärkung oder Dämpfung eines Übertragungsglieds verändert, so spricht man von nichtlinearer Verzerrung.³³ Sie entsteht durch den Betrieb von Bauteilen mit nichtlinearen Kennlinien (z.B. Elektronenröhren oder Transistoren), welche das Signal je nach Grad der Verzerrung mehr oder weniger in der Amplitude beschneiden.³⁴ Aus dieser Beschneidung resultiert die Addition zusätzlicher Frequenzen zum Signal und im Gegenteil zur linearen Verzerrung gibt es keinen einfachen Weg, nichtlineare Verzerrung wieder rückgängig zu machen. Werden die Spitzen des Ausgangssignals scharf abgeschnitten, so dass sie verflachen, spricht man von *hard*

²⁹ Vgl. Hartmann, W. M.: Principles of Musical Acoustics, New York 2013, S. 180.

³⁰ Vgl. Dickreiter: Handbuch der Tonstudientechnik, S. 617.

³¹ Vgl. ebd., S. 617–619.

³² Vgl. Hartmann: Principles of Musical Acoustics, S. 180–181.

³³ Vgl. Weinzierl: Handbuch der Audiotechnik, S. 619.

³⁴ Vgl. ebd., S. 747.

clipping, werden sie sanft abgeflacht handelt es sich um *soft clipping*, dem ein angenehmerer Klang zugesprochen wird.³⁵

Nichtlineare Verzerrung kann in hohem Maße unangenehm klingen, kann die klangliche Charakteristik eines Signals aber auch in deutlich größerem Ausmaß verändern als lineare Verzerrung.³⁶ Die am häufigsten eingesetzte Messgröße für nichtlineare Verzerrung ist der *Klirrfaktor* k , welcher durch ein einfaches Messverfahren ermittelt werden kann. Er gibt an, wie stark die Oberschwingungen U_{nf} , die bei der Verzerrung eines sinusförmigen Signals entstehen, im Vergleich zum Gesamtsignal U_{ges} sind.³⁷

$$k = \frac{\sqrt{U_{2f}^2 + U_{3f}^2 + U_{4f}^2 \dots}}{U_{ges}}$$

Harmonische Verzerrung

Bei nichtlinearen Verzerrungen wird unterschieden zwischen harmonischer Verzerrung und Intermodulationsverzerrung. Wird ein Sinussignal von einem System in der Amplitude beschnitten oder gestreckt, werden harmonische Oberschwingungen generiert, welche das Signal von einem Sinussignal in ein komplexes Signal verwandeln.³⁸ Die entstehenden Oberschwingungen sind hierbei ganzzahlige Vielfache der Grundfrequenz f_1 und werden auch einfach *Harmonische* genannt. Zur Grundfrequenz f_1 , beispielweise 55Hz (erste Harmonische), kommen dann $2f_1=110\text{Hz}$ (zweite Harmonische), $3f_1=165\text{Hz}$ (dritte Harmonische), usw. Dies wird als harmonische Reihe bezeichnet.³⁹ Die Amplituden der einzelnen Harmonischen nehmen dabei mit steigender Ordnungszahl ab. Sogenannte Symmetrische Verzerrungen entstehen durch kubische Kennlinien (eines Systems) und zeichnen sich darin aus, dass sie lediglich ungeradzahlige Harmonische, wie $3f_1$, $5f_1$, usw. beinhalten, während unsymmetrische Verzerrungen durch quadratische Kennlinien erzeugt werden und lediglich geradzahlige Harmonische ($2f_1$, $4f_1$) beinhalten.

Die harmonische Verzerrung eines einzelnen musikalischen Tons stört das Signal relativ wenig, besonders wenn es sich um ein Instrument handelt, welches von sich aus viele Obertöne erzeugt. Wahrgenommen wird sie meist lediglich als Erhöhung der Lautstärke.⁴⁰

Intermodulationsverzerrungen

Bei komplexeren Signalen, wenn mehrere Töne gleichzeitig übertragen werden, wie z.B. ein Akkord, werden nicht nur harmonische Oberschwingungen aus den einzelnen Tönen generiert, sondern zusätzliche Obertöne, die sich aus der Summe und Differenz der Grundfrequenz ergeben, sowie – abhängig vom Verzerrgrad – zwischen den Harmonischen entstehen. Man spricht hierbei von Intermodulationsverzerrung. Bpsw. ein Quintakkord („Powerchord“) C auf einer Gitarre, bestehend aus den Tönen C mit der Grundfrequenz

³⁵ Vgl. Dailey, D. J.: Electronics for Guitarists, New York 2011, S. 141.

³⁶ Vgl. Hartmann: Principles of Musical Acoustics, S. 182.

³⁷ Vgl. Dickreiter: Handbuch der Tonstudientechnik, S. 623–624.

³⁸ Vgl. Hartmann: Principles of Musical Acoustics, S. 183.

³⁹ Vgl. Mynett, M.: Metal Music Manual. Producing, engineering, mixing and mastering contemporary Heavy Music, New York u. a. 2017, S. 10.

⁴⁰ Vgl. Dickreiter: Handbuch der Tonstudientechnik, S. 620–622.

$f_1=65,4\text{Hz}$ und G mit $f_2=98\text{Hz}$ wird - zusätzlich zu den Harmonischen der Grundfrequenzen - ergänzt durch die Summen- und Differenztöne der Grund- und Obertöne. Aus f_1 und f_2 wird also zusätzlich der Summenton $f_2+f_1=163,4\text{Hz}$ und der Differenzton $f_2-f_1=32,6\text{Hz}$ generiert.⁴¹

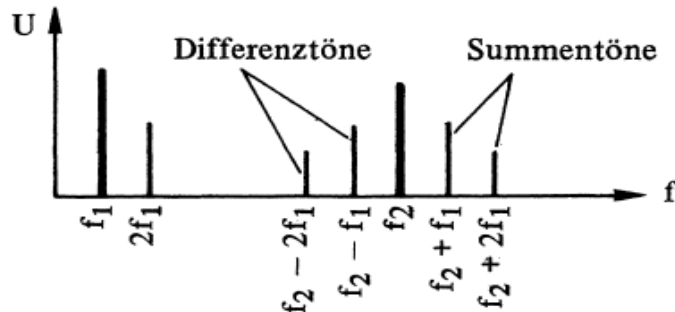


Abbildung 1: Summen- und Differenztöne zweier Sinustöne bei Intermodulationsverzerrung

2.2.3 Verzerrung von Gitarrensensignalen

Für Gitarrensensignale werden Verzerrungen häufig als klangformendes Element eingesetzt. Ein Gitarrenverstärker komprimiert das Gitarrensensignal, wodurch dieses an Dynamik verliert und an Obertönen gewinnt. Das Signal wird geräuschreicher, präsenter und kraftvoller wahrgenommen.⁴² Dieser Effekt erhöhte die allgemeine Präsenz der Gitarre wesentlich und war deshalb essenziell, um die Gitarre zum Lead-Instrument zu entwickeln.⁴³ Das Zustandekommen dieser Änderung der Wahrnehmung des Gitarrensensignals durch Verzerrung soll nachfolgend detailliert erläutert werden.

Es gibt vielfältige Typen von Verzerrern, die für ein Gitarrensensignal zum Einsatz kommen können, wie z.B. *Distortion*, *Overdrive*, *Crunch*, *Fuzz*, etc. Durch die Übersteuerung von Verstärkerstufen, welche bei analogen Verstärkern vorwiegend durch Elektronenröhren oder Transistoren realisiert werden, werden nichtlineare Verzerrungen erzeugt, welche Signalanteile produzieren, die als rau oder geräuschhaft wahrgenommen werden. Der Grad der Verzerrung kann meist gesteuert werden, Bezeichnungen für einen derartigen Regler können *Gain*, *Drive* oder *Boost* sein. Unterschiedliche Verzerrer unterscheiden sich meist durch die Form der genutzten Kennlinie und damit durch unterschiedlich starke Ausprägungen der jeweiligen Obertonklassen, welche wiederum die Klangfarbe maßgeblich beeinflussen.⁴⁴

Durch nichtlineare Verzerrungen werden nicht nur Frequenzen des Ursprungssignals verstärkt oder gedämpft, sondern es werden komplett neue Frequenzen generiert und zum

⁴¹ Vgl. Mynett: Metal Music Manual, S. 10.

⁴² Vgl. Herbst: Distortion and Rock Guitar Harmony, S. 338.

⁴³ Vgl. Herbst, J.: "My setup is pushing about 500 watts - it's all distortion". Emergence, development, aesthetics and intentions of the rock guitar sound, 2019, S. 5.

⁴⁴ Vgl. Weinzierl: Handbuch der Audiotechnik, S. 747.

Signal addiert. Dieser Effekt erhöht einerseits die Dichte an Frequenzen um die Grundschiwingung herum, andererseits wirkt er wesentlich im höheren Bereich des Frequenzspektrums, vor allem im Bereich zwischen 1,5kHz bis 7,5kHz signifikant. Da tiefe Frequenzen feste Objekte leicht durchdringen und somit größere Entfernungen zurücklegen können als hohe Frequenzen, welche sich schnell zerstreuen und absorbieren lassen, erscheinen hochfrequente Geräusche in der Wahrnehmung näher als Tiefe. Dadurch erscheint ein verzerrter Gitarrensound durch seinen erhöhten Gehalt an hohen Frequenzen näher und in Kombination mit der konzentrierten Energie um die Grundschiwingung herum, entsteht ein erhöhter Eindruck von Solidität und Intensität.⁴⁵

Neben dem Einfluss der Verzerrung auf das Frequenzspektrum, hat diese auch einen deutlichen Einfluss auf die Dynamik, welche mit steigendem Verzerrgrad (*Gain*) reduziert wird. Das Signal zeichnet sich durch eine flachere Signalkurve und deutlich mehr Sustain (Nachklingen) aus, wodurch es wiederum lauter erscheint als es tatsächlich ist. Dieser Effekt führt dazu, dass das Ausgangssignal kraftvoller wirkt. Besonders Rhythmus-Gitarren werden hierdurch als härter wahrgenommen.

Die Wirkungen, die Verzerrung auf den Frequenzgehalt und die Dynamik eines Gitarrensignals hat, welche im Allgemeinen mit höherem Verzerrgrad ebenfalls deutlicher werden, bestimmen im Zusammenspiel die Klarheit der gespielten Töne (*Note Definition*). Diese definiert sich dadurch, wie genau der Hörer die Tonhöhe und den Rhythmus von Tönen und Akkorden, welche die Gitarre spielt, im Kontext des gesamten Instrumentariums verorten kann. Sie wird wesentlich vom Verzerrgrad beeinflusst und spiegelt wider, ob der spektrale Gehalt, besonders um die Grundfrequenzen herum, das Auftreten von unmusikalischen Resonanzen und die Dynamik des Signals dem Gespielten und dem entsprechenden Song angemessen sind. Fehlende Note Definition kann durch einen zu hohen Verzerrgrad zustande kommen, welcher einerseits dafür sorgt, dass die Dynamik des Signals zu stark abgeflacht wird. Hierdurch wird der Einsatz der gespielten Noten verwaschen, wodurch es schwierig wird, den Rhythmus präzise zu interpretieren. Zusätzlich kann ein zu hoher Verzerrgrad die Höhen des Gitarrensignals zu stark verstärken, was sich in einem harschen Klangbild manifestiert. Dieses kommt zustande, da die für die Präsenz des Signals wesentlichen Frequenzen, im Allgemeinen zwischen 4kHz-6kHz, überhandnehmen. Ebenfalls kann die Definition unter zu stark ausgeprägten tieffrequenten Anteilen am Signal leiden, welche schnell als Brummen wahrgenommen werden und die gespielten Töne überdecken.⁴⁶

2.3 Gitarrenverstärker

Gitarrenverstärker dienen also bei weitem nicht nur dem Zweck, die wenige Schalleistung, die eine Elektrogitarre abstrahlt zu verstärken, sie nehmen durch ihr frequenzabhängiges Übertragungsmaß und die Verzerrung, die sie generieren, wesentlichen Einfluss auf den Klang der Gitarre. Eben dies macht die Dimensionierung und Analyse der Schaltungen so

⁴⁵ Vgl. Mynett, M.: The Distortion Paradox. In: Brown, A. R.; Spracklen, K.; Kahn-Harris, K.; Scott, N. W. (Hrsg.): Global Metal Music and Culture, New York 2016, S. 68–86, hier S. 74.

⁴⁶ Vgl. ebd., S. 72–77.

kompliziert, da es nicht um objektiv messbare Ergebnisse, sondern um subjektiv wahrgenommene klangliche Charakteristiken geht, welche hier bewertet werden sollen.⁴⁷

Im Allgemeinen bestehen Gitarrenverstärker aus den drei wesentlichen Einheiten Vorverstärker, Klangsteller und Endstufe. Diese können durch weitere Zwischenstufen ergänzt werden. Alle Stufen haben dabei aufeinander Einfluss, ihre Interaktion untereinander und im Einzelnen bestimmt den letztendlichen Klang des Verstärkers.

Man kann grob unterteilen in durch Elektronenröhren betriebene Röhrenverstärker und durch Transistoren betriebene Transistorverstärker. Diese beiden Typen werden ergänzt durch den Modeling-Verstärker, welcher allerdings eine gesonderte Rolle einnimmt, da er digital arbeitet. Dem Röhrenverstärker wird der unübertroffene Klang zugeschrieben, den es mit den alternativen Bauarten nachzueifern gilt. Außerdem werden unterschiedliche Verstärkerklassen unterteilt, wobei es um die Konfiguration der Endstufe geht.⁴⁸ Darauf soll am Ende von Kapitel 2.3.1 näher eingegangen werden.

2.3.1 Röhrenverstärker

Gitarren-Röhrenverstärker sind ein letztes Refugium der Elektronenröhre, deren größter Anteil an der weltweiten Produktion in Gitarrenverstärkern verbaut wird. Die Röhrenverstärker der 50er bis 60er Jahre sind bis heute stilprägend für den Klang der darauffolgenden Verstärkermodelle⁴⁹ und bis heute gelten Röhrenverstärker, trotz zahlreicher Alternativen, als erste Wahl. So hat der Transistor die Röhre zwar in fast allen Bereichen abgelöst, im Bereich Gitarrenverstärker können sie den „vollkommenen“ Klang eines Röhrenverstärkers aber nicht vollständig imitieren.⁵⁰ Aus diesem Grund – dass alle Verstärkertypen versuchen, den Röhrenverstärker zu imitieren - soll dem Aufbau eines Röhrenverstärkers an dieser Stelle besondere Beachtung zugutekommen, indem jede Stufe, also jede um eine Röhre gruppierte Teilschaltung, eines typischen Röhrenverstärkers kurz beleuchtet wird.

Der Vorverstärker

Der Vorverstärker, auch Vorstufe genannt, ist nach dem Gitarrenkabel die erste Station des Signals und dient vor allem der deutlichen Verstärkung des Eingangssignals um das 20-50 fache. Eine typische Vorverstärker-Teilschaltung kann Abbildung 2 entnommen werden.

Typischerweise besteht eine Vorverstärkerstufe aus einer einzelnen Triode in einer Kathoden-Basis-Schaltung. Die Verstärkung wird durch Dimensionierung von R_L unter Berücksichtigung des maximalen Anodenstroms und der angewendeten Betriebsspannung erreicht. Die Regelung der Verstärkung lässt sich an der Verstärkeroberfläche über ein als

⁴⁷ Vgl. Zollner: Physik der Elektrogitarre, S. 765.

⁴⁸ Vgl. Lähdevaara, J.: The science of electric guitars and guitar electronics, Helsinki 2014, S. 479–481.

⁴⁹ Vgl. Barbour, E.: The cool sound of tubes [vacuum tube musical applications]. In: IEEE Spectrum 35 (1998), S. 24–35, hier S. 25.

⁵⁰ Vgl. Dailey: Electronics for Guitarists, S. 195.

Lautstärkereger gekennzeichnetes Potentiometer realisieren, welches in der Schaltung hinter C_L installiert wird.⁵¹

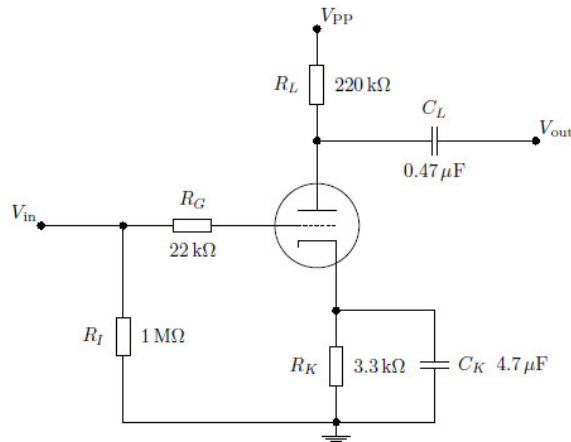


Abbildung 2: Eine simple Vorverstärkerstufe

Der Zwischenverstärker

Für die nach der Vorstufe befindlichen Teilschaltungen existieren unterschiedliche Topologien, welche sich durch die Anordnung von Zwischenverstärker, Klangsteller und Lautstärkereger unterscheiden. Durchgesetzt haben sich vor allem zwei Topologien, bei welchen entweder die Klangsteller und der Lautstärkereger direkt nacheinander vor den Zwischenverstärker geschaltet werden oder bei welcher auf den Lautstärkereger der Zwischenverstärker folgt, an welchen die Klangsteller anschließen.

Der Zwischenverstärker hat die Aufgabe, das durch die Klangsteller oder Lautstärkereger abgeschwächte Signal, erneut zu verstärken. Dies wird standardmäßig durch eine ähnliche Schaltung wie die Vorstufe realisiert. Auch an dieser Stelle wird meist eine einzelne Triode in Kathoden-Basis-Schaltung verwendet.⁵² Ergänzt wird der Zwischenverstärker meist von einer Mischstufe, da die meisten Gitarrenverstärker mehrere Eingänge haben, welche sich durch Klang und Verzerrgrad unterscheiden. Da diese Kanäle alle in derselben Endstufe münden, müssen sie gemischt, das heißt in der Praxis meist addiert, werden.⁵³

Klangsteller

Im Standardfall bestehen die Klangsteller aus separaten Reglern für Höhen, Mitten und Tiefen, welche über die Oberfläche des Verstärkers eingestellt werden können. Die beispielhafte Schaltung aus Abbildung 3 zeigt eine Realisierung, bei welcher drei Hoch- bzw. Tiefpassfilter übereinandergeschichtet sind. Die Kapazität des Kondensators bestimmt jeweils die Grenzfrequenz, so dass das Signal in drei Wege aufgeteilt und am

⁵¹ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 488.

⁵² Vgl. Zollner: Physik der Elektrogitarre, S. 800–801.

⁵³ Vgl. ebd., S. 810.

Ende addiert wird. Wichtig zu beachten ist, dass die einzelnen Abschnitte alle miteinander verbunden sind, wodurch sie sich alle gegenseitig beeinflussen.⁵⁴

Bei Abbildung 3 handelt es sich um einen vielfach eingesetzten passiven Klangsteller mit Höhen-, Mitten- und Tiefenregler. Als passiver Klangsteller zeichnet er sich aus, da er lediglich aus passiven Bauteilen besteht. Aktive Klangsteller sind gänzlich anders strukturiert, indem ihr Netzwerk aus Widerständen und Kondensatoren in die Rückkopplungsschleife einer Röhre eingebunden wird.⁵⁵

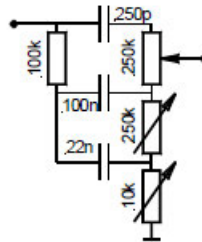


Abbildung 3:
Filterschaltung

Phasenumkehrstufe

An dieser Stelle muss bereits gesagt sein, dass in vielen Endstufen zwei Röhren im sogenannten Gegentaktbetrieb arbeiten, wofür es zwei um 180° phasenverschobene Ansteuersignale benötigt. Der Erzeugung dieser gegenphasigen Signale dient die Phasenumkehrstufe. Bei einer Endstufe mit lediglich einer einzigen Röhre, ist sie nicht nötig.⁵⁶

Bei der in Abbildung 4 gezeigten Phasenumkehrstufe wird der Widerstand R_{L1} in einer Kathoden-Verstärkerschaltung und R_{L2} in einer Anodenbasisschaltung betrieben. Somit fungiert die Schaltung als zwei Verstärker in einer einzigen Schaltung. Da die Kathoden-Verstärkerschaltung nicht überbrückt ist, ist ihre Ausgangsleistung sehr gering mit weniger

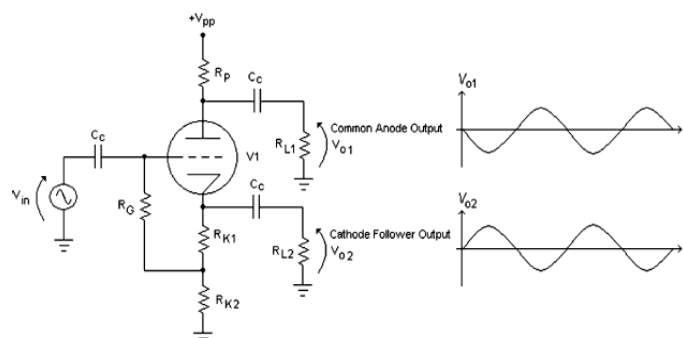


Abbildung 4: Typische Kathodyn-Schaltung mit erzeugten Ausgangsspannungen

⁵⁴ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 492.

⁵⁵ Vgl. Zollner: Physik der Elektrogitarre, S. 813.

⁵⁶ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 495.

als einem Volt. Die Ausgangsspannung der Anodenbasisschaltung ist ähnlich gering, im Idealfall gleich, so dass die Ausgangssignale die gleiche Amplitude aufweisen.⁵⁷

Die erläuterte Schaltung ist nur eines von drei Möglichen Schaltungskonzepten. Sie ist die simpelste Schaltung und gibt zwei Ausgangssignale mit praktisch gleicher Amplitude aus, hat aber den Nachteil, dass sie das Signal abschwächt und die beiden einzelnen Signale mit unterschiedlichen Impedanzen ausgegeben werden.⁵⁸ Eine andere Möglichkeit ist die Paraphase-Schaltung, bei welcher von einer verstärkende Triode die Anodenspannung als Steuersignal einerseits für eine Endröhre, aber auch abgeschwächt als Steuersignal einer weiteren Triode dient, deren gegenphasige Anodenspannung die andere Endröhre ansteuert. Hierbei ist das Problem, dass zwar stark verstärkt wird, die Beträge der Ausgangsspannungen aber nicht exakt gleich sind.⁵⁹ Die dritte Lösung ist der Differenzverstärker, welcher die positiven Eigenschaften der beiden vorigen Schaltungen vereint. Während eine Röhre in Kathoden-Basisschaltung mit Stromgegenkopplung arbeitet, arbeitet die zweite Röhre in Gitter-Basisschaltung und wird von der ersten Röhre über die Kathode angesteuert. Hierdurch führt eine Aussteuerung durch Wechsellspannung der ersten Röhre zu einem Spannungsabfall am Kathoden- und Anodenwiderstand. Mit der sich ändernden Kathodenspannung ändert sich auch die Steuerspannung der rechten Röhre und somit deren Anoden- und Kathodenstrom. Bei Röhren mit identischen Kennlinien resultiert daraus eine beitragsgleiche, aber gegenphasige Anodenspannung.⁶⁰

Endstufe

Die unterschiedlichen Realisierungen von Endstufen werden, bezüglich ihrer elektrischen Charakteristik, in unterschiedliche Klassen unterteilt. Dies betrifft Röhrenverstärker genauso wie Transistorverstärker. Die Einteilung in die unterschiedlichen Klassen (für die Gitarrenverstärkertechnik relevant sind insbesondere Klasse A, B und A/B) repräsentiert die Summe des Ausgangssignals, die bei Anregung durch ein sinusförmiges Eingangssignal über einen Betriebszyklus variiert.

Die besten klanglichen Eigenschaften, bezogen auf Linearität der Übertragung, werden der Klasse A zugeschrieben, welche sich aber nicht unbedingt für den Hochleistungsbetrieb eignet und generell als ineffizient gilt. Verstärker der Klasse A werden lediglich mit einem einzelnen aktiven Bauelement (Röhre, Transistor) betrieben, dessen Arbeitspunkt mittig des Betriebsbereichs vorgespannt wird. Das aktive Bauelement wird also an der Basis/dem Gitter mit einer konstanten Spannung versorgt, so dass es sich die gesamte Zeit im leitenden Zustand befindet. In diesem sogenannten Eintakt-Betrieb werden die vollen 360° des Ausgangssignals übertragen, ohne dass das Signal saturiert oder abgeschnitten wird.

Da der Betrieb von Klasse A-Verstärkern ineffizient ist und starke Hitzeentwicklung verursacht, wurde als Lösung die Klasse B entwickelt, bei welcher zwei aktive Bauelemente zum Einsatz kommen, um jeweils eine Hälfte der Welle zu übertragen, was als Gegentakt-Prinzip bezeichnet wird. Dabei sind die beiden Bauelemente entgegengesetzt zueinander

⁵⁷ Vgl. Dailey: Electronics for Guitarists, S. 211.

⁵⁸ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 498.

⁵⁹ Vgl. Zollner: Physik der Elektrogitarre, S. 826.

⁶⁰ Vgl. ebd., S. 831.

geschaltet, so dass eines die positive und eines die negative Halbwelle überträgt. Diese Betriebsart ist deutlich effizienter, allerdings auch deutlich weniger linear, besonders da der Teil der Halbwellen, welcher unterhalb der nötigen Schwellspannung liegt, um das Bauelement in den leitenden Zustand zu versetzen, nicht adäquat übertragen wird.

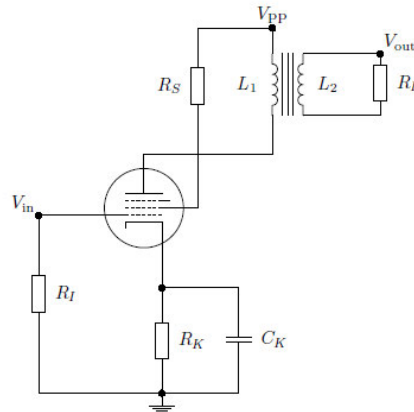


Abbildung 5: Endstufe der Klasse A

Der meistgenutzte Operationsmodus ist der A/B-Betrieb, welcher einen Kompromiss der beiden vorigen Modi darstellt. Im A/B-Betrieb wird eine sehr kleine Vorspannung an die aktiven Bauelemente angelegt, knapp über der nötigen Schwellspannung. Dadurch wird etwas mehr als eine Halbwelle übertragen und das Problem der Verzerrung durch die unerreichte Schwellspannung somit behoben.⁶¹

Der Operationsmodus ist also entscheidend für Klang und Leistung, was auf dieser Stufe entscheidend ist, da die Endstufe die letzte Instanz des Verstärkers ist, bevor das Signal an den Lautsprecher weitergegeben wird. Die Endstufe hat dabei die Aufgabe, die für den Lautsprecher erforderliche Leistung zu liefern.⁶²

Endstufen können also mit einer oder zwei Röhren betrieben werden kann, wobei meistens Pentoden eingesetzt werden, da diese höhere Verstärkungen ermöglichen. In Abbildung 5 ist eine klassisch aufgebaute Endstufe der Klasse A mit einer Pentode zu sehen. Der Widerstand R_L repräsentiert hierbei den Widerstand des Lautsprechers. Das an der Anode abgehende Ausgangssignal ist mit einem Transformator verbunden, der die Impedanz an den Lautsprecher anpasst. Eine Phasenumkehrstufe wird für diese Konfiguration nicht benötigt, dafür kann keine sonderlich hohe Leistung erreicht werden.⁶³

In Abbildung 6 ist ein klassischer Aufbau einer Gegentakt-Endstufe zu sehen. Das Eingangssignal kommt direkt von der Phasenumkehrstufe (Differenzverstärker) durch die Koppelkondensatoren C_{L1} und C_{L2} . Durch Anpassung der Spannung V_{bias} kann der Arbeitspunkt der Röhren eingestellt werden, so dass die Endstufe entweder im

⁶¹ Vgl. Aligarh Muslim University: Amplifier Classes, <https://old.amu.ac.in/emp/studym/100009038.pdf> (21. Juli 2023), S. 1–5.

⁶² Vgl. Zollner: Physik der Elektrogitarre, S. 839.

⁶³ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 503–504.

Operationsmodus B oder A/B arbeitet. Die Verstärkung wird durch die Widerstände einer Gegenkopplung mit der Phasenumkehrstufe bestimmt⁶⁴ (hier nicht eingezeichnet).

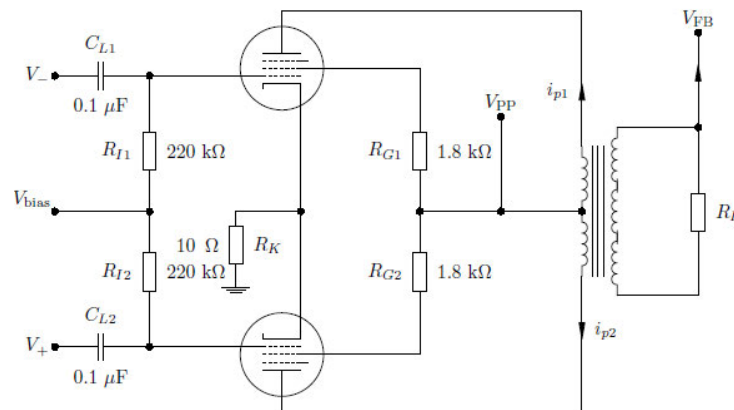


Abbildung 6: Gegentakt-Endstufe

Die letzte Anpassung vor dem Lautsprecher nimmt der Ausgangsüberträger vor. Der Lastwiderstand einer Leistungsröhre ist meist um ein tausendfaches höher als der typische Lautsprecherwiderstand, wodurch fast die gesamte Leistung am Innenwiderstand des Verstärkers abfallen würden. Der Ausgangsüberträger wird in Form eines Transformators am Ende der Endstufe vor dem Lautsprecher installiert und hat als letzte Instanz des Gitarrenverstärkers die Aufgabe, die hochohmige Röhrenschialtung an den niederohmigen Lautsprecher anzupassen.⁶⁵

2.3.2 Transistorverstärker

Eine Alternative zum klassischen Röhrenverstärker stellt der Transistorverstärker dar, welcher viele Nachteile der Röhrenverstärker überwunden hat, da er durch den Einsatz von Transistoren (häufig auch Operationsverstärkern) anstelle von Elektronenröhren deutlich langlebiger, günstiger und leichter ist. Dennoch konnte er sich nie wirklich durchsetzen und gilt seit seiner Einführung als schlechte Alternative,⁶⁶ weshalb dieses Kapitel auch nur der kurzen Erwähnung dienen soll. Ein kurzer Umriss über die klanglichen Unterschiede soll im anschließenden Kapitel gegeben werden.

Vom Prinzip her sind Transistorverstärker gleich aufgebaut wie Röhrenverstärker, in dem Sinne, dass sie aus den gleichen Stufen, die zuvor beschrieben wurden, zusammengesetzt sind. Innerhalb der Stufen, besonders in der Vor- und Endstufe, gibt es selbstverständlich schaltungstechnische Unterschiede.⁶⁷

⁶⁴ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 505.

⁶⁵ Vgl. Zollner: Physik der Elektrogitarre, S. 907.

⁶⁶ Vgl. Special Project Röhren 2. Röhren in Gitarrenverstärkern. In: Elektor (2006).

⁶⁷ Vgl. Lähdevaara: The science of electric guitars and guitar electronics, S. 509.

2.3.3 Der „Röhrensound“

Ziel der Transistor- und später auch der Modeling-Verstärker war es stets, den Klang eines Röhrenverstärkers zu imitieren, ohne die praktischen Nachteile desselbigen, hinsichtlich Preises, Fragilität etc. mitzubringen. Der „Röhrensound“ kann mittlerweile als schwammige Definition angesehen werden, wie ein Verstärker optimalerweise klingen soll. Den Röhrenverstärkern wird ein *warmer, runder* Klang nachgesagt, der Transistorverstärker wird antagonistisch als *steril, leblos* oder *kratzend* beschrieben.⁶⁸ Auch wenn sie eine ungefähre Vorstellung der klanglichen Eigenschaften zulassen, sind diese Worte dennoch ungenau und subjektiv. Technisch zu fassen ist der „Röhrensound“ in einigen Grundzügen aber durchaus, gerade wenn man ihn dem „Transistorsound“ gegenüberstellt.

Eine wesentliche Kenngröße ist der Ausgangswiderstand der Endstufe. Eine moderne Transistorendstufe hat eine nahezu lineare Übertragungskennlinie, was für HiFi-Anwendung oft gewünscht ist. Hier unterscheiden sich Transistor- und Röhrenendstufe wesentlich. Während die Transistorendstufe einen sehr niedrigen Ausgangswiderstand wegen der starken Gegenkopplung vom Aus- zum Eingang und damit einen hohen Dämpfungsfaktor (Verhältnis von Lautsprecherimpedanz zu Ausgangswiderstand) aufweist, hängt der Ausgangswiderstand eines Röhrenverstärkers stark von der Gegenkopplung ab. Generell ist er aber deutlich höher und der Dämpfungsfaktor entsprechend kleiner. Der Frequenzgang eines Röhrenverstärkers unterscheidet sich durch den daraus resultierenden Impedanzverlauf, in welchem Eigenresonanz und Höhen stark hervortreten, deutlich vom linearen Transistor-Frequenzgang. Eben dieses Eigenleben, das der Lautsprecher bei hohem Ausgangswiderstand der Endstufe entwickelt, ist aber beliebt bei Gitarristen⁶⁹ und somit Teil des „Röhrensounds“.

Ein weiterer wesentlicher Unterschied ist die Art, wie verzerrt wird, wobei es wesentliche Unterschiede zwischen Röhren- und Transistorverstärkern gibt. In einer üblichen Röhrenvorstufe verstärkt die Triode relativ unkontrolliert, wodurch die negativen Halbwellen einer Schwingung stärker verstärkt werden als die Positiven und somit auch ohne Übersteuerung der Triode bereits nichtlineare Verzerrungen erzeugt werden. Transistorverstärker erzeugen diese Verzerrung typischerweise nicht, da normalerweise mehrere Transistoren hintereinandergeschaltet und das Signal über eine Gegenkopplung zum Eingang zurückgeführt wird, was die Verzerrung vermindert und die Übertragung linearisiert.⁷⁰ Dieser Effekt und das Verhalten von Röhren, im Gegensatz zu Transistoren, *soft clipping* zu induzieren, im Gegensatz zum Transistor-typischen *hard clipping*, führen dazu, dass besonders viele geradzahlige harmonische Oberschwingungen erzeugt werden, welche im Gegensatz zu den als rau, unangenehm und Transistor-typisch beschriebenen ungeradzahligen Oberschwingungen, als deutlich angenehmer wahrgenommen werden.⁷¹

Hierzu ist allerdings eine wichtige Bemerkung zu machen: Die vorhergehenden Absätze dienten dazu, einen groben Eindruck zu vermitteln, was den „Röhrensound“ ausmacht und - ergänzt durch das Kapitel über den Aufbau des Röhrenverstärkers - wodurch dieser

⁶⁸ Vgl. Zollner: Physik der Elektrogitarre, S. 989.

⁶⁹ Special Project Röhren 2.

⁷⁰ Ebd.

⁷¹ Vgl. Dailey: Electronics for Guitarists, S. 141.

erreicht wird. Diese Frage bleibt nur unvollständig beantwortet, da ein Röhrenverstärker als Gesamtsystem, mit all den miteinander interagierenden linearen und nicht-linearen Bauelementen nahezu undurchschaubar ist.⁷² Zum Transistorverstärker sei gesagt, dass im obenstehenden Vergleich von einem höchst primitiven Transistorverstärker ausgegangen wurde. Der schlechte Ruf der Transistorverstärker liegt darin begründet, dass die zentralen Bauelemente in Röhren- und Transistorschaltungen in ihrer Topologie nicht einfach gegenseitig austauschbar sind. Transistorschaltungen sind, wenn in durchdachten Schaltungskonzepten realisiert, durchaus in der Lage, die oben beschriebenen Röhrentypischen Effekte klanglich zu imitieren.⁷³

2.3.4 Modeling-Verstärker

Es mag an dieser Stelle nicht verwundern, dass nach dem Versuch, Röhrenverstärker mit Transistoren nachzubauen, der nächste logische Schritt die Simulation mittels DSP und Computern war. Eine gelungene Umsetzung hätte zahlreiche Vorteile. Nicht nur, dass die klassischen Probleme eines Röhrenverstärkers überwunden wären, könnte man digitale Nachbildungen eines Verstärkers auf einem PC speichern und abrufen, würde dies den Wartungsbedarf eliminieren, den Aufnahmeprozess eines Gitarrenspiels deutlich vereinfachen und die Möglichkeiten der klanglichen Gestaltung vervielfachen.⁷⁴

Generell unterscheidet man zwischen Blackbox- und Whitebox-Modeling, wobei letzteres voraussetzt, dass man Informationen über den inneren Aufbau des Referenzverstärkers besitzt. Dieser Aufbau wird analysiert, das Verhalten der nicht-linearen Bauelemente wird messtechnisch erfasst und aufgrund dieser Informationen wird anschließend das digitale Modell konstruiert.⁷⁵ Diese digitalen Nachbildung, welche oft mit Programmen wie SPICE, welche elektronische Schaltungen analysieren und numerische Lösungen von Schaltungsgleichungen berechnen können, realisiert werden, haben den Nachteil, dass sie rechnerisch sehr aufwendig sind, weshalb ein Verständnis über die Wirkweise des Referenzverstärkers notwendig ist, um die digitale Nachbildung zu vereinfachen, ohne zu große klangliche Abstriche machen zu müssen.⁷⁶ Häufig werden die Referenzschaltungen zur Vereinfachung in Stufen unterteilt, um lineare Stufen mit Filtern nachzubilden, um die Menge an rechnerisch aufwendigen nicht-linearen Differenzialgleichungen gering zu halten.⁷⁷

Beim Blackbox-Modeling dagegen ist der innere Aufbau nicht bekannt. Das Verhalten des Referenzverstärkers wird lediglich durch die Messung der Relation zwischen Eingangs- und Ausgangssignal erfasst und die Simulation des nicht-linearen Verhaltens mittels Gleichungen wie den Volterra-Reihen-Gleichungen – ein Modellansatz für nichtlineare

⁷² Vgl. Zollner: Physik der Elektrogitarre, S. 990.

⁷³ Vgl. Teemu Kyttälä: Solid-State Guitar Amplifiers, 2008, S. 21–25.

⁷⁴ Vgl. Pakarinen, J.; Yeh, D. T.: A Review of Digital Techniques for Modeling Vacuum-Tube Guitar Amplifiers. In: Computer Music Journal 33 (2009), S. 85–100, hier S. 85.

⁷⁵ Vgl. Felix Eichas; Udo Zölzer: Virtual Analog Modeling of Guitar Amplifiers with Wiener-Hammerstein Models, 2018, S. 1.

⁷⁶ Vgl. Mačák, J.: Real-time Digital Simulation of Guitar Amplifiers as Audio Effects. Doktorarbeit, Brno University of Technology 2012 2012, S. 20.

⁷⁷ Vgl. Yeh, D. T.-M.: Digital Implementation of Musical Distortion Circuits by Analysis and Simulation. Doktorarbeit, 2009 2009, S. 3.

Systeme mit Ein- und Ausgang, eine Erweiterung der zeitdiskreten Faltungssumme⁷⁸ - realisiert. Auch diese Methodik hat ihre Limitierungen, da die meisten Verstärker variable Parameter innerhalb der Schaltung aufweisen, deren Verarbeitung zusätzlich großen Aufwand erzeugt.⁷⁹

Ergänzend zu beiden grundlegenden Vorgehensweisen ist zu erwähnen, dass häufig eine Mischung aus beiden Varianten angewandt wird, das sogenannte Graybox-Modeling, bei welchem Whitebox-Modelle mittels Eingangs-/Ausgangsmessungen ergänzt werden.⁸⁰

2.4 Lautsprecherkabinette und Mikrofonierung

Das Lautsprecherkabinett hat, anders als gewöhnliche Lautsprecher, die Aufgabe, aktiv an der Klangformung des Gitarrensounds mitzuwirken. Eine Verzerrung des Signals ist also ausdrücklich erwünscht.

Für die Kabinette kommen generell nur Lautsprecher mit dynamischen Wandlerprinzip zum Einsatz, meist mit einem Durchmesser zwischen 8 bis 15 Zoll.⁸¹ Eine Besonderheit der Lautsprecher ist, dass diese mit Rippen versehen werden, um möglichst viele Partialschwingungen zu erzeugen, welche bei HiFi-Lautsprechern unerwünscht sind, den Gitarrenklang aber mit Hochtönen anreichern. Ein Effekt, der durch das Schwingen der Membran in Eigenmoden bei höheren Frequenzen (>600Hz) auftritt.⁸² Die Lautsprecher werden in ein Gehäuse aus Holz eingebaut, welches ebenfalls Anteil an der Schallerzeugung hat, indem es den Schall führt und Hohlraumresonanzen ausbildet. Die Gehäuse gibt es in Ausführung mit geschlossener und offener Rückwand. Bei geschlossener Rückwand, erhöht sich der Druck im Inneren des Gehäuses, wenn die Membran sich nach innen bewegt, wodurch das Gehäuse als Luftfeder wirkt. Bei geschlossener Rückwand kann der Druck entweichen und das Gehäuse wird durch den Luftdruck kaum belastet.⁸³ Durch diese, die maximale Auslenkung der Lautsprecherkone limitierende Eigenschaft, wird der Ausführung mit geschlossener Rückwand ein definierterer, präziserer Klang nachgesagt als der Ausführung mit offener Rückwand.⁸⁴

Zur Aufnahme des gesamten Gitarrensounds (Gitarre mit Verstärker und Lautsprecher) wird das Kabinett mikrofoniert. Hierfür können unterschiedlichste Mikrofone und Konzepte zum Einsatz kommen. Da die Position und die Ausrichtung des Mikrofons am Lautsprecher wesentlichen Einfluss auf den Klang haben, wird oft mit mehreren – ggf. auch mit unterschiedlichen - Mikrofonen an verschiedenen Stellen des Kabinetts aufgenommen, um sich für die beste Option entscheiden zu können oder diese miteinander zu kombinieren.⁸⁵ Grundlegend kann man sagen, dass das Mikrofon meist mit einem Abstand von einem Zoll

⁷⁸ Vgl. Hofmann, S.: Identifikation von nichtlinearen mechatronischen Systemen auf der Basis von Volterra-Reihen. Dissertation, 2003 2003, S. 50.

⁷⁹ Vgl. Mačák: Real-time Digital Simulation of Guitar Amplifiers as Audio Effects, S. 21.

⁸⁰ Vgl. Felix Eichas/ Udo Zölzer: Virtual Analog Modeling of Guitar Amplifiers with Wiener-Hammerstein Models, S. 1.

⁸¹ Vgl. Zollner: Physik der Elektrogitarre, S. 1077.

⁸² Vgl. ebd., S. 1182.

⁸³ Vgl. ebd., S. 1147–1148.

⁸⁴ Vgl. Mynett: Metal Music Manual, S. 119.

⁸⁵ Vgl. Rosiński, A.: Microphone techniques in stereo and surround recording, Kraków 2022, S. 151.

zur Lautsprecherfront in einem Winkel von 90° aufgestellt wird. Näher an der Lautsprecherkone, wird die Klangqualität heller und höhenreicher, näher am Rand der Membran, wird sie dunkler und matter.⁸⁶ Dies versteht sich selbstverständlich nur als grober Leitfaden, hinsichtlich der Vielzahl an möglichen Strategien. Ebenso vielfältig ist die Mikrofonauswahl. Erwähnt werden sollten hier als häufig gewählte Optionen aber zumindest das Shure Sm57 und das Sennheiser MD421.⁸⁷

⁸⁶ Vgl. Mynett: Metal Music Manual, S. 125–128.

⁸⁷ Vgl. ebd., S. 118.

3 Künstliche Intelligenz und deren Anwendung zur Verstärker-Emulation

Nachdem ein Überblick über die Funktionsweise verschiedener Gitarrenverstärker gegeben wurde, sollen in diesem Kapitel Grundlagen der künstlichen Intelligenz beleuchtet werden, um nachvollziehen zu können, wie KI-basierte Emulationen von Gitarrenverstärkern konzipiert sein können.

Laut dem *Fraunhofer-Institut für Kognitive Systeme IKS* wird künstliche Intelligenz wie folgt definiert:

„Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik. Sie imitiert menschliche kognitive Fähigkeiten, indem sie Informationen aus Eingabedaten erkennt und sortiert. Diese Intelligenz kann auf programmierten Abläufen basieren oder durch maschinelles Lernen erzeugt werden.“

(...) Bei maschinellen Lernverfahren erlernt ein Algorithmus durch Wiederholung selbstständig eine Aufgabe zu erfüllen. [...] Anders als bei herkömmlichen Algorithmen wird kein Lösungsweg modelliert. Der Computer lernt selbstständig die Struktur der Daten zu erkennen.“⁸⁸



Abbildung 7: Ordnung der KI-betreffenden Begriffe nach Ralf T. Kreuzer und Marie

KI kann dabei Anwendung in beispielweise der Bilderkennung, Spracherkennung oder Mustererkennung finden, jeweils mit unterschiedlichen möglichen Zielen. *Künstliche Intelligenz* ist als Begriff mittlerweile inflationär genutzt und schwimmt mit Begriffen wie *Deep Learning* und *Neuronalen Netzen*. Abbildung 7 dient dazu, die Begriffe richtig einzuordnen. In den nachfolgenden Kapiteln sollen grundlegende Informationen über das Thema künstliche Intelligenz und die zugehörigen Unterthemen maschinelles Lernen, Neuronale Netze und Deep Learning bereitgestellt werden, um abschließend zu erläutern, wie diese Mittel eingesetzt werden können, um Gitarrenverstärker zu emulieren.

⁸⁸ Fraunhofer-Institut für Kognitive Systeme IKS: Künstliche Intelligenz (KI) und maschinelles Lernen, <https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html> (29. September 2023).

3.1 Grundlagen maschinellen Lernens

Das maschinelle Lernen ist ein Teilgebiet der KI, welches als Schlüsseltechnologie selbiger verstanden wird, da es bezweckt, jenes „Wissen“ zu generieren, auf dessen Basis ein intelligentes System später als solches attribuiert wird.⁸⁹ Es wird „sich mit Algorithmen befasst, mit denen auf Basis von Beispielen (Trainingsdaten) erwünschtes Verhalten automatisiert vom System erlernt werden kann“.⁹⁰ Es geht also darum, dem Computer beizubringen, wie er bestimmte Aufgaben löst, ohne dass konkrete Algorithmen zur Lösung dieser bestimmten Aufgaben geschrieben werden müssen. Die Algorithmen beschreiben stattdessen das Lernverfahren und die nachfolgende Verwendung, wobei zahlreiche frei einzustellende Parameter zur mathematischen Reduktion einer Aufgabe eingesetzt werden, welche im Lernprozess automatisiert eingestellt werden. Beim Entwurf eines maschinellen Lernsystems wird also ein Lernverfahren entworfen, welches mit Beispieldaten lernt, Parameter zur möglichst exakten Aufgabebearbeitung einzustellen.⁹¹

3.1.1 Entwurf eines maschinellen Lernsystems

Um den Prozess des maschinellen Lernens verständlich zu machen, soll das Entwerfen eines beispielhaften maschinellen Lernsystems nach Tom Mitchell erklärt werden, welcher dieses in seinem Buch *Machine Learning*⁹² anschaulich am Beispiel erklärt, einem Computer das Spiel *Checkers*, einer Variante des Brettspiels *Dame* beizubringen.

Zuallererst soll festgelegt werden, worin die Aufgabe besteht und wodurch die Leistung bei der Lösung der Aufgabe gemessen wird. Mitchell definiert Ersteres als *task T*: Checkers spielen und Letzteres als *performance Measure P*: Prozentzahl an gewonnenen Spielen. Im ersten Schritt geht es darum, eine Trainingserfahrung *E* festzulegen. Diese kann indirekte oder direkte Rückmeldung über *P* liefern. So könnte sie hier darin liegen, aus zahlreichen unterschiedlichen Spielständen zusammen mit dem jeweils korrekten Spielzug zu lernen (direkte Rückmeldung) oder aus den Zug-Sequenzen und Ergebnissen kompletter Spielverläufe. Letzteres indirekte Verfahren sieht sich mit dem Problem der Glaubwürdigkeitsvergabe (*credit assignment*) konfrontiert, also der Gewichtung einzelner Spielzüge und ihrem Beitrag zum Ergebnis.

Weitere Entscheidungen liegen darin, zu welchem Grad das Lernsystem mit Kontrolle über die Trainingsbeispiele ausgestattet wird. So könnte das System einfach mit Spielständen und den korrekten Spielzügen ausgestattet werden oder selbst Spielstände vorschlagen, die es noch nicht versteht. Zudem sollte überlegt werden, wie präzise die Trainingserfahrung die unterschiedlichen Beispiele repräsentiert, durch welche die Leistung *P* des finalen Systems gemessen wird.

⁸⁹ Vgl. Döbel, I., et al.: Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung, 2018, S. 8.

⁹⁰ Ernst, H.; Schmidt, J.; Beneken, G. H.: Grundkurs Informatik. Grundlagen und Konzepte für die erfolgreiche IT-Praxis – eine umfassende, praxisorientierte Einführung, 7. Aufl., Wiesbaden 2020, S. 801.

⁹¹ Vgl. ebd., S. 802.

⁹² Mitchell, T. M.: Machine learning, New York, NY 2010.

Mitchell entscheidet, sein System lernen zu lassen, indem es gegen sich selbst spielt. Er definiert die Trainingserfahrung als *training experience E*: Gegen sich selbst gespielte Spiele. Dies hat den Vorteil, dass keine externen Trainingsdaten nötig sind und das System so viele Daten wie nötig selbst generieren kann.

Im zweiten Schritt wird die Zielfunktion festgelegt, wofür bestimmt werden muss, welches exakte Wissen erlernt werden soll und wie dieses verwendet wird. Grundlegend sollte mit einem Programm begonnen werden, das aus jedem Spielzustand alle erlaubten Spielzüge generiert, um das System mit den Spielregeln auszustatten. Danach gibt es unterschiedliche mögliche grundlegende Gestaltungsmöglichkeiten. Mitchell entscheidet sich dafür, die Zielfunktion als Evaluierungsfunktion V zu gestalten, welche jedem möglichen Spielzustand B einen numerischen Punktestand zuschreibt, welcher bei besseren Spielständen höhere Werte vergibt. Kann das System diese Funktion V erfolgreich lernen, ist es einfach, den bestmöglichen Spielzug b aus jedem aktuellen Spielstand zu wählen, indem zuerst alle erlaubten Spielzüge berechnet und dann der beste nachfolgende mögliche Spielzustand anhand von V errechnet wird. Da es schwierig ist, einen numerischen Wert $V(b)$ zuzuweisen für einen Spielstand, mit welchem das Spiel nicht beendet ist, ist es notwendig vorausschauend eine optimale Zugabfolge bis zum Ende des Spiels zu berechnen. Da diese Berechnungen zu lange dauern würden, geht es darum – und damit ist die Zielfunktion definiert – eine funktionale Annäherung \hat{v} an die ideale Funktion V zu finden, welche Spielzüge in realistischen zeitlichen Begrenzungen berechnet.

Nachdem die ideale Zielfunktion V definiert wurde, gilt es im dritten Schritt, eine geeignete Repräsentierung zu finden, die das Programm nutzen wird, um \hat{v} zu beschreiben. Diese sollte möglichst aussagekräftig sein, aber gleichzeitig auch effizient hinsichtlich Berechnungszeiten und benötigten Trainingsdaten. Mitchell wählt eine simple Repräsentierung, welche eine Linearkombination der relevanten Eigenschaften des Spielstands sind: Anzahl schwarzer (x_1) und roter (x_2) Steine, Anzahl schwarzer (x_3) und roter (x_4) Könige, sowie die Anzahl schwarzer Steine, welche im nächsten Zug von Rot geschlagen werden können (x_5) und selbiges umgekehrt (x_6).

Diese Größen werden jeweils mit numerischen Koeffizienten (Gewichtungen) versehen, welche ihre Relevanz kennzeichnen und vom Lernsystem während des Lernprozesses automatisch eingestellt werden. Die Repräsentierung der Zielfunktion sieht also wie folgt aus:

$$\hat{v}(b) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

Im letzten Schritt zum Entwurf des Lernsystems, gilt es, einen Algorithmus zur Funktionsnäherung zu bestimmen, indem durch Trainingsbeispiele mit spezifischen Spielzuständen und zugehörigem Trainingswert $V_{\text{train}}(b)$ die Zielfunktion \hat{v} mit ihren Gewichtungen gelernt wird. Hierfür ist es einerseits von Bedeutung, die Trainingswerte $V_{\text{train}}(b)$ zu schätzen, was bei den finalen Spielständen leichtfällt, bei den nichtfinalen Spielzügen in der Mitte des Spielverlaufs allerdings schwer zu bestimmen ist, da diese weniger eindeutig auf ihren Beitrag zum Sieg oder der Niederlage des Spiels zurückzuführen sind. In diesem Fall kann ein Vorgehen erfolgreich eingesetzt werden, bei welchem $V_{\text{train}}(b)$ für jeden nichtfinalen Spielzustand b als $\hat{v}(\text{successor}(b))$ beschrieben wird,

wobei \hat{v} die aktuelle Approximierung an V ist und $\text{Successor}(b)$ den nächsten Spielzustand nach b beschreibt, an welchem das System wieder am Zug ist.

Abschließend müssen lediglich die Gewichtungen eingestellt werden, was häufig nach einer Formel geschieht, welche den quadrierten Fehler E zwischen den Trainingswerten $V_{\text{train}}(b)$ und den vorhergesagten Werten $\hat{v}(b)$ minimiert. Das Einstellen der Gewichtungen funktioniert dann nach einem Algorithmus, welcher sukzessive die Gewichtungen an neue Trainingsbeispiele anpasst, wie der Last-Mean-Square-Algorithmus (LMS), welcher nach jedem Trainingsbeispiel die Gewichtungen ein kleines Stück in die Richtung anpasst, die den Fehler E für dieses Beispiel reduziert.⁹³

Damit ist das maschinelle Lernsystem fertig entworfen. Durch die Reduzierung auf die Repräsentierung der Zielfunktion $\hat{v}(b)$, die Auswahl von Trainingserfahrung- und Beispielen und einen Algorithmus zur Fehlerreduktion, ist das System in der Lage, sich selbstständig beizubringen, Checkers zu spielen und zu gewinnen. Dies zeigt beispielhaft, an einem einfachen Entwurf, wie ein maschinelles Lernsystem konzipiert werden kann. Selbstverständlich gibt es ein breites Spektrum an Konzepten und Anwendungen.

3.1.2 Unterscheidung maschineller Lernsysteme

Ein verbreiteter Ansatz, Ordnung in das breite Spektrum an möglichen Konzepten maschineller Lernsysteme zu bringen, ist es, sie anhand ihrer Lernstrategie zu unterscheiden. Dabei spielt das Verhältnis zwischen dem lernenden System und dem „Lehrenden“ die entscheidende Rolle. Während Systeme durch Auswendiglernen, Anweisungen, oder aus Analogien lernen können, Aufgaben zu bewältigen, spielt das Lernen aus Beispielen bei heutigen maschinellen Lernsystemen, gerade den Deep-Learning-Systemen die entscheidende Rolle. Hierbei soll das Lernsystem aus einem Satz von Beispielen ein generelles Konzept zur Beschreibung derartiger Beispiele erlernen.⁹⁴ Den Prozess, das Konzept hinter Beispielen zu erlernen, nennt man Klassifikation. Diese lässt sich in drei grundlegende Typen unterscheiden:

Überwachtes Lernen ist der am weitesten verbreitete Typus. Er zeichnet sich darin aus, dass die Beispiele, aus denen das System lernt, annotiert, also durch Menschen im Vorfeld klassifiziert worden sind. Auf Basis dieser bereits klassifizierten Beispiele lernt das System das generelle Konzept hinter den Trainingsbeispielen. Beim **unüberwachten Lernen** sind die Beispiele zuvor nicht klassifiziert worden, lediglich die Anzahl an unterschiedlichen vorkommenden Klassen werden dem System üblicherweise mitgeteilt. Das System soll automatisiert lernen, die Klassifizierung vorzunehmen. Dies gestaltet sich deutlich schwieriger als das überwachte Lernen und ist bislang deutlich weniger leistungsfähig, findet aber dennoch bereits Anwendung. Der Typus des **verstärkenden Lernens** ist bereits im vorhergehenden Kapitel beleuchtet worden, beim exemplarischen Entwurf eines maschinellen Lernsystems. Beim verstärkenden Lernen werden ebenfalls keine klassifizierten Stichproben benötigt, stattdessen wird durch Konditionierung gelernt. Das

⁹³ Vgl. Mitchell: Machine learning, S. 5–11.

⁹⁴ Vgl. Camastra, F.; Vinciarelli, A.: Machine learning for audio, image and video analysis. Theory and applications, 2. Aufl., London u. a. 2015, S. 100–102.

System erhält eine Rückmeldung über die Wirksamkeit seiner getroffenen Entscheidungen, aus denen es für zukünftige Entscheidungen Rückschlüsse zieht. Dadurch kann, wie im Beispiel aus 3.1.1, ein System ein Spiel erlernen, indem es gegen sich selbst spielt.⁹⁵

3.2 Deep Learning

Auf welchem Klassifikations-Typus ein Lernsystem auch beruht, das simple Beispiel aus Kapitel 3.1.1 zeigt, wie sehr es von der Repräsentation der ihm zugeführten Daten abhängt. Im Beispiel war die Reduktion des zu lernenden Sachverhaltes auf eine Zielfunktion simpel, dies kann mit komplexeren Aufgaben allerdings deutlich schwieriger werden. Geht es bspw. darum, Objekte auf Fotografien automatisiert zu erkennen, ist es schwer, zu bestimmen, welche Eigenschaften extrahiert und in einen Lernalgorithmus übertragen werden sollen. Dieses Problem kann durch Deep Learning gelöst werden, indem hierarchisch nach Simplizität geordnete Konzepte eingeführt werden, durch welche komplexe Konzepte durch mehrere Konzept-Schichten repräsentiert und erlernt werden. Wegen dieses Prinzips der viellagigen Schichten wird dieser Ansatz Deep Learning genannt.⁹⁶

Anders als traditionelle Algorithmen zum maschinellen Lernen, lernen derartige Algorithmen selbstständig, Eigenschaften zu extrahieren, weshalb es keiner zeitaufwendigen Supervision des Lernprozesses durch menschliche Kräfte bedarf. Die Anwendung hat sich in den letzten Jahren auf viele Felder ausgebreitet. So wird Deep Learning beispielweise für automatische Spracherkennung, sprachliches Interpretieren, Bilderkennung und viele andere Disziplinen z.B. in der Medizin, Biologie und Finanzbranche eingesetzt. Die meisten Algorithmen des Deep Learning basieren auf dem Einsatz von künstlichen neuronalen Netzen,⁹⁷ einem Konzept, das nachfolgend betrachtet werden soll.

3.3 Künstliche Neuronale Netze

Die Aufgaben, für deren Bewältigung Deep Learning und neuronale Netze eingesetzt werden, stellen klassische Algorithmen zur Informationsbewältigung vor große Schwierigkeiten, da diese auf typisch menschlichen Fähigkeiten (z.B. Gesichts- oder Spracherkennung) beruhen, welche in hohem Maße indeterminiert und unscharf sind.⁹⁸ Für derartige Probleme, bei denen komplexe Daten, z.B. Sensordaten einer Kamera oder eines Mikrofons, aus der Umwelt interpretiert werden sollen, ist der Einsatz künstlicher neuronaler Netze eine der wirkungsvollsten Methoden.

Die Funktionsweise der Informationsverarbeitung ist dabei inspiriert von den im menschlichen Gehirn ablaufenden Prozessen. So wie das Gehirn ein Netz aus zahlreichen (schätzungsweise 10^{11}) Nervenzellen (Neuronen) innehat, von welchen jedes Einzelne durch zahlreiche (ca. 10^4) Verbindungen mit anderen Neuronen verbunden ist, bestehen auch neuronale Netze aus eng miteinander verbundenen Einzeleinheiten. Jede dieser

⁹⁵ Vgl. Ernst/ Schmidt/ Beneken: Grundkurs Informatik, S. 802–803.

⁹⁶ Vgl. Goodfellow, I.; Bengio, Y.; Courville, A.: Deep learning, Cambridge, Massachusetts u. a. 2016, S. 1–3.

⁹⁷ Vgl. Chen, L.: Deep Learning and Practice with MindSpore, Singapore 2021, S. 3–4.

⁹⁸ Vgl. Strecker, S.: Künstliche neuronale Netze - Aufbau und Funktionsweise, Mainz 1997, S. 3.

Einheiten produziert aus mehreren realwertigen Eingängen einen einzelnen realwertigen Ausgang, der dann als Eingangswert an andere Einheiten weitergegeben wird.⁹⁹ Die Leitungen zwischen den Einheiten sind dabei mit Gewichtungen versehen, wie sie in Kapitel 3.1.1 bereits Erwähnung fanden. Während des Trainingsprozess werden diese Gewichtungen eingestellt, um bestimmte Ausgangswerte der jeweiligen Aufgabe auf bestimmte Eingangswerte zurückzuführen und dadurch zu lernen, Ausgangswerte aus bisher ungesehenen Eingangswerten zu bestimmen.¹⁰⁰

3.3.1 Aufbau und Funktionsweise

Die kleinen Einheiten, werden auch (künstliche) Neuronen, Knoten oder Units genannt. Sie dienen der Aufnahme von Informationen in Form von Zahlenwerten – aus der Umwelt oder von anderen Einheiten – und deren Weiterleitung in modifizierter Form. Weitergeleitet wird über sogenannte Links, also mit Gewichtungen versehene Verbindungen.¹⁰¹

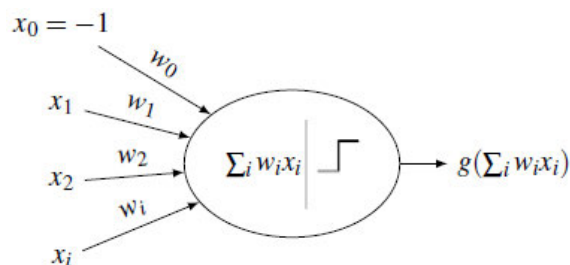


Abbildung 8: Schematische Darstellung einer Einheit.

Die Informationsverarbeitung erfolgt dabei durch Berechnung der einzelnen Eingangswerte mit ihren Gewichtungen und anschließende Summierung. Die Summe bildet den Inputwert der Einheit und wird im anschließenden Schritt einer Aktivierungsfunktion zugeordnet, bevor der Ausgangswert erzeugt wird.¹⁰² Dieses Schema wird in Abbildung 8 dargestellt.

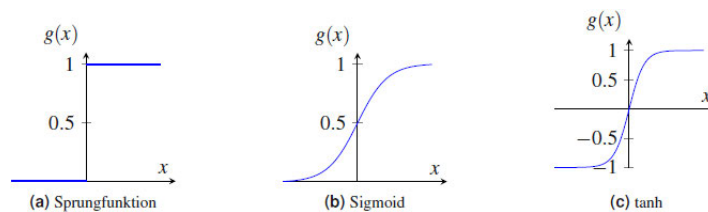


Abbildung 9: Einige Aktivierungsfunktionen

Die Aktivierungsfunktion bestimmt das Aktivitätslevel einer Einheit, definiert also ein Verhältnis zwischen Ausgang und Eingang. Es gibt zahlreiche unterschiedliche Aktivierungsfunktionen. Typischerweise wird für ein neuronales Netz eine einheitliche

⁹⁹ Vgl. Mitchell: Machine learning, S. 81–82.

¹⁰⁰ Vgl. Camastra/ Vinciarelli: Machine learning for audio, image and video analysis, S. 193.

¹⁰¹ Vgl. Rey, G. D.; Wender, K. F.: Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung, 3. Aufl., Bern 2018, S. 16–17.

¹⁰² Vgl. ebd., S. 18.

Aktivierungsfunktion für alle Einheiten festgelegt.¹⁰³ Klassische Beispiele sind die Sprungfunktion, welche nur den Wert 1 oder 0 ausgeben kann (wie in Abbildung 8), die Sigmoidfunktion, mit Werten zwischen 0 und 1 oder die Funktion Tangens Hyperbolicus, welche Werte von -1 bis 1 annehmen kann.¹⁰⁴ Abbildung 9 können einige Visualisierungen unterschiedlicher Aktivierungsfunktionen entnommen werden.

Ein neuronales Netz besteht häufig aus mehreren Schichten. Bei solchen Netzwerken, sogenannten Mehrschichtperzeptronen, besteht eine Schicht aus Einheiten, welche je nach Netztopologie nach bestimmten Mustern untereinander verbunden sind. Jede Schicht ist als Gesamteinheit mit anderen Schichten vernetzt.¹⁰⁵ An diesem Punkt ist eine Unterscheidung sinnvoll in Eingangs-Einheiten, welche Signale von der Außenwelt in Form von Zahlen erhalten, Ausgangs-Einheiten, welche Signale als Zahlenwerte an die Außenwelt abgeben und versteckte Einheiten, die sich zwischen Eingang und Ausgang befinden und ausschließlich Informationen von anderen Einheiten bekommen und an andere Einheiten weitergeben. Eine äquivalente Unterteilung kann für ganze Schichten vorgenommen werden.¹⁰⁶ Eine Schicht kann dabei als eine Sinneinheit des Gesamtnetzwerks betrachtet werden, in welcher die Einheiten „eine gleichartige Verhaltensweise hinsichtlich der Datenverarbeitung aufweisen und in der Netzwerkstruktur eine identische Funktion einnehmen“.¹⁰⁷

3.3.2 Netztopologien

Die Verbindungen zwischen den Einheiten eines neuronalen Netzes regeln den Informationsfluss und sind zu unterscheiden in uni- und bidirektionale Verbindungen. Erstere ermöglichen den Informationsfluss nur in eine Richtung, bei Letzteren ist ein wechselseitiger Informationsfluss möglich. Zudem können die Verbindungen unterteilt werden in exzitatorische (verstärkende) und inhibitorische (hemmende) Verbindungen.¹⁰⁸

Daraus resultieren zwei unterschiedliche grundlegende Topologien. Die Erste ist das FeedForward-Netz, welches sich aus Eingabeschicht, Ausgabeschicht und versteckten Verarbeitungsschichten dazwischen zusammensetzt. Die Verbindungen zwischen den Neuronen dürfen dabei ausschließlich in Richtung der Ausgabeschicht verlaufen. Häufig bestehen die Verbindungen von den Neuronen einer Schicht zu den Neuronen der jeweils darauffolgenden Schicht, aber auch sogenannte ShortCut-Connections sind möglich, bei denen die Verbindungen eine oder mehrere Schichten überspringen.¹⁰⁹

Die zweite grundlegende Netztopologie ist das FeedBackward-Netz, bei welchem die Einheiten sich durch Rückkopplungen selbst beeinflussen können. Dabei unterscheidet man zwischen direkter, indirekter und lateraler Rückkopplung. Direkte Rückkopplung beschreibt die Verbindung von Einheiten zu sich selbst, wodurch diese sich selbst stärken

¹⁰³ Vgl. Rey/ Wender: Neuronale Netze, S. 20.

¹⁰⁴ Vgl. Ernst/ Schmidt/ Beneken: Grundkurs Informatik, S. 806–807.

¹⁰⁵ Vgl. ebd., S. 809.

¹⁰⁶ Vgl. Rey/ Wender: Neuronale Netze, S. 17.

¹⁰⁷ Strecker: Künstliche neuronale Netze - Aufbau und Funktionsweise, S. 15.

¹⁰⁸ Vgl. ebd., S. 16.

¹⁰⁹ Vgl. David Kriesel: Ein kleiner Überblick über Neuronale Netze, 2007, S. 50.

oder hemmen können. Unter indirekter Rückkopplung versteht man Verbindungen in Richtung der Eingabeschicht, wodurch eine Einheit sich nur durch Umwege selbst beeinflussen kann. Laterale Rückkopplung meint die Verbindung von Einheiten innerhalb einer Schicht zueinander.¹¹⁰

Ausgehend von diesen Topologien lassen sich Netzstrukturen in vier unterschiedliche Gruppen unterteilen:

Die einschichtigen FeedForward-Netze, welche auf lineare Problemstellungen begrenzt sind, da die Eingabeschicht vollständig mit der Ausgabeschicht verbunden ist. Die Weiterentwicklung hiervon sind die mehrschichtigen FeedForward-Netze, die mit versteckten Schichten arbeiten und somit nichtlineare Probleme lösen können. Sie gelten als sehr flexibel, da sie beliebige mathematische Funktionen approximieren können, allerdings setzt die Entwicklung meist einen intensiven „Trial-and-Error“-Prozess voraus. Die dritte Gruppe bilden die Netze mit direkten und indirekten Rückkopplungen. Ihre Informationsverarbeitung unterscheidet sich grundlegend von den FeedForward-Netzen, da die Ausgabewerte der rückgekoppelten Einheiten immer wieder in den Verarbeitungsprozess zurückgeführt werden. Unterscheiden kann man hier zusätzlich zwischen Netzen mit lediglich direkter Rückkopplung, indirekten Rückkopplungen und Netze, welche vollständig verbunden sind, so dass jede Einheit mit jeder anderen Einheit (außer sich selbst) verbunden ist. Den Abschluss bildet die vierte Gruppe: Die Netze mit lateralen Rückkopplungen und Gitterstrukturen. Hierbei sind die Einheiten der Ausgabeschicht geometrisch als n-dimensionales Gitter mit lateraler Rückkopplung angeordnet, während die Eingabeschicht vollständig unidirektional mit der Gitterstruktur verbunden ist. Diese Anordnung ist die Basis für Netzwerkmodelle, welche als selbstorganisierende Karten bezeichnet werden. Sie besitzen die Fähigkeit, Klassen in den Eingabedaten selbstständig abzubilden.¹¹¹

3.4 Emulation von Röhrenverstärkern mittels künstlicher neuronaler Netze

Wie in Kapitel 2.3.4 beschrieben, gibt es bereits zahlreiche digitale Emulationen von Röhrenverstärkern, welche die besprochenen wesentlichen Vorteile mit sich bringen, klanglich aber noch immer das Nachsehen haben. Die Anwendung neuronaler Netze scheint angemessen, die Blackbox-Methode zur Verstärkeremulation weiter nach vorne zu bringen. Da es beim Blackbox-Modeling darum geht, das Verhalten eines Verstärkers - also das aus einem Eingangssignal resultierende Ausgangssignal - an einen Referenzverstärker durch den Gebrauch mathematischer Funktionen anzupassen, ohne Kenntnisse über den Aufbau des Referenzverstärkers zu besitzen, scheinen neuronale Netze durch ihr nichtlineares Verhalten und ihre Architektur geeignet zu sein, das nichtlineare Verhalten und die Dynamik eines Röhrenverstärkers zu erfassen und zu replizieren.¹¹²

¹¹⁰ Vgl. David Kriesel: Ein kleiner Überblick über Neuronale Netze, S. 50–55.

¹¹¹ Vgl. Strecker: Künstliche neuronale Netze - Aufbau und Funktionsweise, S. 17–21.

¹¹² Vgl. Vanhatalo, T.; Legrand, P.; Desainte-Catherine, M.; Hanna, P.; Brusco, A.; Pille, G.; Bayle, Y.: A Review of Neural Network-Based Emulation of Guitar Amplifiers. In: Applied Sciences 12 (2022), hier S. 2.

Tatsächlich handelt es sich hierbei um eine junge Disziplin, zu der lediglich eine Handvoll Studien – größtenteils innerhalb der vergangenen fünf Jahre veröffentlicht - vorliegen, welche unterschiedliche Herangehensweisen nutzen, um neuronale Netze das komplexe Verhältnis zwischen Aus- und Eingangssignal von Gitarrensensoren lernen zu lassen.

Zu nennen sind hierbei die Arbeiten^{113, 114, 115} von Damskägg et al., in welchen der Einsatz von sog. Faltungsgewichten gewählt wurde, um zuerst die generelle Eignung zu prüfen und in darauf aufbauenden Studien auch den Einsatz zur Echtzeit-Anwendung zu untersuchen.

Des Weiteren wurden Modelle, die auf rückgekoppelten Netzen basieren untersucht. So haben David L. Livingston und John Covert in einer Studie¹¹⁶ die Eignung eines simplen rückgekoppelten Netzes untersucht. Zudem haben Thomas Schmitz und Jean-Jacques Embrechts Untersuchungen^{117, 118} zur Anwendung von LSTM-Netzen durchgeführt.

Darüber hinaus existieren hybride Methoden, welche Autoencoder-Architekturen miteinbeziehen, wie in den Arbeiten^{119, 120, 121} von Marco Martínez et al. beleuchtet wird.

Die nachfolgenden Kapitel dienen dazu, einen Überblick zu geben, in welcher Weise neuronale Netze eingesetzt werden können, um Röhrenverstärker zu emulieren und wie erfolgreich unterschiedliche Methoden sind. Dazu sollen die hierfür relevantesten Modelle neuronaler Netze kurz erklärt werden, zudem wie das Training abläuft und zu welchen Ergebnissen die Studien gekommen sind. Anschließend sollen kommerzielle Software-Angebote von Verstärker-Emulation mittels künstlicher Intelligenz beleuchtet werden.

3.4.1 Relevante Modelle

Die in 3.3.2 vorgestellten Topologien sind als grundlegendste Unterscheidung zu verstehen, auf deren Basis weitere Modelle aufbauen, deren spezifische Architekturen sie mit bestimmten Attributen versehen, welche sie für den Einsatz in bestimmten Bereichen besonders effektiv machen.

¹¹³ Damskägg, E.-P., et al.: Deep Learning for Tube Amplifier Emulation, 2019.

¹¹⁴ Eero-Pekka Damskägg; Lauri Juvela; Vesa Välimäki: Real-Time Modeling of Audio Distortion Circuits with Deep Learning, 2019.

¹¹⁵ Wright, A.; Damskägg, E.-P.; Juvela, L.; Välimäki, V.: Real-Time Guitar Amplifier Emulation with Deep Learning. In: Applied Sciences 10 (2020).

¹¹⁶ Covert, J.; Livingston, D. L.: A vacuum-tube guitar amplifier model using a recurrent neural network. In: 2013 Proceedings of IEEE Southeastcon, 2013, S. 1–5.

¹¹⁷ Schmitz, T.; Embrechts, J. J.: Nonlinear Real-Time Emulation of a Tube Amplifier with a Long Short Term Memory Neural-Network, 2018.

¹¹⁸ Dies.: Real Time Emulation of Parametric Guitar Tube Amplifier With Long Short Term Memory Neural Network, 2018.

¹¹⁹ Martínez, M.; Reiss, J.: Modeling of nonlinear audio effects with end-to-end deep neural networks. In: Proceedings of the ICASSP, 2019.

¹²⁰ Martínez, M.; Benetos, E.; Reiss, J.: Deep Learning for Black-box Modeling of Audio Effects. In: Applied Sciences (2020).

¹²¹ Martínez, M.; Reiss, J.: A General-purpose Deep Learning Approach to Model Time-varying Audio Effects, 2019.

Neuronale Faltungsnetze

Sogenannte Faltungsnetze, auch bekannt als *Convolutional Neural Networks (CNN)* haben sich als besonders effektiv herausgestellt, wenn es um das Verarbeiten von Bildern oder zeitabhängigen Daten wie Audiodaten geht.¹²² Im Gegensatz zu einem gewöhnlichen Mehrschichtperzeptron besteht ein CNN aus einer Faltungsschicht, gefolgt von einer Pooling-Schicht. Eine Anordnung, die sich mehrfach direkt hintereinander wiederholen kann und schließlich in einem vollvernetzten Mehrschichtperzeptron mündet.¹²³

Die Faltungsschicht besteht aus einzelnen linearen Faltungsfiltern (*rezeptive Felder*), welche bei der Faltung mit einer bestimmten Fenstergröße über die Eingangsdaten (z.B. Bild oder Audio) geschoben werden und jeweils die gewichtete Summe für das mittlere Feld des Fensters berechnen (Siehe Abb. 10). Durch dieses Vorgehen wird es möglich, ohne die Gesamtheit der Daten verarbeiten zu müssen, auftretende Muster zu betrachten und die räumliche Struktur der Daten zu erhalten.¹²⁴

$$\begin{pmatrix} 3 & 2 & 3 & 0 & 2 & 3 & 3 \\ 3 & 3 & 0 & 3 & 0 & 2 & 3 \\ 0 & 3 & 1 & 3 & 2 & 1 & 0 \\ 3 & 0 & 3 & 2 & 0 & 3 & 1 \\ 2 & 3 & 3 & 3 & 1 & 0 & 1 \\ 0 & 3 & 3 & 2 & 0 & 1 & 2 \\ 1 & 1 & 2 & 1 & 0 & 1 & 2 \end{pmatrix} * \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 35 & 28 & 26 & 25 & 29 \\ 29 & 30 & 31 & 26 & 22 \\ 30 & 36 & 33 & 23 & 20 \\ 37 & 43 & 34 & 18 & 15 \\ 34 & 40 & 28 & 13 & 14 \end{pmatrix}$$

Bild B Maske M $B * M$

Abbildung 10: Beispiel einer Faltung. B sind numerische Eingabedaten eines Bildes, M der Faltungsfilter und $B * M$ die resultierende sogenannte Aktivierungskarte

In der sich anschließenden Pooling-Schicht wird die vorherige Schicht durch Unterabtastung verkleinert, um die Anzahl der zu trainierenden Parameter zu reduzieren. Hierfür wird ebenfalls ein Faltungsfilter angewendet, welcher die Daten, über die er bewegt wird, innerhalb des Fensters zu einem einzigen Wert zusammenfasst. Die Faltungs- und Pooling-Schichten wirken auf diese Weise wie ein Merkmalsextraktor, der spezifische Merkmale aus den Eingabedaten generiert und diese an das anschließende vollvernetzte Mehrschichtperzeptron weiterleitet, welches diese dann klassifiziert.¹²⁵

Autoencoder

Bei der Autoencoder-Architektur handelt es sich um eine beliebte Wahl für unüberwachtes Lernen. Grundlegend besteht er aus zwei Einheiten: Dem Encoder und dem Decoder, zwischen welchen sich die latente Darstellung befindet. Der Encoder hat zur Aufgabe, die Eingabedaten zu einer vereinfachten Darstellung umzuwandeln, woraus die latente Darstellung resultiert. Der Decoder hingegen hat die Aufgabe, ohne die Eingabedaten zu

¹²² Vgl. Goodfellow/ Bengio/ Courville: Deep learning, S. 326.

¹²³ Vgl. Ernst/ Schmidt/ Beneken: Grundkurs Informatik, S. 816.

¹²⁴ Vgl. ebd., S. 817.

¹²⁵ Vgl. ebd., S. 820–821.

kennen, aus der latenten Darstellung eine Ausgabe zu ermitteln, welche der Eingabe möglichst nahekommen soll.

Diese Vorgehensweise hat zum Ziel, Repräsentierungen der Daten zu finden, welche die enthaltenen Informationen bewahren, aber die Daten simplifizieren und komprimieren. Indem Encoder und Decoder parallel unüberwacht trainiert werden, wird erreicht, dass die latente Darstellung die relevanten Informationen beinhaltet. Von einem trainierten Autoencoder-Netzwerk können die einzelnen Komponenten, also Encoder und Decoder, aber besonders die latente Darstellung für unterschiedliche Aufgaben genutzt werden.¹²⁶

Long short-term memory-Netze (LSTM)

Sogenannte LSTM-Netze erweitern gewöhnliche rückgekoppelte Netze um ein Langzeitgedächtnis. Diese verfügen zwar, bedingt durch die Rückkopplung, über ein Kurzzeitgedächtnis, dieses bewertet Informationen allerdings abfallend relevant, desto weiter deren Eingabe zurückliegt, also desto mehr Wege es durch die verdeckten Schichten genommen hat. Dieses Problem wird das *vanishing gradient problem* genannt.¹²⁷

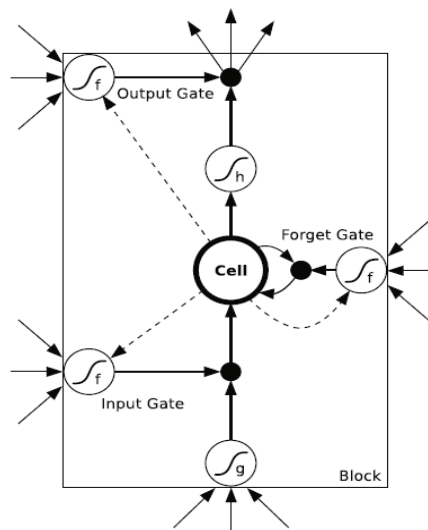


Abbildung 11: LSTM-Block mit einer Zelle. Die schwarz gefüllten Kreise stellen Gewichtungen dar. In dieser Form wird der Block zwischen die Einheiten zweier Schichten geschaltet.

Zur Lösung des Problems wird der rückgekoppelten Architektur eine zusätzliche FeedForward-Schicht hinzugefügt, welche spezielle Einheiten beinhaltet, die mithilfe von Schranken, bzw. *Gates*, den Informationsfluss zu und von den Einheiten beschränken. Die Schranken verfügen über zusätzliche Gewichtungen, und eine sigmoid-Aktivierungsfunktion, wodurch sich ein binärer Ausgangswert ergibt. Dadurch kann das *vanishing gradient problem* gelöst werden, indem Informationen entfernt werden, welcher

¹²⁶ Vgl. Lempitsky, V.: Autoencoder. In: Computer Vision: A Reference Guide, Cham 2019, S. 1–6, hier S. 1.

¹²⁷ Vgl. Graves, A.: Supervised sequence labelling with recurrent neural networks, Berlin u. a. 2012, S. 37.

es nicht länger bedarf, und jene gespeichert werden, welche für spätere Entscheidungen relevant sein könnten. Hierbei spricht man von Gates, welche jeweils eine spezielle Aufgabe haben. Das *forget gate* löscht Informationen aus dem Kontext, die nicht mehr benötigt werden, das *input gate*, welches die Informationen auswählt, welche dem Kontext hinzugefügt werden und das *output gate*, welches entscheidet, welche Informationen für den aktuellen Zustand benötigt werden.¹²⁸ Siehe Abbildung 11.

3.4.2 Konzepte und Training

Die in 3.4 erwähnten Studien basieren alle auf zuvor erläuterten Modellen und Topografien. Das wahrscheinlich simpelste Vorgehen stammt von Covert und Livingston.¹²⁹ Sie nutzten ein Netzwerk, welches auf dem NARX-Netzwerk basiert, einem Feedforward-Netz, welches um Verzögerungsleitungen an den Eingangseinheiten und Rückkopplungen der Ausgangsan die Eingangseinheiten erweitert wird, wie Abbildung 12 zu entnehmen ist. Das Netz besteht aus einer Eingangsschicht, einer versteckten Schicht aus Einheiten mit sigmoidaler Aktivierungsfunktion und einer linearen Ausgangsschicht. Es arbeitet nach der Funktion:

$$y_k = f(y_{k-1}, y_{k-2}, \dots, y_m, x_k, x_{k-1}, \dots, x_n)$$

Hierbei ist y_k der Ausgang, x_k der Eingang, y_{k-i} sind verzögerte Ausgänge, x_{k-i} verzögerte Eingänge, die Indizes m und n stehen jeweils für die Nummer der verzögerten Versionen und f ist eine nichtlineare Funktion, die von einem mehrschichtigen Feedforward-Netz bestimmt wird.

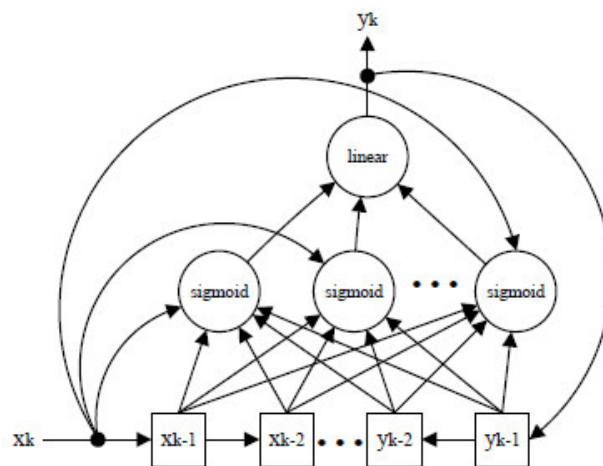


Abbildung 12: Struktur des verwendeten NARX-Netzwerks

Die Trainingsdaten bestehen aus unverarbeiteten Signalen einer Gitarre und eines Funktionsgenerators, sowie den zugehörigen zeitlich synchronisierten Ausgangssignalen eines Gitarrenverstärkers (Vox AC4TV), welcher mit eben jenen Signalen gespeist wurde.

¹²⁸ Vgl. Jurafsky, D.; Martin, J. H.: Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2023, S. 198–199.

¹²⁹ Covert/ Livingston: A vacuum-tube guitar amplifier model using a recurrent neural network.

Über die Länge an Audiomaterial, das zum Training verwendet wurde und über die Dauer des Trainings werden leider keine Angaben gemacht.

Ein weiterer Ansatz, der auf rückgekoppelten Netzen basiert, ist derer von Schmitz und Embrechts,^{130, 131} welche zusätzlich LSTM-Zellen implementieren. Den Studien kann leider nur der prinzipielle Aufbau entnommen werden, welcher in Abbildung 13 zu sehen ist.

Hierbei ist $x(n)$ ein Gitarrensinal, $y(n)$ ist ein resultierendes Signal von $x(n)$, das durch einen Röhrenverstärker gesendet wurde. Die Signale $x(n)$ werden an die LSTM-Zellen und anschließend an ein vollständig vernetztes neuronales Netz übertragen. N steht für einen Eingangs-Puffer, der benötigt wird, um im vollständig vernetzten Netz den Ausgang $pred(n)$ zu berechnen. $pred(n)$ und $y(n)$ werden anschließend zusammengeführt, um durch eine sogenannte Verlustfunktion (siehe 3.4.3), den Unterschied zwischen den beiden Signalen zu berechnen. Die Studienurheber haben diesen Ansatz erweitert, um einer Echtzeit-Anwendung möglichst nahe zu kommen, wofür das Gesamtsystem durch möglichst geringen Rechenaufwand eine kleinstmögliche Latenz aufweisen muss. Hierfür wurde eine Faltungsschicht nach $x(n)$ implementiert, um die Anzahl an LSTM-Zellen so gering wie möglich zu halten.

Für das Training wurden ebenfalls Audiodaten von Gitarrensinalen, jeweils roh und synchron durch einen Röhrenverstärker gespeist, verwendet. Die Daten ergeben, dass nur

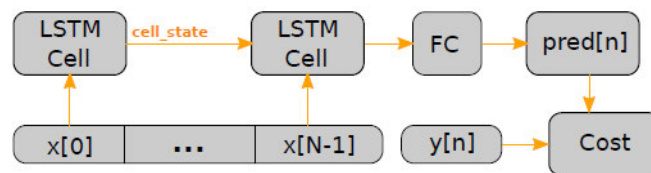


Abbildung 13: Topologie der LSTM-Methode

wenige Sekunden bis Minuten an Trainingsdaten genügen können („*In fact, we experimentally found out that a data-set of twenty seconds is already long enough to bring interesting results*“).¹³² Dargestellt wurde zudem die Genauigkeit der Ergebnisse in Abhängigkeit von der Trainingsdauer, woraus sich ergibt, dass nach ungefähr 13 Stunden Training keine Verbesserung mehr eintrat.

In einer Reihe von Studien^{133, 134, 135} untersuchen Damskägg et al. Den Einsatz von Faltungsnetzen zur Verstärker-Emulation. Emuliert wurden, je nach Studie ein Röhrenverstärker und unterschiedliche Verzerr-Pedale. Eingesetzt wurde eine Variante von

¹³⁰ Schmitz/ Embrechts: Nonlinear Real-Time Emulation of a Tube Amplifier with a Long Short Term Memory Neural-Network.

¹³¹ Schmitz/ Embrechts: Real Time Emulation of Parametric Guitar Tube Amplifier With Long Short Term Memory Neural Network.

¹³² Schmitz/ Embrechts: Real Time Emulation of Parametric Guitar Tube Amplifier With Long Short Term Memory Neural Network, S. 3.

¹³³ Damskägg/ Juvela/ Thuillier/ Välimäki: Deep Learning for Tube Amplifier Emulation.

¹³⁴ Eero-Pekka Damskägg/ Lauri Juvela/ Vesa Välimäki: Real-Time Modeling of Audio Distortion Circuits with Deep Learning.

¹³⁵ Wright/ Damskägg/ Juvela/ Välimäki: Real-Time Guitar Amplifier Emulation with Deep Learning.

WaveNet, einem neuronalen Netz von Google zur Audiogenerierung, das aus mehreren Faltungsschichten und einem Nachbearbeitungs-Modul, bei dem es sich um ein vollvernetztes neuronales Netz handelt, besteht. Der Aufbau kann Abbildung 14 entnommen werden.

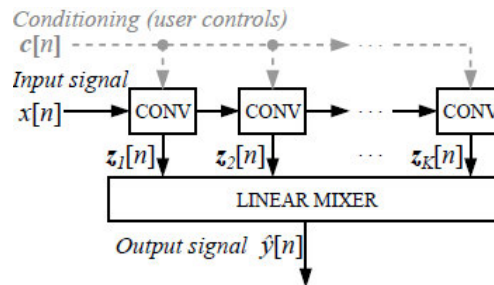


Abbildung 14: Modell des Faltungsnetzes

Das rohe Signal $x(n)$ wird in eine Serie von Faltungsschichten $conv$ gespeist, welche eine lineare Filterung und die Weitergabe durch eine nichtlineare Aktivierungsfunktion $Z_k(n)$ anwenden. Optional kann das Netzwerk zudem durch die Implementierung der Emulation von Nutzereinstellungen (z.B. Gain, EQ) am Referenzgerät durch $c[n]$ erweitert werden. In dem aus aus drei Schichten bestehende vollvernetztes Nachbearbeitungsmodul wird das Ausgangssignal \hat{y} generiert. Das Ausgangssignal wird nach folgender Formel approximiert:

$$\hat{y}[n, \theta] = f(x[n], \dots, x[n - N], \theta)$$

Dabei ist N das rezeptive Feld des Modells, f ist eine nichtlineare Transformation, die das Modell erlernt, $x[n]$ ist das rohe Eingangssignal zum Zeitpunkt n und θ stellt die gelernten Netzwerkparameter dar. Das rezeptive Feld N hängt von der Anzahl Faltungsschichten und deren Tiefe ab.

Für das Training wurden in den Studien jeweils unterschiedliche Datensatzursprünge gewählt. Bei den Emulationen der Verzerr-Pedale wurden die Rohsignale der Gitarre in die tatsächlichen Referenzgeräte gespeist, um die synchronen Referenzaudios zu erhalten, bei der Emulation des Röhrenverstärkers wurde mittels der Software SPICE eine Modellierung des Referenzverstärkers erstellt, durch welche die Referenzaudios generiert wurden. Wenige Minuten an Trainingsdaten wurden verwendet und mehrere Stunden an Trainingszeit.

In weiterführenden Studien wurde eine Echtzeit-Anwendung untersucht, indem unterschiedliche Aktivierungsfunktionen, sowie eine variierende Anzahl an Schichten bei der Faltung in Abhängigkeit zur Rechengeschwindigkeit und der Abweichung von \hat{y} zum Referenzsignal gestellt wurden. Somit sollte der optimale Kompromiss aus Klangqualität und Geschwindigkeit gefunden werden, um eine Echtzeit-Anwendung zu ermöglichen.

Bei den Arbeiten^{136, 137, 138} von Martínez et al. kommen hybride Methoden zum Einsatz, welche auf eine Autoencoder-Architektur mit Faltungsschichten und umgekehrten Faltungsschichten jeweils im Encoder und Decoder arbeiten. Hierbei bezogen sie sich auf die Modellierung genereller nichtlinearer Audioeffekte und nicht spezifisch auf Gitarrenverstärker.

Der Encoder besteht bei ihrem Modell grundlegend aus zwei Faltungsschichten, gefolgt von einer Pooling-Schicht, welche eine zeitabhängige Faltung des rohen Audiosignals durchführen. Die resultierende Darstellung im latenten Raum besteht aus zwei Schichten: Einer Schicht mit begrenzt vernetzten Einheiten und einer vollvernetzten Schicht. Der Decoder setzt sich äquivalent umgekehrt zusammen. Er besteht aus einer Schicht, die eine umgekehrte Pooling-Operation ausführt und anschließend aus zwei umgekehrten Faltungsschichten. Diese Struktur wurde in den darauf aufbauenden Arbeiten erweitert, indem die latente Darstellung um LSTM-Zellen ergänzt wurde, was eine noch komplexere Modellierung der Nonlinearitäten ermöglichte.

Trainiert wurde das System mittels roher und zugehörigen synchronen Gitarrensensoren, welche mit nichtlinearen Audioeffekten bearbeitet wurden. Das Training fand dabei in zwei Schritten statt. Zuerst wurden lediglich die Faltungsschichten im Encoder trainiert, um bereits im Voraus die Gewichtungen so einzustellen, dass die Daten der Eingangs- und Zieldaten besser rekonstruiert werden können. Anschließend wurde im zweiten Schritt das gesamte Netzwerk trainiert. Über die Länge der Trainingsdaten und des Trainingsprozesses werden keine Angaben gemacht.

3.4.3 Evaluation der Ergebnisse

Ein System, wie eines der zuletzt vorgestellten zu bewerten, meint zwei Dinge: Zum einen ist der Funktionsweise neuronaler Netze zwangsweise eine Evaluation immanent, welche entscheidenden Beitrag an der Wirkweise des Netzes leistet. Wie bereits in den Modell-Beschreibungen von 3.4.2 angedeutet wurde, gibt es während des Trainingsprozesses meist ein rohes Eingangssignal, ein Referenzsignal vom zu emulierenden Verstärker und ein vom Netzwerk erzeugtes Ausgangssignal. Das Netzwerk trainiert, den Unterschied zwischen Ausgangssignal und Referenzsignal zu minimieren. Die Unterschiedlichkeit muss hierfür objektiv messbar gemacht werden, wofür sogenannte Verlustfunktionen eingesetzt werden, welche den Abstand zwischen Ausgangs- und Referenzsignal messen und somit die Basis für die Fehlerrückführung sind, also dem Lernprozess, bei dem die Gewichtungen des Netzwerks angepasst werden.¹³⁹

¹³⁶ Martínez/ Reiss: A General-purpose Deep Learning Approach to Model Time-varying Audio Effects.

¹³⁷ Martínez/ Reiss: Modeling of nonlinear audio effects with end-to-end deep neural networks.

¹³⁸ Martínez/ Benetos/ Reiss: Deep Learning for Black-box Modeling of Audio Effects.

¹³⁹ Vgl. Vanhatalo/ Legrand/ Desainte-Catherine/ Hanna/ Brusco/ Pille/ Bayle: A Review of Neural Network-Based Emulation of Guitar Amplifiers, S. 14.

Die Verlustfunktion ist damit substanzieller Bestandteil der Wirkweise und kann unterschiedlich gestaltet werden. Die meistverwendete Funktion ist der *Mean Squared Error* (MSE), definiert als:

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=0}^{N-1} (y[i] - \hat{y}[i])^2$$

Dabei werden allerdings lediglich die Wellenformen der Signale miteinander verglichen, was nicht direkt mit der menschlichen Hörwahrnehmung korrespondiert, da bspw bestimmte Frequenzen vom Menschen lauter wahrgenommen werden als andere. Häufig eingesetzt wird auch der *Root Mean Square Error* (RMSE):

$$\mathcal{L}_{RMSE} = \sqrt{\frac{\sum_{i=0}^{N-1} (y[i] - \hat{y}[i])^2}{N}}$$

Diese Funktion wurde in der Studie von Covert und Livingston, sowie denen von Schmitz und Embrechts verwendet. Letztere (LSTM-Methode) erzielten Ergebnisse eines RMSE von unter 1%, selbst bei Echtzeitanwendung,¹⁴⁰ Erstgenannte erreichten zwar ähnliche Werte im Training, bei der tatsächlichen Anwendung außerhalb des Trainings schnitt das Netzwerk (rückgekoppeltes Netz) aber deutlich schlechter ab als erwartet.¹⁴¹

Martínez et al. Nutzen für ihre Autoencoder-Methode den sogenannten *Mean Absolute Error* (MAE), welcher wie folgt definiert ist:

$$\mathcal{L}_{MAE} = \frac{1}{N} \sum_{i=0}^{N-1} |\hat{y}[i] - y[i]|$$

Sie erreichten während des Trainings Werte weit unter 0,01%. Im Test des trainierten Netzwerks ergaben sich Werte zwischen 0,02% und 0,35%.¹⁴²

Die Studien von Damskägg et al. (CNN-Methode) nutzten alle die sogenannte *Error-to-Signal Ratio* (ESR):

$$\mathcal{L}_{ESR} = \frac{\sum_{i=0}^{N-1} (y[i] - \hat{y}[i])^2}{\sum_{i=0}^{N-1} y[i]^2}$$

In den besagten Studien wurde die ESR zudem um Filter erweitert, welche die menschliche Hörwahrnehmung implementieren.¹⁴³ Sie erreichten ebenfalls sehr gute Ergebnisse von 0,32%-0,49%¹⁴⁴ und bei Echtzeitanwendung 0,46%-0,6%.¹⁴⁵

¹⁴⁰ Vgl. Schmitz/ Embrechts: Real Time Emulation of Parametric Guitar Tube Amplifier With Long Short Term Memory Neural Network, S. 5.

¹⁴¹ Vgl. Covert/ Livingston: A vacuum-tube guitar amplifier model using a recurrent neural network, S. 3–4.

¹⁴² Vgl. Martínez/ Reiss: Modeling of nonlinear audio effects with end-to-end deep neural networks, S. 4.

¹⁴³ Vgl. Vanhatalo/ Legrand/ Desainte-Catherine/ Hanna/ Brusco/ Pille/ Bayle: A Review of Neural Network-Based Emulation of Guitar Amplifiers, S. 15.

¹⁴⁴ Damskägg/ Juvela/ Thuillier/ Välimäki: Deep Learning for Tube Amplifier Emulation, S. 5.

¹⁴⁵ Eero-Pekka Damskägg/ Lauri Juvela/ Vesa Välimäki: Real-Time Modeling of Audio Distortion Circuits with Deep Learning, S. 12.

Dies gesagt, steht auf der anderen Seite der Evaluation die Bewertung der Ergebnisse als Gesamtheit. Da objektive Evaluationen zwar relevanter Bestandteil eines neuronalen Netzes sind, die menschliche Hörwahrnehmung aber nicht angemessen widerspiegeln, sind zusätzliche Evaluationen nötig, um die Ergebnisse des trainierten Netzes zu testen. Hierfür sind Hörtests mit menschlicher Partizipation das geeignete Mittel, sogenannte *Multiple stimuli with hidden reference and anchor (MUSHRA)*-Tests werden durchgeführt.¹⁴⁶ Bei dieser Methode wird den Probanden ein gekennzeichnetes Referenzsignal präsentiert und mehrere ungekennzeichnete prozessierte Varianten des Signals, sowie eine ungekennzeichnete Kopie des Referenzsignals und mehrere ungekennzeichnete Anker-Signale, also bandbegrenzte Signalversionen. Die Probanden sollen auf einer Skala von 0 bis 100 alle ungekennzeichneten Signale daraufhin bewerten, wie ähnlich sie dem gekennzeichneten Referenzsignal sind, wobei 100 *exakt gleich* bedeuten würde.¹⁴⁷

Nicht alle erwähnten Studien haben MUSHRA-Tests durchgeführt. Covert und Livingston (rückgekoppeltes Netz), sowie Schmitz und Embrechts (LSTM-Methode) haben darauf verzichtet. Martínez et al. haben einen solchen Test ausgeführt für mehrere Varianten ihrer Autoencoder-Modelle, wobei die Emulation unterschiedlicher Audioeffekte untersucht wurde, unter anderem auch die eines Röhren-Vorverstärkers. Tatsächlich ergab der Test, dass die Ergebnisse bestimmter Varianten ihres Netzes kaum von der ungekennzeichneten Referenz zu unterscheiden waren.¹⁴⁸ Die Arbeiten von Damskäg et al. (CNN-Methode) enthalten ebenfalls Ergebnisse durchgeführter MUSHRA-Tests. In ihrer früheren Arbeit ergaben die Tests zwar gute Ergebnisse, die Ausgangssignale der neuronalen Netze waren allerdings immer noch von der Referenz zu unterscheiden.¹⁴⁹ In ihrer späteren aufbauenden Arbeit gelang es ihnen, die neuronalen Netze so zu optimieren, dass sie selbst bei Echtzeit-Anwendung Ergebnisse liefern konnten, die nahezu nicht mehr von der Referenz unterschieden werden konnten.¹⁵⁰

3.4.4 Kommerzielle Angebote

Die vorliegenden Studien legen nahe, dass es bereits möglich ist, mittels neuronalen Netzen gute Ergebnisse bei der Verstärker-Emulation zu erzielen. Tatsächlich gibt es auch schon kommerzielle Anbieter, die solche Lösungen als Soft- und Hardware verkaufen. Im Zentrum stehen dabei *IK Multimedia* mit *ToneX* und *Neural DSP* mit der Hardware-Lösung *Quad Cortex*. Beide setzen auf nicht weiter spezifizierte deep-learning-Technologie.

ToneX spricht dabei vom *AI Machine Modeling*, einer Technologie, mit welcher Verstärker, Effekte und gesamte Rigs (Verstärker, Kabinett, Mikrofon, ggf. Effekte) modelliert werden können, mit Ergebnissen, die „ununterscheidbar vom Original“¹⁵¹ sehen sollen. Bei *ToneX*

¹⁴⁶ Vgl. Vanhatalo/ Legrand/ Desainte-Catherine/ Hanna/ Brusco/ Pille/ Bayle: A Review of Neural Network-Based Emulation of Guitar Amplifiers, S. 16.

¹⁴⁷ Vgl. ITU-R. ITU-R BS.1534-3. Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems. In: Intern. Telecom. Union Radiocom, 34. Aufl., 2015, hier S. 7–9.

¹⁴⁸ Martínez/ Benetos/ Reiss: Deep Learning for Black-box Modeling of Audio Effects, S. 14, Abb. 6.

¹⁴⁹ Vgl. Damskäg/ Juvela/ Thuillier/ Välimäki: Deep Learning for Tube Amplifier Emulation, S. 5.

¹⁵⁰ Wright/ Damskäg/ Juvela/ Välimäki: Real-Time Guitar Amplifier Emulation with Deep Learning, S. 14, Abb. 13, Abb. 14.

¹⁵¹ IK Multimedia: AI Machine Modeling, <https://www.ikmultimedia.com/products/aimachinemodeling/> (26. Oktober 2023), [Übersetzung].

handelt es sich um eine Software-Lösung, die als Plugin und Standalone-Variante verfügbar ist und auch zur Echtzeit-Anwendung dient. Mittels der Software können beliebige Geräte und Rigs modelliert und ihr Klang als sogenanntes *Tone Model* gespeichert werden.

Der Konkurrent Neural DSP bietet hingegen mit dem Quad Cortex eine Hardware-Lösung an. Bei dem Gerät handelt es sich um einen Modeling-Verstärker, welcher mit sogenannten *Captures* arbeitet, dem firmeneigenen Äquivalent der *Tone Models*. Diese können ebenfalls selbst erzeugt werden und sollen zu „beispiellosen Ergebnissen“¹⁵² führen.

Beide bieten zudem Plattformen an, auf welchen die digitalen Abbilder ausgetauscht werden können, so dass es nicht zwingend nötig ist, eigene Abbilder zu erstellen.

¹⁵² Neural DSP: Quad Cortex, <https://neuraldsp.com/quad-cortex> (26. Oktober 2023).

4 Parameter der musikalischen Audioanalyse

In Vorbereitung auf die qualitative Analyse der klanglichen Eigenschaften von Verstärker-Emulationen sollen an dieser Stelle einige Klangeigenschaften, zugehörige Begriffsdefinitionen und geeignete Darstellungsformen erläutert werden. Die relevantesten Größen sind dabei Dynamik, Intensität, Lautstärke, Frequenz, Tonhöhe und das Timbre.¹⁵³

4.1 Dynamik

Der Begriff Dynamik ist als Oberbegriff zu verstehen, unter welchem alle Vorgänge und Begriffe gebündelt werden, welche die Lautstärke und deren Änderung betreffen.¹⁵⁴ Die die Dynamik betreffenden Eigenschaften können einem Wellenformdiagramm entnommen werden, in welchem die Amplitude in Abhängigkeit der Zeit nach $x=f(t)$ dargestellt wird.¹⁵⁵

4.1.1 Lautstärke

Um den Begriff der Lautstärke zu klären, gilt es einige Begrifflichkeiten zu definieren. Zum einen die **Schalleistung** P_{ak} in *Watt (W)*, welche angibt, wie viel Energie von einer Schallquelle innerhalb einer bestimmten Zeit in alle Richtungen an die Luft abgegeben wird. Die **Schallintensität** I wiederum definiert die Schalleistung auf einer Fläche. Sie wird in W/m^2 angegeben. Die **Lautstärke** hingegen ist eine komplexere Größe, die von mehreren Faktoren abhängig sein kann. So kann derselbe Klang von unterschiedlichen Personen unterschiedlich laut wahrgenommen werden, zudem spielt die Länge und die Frequenz des Klangs eine Rolle für die wahrgenommene Lautstärke. Als Einheit für die wahrgenommene Lautstärke, hier **Lautstärkepegel** genannt kommen Phon L_N zum Einsatz, welche die Lautstärke in Abhängigkeit vom frequenzabhängigen Hören des Menschen darstellen.¹⁵⁶

4.1.2 Schwingungsphasen

Ein Klang kann in mehrere Abschnitte geteilt werden, von denen jeder bestimmte Klangeigenschaften innehat. M. Dickreiter unterscheidet diese im *Handbuch der Tonstudioteknik* in Klangeinsatz, quasistationären Abschnitt und Ausklingvorgang. Der Klangeinsatz ist dabei der Einschwingvorgang, währenddessen sich das Signal aus der Ruhe zu seinem eingeschwungenen Zustand entwickelt. Dieser resultiert bei besonders kurzen Klangeinsätzen, wie denen einer Trommel, in einem Knack-Geräusch, bei den komplexen Schwingsystemen der meisten Instrumente verläuft er aber sanfter. Der quasistationäre Klangabschnitt ist der Abschnitt, in dem der Klang nur von kleinen Schwankungen geprägt ist und vereinfacht als unverändert betrachtet werden kann. In

¹⁵³ Vgl. Müller, M.: Fundamentals of Music Processing, Cham 2015, S. 19.

¹⁵⁴ Vgl. ebd., S. 24.

¹⁵⁵ Dickreiter: Handbuch der Tonstudioteknik, S. 74, Abb. 2/9.

¹⁵⁶ Vgl. Müller: Fundamentals of Music Processing, S. 24–26.

diesem Abschnitt werden die wesentlichen Komponenten des Klangcharakters eines Musikinstruments transportiert. Er mündet im Ausklingvorgang, der die Zeit nach dem Ende der Anregung bis zur völligen Ruhe beschreibt. Er verläuft desto langsamer, je weniger ein Resonanzsystem bedämpft ist. Meist sind die höheren Teiltöne stärker bedämpft, wodurch die Klangfarbe im Verlauf des Ausklingens zunehmend dunkler und weicher gefärbt wird.¹⁵⁷

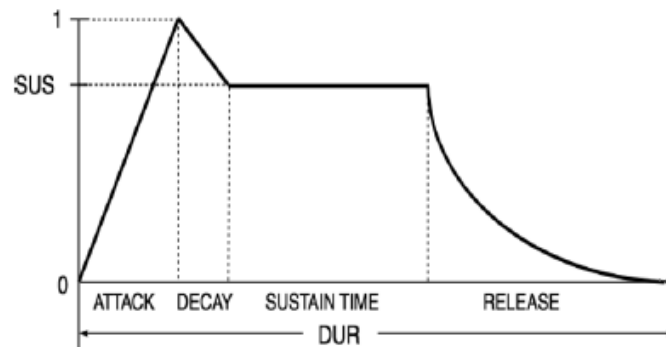


Abbildung 15: Die Amplitude eines Signals in Abhängigkeit von der Zeit unterteilt in Attack, Decay, Sustain und Release

In anderer Literatur werden diese Phasen auch auf Englisch als *Attack* (Klangeinsatz), *Sustain* (quasistationärer Klangabschnitt) und *Release* (Ausklingvorgang) bezeichnet, also mit Begriffen, die auch von den Hüllkurven bspw. eines Kompressors bekannt sind. Zudem werden die drei Phasen noch um eine vierte, den *Decay* erweitert, eine Phase zwischen Attack und Sustain, in der die Amplitude des Klangeinsatzes abfällt bis zum konstanten Level der Sustain-Phase.¹⁵⁸

4.2 Spektralverhalten

Stellt man einen einzelnen Sinuston in der Form $x = a \cdot \sin(2\pi ft)$ dar, kann aus dieser Darstellung sowohl der zeitliche Verlauf als auch die Frequenz entnommen werden.¹⁵⁹ Da es sich bei musikalischen Signalen um komplexe Klänge handelt, welche aus zahlreichen Komponenten zusammengesetzt sind, stellt es eine große Schwierigkeit dar, alle musikalisch relevanten Informationen allein aus dem Wellenformdiagramm zu gewinnen. Daher kann die Darstellung des Signals in Abhängigkeit von der Frequenz in einem sogenannten Spektrogramm ein nützliches Werkzeug sein.¹⁶⁰

¹⁵⁷ Vgl. Dickreiter: Handbuch der Tonstudioteknik, S. 73–79.

¹⁵⁸ Vgl. Devi, S.; Srinivas, Y.; Namana, D. k.: A Study: Analysis of Music Features for Musical Instrument Recognition and Music Similarity Search. In: International Journal of Computer Science and Informatics (2011), hier S. 182.

¹⁵⁹ Vgl. Dickreiter: Handbuch der Tonstudioteknik, S. 660.

¹⁶⁰ Vgl. Müller: Fundamentals of Music Processing, S. 39.

4.2.1 Tonhöhe und Frequenz

Die Frequenz ist der Faktor, der die menschliche Wahrnehmung von Tonhöhe im Wesentlichen bestimmt. Während bei reinen Sinussignalen die Frequenz direkt mit der empfundenen Tonhöhe korrespondiert, ist es bei Signalen, welche aus Grund- und Obertönen bestehen, der Grundton, der die empfundene Tonhöhe bestimmt.¹⁶¹ Musikalische Klänge, auch der Klang einer einzelnen Saite sind komplex und bestehen aus einer Mixtur vieler verschiedener Frequenzen, die sich über die Zeit verändern. Ein solcher Klang kann als Superposition einzelner Sinustöne beschrieben werden, welche in Summe das komplexe Signal bilden. Im Allgemeinen werden zwei Tonhöhen vom Menschen als gleich empfunden, wenn die Frequenzen $f_2 / f_1 = 2$ ergeben. Daraus folgt, dass die menschliche Hörwahrnehmung logarithmischer Natur ist.¹⁶²

4.2.2 Fourier-Transformation

Um ein im Zeitbereich beschriebenes Signal in die spektrale Darstellung zu konvertieren, wird es mittels der sogenannten Fourier-Analyse in seine spektralen Bestandteile zerlegt. Für zeitlich periodische Signale geschieht dies durch die Fourier-Reihenentwicklung. Diese ist die Linearkombination aller vorkommenden Sinus- und Cosinusschwingungen eines periodischen Tonsignals in folgender vereinfachter analytischer Form:

$$x(t) = A_0 + \sum_{n=1}^{\infty} A_n \cdot \cos(2\pi n f_0 + a_n)$$

A_0 beschreibt hierbei die Verschiebung des Signals aus der Amplituden-Nulllinie, A_n sind die Amplituden der einzelnen Teilschwingungen, auch *Harmonische* genannt, mit der jeweiligen Frequenz f . Die erste Harmonische ist der Grundton. Dieser hat die Amplitude A_1 und die Frequenz f_0 . Die weiteren Harmonischen sind die Oberschwingungen.

Will man ein aperiodisches Signal in die Darstellung im Frequenzbereich konvertieren, geschieht dies mittels der sogenannten Fourier-Transformation, welche wie folgt definiert ist:

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j2\pi f t} dt$$

Als Resultat der Fourier-Transformation erhält man das Fourier-Spektrum, also die spektrale Darstellung des Signals mit den Amplituden in Abhängigkeit der Frequenz. Umgekehrt kann man aus einem Signal in spektraler Darstellung mithilfe der inversen Fourier-Transformation die zugehörige Zeitfunktion mit folgender Funktion rückgerechnet werden:¹⁶³

¹⁶¹ Vgl. Dickreiter: Handbuch der Tonstudioteknik, S. 124.

¹⁶² Vgl. Müller: Fundamentals of Music Processing, S. 21–23.

¹⁶³ Vgl. Dickreiter: Handbuch der Tonstudioteknik, S. 660–661.

$$x(t) = \int_{-\infty}^{+\infty} X(f) \cdot e^{j2\pi ft} df$$

4.3 Timbre

Ein von größerer Subjektivität, aber dennoch relevanter Begriff ist das Timbre, bzw. die Klangfarbe eines Klangs. Timbre wird vage definiert als Attribut, auf dessen Basis zwei Klänge als unterschiedlich beschrieben werden können, unabhängig von den Kriterien Tonhöhe und Lautstärke. Zur Beschreibung kommen Worte wie *warm*, *hell* oder *rau* zum Einsatz.¹⁶⁴ Timbre ist ein mehrdimensionales Konzept, welches im Wesentlichen von folgenden Faktoren abhängt: Dem Abstand zwischen tonalem und Geräusch-artigem Charakter, dem spektralen Schwingungsverhalten, dem zeitlichen Schwingungsverhalten, den Änderungen des spektralen Schwingungsverhaltens und der Grundfrequenz, sowie dem Einsatz von Klängen mit hörbarem Unterschied vom quasistationären Klangabschnitt.¹⁶⁵ Um das Timbre objektiv zu vergleichen, werden häufig die *Mel-Frequenz Cepstral Koeffizienten* (MFCC) verwendet. Diese wurden ursprünglich für die automatische Spracherkennung entwickelt, haben sich aber als nützlich erwiesen, das Timbre von Signalen zu messen.¹⁶⁶

4.3.1 Mel-Spektrum

Den MFCC voraus geht der Begriff der Tonheit h . Diese beschreibt die empfundene Höhe eines Tones, welche sich nicht linear, sondern logarithmisch verhält. Zudem haben Experimente gezeigt, dass das menschliche Gehör Schallintensitäten von Einzelsignalen innerhalb eines schmalen Frequenzbereichs zusammenfasst zu einer Gesamtintensität. Diese sog. Frequenzgruppen sind über den gesamten Hörbereich hinweg ausgebildet.

Diese Erkenntnisse führten dazu, anstelle der Betrachtung des Spektrums auf einer linearen Frequenzskala, stattdessen die *Mel-Skala* einzuführen, bei welcher die Darstellung der Tonheit h in Abhängigkeit der Frequenz f nach folgender Skalentransformation erfolgt:¹⁶⁷

$$h(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700\text{Hz}} \right)$$

4.3.2 Mel-Frequenz-Cepstral-Koeffizienten (MFCC)

Die MFCC sind ein verbreitetes Werkzeug in der Audio-Analyse. Sie basieren auf der Mel-Skala und haben die Fähigkeit, kleine Veränderungen in der Wahrnehmung eines Klangs

¹⁶⁴ Vgl. Müller: Fundamentals of Music Processing, S. 26.

¹⁶⁵ Muller, M.; Ellis, D. P.; Klapuri, A.; Richard, G.: Signal Processing for Music Analysis. In: IEEE Journal of Selected Topics in Signal Processing 5 (2011), S. 1088–1110, hier S. 1097.

¹⁶⁶ Vgl. Müller: Fundamentals of Music Processing, S. 177.

¹⁶⁷ Vgl. Pfister, B.; Kaufmann, T.: Sprachverarbeitung. Grundlagen und Methoden der Sprachsynthese und Spracherkennung, Berlin, Heidelberg 2008, S. 94–95.

als kleine numerische Änderung eines Koeffizienten darzustellen.¹⁶⁸ Sie werden in der Spracherkennung, aber auch in der musikalischen Analyse, wie zur Erkennung von Musikinstrumenten,¹⁶⁹ Timbre-Analyse von Musikinstrumenten¹⁷⁰ oder zum Vergleich vom Modeling-Verstärker *Kemper* mit den emulierten Referenzverstärkern eingesetzt.^{171, 172}

Die Berechnung der MFCC gliedert sich in mehrere Schritte. Zuerst wird das Signal $x(n)$ durch einen Hochpassfilter gesendet, welcher das Ausgangssignal $y(n)$ wie folgt ergibt:

$$y(n) = x(n) - a \cdot x(n - 1)$$

Anschließend wird das Signal in kleine Einzelteile zerlegt mit jeweils einer Länge von 20-30ms, um anschließend eine Fensterfunktion in der Form $Y(n) = X(n) \cdot W(n)$ anzuwenden. Diese ist eine Funktion, die an den Extremstellen Null ergibt und somit das Signal in temporäre Fragmente zerlegt. Anschließend wird eine Fourier-Transformation auf jedes Fragment angewendet. Das resultierende Spektrogramm wird nach folgender Formel mit rechteckigen Bandpass-Filtern multipliziert und ergibt eine logarithmierte Skala des Spektrums, entsprechend der Mel-Skala:

$$Mel(f) = 1125 \cdot \ln\left(1 + \frac{f}{700}\right)$$

Schlussendlich wird eine diskrete Cosinus-Transformation vollzogen, definiert als:

$$C_i = \sum_{n=1}^{N_f} S_n \cos\left[i(n - 0,5) \cdot \left(\frac{\pi}{N_f}\right)\right], i = 1, 2, \dots, L,$$

Dabei ist C_i der i -te Koeffizient, N_f die Anzahl Bandpass-Filter, S_n die logarithmische Energie des n -ten Bandpass-Filters und L die Anzahl an zu berechnenden Koeffizienten. Durch diese Transformation wird das Mel-Spektrum zurück in den Zeitbereich konvertiert, woraus man die Mel-Frequenz Cepstral Koeffizienten als Ergebnis erhält.^{173, 174}

¹⁶⁸ Vgl. Muller/ Ellis/ Klapuri/ Richard: Signal Processing for Music Analysis, S. 1097–1098.

¹⁶⁹ Roisin Loughran; Jacqueline Walker; Michael O'Neill; Marion O'Farrell (Hrsg.): The Use of Mel-frequency Cepstral Components in Musical Instrument Identification, 2008.

¹⁷⁰ Ronchetti, V.: Violin Timbre Analysis with Mel-Frequency Cepstral Coefficients, New York 2020.

¹⁷¹ Düvel/ Kopiez/ Wolf/ Weihe: Confusingly Similar: Discerning between Hardware Guitar Amplifier Sounds and Simulations with the Kemper Profiling Amp.

¹⁷² Jan-Peter Herbst, Christoph Reuter, Isabella Czedik-Eysenberg: Guitar profiling technology in metal music production: public reception, capability, consequences and perspectives.

¹⁷³ Vgl. Ali, S.; Tanweer, S.; Khalid, S.; Rao, Naseem (Hrsg.): Mel Frequency Cepstral Coefficient: A Review, 2021, S. 4–6.

¹⁷⁴ Vgl. Saksamudre, S.; Deshmukh, R.: Comparative Study of Isolated Word Recognition System for Hindi Language. In: International Journal of Engineering Research & Technology 4 (2015), S. 536–540, hier S. 537–538.

5 Methodik

Da es sich bei heutigen Verstärker-Emulationen um eine fortgeschrittene Technik handelt, die KI-basierten Modelle in den dargelegten Studien schon teilweise kaum vom Original zu unterscheiden waren und Vorgänger wie der Kemper Profiling Amp bereits sehr gute Ergebnisse erzielen konnten, liegt den Versuchen die Annahme zugrunde, dass die KI-basierten Emulationen wie ToneX zu sehr guten Ergebnissen in der Lage sind. Auch aufgrund der Aussagen der Anbieter, dass das Ergebnis nicht vom Original unterscheidbar wäre, soll also nicht die Frage beantwortet werden, ob die Emulationen gute Ergebnisse erzielen, was sie sicherlich tun, sondern, ob überhaupt noch ein Unterschied existiert und wenn ja, worin er besteht. Deshalb sollen keine simplen Testsignale, wie kurze Impulse oder Sweeps verwendet werden, sondern tatsächliche Gitarrensигнале.

5.1 Versuchsaufbau

Für den Versuch werden zwei unterschiedliche Röhrenverstärker-Box-Kombinationen gewählt, durch welche jeweils unterschiedliche Stimuli geschickt werden sollen. Diese werden als Referenzsignale dienen. Zudem wird mit der Software *ToneX* das sogenannte *Capturing* durchgeführt, bei welchem eine digitale Emulation der Referenzverstärkers erstellt wird. Diese Emulation wird anschließend ebenfalls mit den Stimuli gespeist und die entstehenden Signale sollen verglichen werden.

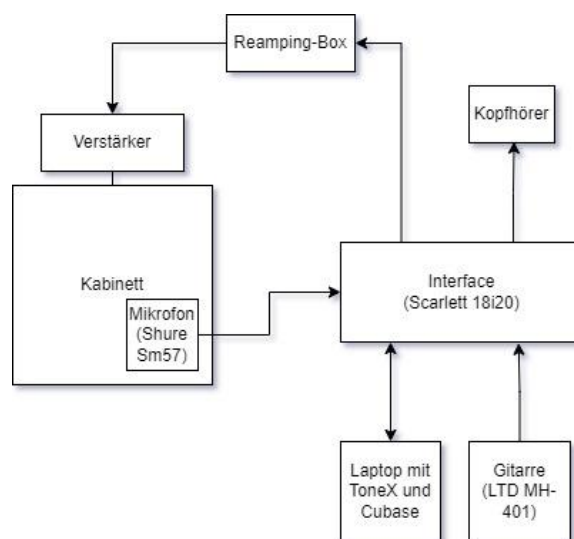


Abbildung 16: Versuchsaufbau

Zum Einsatz kommen ein *Marshall JCM900* und ein *Revv The Generator 7-40*. Beide werden jeweils mit dem Kabinett *Harley Benton G212 Celestion V30* betrieben. Verwendet werden zudem das Audiointerface *Scarlett 18i20*, die Reamping-Box *Palmer RIVER trave*, ein Laptop mit *ToneX* und der DAW *Cubase 11*, sowie die Gitarre *LTD MH-401*. Für den Versuch wird der Laptop per USB mit dem Interface verbunden, die Gitarre wird an den

hochohmigen Eingang 1 geschaltet, das vor dem Kabinett platzierte Mikrofon *Shure Sm57* an den Eingang 2. Line Out 1 der Interfaces wird mit dem Eingang der Reamping-Box und deren Ausgang mit dem Eingang des Verstärkers verbunden. Dieser wiederum ist mit dem Kabinett verbunden. Zum Abhören werden zusätzlich Kopfhörer mit dem Interface verbunden. Die Klangsteller der beiden Verstärker werden alle in neutrale Position gestellt, die Lautstärke auf einen mittleren Wert gebracht und der Gain so eingestellt, dass das Signal verzerrt, aber noch einen hohen Grad an Note Definition behält. Der Versuch findet in einem ca. 30m² großem Raum statt, welcher mit Absorbern akustisch behandelt wurde.

5.1.1 Wahl der Stimuli

Die Stimuli werden im Vorfeld des Experiments aufgenommen. Es sind DI-Signale einer Gitarre, welche unterschiedliche Spielweisen einer Gitarre widerspiegeln sollen. Das DI-Signal ist das reine Ausgangssignal der Gitarre, welches lediglich mithilfe einer DI-Box von einem hochohmigen unsymmetrischen Signal in ein niederohmiges symmetrisches Signal konvertiert wird. Dieser Vorgang ermöglicht es, das Gitarrensiegel zu speichern und bei Bedarf erneut durch einen anderen Signalweg zu leiten. Hierfür muss lediglich das gespeicherte DI-Signal, mithilfe einer Reamping-Box zurück in ein hochohmiges unsymmetrisches Signal gewandelt und in den gewünschten Signalweg geführt werden.¹⁷⁵

Es werden fünf unterschiedliche Stimuli aufgenommen, welche unterschiedliche Spielweisen beinhalten: Ein einzelner ausklingender Akkord, einige Akkordwechsel, ein kurzes Arpeggio, ein Metal-typisches Riff und ein kurzes Solo. Die Signale werden in Cubase 11 mit einer Abtastrate von 44,1kHz und 24 Bit aufgenommen. Am Anfang und am Ende der Signale wird jeweils ein kurzer Impuls eingebaut, welcher später die exakte zeitliche Synchronisation der Signale ermöglichen soll.

5.1.2 Versuchsdurchführung

Ist der Aufbau wie in 5.1 erfolgt, gliedert sich die Durchführung des Versuchs in drei Schritte. Im ersten Schritt wird das Tone Model des jeweiligen Setups mit ToneX erstellt. Hierfür wird die Funktion *Modeler* in ToneX gestartet und einige Grundeinstellungen getroffen, bei welchen festgelegt wird, ob es sich um Gitarre oder Bass handelt und was für ein Setup modelliert wird (in diesem Fall Verstärker und Box). Anschließend werden Eingangs- und Ausgangslautstärken passend eingestellt, wonach bereits alles bereit für das Capturing ist. Das Capturing dauert ca. 7 Minuten und macht sich durch Testsignale bemerkbar, welche von ToneX an den Verstärker geschickt und mit dem Mikrofon wieder abgenommen werden. Die Testsignale bestehen dabei zum allergrößten Teil aus tatsächlichen Gitarrensiegeln.

Ist das Capturing abgeschlossen, wird im Anschluss das neuronale Netz trainiert. Hierfür kann ein Grad der Genauigkeit festgelegt werden, von welchem wiederum die Dauer des Trainings abhängt. Für diese Arbeit wurde der Grad *Advanced*, die maximale Genauigkeit, gewählt. Daraufhin beginnt der Trainingsprozess, welcher ca. eine Stunde beansprucht. Ist

¹⁷⁵ Vgl. Mynett: Metal Music Manual, S. 113.

dieser abgeschlossen, kann das Tone Model mit dem Signal des Referenzverstärkers verglichen und der Verzerrgrad mit diesem feinabgestimmt werden. Wird im Anschluss ein *Skin*, also ein optisches Erscheinungsbild des Tone Models, so wie es in der Software dargestellt wird gewählt, ist der Prozess abgeschlossen.



Abbildung 17: Versuchsaufbau mit JCM900

Der zweite Schritt des Versuchs erfolgt direkt im Anschluss an das Capturing, wobei es wichtig ist, dass absolut nichts am Aufbau verändert wird. Dies ist wesentlich, da im zweiten Schritt die Stimuli in den Verstärker gespeist und durch das Mikrofon wieder aufgenommen werden. Die entstehenden Referenzsignale müssen mit einem identischen Aufbau erfolgen, um wirklich mit dem Capture vergleichbar zu sein. Die Referenzsignale werden in Cubase aufgenommen und gespeichert.

Der Dritte Schritt ist unabhängig vom Versuchsaufbau und kann zu einem beliebigen späteren Zeitpunkt erfolgen. Die Stimuli dienen hierbei als Eingangssignal für ToneX mit dem geladenen Tone Model des Referenzverstärkers. Somit werden die Vergleichssignale erzeugt, welche später mit den Referenzsignalen verglichen werden sollen.

5.1.3 Nachbereitung

Um die Signale vergleichbar zu machen, ist eine Nachbearbeitung der Aufnahmen notwendig. Die Referenz- und Vergleichssignale werden zeitlich synchronisiert und die Phasen exakt aneinander angeglichen. Die Signale werden in der Lautstärke normalisiert und so geschnitten, dass Referenz- und Vergleichssignal jeweils den gleichen Start- und Endpunkt haben.

Zudem wird ein zusätzliches Signal erzeugt, welches nachfolgend Differenzsignal genannt werden soll. Dieses wird erzeugt, indem die deckungsgleichen Referenz- und Vergleichssignale miteinander addiert werden, nachdem eines der beiden in der Phase

invertiert wird. Durch den Effekt der destruktiven Interferenz werden somit gleiche Signalanteile ausgelöscht und es entsteht ein Signal, welches lediglich jene Signalanteile enthält, in welchen sich Referenz- und Vergleichssignal unterscheiden. Dieses soll dazu dienen, Unterschiede konkret zu identifizieren. Pro Stimulus und Verstärker resultieren also drei unterschiedliche Signale: Referenzsignal, Vergleichssignal und Differenzsignal.

5.2 Parameter und deren Darstellung

Die Entstehenden Signale sollen anschließend einem qualitativen Vergleich unterzogen werden, mit dem Ziel, klangliche Unterschiede zwischen Verstärker und Emulation zu identifizieren. Für den klanglichen Unterschied als wesentliche Parameter gelten Spektralverhalten, Dynamik und Timbre. Diese können durch Spektrogramm, Wellenformdiagramm und Mel-Frequenz-Cepstral-Koeffizienten dargestellt werden.

Für den Vergleich wird pro Stimulus und Verstärker ein Wellenformdiagramm des Gesamtsignals erstellt, in welchem Referenz- und Vergleichssignal zusammen abgebildet sind, sowie ein Einzelnes Diagramm des Differenzsignals. Zudem werden ggf. stark vergrößerte Ansichten erstellt, welche das Untersuchen des genauen Wellenverlaufs von einzelnen Zeitpunkten der Signale zulassen, sollten bestimmte Elemente von besonderem Interesse sein. Als Einheit der Lautstärke wird hierfür dBFS gewählt. Für jedes Signal wird pro Verstärker zudem ein Spektrogramm erstellt, welches den spektralen Verlauf im relevanten Frequenzbereich zwischen 80Hz und 20kHz darstellt.

Für den komplexen Parameter Timbre kommen die MFCC zum Einsatz. Diese lassen zwar im Vergleich zum Wellenformdiagramm oder Spektrogramm keine präzisen Rückschlüsse über das Signal zu, dienen aber als Maß für die Unterschiedlichkeit der beiden Signale über Dynamik und Spektralverhalten hinaus. Extrahiert werden die ersten zwölf Koeffizienten. Da hierbei mehrere zehntausende einzelne Werte entstehen, werden diese weiter aufbereitet. Erstellt wird pro Stimulus und Verstärker ein Kastendiagramm, welchem die Verteilung, Minimal- und Maximalwerte, sowie Mittelwert und Median entnommen werden können. Die MFCC des Differenzsignals werden nicht mit einbezogen, da aufgrund des hohen Abstraktionsgrads des Parameters die Aussagekraft fraglich wäre.

Die Erstellung der Diagramme, sowie die Extraktion der MFCC erfolgt mittels der Software *Sonic Visualiser*,¹⁷⁶ die Kastendiagramme werden mit Microsoft Excel erstellt.

¹⁷⁶ Cannam, C.; Landone, C.; Sandler, Mark (Hrsg.): *Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files*, 2010.

6 Ergebnisse

Nachfolgend sollen die Ergebnisse der Messungen, durch die in 5.2 vorgestellten Darstellungsformen präsentiert und analysiert werden. Gegliedert wird dieses Vorgehen nach den Stimuli, zu welchen jeweils die unterschiedlichen Parameter der beiden Verstärker im Vergleich zu ihren Tone Models und das resultierende Differenzsignal dargestellt werden. Diese sollen analysiert und wesentliche Unterschiede herausgearbeitet werden. Anschließend wird ein Fazit gezogen, ob KI-Modeler wie ToneX halten, was sie versprechen, ununterscheidbar vom Original zu sein, und es wird ein Ausblick auf weitere mögliche Forschungen gegeben.

Die Wellenformdiagramme sind immer gleich definiert: Blau ist das Signal des Röhrenverstärkers (Referenzsignal), Grün das Signal von ToneX (Vergleichssignal) und Lila das Differenzsignal. Die beiden Verstärkersignale werden in einem Diagramm zusammen dargestellt, das Differenzsignal einzeln. Das oberste Spektrogramm ist stets das Signal des Röhrenverstärkers, das Mittlere das Signal von ToneX und das Untere das des Differenzsignals. Die Diagramme der MFCC sind einzeln beschriftet. Aufgrund der Menge an aus der Methodik resultierenden Graphen und Diagrammen werden diese in den Anhang eingefügt und nur einzeln an geeigneten Stellen in den Fließtext eingearbeitet. Die dieser Arbeit beigelegte CD enthält alle aus dem Versuch resultierenden Signale, sowie hochauflösende digitale Versionen, der im Anhang befindlichen Graphen.

6.1 Analyse der Messergebnisse

6.1.1 Stimulus 1: Ausklingender Akkord

Das Signal eines einzelnen Akkordes besitzt eine hohe Komplexität, da mehrere Saiten gleichzeitig schwingen, woraus viele Intermodulationsverzerrungen resultieren. Das Signal ist zu Beginn stark verzerrt, ab Sekunde fünf verliert es dann zunehmend an Verzerrung und auch an tiefem Frequenzgehalt. Aber Sekunde zwölf ist quasi nur noch ein Ausklingen von Obertönen zu hören, welches einen leicht Feedback-artigen Charakter besitzt.

Die Wellenformen des Gesamtsignals (Anhang 1.1) von Referenz- und Vergleichssignal decken sich im ersten Drittel des zeitlichen Verlaufs nahezu komplett, sowohl beim JCM900 als auch beim Revv Amp. Je mehr der Akkord ausklingt und an Feedback-Charakter gewinnt, desto mehr unterscheidet sich das Vergleichssignal des JCM900 allerdings von der Referenz. Ab Sekunde fünf sieht man eine zunehmende Abweichung, da das Vergleichssignal schneller und mehr an Pegel verliert. Dieser Effekt tritt in dieser Deutlichkeit allerdings nicht beim Revv Amp auf.

Das Differenzsignal verdeutlicht diese Tendenz. Es verliert deutlich später an Pegel als die Verstärkersignale, was auf eine zunehmende Unterschiedlichkeit beim Ausklingen schließen lässt. Dieser Effekt ist bei beiden Verstärkern zu sehen, beim JCM900 ist er

allerdings zu Beginn des Signalabfalls deutlicher, während er beim Revv Amp gegen Ende stärker wird, wo sich das Differenzsignal trotz weiter abfallenden Pegels von ToneX und Verstärker ab Sekunde elf verstetigt.

Zudem bringt das Differenzsignal etwas zu Tage, was aus den beiden Verstärker-Graphen nicht hervorginge: Die größte Differenz der Signale tritt ganz am Anfang beim Anschlagen des Akkordes auf, wobei der Differenzpegel beim JCM900 länger nach oben ausschlägt, was vermuten lässt, dass beim Revv Amp vor allem beim Attack der einzelnen Saiten die Differenz kulminiert, während beim JCM900 auch die Decay-Phase deutlich abweicht.

Generell hat das Differenzsignal einen deutlich geringeren Pegel als die Verstärker-Signale, wobei es auch hier Unterschiede gibt. Während das Differenzsignal beim JCM900 im Schnitt ca. 10dB leiser ist, ist es beim Revv Amp meist 20dB leiser, teils sogar noch mehr. Referenz- und Vergleichssignal sind sich beim Revv Amp hier also deutlich ähnlicher.

Dies verdeutlicht auch die Vergrößerung auf den Beginn der zwölften Sekunde (Anhang 1.1), in der die Saiten nur noch schwach nachschwingen. Bei beiden Verstärkern sehen sich die Kurvenformen der Verstärker-Signale sehr ähnlich, während beim Revv Amp die Kurven aber nahezu identisch sind, ist beim JCM900 die Amplitude des Vergleichssignal stets etwas geringer als die der Referenz. Dies erklärt auch die größere Abweichung des Differenzsignals.

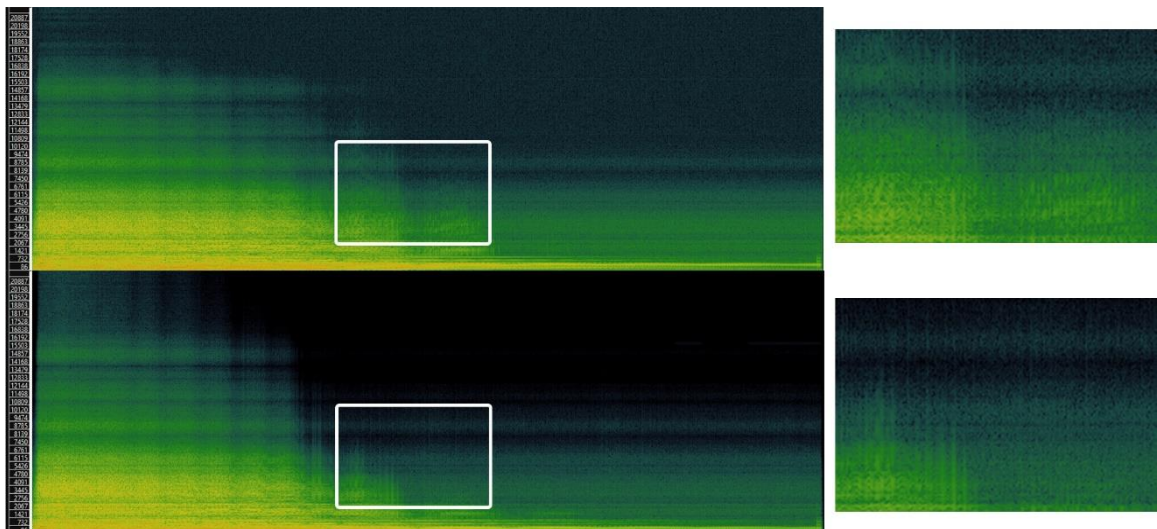


Abbildung 18: Obertonverlust während Releasephase beim JCM900 (Oben Röhrenverstärker, unten ToneX)

Im Spektrogramm (Anhang 1.2) sieht man zuerst einen deutlichen Unterschied: Einen Gehalt an Obertönen des Referenzsignals bei abfallendem Pegel, der beim Vergleichssignal nicht existiert. Dieser ist allerdings mit größter Wahrscheinlichkeit auf das Grundrauschen des Lautsprecherkabinetts zurückzuführen, welches bei ToneX nicht existiert und somit zu vernachlässigen ist. Dennoch ist darüber hinaus bei beiden Verstärkern im Referenzsignal ein ausgeprägter Obertongehalt zu vermerken, welcher vor allem bei zunehmenden Signalabfall deutlich sichtbar ist (siehe Abbildung 18). Offenbar verlieren die Signale von ToneX beim Ausklingen schneller an Obertönen, besonders zwischen 700Hz-10kHz. Dieser Effekt ist beim JCM900 deutlich stärker ausgeprägt.

Zudem sieht man bei beiden Verstärkern, dass der Obertongehalt ab ca. 16kHz deutlicher voneinander differenziert ist. Wo beim Referenzsignal einzelne horizontale Linien in bestimmten Frequenzbereichen zu erkennen sind, sieht man beim Vergleichssignal keine Abgrenzungen.

Die MFCC (Anhang 1.3) sind sich bei beiden Verstärkern sehr ähnlich. Bis auf einzelne Ausreißer beim ersten Koeffizienten, liegen lediglich Abweichungen im unteren einstelligen Bereich vor, was auf ein stark ähnliches Timbre von Vergleichs- und Referenzsignalen schließen lässt.

6.1.2 Stimulus 2: Akkordwechsel

Das zweite Signal beinhaltet mehrere Akkorde, welche teils mehrfach hintereinander angeschlagen und danach eine Weile ausgehalten werden. Der letzte Akkord klingt wie auch zuvor vollständig aus. Das Signal verhält sich somit ähnlich wie das Vorige, in dem Punkt, dass es während der Akkordanschläge sehr verzerrt ist und ab dem Ausklingen des letzten Akkordes bei Sekunde zehn an Verzerrung und tiefen Frequenzen verliert.

Mit Blick auf die Wellenformdiagramme (Anhang 2.1) bestätigt sich, was bereits zuvor geschildert wurde: Das Vergleichssignal deckt sich zuerst mit dem Referenzsignal, verliert beim Ausklingen ab Sekunde 19 allerdings wieder schneller an Pegel als die Referenz. Die Differenzsignale fallen erneut später ab als die Verstärker-Signale und die größten Auslenkungen des Differenzsignals sind beim Anschlagen der Akkorde zu finden. Auch hier ist der Decay nach einem Akkordanschlag beim JCM900 mehr betroffen als beim Revv Amp, wie z.B. bei Sekunde 10,4 sichtbar wird. Dieser Effekt wird hier tatsächlich nochmal deutlich, da im Wellenformdiagramm des Differenzsignals vom Revv Amp ziemlich genau ablesbar ist, wann ein Akkord angespielt wird, anhand der deutlichen kurzen und spitzen Ausschläge, wie in Abbildung 19 zu sehen ist.

Obwohl der Pegelunterschied zwischen Verstärker- und Differenzsignalen wieder ähnlich wie zuvor ist, nämlich 10dB Unterschied beim JCM900 und 20dB beim Revv Amp, betragen die Pegelunterschiede bei den starken Auslenkungen beim Akkordanschlag tatsächlich bei beiden Signalen nur noch wenige dB. Es besteht also eine hohe Unterschiedlichkeit der beiden Signale, besonders beim Attack und Decay. Zudem weist das Differenzsignal des Revv Amp eine höhere Pegelvarianz auf, die Unterschiedlichkeit ist also generell geringer, unterliegt aber größeren Schwankungen als die des JCM900.

Vergrößert man einen Akkordanschlag (Anhang 2.1), sieht man trotz genereller Ähnlichkeit doch einige Unterschiede. Gerade beim JCM900 sieht man, dass es häufig kleine Phasenunterschiede zwischen den beiden Verstärkersignalen gibt, zudem ist das Vergleichssignal an Stellen, an denen das Referenzsignal starke Auslenkungen hat, öfters verflacht und deutet die Auslenkung nur minimal an, so wie bei 8,918s, wo sowohl Amplitude als auch Phasenlage leicht abweichen, woraus ein hoher Differenzpegel resultiert. Beim Revv Amp ist dieses Phänomen deutlich weniger ausgeprägt, besonders die Phasen passen hier besser und bei den Extremstellen ist es im Gegensatz zum JCM900

eher das Vergleichssignal das zu hohe Amplituden aufweist, was hier die Maxima des Differenzsignals erklärt. Generell sind sich die Signale des Revv Amp ähnlicher.

Die Spektrogramme (Anhang 2.2) zeigen ebenfalls ein ähnliches Bild wie bei 6.1.1, besonders was den Obertongehalt beim Ausklingen angeht, auch wenn der Verlust an Obertönen hier etwas geringer ausfällt. Das Differenzsignal zeigt, dass beim JCM900 die größte Differenz in den tiefen Höhen zwischen 2kHz-5kHz liegt. Hier hat das Differenzsignal die größte Intensität. Beim Revv Amp geht dies nicht hervor, hier sind die Intensitäten der

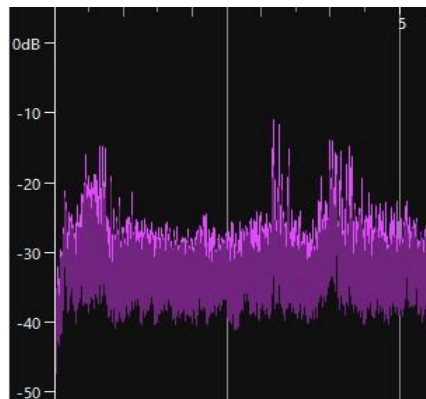


Abbildung 19: Differenzsignal des Revv Amp. Hohe Amplituden bei den Akkordanschlägen sind zu erkennen.

Frequenzen zwischen 80Hz-6kHz des Differenzsignals relativ gleichmäßig betroffen, mit einer leichten Tendenz zu höherer Intensität bei den Frequenzen zwischen 80Hz-700Hz, allerdings liegt hier auch die höchste Intensität der Verstärkersignale generell.

Die MFCC (Anhang 2.3) sind beim JCM900 nahezu identisch, lediglich minimale Abweichungen liegen vor. Überraschend ist, dass beim Revv Amp im Gegensatz dazu recht große Abweichungen zu verzeichnen sind. Teilweise um einen Wert von zehn oder sogar höher. Möglicherweise resultieren diese Unterschiede aus den großen Pegelunterschieden beim Akkordanschlag.

6.1.3 Stimulus 3: Arpeggio

Das Arpeggio ist ein deutlich kürzeres Signal aus einzeln gespielten Tönen von der tiefsten bis zur höchsten Saite, während die hohen Töne in der Mitte des Signalverlaufs teils übereinander ausgehalten werden. Somit entsteht in der Mitte des Signals eine stärkere Verzerrung als zu Beginn und gegen Ende.

Den Wellenformdiagrammen (Anhang 3.1) der Gesamtsignale der Verstärker kann hier keine deutliche Information entnommen werden. Leichte Unterschiede werden lediglich in der Mitte des Signalverlaufs sichtbar. Bei den gemeinsam klingenden, stärker verzerrten hohen Tönen ist bei beiden Verstärkern eine minimal stärkere Auslenkung des Vergleichssignals zu sehen. Den Differenzsignalen sieht man teils deutlich an, wenn eine

Note angespielt wird, da diese dann deutlich ausschlagen. Beim Revv Amp ist dieser Effekt wieder etwas stärker ausgeprägt.

Generell besteht wieder eine Differenz von 10dB Lautstärkeunterschied zwischen den beiden Differenzsignalen. Die hohen Amplituden der Differenzsignale bleiben bei den tieferen Saiten zu Beginn und am Ende des Signals länger oben und sind weniger eindeutig als die spitzen Auslenkungen bei den höheren Saiten in der Mitte des Verlaufs. Es liegt also die Vermutung nahe, dass bei tieferen Tönen der Decay einer Note sich mehr unterscheidet als bei einem höheren Ton, wo vordergründig der Attack den Unterschied ausmacht.

Zoomt man auf den Anschlag einer tiefen Saite (Anhang 3.1), sieht man beim JCM900, dass beim Vergleichssignal die spitzen Verläufe des Referenzsignals an vielen Stellen, besonders bei 0,91s-0,9105 und 0,918s-0,9185s, deutlich verflacht und lediglich grob angedeutet dargestellt werden (Siehe Abbildung 20). Dies sind auch die Stellen, an denen die größten Maxima des Differenzsignals entstehen. Beim Revv Amp ist dieser Effekt deutlich weniger ausgeprägt. Die Differenzmaxima entstehen vor allem durch unterschiedliche Amplituden der Minima der beiden Verstärkersignale. Bei den Maxima der Verstärkersignale erreicht das Differenzsignal tatsächlich häufig einen Pegel von $-\infty$.

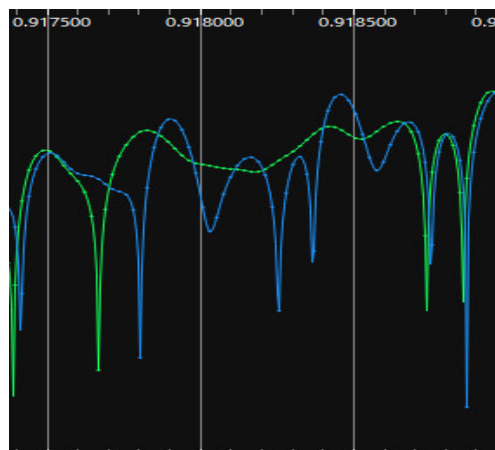


Abbildung 20: Verflachte Kurve des Vergleichssignals im Kontrast zum Referenzsignal (JCM900)

Die Spektrogramme (Anhang 3.2) sehen sich jeweils ziemlich ähnlich, es kann kaum ein Unterschied festgestellt werden. Dem Differenzsignal kann entnommen werden, dass zumindest beim JCM900 eine etwas größere Differenz der Signale bei den tiefen Saiten zu Beginn des Signals besteht, besonders zwischen 80Hz-6kHz. Bei den höheren Saiten ist die Differenz im Bereich von 2,5kHz-5kHz wieder deutlicher. Beim Revv Amp sieht man, dass eher die Signalanteile über 8kHz etwas unterschiedlicher sind, auch wenn der Unterschied nicht sonderlich gravierend ausfällt.

Bei den MFCC (Anhang 3.3) zeigt sich wieder ein ähnliches Bild wie bei 6.1.2. Während sich die MFCC des JCM900 erneut sehr ähnlich sind, mit lediglich minimalen Abweichungen, zeigen sich beim Revv Amp wieder Abweichungen mit einem Wert von zehn und höher. Es bleibt die Vermutung, dass sich dies aus den Pegelunterschieden beim

Notenanschlag ergibt, da es sich hier um ein Signal mit viel Attack und Decay und wenig Sustain und Release handelt.

6.1.4 Stimulus 4: Kurzes Riff

Das kurze Riff beinhaltet viele unterschiedliche Elemente, welche größtenteils einzeln, aber schnell hintereinander weg erklingen. Es beinhaltet ausgespielte Noten, sowohl Hohe als auch Tiefe, teils auch überlappende Töne, sowie Palm Mutes (Dämpfung der Saite mit dem Handballen) und im Generellen viel Attack.

Den Wellenformdiagrammen (Anhang 4.1) des Gesamtsignals der beiden Verstärker kann man recht deutlich entnehmen, wo Palm Mutes gespielt werden, da die allgemeine Auslenkung nach oben und unten hier geringer wird, wie bei 0,5s-0,8s oder bei 5,2s-5,6s. Den Wellenformdiagrammen können keine konkreten Abweichungen der Verstärkersignale entnommen werden, die Differenzsignale zeigen ebenfalls eine deutlich geringere generelle Pegelabweichung zueinander als bei den Signalen zuvor.

In der vergrößerten Darstellung (Anhang 4.1) kann man einen einzelnen tiefen Ton sehen, der als Palm Mute gespielt wird. Bei beiden Verstärkern zeigt sich ein ähnliches Verhalten. Während des Attack zwischen 5,32s-5,35s sind einige Unterschiede im Wellenverlauf zu sehen und auch das Differenzsignal hat teils hohe Ausschläge. Zwischen 5,36s-5,4s kann von einer kurzen Sustain-Phase gesprochen werden (Sustain und Decay sind hier schwer zu unterscheiden), in welcher die Verstärker-Signale ziemlich identisch verlaufen, bis auf die unterschiedlichen Amplituden der Minima, welche besonders beim JCM900 zu sehen sind. Mit dem Ausklingen von 5,4s-5,45s, unterscheiden sich die Amplituden der Maxima zunehmend, da das Vergleichssignal erneut stärker abfällt als die Referenz.

Den Spektrogrammen (Anhang 4.2) sind nur wenig Unterschiede zu entnehmen. Lediglich beim Revv Amp fällt auf, dass im Vergleichssignal um 13,5kHz kaum Frequenzgehalt existiert. Eine deutliche horizontale schwarze Linie ist zu erkennen. Die Spektrogramme der beiden Differenzsignale sind sich ebenfalls sehr ähnlich, besonders zwischen 80Hz-7kHz ist bei beiden eine stetige leichte Differenz zu erkennen.

Die MFCC (Anhang 4.3) weisen das erste Mal Unterschiede beim JCM900 auf, auch wenn diese nicht gravierend sind. Die Unterschiede bewegen sich jeweils um einen Wert von fünf. Das gleiche Bild ergibt sich für den Revv Amp, was der Vermutung widerspricht, dass die MFCC beim Revv Amp besonders dann große Unterschiede aufweisen, wenn ein Signal viel Attack enthält.

6.1.5 Stimulus 5: Kurzes Solo

Das Solo besteht aus ausschließlich einzeln angespielten Tönen. Hohe Saiten machen den größeren Anteil aus, aber auch mittlere und tiefe Saiten kommen zum Einsatz. Von 0s-5,5s werden die Töne länger ausgehalten, von 5,5s-8s wird schneller gespielt, von 8s bis zum Schluss werden die Töne wieder länger ausgehalten. Die Verzerrung bleibt einheitlich.

Während die Wellenformdiagramme der Gesamtdarstellungen der Verstärkersignale (Anhang 5.1) wenig aufschlussreich sind, lassen die Differenzsignale wieder deutlich das Anspielen der Noten erkennen. Die Differenzsignale schlagen besonders beim Anschlagen von hohen Tönen aus und wieder zeigt sich eine ungefähre Differenz von 10dB zwischen beiden Differenzsignalen, auch wenn die Ausschläge beim Attack häufig einen ähnlichen Spitzenpegel erreichen. Zudem zeigt sich eine hohe Varianz des Ausschlages bei der letzten Note ab 11,5s des Signals, welche einen längeren Moment ausgehalten wird.

Vergößert man die Sustainphase dieser Note (Anhang 5.1), sieht man deutlich, dass der Kurvenverlauf zwar von beiden Verstärkersignalen nahezu perfekt identisch ist, die Signalminima aber wieder einmal deutlich unterschiedlich ausfallen. Dabei schlägt scheinbar willkürlich mal die Amplitude des Referenzsignals und mal die des Vergleichssignals deutlich weiter nach unten aus. Dadurch ergeben sich bei den Minima der Verstärkersignale häufig Maxima beim Differenzsignal. Dieses Phänomen ist beim JCM900 etwas deutlicher ausgeprägt und erklärt die hohe Varianz beim letzten Ton.

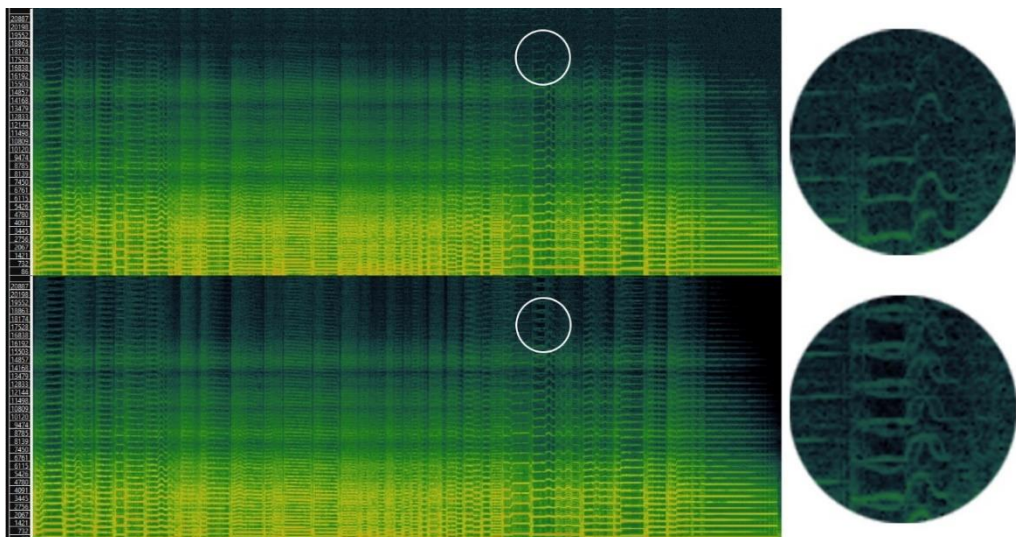


Abbildung 21: Zusätzliche Obertöne im Vergleichssignal (unten) beim JCM900.

Die Spektrogramme (Anhang 5.2) bestätigen erneut, was schon zuvor gesagt wurde. Die Differenzsignale offenbaren wieder einmal, dass die größte Abweichung beim JCM zwischen 2,5kHz-5kHz liegt und beim Revv Amp zwischen 80hZ-6kHz. Generell ist die Intensität des Spektrogramms vom Differenzsignal des Revv Amp im Bereich seiner größten Abweichung deutlich geringer als die Intensität des Differenzsignals des JCM900 bei seiner größten Abweichung. Das Tone Model des JCM900 weist allgemein eine stärkere spektrale Abweichung auf als das des Revv Amp, was sich durch die höhere Intensität des Differenzsignals zeigt. Zudem verliert das Vergleichssignal bei der letzten kurz ausgehaltenen Note schneller an Obertönen, was sich besonders beim JCM900 zeigt.

Zusätzlich zeigen die Spektrogramme etwas, was zuvor nicht sichtbar war, nämlich dass ToneX scheinbar einige zusätzliche Obertöne generiert. Dies fällt besonders im letzten Drittel der Spektrogramme der Verstärkersignale beider Verstärker ab ca. 13kHz auf. Die Intensität ist allerdings recht gering (Siehe Abbildung 21).

Bei den MFCC (Anhang 5.3) ergeben sich beim JCM900 Abweichungen um die Werte 5-8 für nahezu jeden Koeffizienten. Die letzten Koeffizienten weichen deutlich mehr ab als die ersten. Ähnliches ergibt sich für den Revv Amp, nur dass die Abweichung hier generell sehr viel geringer ist. Die Abweichung steigt zwar bei den hinteren Koeffizienten bis zu einem Wert von 5, die ersten sechs Koeffizienten sind dafür nahezu exakt identisch.

6.2 Auswertung der Ergebnisse

6.2.1 Zusammenfassung der Analyse

Um ein generelles Fazit aus den Messungen und deren Analyse zu ziehen, soll diese zuerst zusammengefasst und auf einen allgemeinen Nenner gebracht werden. Anschließend soll ein Fazit zu der Qualität der erstellten Tone Models gezogen werden und ob diese das von IK Multimedia gegebene Versprechen halten.

Generell lässt sich sagen, dass aus den Wellenformdiagrammen deutlich wird, dass das Signal von ToneX dem des Röhrenverstärkers wirklich sehr nahekommt. Besonders in der vergrößerten Ansicht wird deutlich, dass die Kurvenformen nahezu identisch sind und dass die Abweichungen, die man besonders in der Releasephase der ausgehaltenen Töne im Wellenformdiagramm der Gesamtsignale sehen kann, nicht auf eine unterschiedliche Kurvenform, sondern lediglich auf eine leicht geringere Amplitude des Vergleichssignal zurückzuführen sind. Zudem entstehen Unterschiede, besonders beim JCM900, häufig da, wo die Kurvenform zwar sehr ähnlich, die Phase aber minimal verschoben ist.

Unterschiede, die man im Gegensatz zu den Vorherigen als nicht zu vernachlässigend betrachten könnte, entstehen da, besonders beim Attack, wo ToneX den Kurvenverlauf im Vergleich zum Röhrenverstärker deutlich verflacht und tatsächlich eine unterschiedliche Wellenform resultiert. Ein Phänomen im Übrigen, das auch schon in der Untersuchung des Verhaltens des *Kemper Profiling Amps* von Terence Dau auffiel,¹⁷⁷ und ein gängiges Verhalten digitaler Emulationen zu sein scheint. Allerdings muss man sagen, dass das Phänomen bei ToneX sehr viel weniger ausgeprägt ist als beim Kemper. Dafür hat ToneX scheinbar Probleme mit den Signalminima, welche in ihrer Amplitude immer wieder deutlich vom Referenzsignal variieren und zu hohen Ausschlägen beim Differenzsignal führen. Da die Kurvenform aber dennoch fast identisch ist, nur der Ausschlag weniger stark ist, kann man davon ausgehen, dass auch dies zu vernachlässigen ist.

Zwischen beiden Verstärkern liegen zudem Unterschiede. Generell ist das Differenzsignal des JCM900 meist 10dB leiser als das Referenzsignal, was durch die logarithmische Skala bereits signifikant ist und als Halbierung betrachtet werden kann. Das Differenzsignal des Revv Amp ist aber im Schnitt nochmal weitere 10dB leiser als das des JCM900. Der Revv Amp hat weniger Probleme mit den Amplituden beim ausklingenden Signal als der JCM900, dafür hat das Differenzsignal sehr viel deutlichere Ausschläge in der Attackphase, beim Anschlagen einer Saite. Hier sind die Ausschläge beider Differenzsignale fast gleich hoch,

¹⁷⁷ Vgl. Terence Dau: Qualitativer Vergleich der Performance verschiedener Gitarrenamps mit den entsprechenden Emulationen des Kemper Profiling Amps, S. 76.

obwohl das Differenzsignal des Revv Amp sonst deutlich leiser ist. Der Revv Amp weist also generell eine größere Varianz in der Unterschiedlichkeit der Verstärkersignale auf. Der höhere Ausschlag des Differenzsignals beim Anschlag einer Note ist zwar auch beim JCM900 zu sehen, der Unterschied zum sonstigen Signalverlauf ist aber geringer und zudem halten die Ausschläge länger an. Während beim Revv Amp also vor allem die kurze Attackphase betroffen ist, ist es beim JCM900 Attack- als auch Decayphase.

Bei beiden Verstärkern entstehen die deutlichsten Unterschiede beim Anschlagen einer Note. Unterschiede liegen darin, dass der JCM900 größere Unterschiede als der Revv Amp beim Ausklingen aufweist, die Amplituden und Phase betreffend und dass der Revv Amp beim Attack stärker variiert. Bei beiden Verstärkern verflacht ToneX teils die Kurvenverläufe und prägt die Amplituden der Signalminima mal zu stark und mal zu schwach aus.

Auch bei den Spektrogrammen sieht man bis auf das Rauschen beim Referenzsignal auf den ersten Blick kaum einen Unterschied, beim genauen Hinsehen fallen aber doch Unterschiede auf. Gerade bei ausklingenden Noten, verliert ToneX schneller an Obertönen als der Röhrenverstärker. Beim JCM900 ist dieser Effekt sogar deutlich sichtbar in jedem Spektrogramm (mit ausklingenden Tönen) ausgeprägt. Beim Revv Amp ist es weniger stark sichtbar, allerdings hat er generell einen weniger brillanten Klang als der JCM900.

Des Weiteren sieht man Unterschiede bei den Obertönen ab 16kHz, welche beim Röhrenverstärker klarer differenziert sind. Es sind klare voneinander abgetrennte horizontale Linien in den Spektrogrammen der Röhrenverstärker zu erkennen, während in den Spektrogrammen der Vergleichssignale ein großer „Block“ an Frequenzgehalt über alle Frequenzen ab 16kHz hinweg zu sehen ist.

Allgemein lässt sich aus den Spektrogrammen der Differenzsignale entnehmen, dass das Tone Model des Revv Amp auch hier etwas besser abschneidet. Die Spektrogramme der Differenzsignale sind weniger intensiv als beim JCM900 und zeigen generell weniger Verlust von Obertönen während Releasephasen. Die größten Abweichungen finden beim Revv Amp zwischen 80Hz-6kHz statt, das ist allerdings auch der Frequenzbereich, in dem das Signal generell am stärksten ausgeprägt ist, mit einem höheren Ausschlag ist also sowieso zu rechnen. Beim JCM900 ist dagegen zwischen 2,5kHz-6kHz eine stärkere Ausprägung der Differenz zu sehen.

Bei den MFCC gibt es bei den Stimuli 2 und 3 Unterschiede, für deren Erklärung man nur mutmaßen kann. Der subjektive Eindruck beim Hören der Signale führt jedenfalls nicht zum Eindruck einer starken Abweichung des Timbres zwischen Röhrenverstärker und ToneX. Bei beiden Verstärkern nicht und bei keinem der Stimuli. Generell kann als Ankerpunkt für die Werte der MFCC ein Vergleich der Werte zwischen den Referenzsignalen beider Verstärker dienen. Diese haben, vom subjektiven Hören her geurteilt, tatsächlich ein recht unterschiedliches Timbre. Somit kann aus der Differenz der Werte der beiden Referenzsignale ein Schluss über die Signifikanz der Wertedifferenz zwischen Referenz- und Vergleichssignal gezogen werden. Vergleicht man bspw. die MFCC der Referenzsignale des Arpeggios (Anhang 3.3) miteinander, sieht man bei den ersten drei Koeffizienten Abweichungen um Werte von fünf bis zehn. Beim JCM900 weichen die MFCC des Vergleichssignals kaum ab. Dies wird durch den subjektiven Eindruck bestätigt, da sich

kein Unterschied des Timbres bemerkbar macht. Den gleichen Subjektiven Eindruck gewinnt man auch beim Hören der Signale des Revv Amp, allerdings weichen die MFCC des Vergleichssignals hier deutlich von denen der Referenz ab, sogar erheblich mehr als die MFCC der beiden Referenzsignale voneinander abweichen. Dies müsste bedeuten, dass das Timbre des Vergleichssignals des Revv Amp deutlich mehr vom Timbre des Referenzsignals abweicht als das des Referenzsignals des JCM900 vom Referenzsignal des Revv Amp. Ein Eindruck, der sich beim Hören überhaupt nicht bestätigen lässt. Aufgrund des hohen Abstraktionsgrads der MFCC lässt sich dieses Phänomen leider nicht auf eine Ursache zurückführen. Es tritt allerdings nur beim zweiten und dritten Stimulus auf. Bei den anderen Messungen schneidet der Revv Amp ähnlich gut ab wie der JCM900.

Zusammenfassend lässt sich also sagen, dass das Tone Model des Revv Amp bei den Wellenformdiagrammen und auch in den Spektrogrammen besser abschneidet als das des JCM900. Die wesentlichen Unterschiede zwischen ToneX und Röhrenverstärker sind kleine Abweichungen in der Phasenlage, Abweichungen der Amplitude beim Ausschwingen, Abweichungen der Wellenform in Attack- und Decayphase und schnellerer Verlust an Obertönen beim Ausschwingen. Bei den MFCC schneidet das Tone Model des JCM900 sehr gut ab. Hier liegen kaum Unterschiede zur Referenz vor. Bis auf die ungeklärten starken Abweichungen bei den Stimuli zwei und drei, schneidet das Capture des Revv Amp hier ebenfalls gut ab.

6.2.2 Fazit

Auch wenn die Tone Models von ToneX sehr nah an den Röhrenverstärker herankommen und die tatsächlichen Abweichungen nur sehr gering sind, sind sie dennoch vorhanden. Sowohl bei in den Wellenformdiagrammen, den Spektrogrammen als auch bei den MFCC sind Abweichungen des Signals von ToneX zu erkennen. Doch was folgt daraus?

Diese Arbeit wollte untersuchen, ob die modernen KI-Modeler halten, was sie versprechen, indem untersucht wird, ob Unterschiede bestehen und wenn ja, worin diese bestehen. Sie hat gezeigt, dass es messbare Unterschiede gibt und wo diese zu finden sind. Allerdings hat ToneX nicht behauptet, dass ihre Tone Models keine Unterschiede aufweisen, wenn man sie unters Mikroskop legt und wissenschaftlichen Messungen unterzieht. Die Behauptung, sie wären „ununterscheidbar“ vom Original ist eine Aussage, die sich auf die menschliche Fähigkeit bezieht, Unterschiede wahrnehmen zu können. Wesentlich ist nicht, ob die Tone Models noch Unterschiede aufweisen, sondern ob Unterschiede existieren, die oberhalb der Wahrnehmungsgrenze liegen. Ein Großteil der gemessenen Unterschiede ist generell zu vernachlässigen. Das Differenzsignal schlägt zwar aus, wenn sich die Phase oder die Amplitude des ToneX-Signals leicht von dem des Röhrenverstärkers unterscheidet, ein Unterschied im Klang entsteht dadurch allerdings nicht wirklich, lediglich ein Unterschied in der Messung.

Tatsächlich scheint es größere Abweichungen, beim Notenanschlag, eventuell auch beim Ausklingen zu geben, hinsichtlich Kurvenverlauf und Obertongehalt. Doch auch bei den deutlicheren Differenzen ist nicht klar, ob diese überhaupt wahrnehmbar sind.

Der subjektive Eindruck beim vergleichenden Hören der Signale ist, dass kein wahrnehmbarer Unterschied existiert. Doch subjektives Hören ist personenabhängig und Wahrnehmungsschwellen variieren. Möglicherweise gibt es Menschen, die in der Lage sind, das ToneX-Signal vom Original zu unterscheiden.

Diese Arbeit hat sicherlich gezeigt, dass es noch Unterschiede gibt, aber auch dass diese minimal sind, selbst unter dem Mikroskop. Und ob diese wahrnehmbar sind, ob es irgendwo Menschen gibt, die diese Unterschiede über ihr Gehör feststellen können, ist wissenschaftlich kaum zu falsifizieren.

Letztendlich bleibt allerdings auch die Frage, warum es dies überhaupt brauchen sollte. Digitale Verstärkermodelle sind längst in der Lage, sehr gute Ergebnisse zu erzielen. Das waren sie auch schon bevor KI-Modeler wie ToneX die Weltbühne betreten haben. Vielleicht haben KI-Modeler die Technik abschließend verfeinert und es ist tatsächlich ab diesem Punkt niemandem mehr möglich, einen Unterschied wahrzunehmen. Vielleicht wird es auch noch weitere Verbesserungen des Klangs geben. Doch ist das relevant? Selbst wenn es Menschen gibt, die hier noch Unterschiede hören können, die Vermutung liegt nah, dass die breite Masse dies nicht kann. Aber selbst die Menschen, die dazu in der Lage sind, können dies bei einer Gitarre in einem Mix mit anderen Instrumenten wahrscheinlich nicht mehr. Worin liegt also der Gewinn, eine Technik weiter zu verfeinern, wenn die Verbesserung garnicht mehr wahrnehmbar ist?

Schlussendlich kann man sagen, dass die Qualität der Signale hervorragend ist und auch die des Kemper Profiling Amp nochmal übertrifft. Es bleiben messbare Unterschiede zurück, ob diese hörbar sind, bleibt allerdings fraglich und ob es wirklich von Relevanz wäre, wenn irgendwer irgendwo in einem Vergleich doch noch einen Unterschied hören kann, auch.

6.3 Ausblick auf weitere Forschungen

Auch wenn die genutzten Parameter die wesentlichen Eigenschaften eines Audiosignals darzustellen vermögen, gibt es noch sehr viele weitere Parameter, auf deren Basis man einen noch detaillierteren Vergleich anstellen könnte. Auch könnte man noch weitere unterschiedliche Röhrenverstärker, Lautsprecherkabinette und Mikrofone testen. Man könnte Versuche mit lediglich Verstärkern ohne Kabinett anstellen, Versuche nur mit Verzerr-Pedalen oder Versuche mit Verzerr-Pedalen, die vor einen Röhrenverstärker mit Kabinett geschaltet sind, welcher von mehreren Mikrofonen gleichzeitig mikrofoniert wird. Weitere Möglichkeiten existieren zur Genüge

Wie in 6.2.2. dargelegt, ist aber fraglich, ob dabei Ergebnisse zustande kämen, die wirklich von Relevanz sind oder die bereits betriebenen Forschungen entscheidend ergänzen würden. Klar ist mittlerweile: Digitale Lösungen können extrem gute Klangqualität erreichen.

Wenn dieser Punkt geklärt ist, gibt es aber tatsächlich offene Fragen, die wohl weniger für eine Konzert-Audienz, dafür aber für die Gitarristen von großem Interesse sein könnten, ebenso für Livetechniker oder Tonmeister im Studio. Das Spielgefühl zum Beispiel wurde

in dieser Arbeit komplett außer Acht gelassen. Ob es, wenn es keine hörbaren Unterschiede mehr gibt, spürbare Unterschiede gibt?

Darüber hinaus gibt es wesentliche praktische Fragen, welche interessant sein könnten. Zum Beispiel die Notwendigkeit, Lösungen wie ToneX lediglich über die PA einer Spielstätte amplifizieren zu können. Im Studio oder Live in einer Arena, also in Situationen, in denen beim Einsatz eines Röhrenverstärkers nur das mikrofonierte Gitarrensinal von Interesse ist, mag es für das Publikum keinen Unterschied machen, ob man einen echten oder einen digitalen Verstärker nutzt. In z.B. kleineren Clubs allerdings kann das von dem Kabinett abgestrahlte Signal einen wesentlichen Anteil von dem Gitarrensinal ausmachen, das das Publikum erreicht. Teilweise wird das Signal garnicht zusätzlich mikrofoniert, da es nicht nötig ist oder es die vorhandene PA sogar überlasten würde. Da es bei digitalen Lösungen wie ToneX aber in so einem Falle nötig ist, das Signal über die PA auszugeben, da gar kein Kabinett mehr im Gitarrensetup vorhanden ist, kann es passieren, dass trotz theoretisch guter Klangqualität praktisch am Ende kaum etwas von der Gitarre zu hören ist. Die meisten Gitarristen treten wahrscheinlich eher in solchen kleinen Clubs auf, das Phänomen ist also von deutlichem Interesse, sowohl für Gitarristen als auch Tontechniker.

Gitarristen würden wahrscheinlich deutlich mehr davon profitieren, wenn anstatt weiterer Untersuchungen zur Klangqualität Forschungen in den Vordergrund rücken, die solche praktischen Probleme unter die Lupe nehmen und Lösungen erarbeiten. Denn dass ein Gerät theoretisch in der Lage ist, hohe Klangqualität zu produzieren, heißt nicht, dass in unterschiedlichen Szenarien, im tatsächlichen praktischen Einsatz auch wirklich alle Notwendigkeiten dafür gegeben sind. Solange dies nicht stärker untersucht wird, hat der Röhrenverstärker hier einfach aufgrund seiner Simplizität weiterhin die Nase vorn.

IV. Literaturverzeichnis

- Ali, S.; Tanweer, S.; Khalid, S.; Rao, Naseem (Hrsg.): Mel Frequency Cepstral Coefficient: A Review, 2021.
- Aligarh Muslim University: Amplifier Classes, <https://old.amu.ac.in/emp/studym/100009038.pdf> (21. Juli 2023).
- Bähring, H.: Anwendungsorientierte Mikroprozessoren. Mikrocontroller und Digitale Signalprozessoren, 4. Aufl., Berlin 2010.
- Barbour, E.: The cool sound of tubes [vacuum tube musical applications]. In: IEEE Spectrum 35 (1998), S. 24–35.
- Böhmer, E.; Ehrhardt, D.; Oberschelp, W.: Elemente der angewandten Elektronik. Kompendium für Ausbildung und Beruf : mit umfangreichem Bauteilekatalog, 17. Aufl., Wiesbaden u. a. 2018.
- Camastra, F.; Vinciarelli, A.: Machine learning for audio, image and video analysis. Theory and applications, 2. Aufl., London u. a. 2015.
- Cannam, C.; Landone, C.; Sandler, Mark (Hrsg.): Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files, 2010.
- Chen, L.: Deep Learning and Practice with MindSpore, Singapore 2021.
- Covert, J.; Livingston, D. L.: A vacuum-tube guitar amplifier model using a recurrent neural network. In: 2013 Proceedings of IEEE Southeastcon, 2013, S. 1–5.
- Dailey, D. J.: Electronics for Guitarists, New York 2011.
- Damskäg, E.-P., et al.: Deep Learning for Tube Amplifier Emulation, 2019.
- David Kriesel: Ein kleiner Überblick über Neuronale Netze, 2007.
- Devi, S.; Srinivas, Y.; Namana, D. k.: A Study: Analysis of Music Features for Musical Instrument Recognition and Music Similarity Search. In: International Journal of Computer Science and Informatics (2011).
- Dickreiter, M.: Handbuch der Tonstudioteknik, 7. Aufl., München 2008.
- Döbel, I., et al.: Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung, 2018.
- Düvel, N.; Kopiez, R.; Wolf, A.; Weihe, P.: Confusingly Similar: Discerning between Hardware Guitar Amplifier Sounds and Simulations with the Kemper Profiling Amp. In: Music & Science (2020), 1-16.
- Eero-Pekka Damskäg; Lauri Juvela; Vesa Välimäki: Real-Time Modeling of Audio Distortion Circuits with Deep Learning, 2019.
- Elflein, D.: Schwermetallanalysen. Die musikalische Sprache des Heavy Metal, Bielefeld 2010.
- Ernst, H.; Schmidt, J.; Beneken, G. H.: Grundkurs Informatik. Grundlagen und Konzepte für die erfolgreiche IT-Praxis – eine umfassende, praxisorientierte Einführung, 7. Aufl., Wiesbaden 2020.
- Felix Eichas; Udo Zölzer: Virtual Analog Modeling of Guitar Amplifiers with Wiener-Hammerstein Models, 2018.

- Fraunhofer-Institut für Kognitive Systeme IKS: Künstliche Intelligenz (KI) und maschinelles Lernen, <https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html> (29. September 2023).
- Goodfellow, I.; Bengio, Y.; Courville, A.: Deep learning, Cambridge, Massachusetts u. a. 2016.
- Graves, A.: Supervised sequence labelling with recurrent neural networks, Berlin u. a. 2012.
- Hahn, W.: Elektronik-Praktikum für Informatiker, Berlin 1971.
- Hartmann, W. M.: Principles of Musical Acoustics, New York 2013.
- Herbst, J.: "My setup is pushing about 500 watts - it's all distortion". Emergence, development, aesthetics and intentions of the rock guitar sound, 2019.
- Herbst, J.-P.: Distortion and Rock Guitar Harmony. In: Music Perception 36 (2019), S. 335–352.
- Hering, E.; Bressler, K.; Gutekunst, Jürgen (Hrsg.): Elektronik für Ingenieure und Naturwissenschaftler, 7. Aufl., Berlin u. a. 2017.
- Hofmann, S.: Identifikation von nichtlinearen mechatronischen Systemen auf der Basis von Volterra-Reihen, 2003.
- IK Multimedia: AI Machine Modeling, <https://www.ikmultimedia.com/products/aimachinemodeling/> (26. Oktober 2023).
- ITU-R. ITU-R BS.1534-3. Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems. In: Intern. Telecom. Union Radiocom, 34. Aufl., 2015.
- Jan-Peter Herbst, Christoph Reuter, Isabella Czedik-Eysenberg: Guitar profiling technology in metal music production: public reception, capability, consequences and perspectives. In: Metal Music Studies (2018), S. 481–506.
- Jurafsky, D.; Martin, J. H.: Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2023.
- Kemper GmbH: Profiler Overview, <https://www.kemper-amps.com/profiler/overview>.
- Lähdevaara, J.: The science of electric guitars and guitar electronics, Helsinki 2014.
- Lempitsky, V.: Autoencoder. In: Computer Vision: A Reference Guide, Cham 2019, S. 1–6.
- Mačák, J.: Real-time Digital Simulation of Guitar Amplifiers as Audio Effects, Brno University of Technology 2012.
- Maloof, R.: Jim Marshall - The Father of Loud. The Story of the Man Behind the World's Most Famous Guitar Amplifiers, Milwaukee 2004.
- Martínez, M.; Benetos, E.; Reiss, J.: Deep Learning for Black-box Modeling of Audio Effects. In: Applied Sciences (2020).
- Martínez, M.; Reiss, J.: A General-purpose Deep Learning Approach to Model Time-varying Audio Effects, 2019.
- Martínez, M.; Reiss, J.: Modeling of nonlinear audio effects with end-to-end deep neural networks. In: Proceedings of the ICASSP, 2019.
- Mitchell, T. M.: Machine learning, New York, NY 2010.
- Muller, M.; Ellis, D. P.; Klapuri, A.; Richard, G.: Signal Processing for Music Analysis. In: IEEE Journal of Selected Topics in Signal Processing 5 (2011), S. 1088–1110.
- Müller, M.: Fundamentals of Music Processing, Cham 2015.

- Mynett, M.: The Distortion Paradox. In: Brown, A. R.; Spracklen, K.; Kahn-Harris, K.; Scott, N. W. (Hrsg.): *Global Metal Music and Culture*, New York 2016, S. 68–86.
- Mynett, M.: *Metal Music Manual. Producing, engineering, mixing and mastering contemporary Heavy Music*, New York u. a. 2017.
- Neural DSP: Quad Cortex, <https://neuraldsp.com/quad-cortex> (26. Oktober 2023).
- Pakarinen, J.; Yeh, D. T.: A Review of Digital Techniques for Modeling Vacuum-Tube Guitar Amplifiers. In: *Computer Music Journal* 33 (2009), S. 85–100.
- Pfister, B.; Kaufmann, T.: *Sprachverarbeitung. Grundlagen und Methoden der Sprachsynthese und Spracherkennung*, Berlin, Heidelberg 2008.
- Rey, G. D.; Wender, K. F.: *Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*, 3. Aufl., Bern 2018.
- Roisin Loughran; Jacqueline Walker; Michael O'Neill; Marion O'Farrell (Hrsg.): *The Use of Mel-frequency Cepstral Components in Musical Instrument Identification*, 2008.
- Ronchetti, V.: *Violin Timbre Analysis with Mel-Frequency Cepstral Coefficients*, New York 2020.
- Rosiński, A.: *Microphone techniques in stereo and surround recording*, Kraków 2022.
- Saksamudre, S.; Deshmukh, R.: Comparative Study of Isolated Word Recognition System for Hindi Language. In: *International Journal of Engineering Research & Technology* 4 (2015), S. 536–540.
- Schmitz, T.; Embrechts, J. J.: *Nonlinear Real-Time Emulation of a Tube Amplifier with a Long Short Term Memory Neural-Network*, 2018.
- Schmitz, T.; Embrechts, J. J.: *Real Time Emulation of Parametric Guitar Tube Amplifier With Long Short Term Memory Neural Network*, 2018.
- Special Project Röhren 2. Röhren in Gitarrenverstärkern. In: *Elektor* (2006).
- Stiny, L.: *Aktive elektronische Bauelemente. Aufbau, Struktur, Wirkungsweise, Eigenschaften und praktischer Einsatz diskreter und integrierter Halbleiter-Bauteile*, 2. Aufl., Wiesbaden 2015.
- Strecker, S.: *Künstliche neuronale Netze - Aufbau und Funktionsweise*, Mainz 1997.
- Tan, L.: *Digital Signal Processing. Fundamentals and Applications*, s.l. 2007.
- Teemu Kyttälä: *Solid-State Guitar Amplifiers*, 2008.
- Terence Dau: *Qualitativer Vergleich der Performance verschiedener Gitarrenamps mit den entsprechenden Emulationen des Kemper Profiling Amps*, Mittweida 2022.
- ToneX, https://www.ikmultimedia.com/products/tonex/#nl_form_area (1. Juni 2023).
- Vanhatalo, T.; Legrand, P.; Desainte-Catherine, M.; Hanna, P.; Brusco, A.; Pille, G.; Bayle, Y.: A Review of Neural Network-Based Emulation of Guitar Amplifiers. In: *Applied Sciences* 12 (2022).
- Weinzierl, S. (Hrsg.): *Handbuch der Audiotechnik*, Berlin u. a. 2008.
- Wright, A.; Damskägg, E.-P.; Juvela, L.; Välimäki, V.: Real-Time Guitar Amplifier Emulation with Deep Learning. In: *Applied Sciences* 10 (2020).
- Yeh, D. T.-M.: *Digital Implementation of Musical Distortion Circuits by Analysis and Simulation*, 2009.
- Zollner, M.: *Physik der Elektrogitarre*, Regensburg 2014.

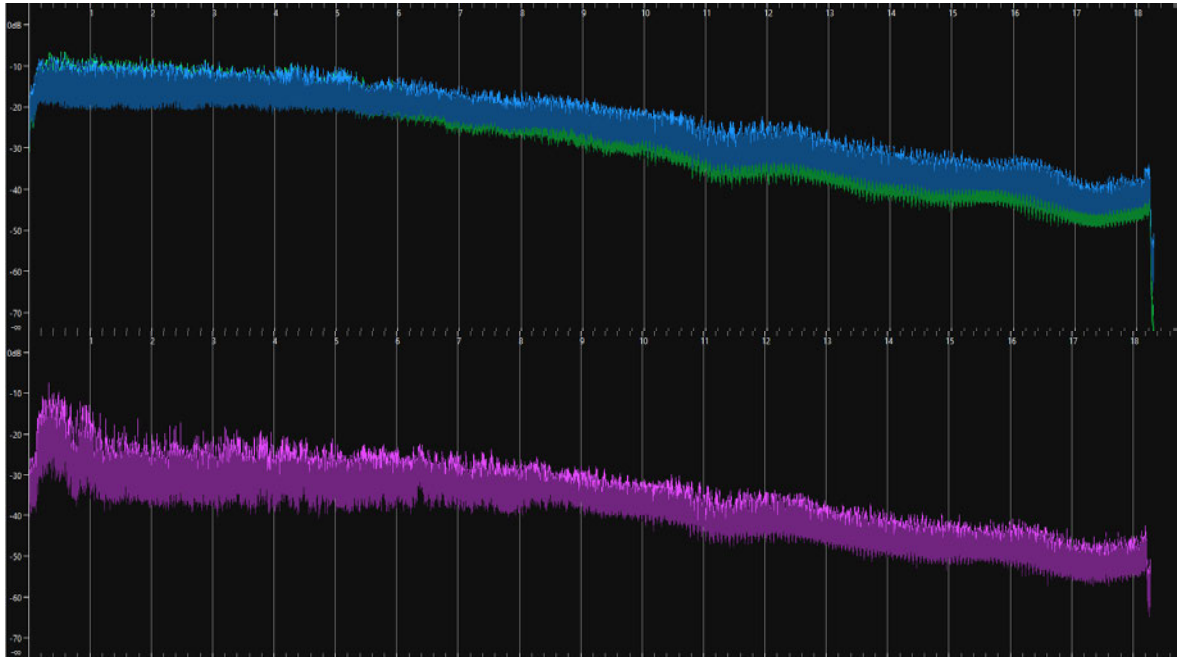
V. Anhang

Anhangsverzeichnis

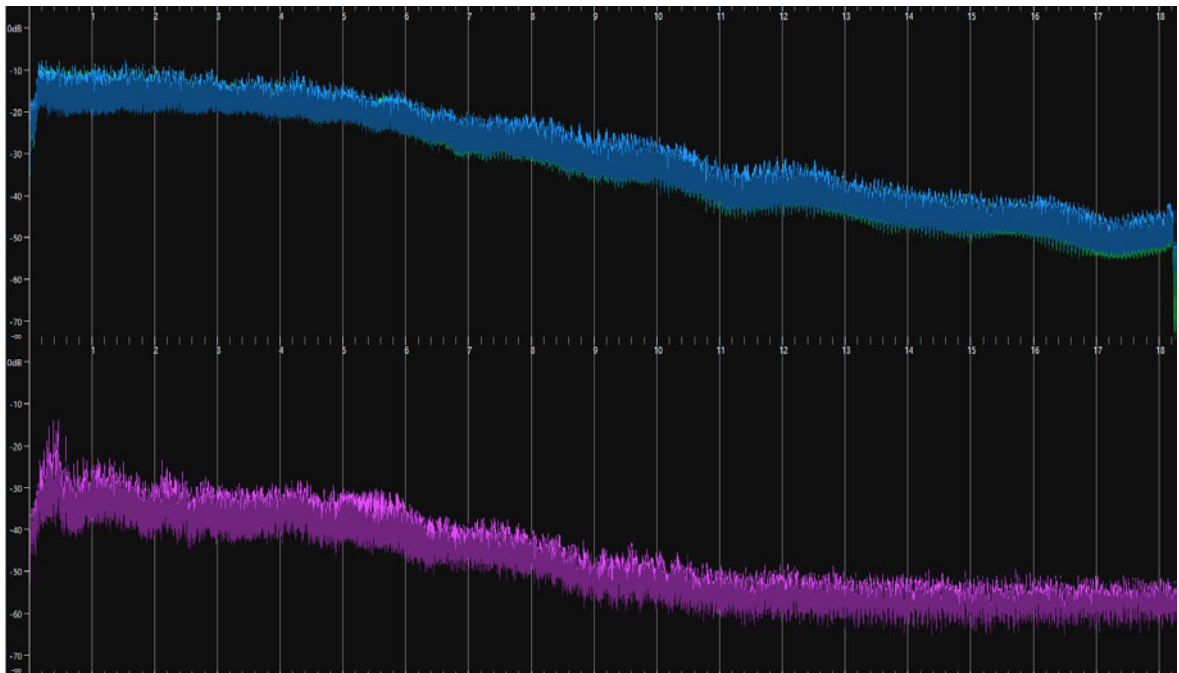
Anhang 1: Ausklingender Akkord	X
Anhang 1.1: Wellenformdiagramme	X
Anhang 1.2: Spektrogramme.....	XII
Anhang 1.3: MFCC	XIV
Anhang 2: Akkordwechsel	XV
Anhang 2.1: Wellenformdiagramme	XVI
Anhang 2.2: Spektrogramme.....	XVII
Anhang 2.3: MFCC	XXIX
Anhang 3: Arpeggio	XX
Anhang 3.1: Wellenformdiagramme	XX
Anhang 3.2: Spektrogramme.....	XXII
Anhang 3.3: MFCC	XXIV
Anhang 4: Kurzes Riff	XXV
Anhang 4.1: Wellenformdiagramme	XXV
Anhang 4.2: Spektrogramme.....	XXVII
Anhang 4.3: MFCC	XXIX
Anhang 5: Kurzes Solo	XXX
Anhang 5.1: Wellenformdiagramme	XXX
Anhang 5.2: Spektrogramme.....	XXXII
Anhang 5.3: MFCC	XXXIV

Anhang 1: Ausklingender Akkord

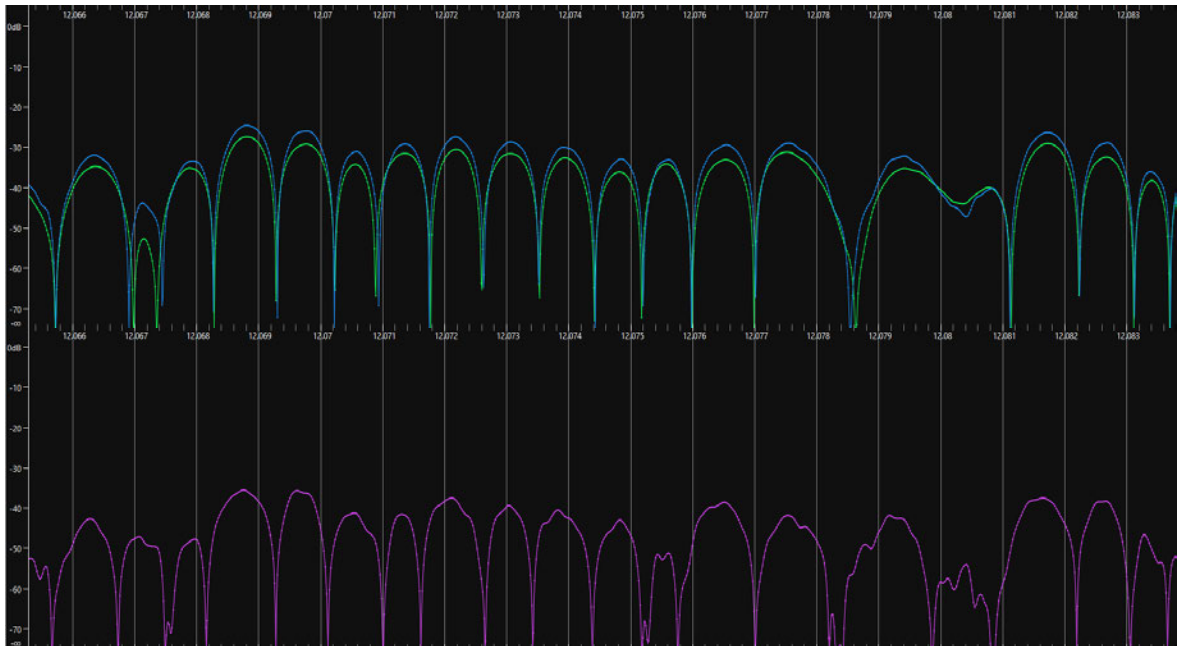
Anhang 1.1: Wellenformdiagramme



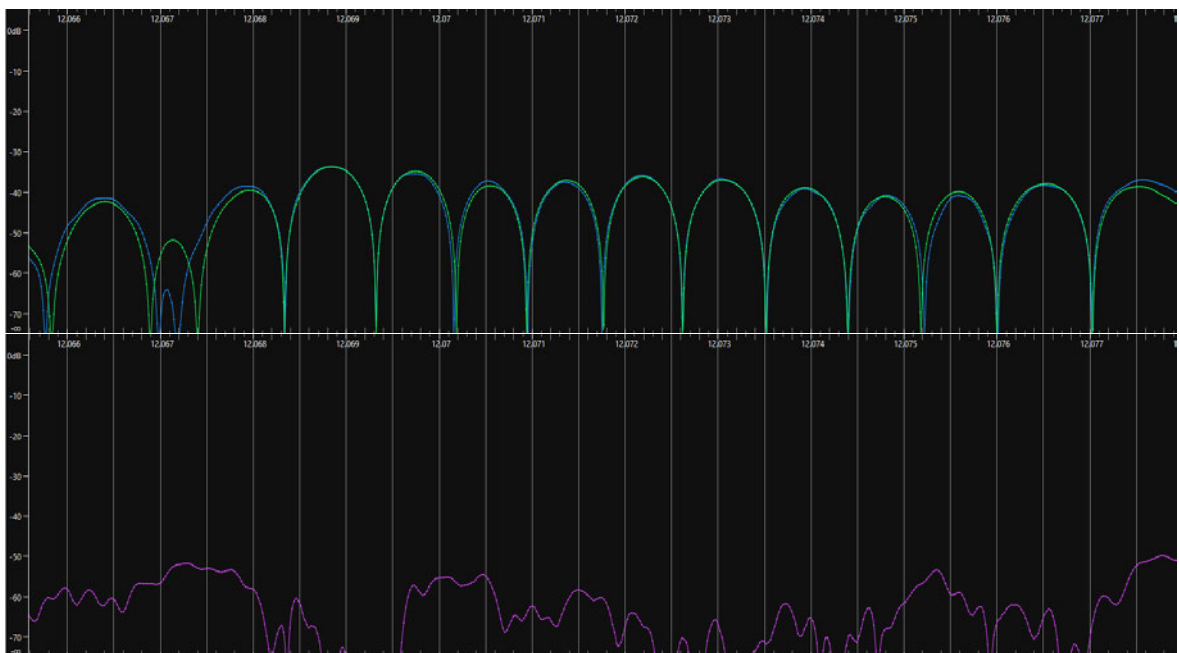
Gesamtsignal JCM900 (Blau: Röhrenverstärker; Grün: ToneX; Lila: Differenzsignal)



Gesamtsignal Revv Amp

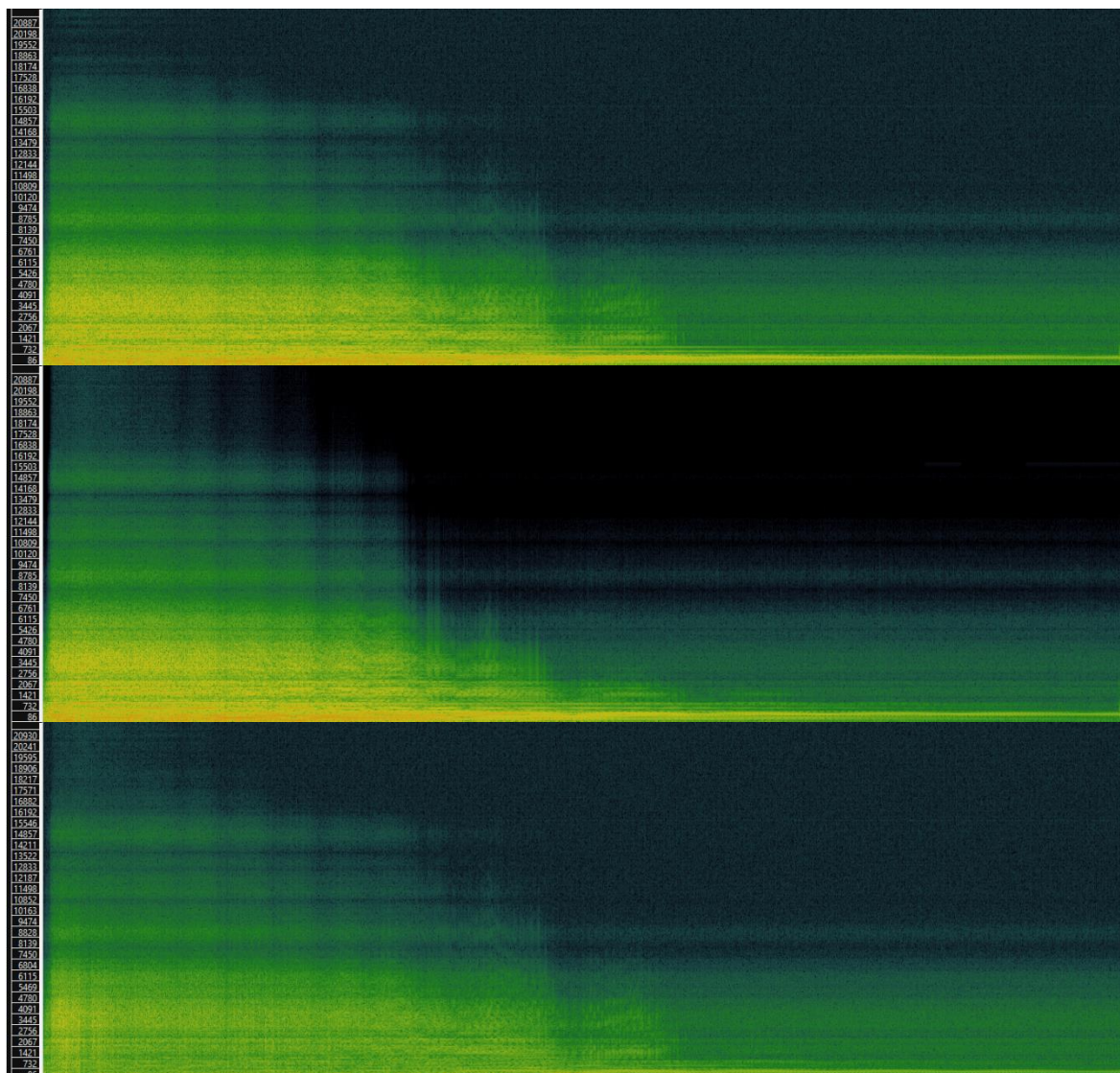


Zoom auf Ende des Ausschwingens der Saiten JCM900

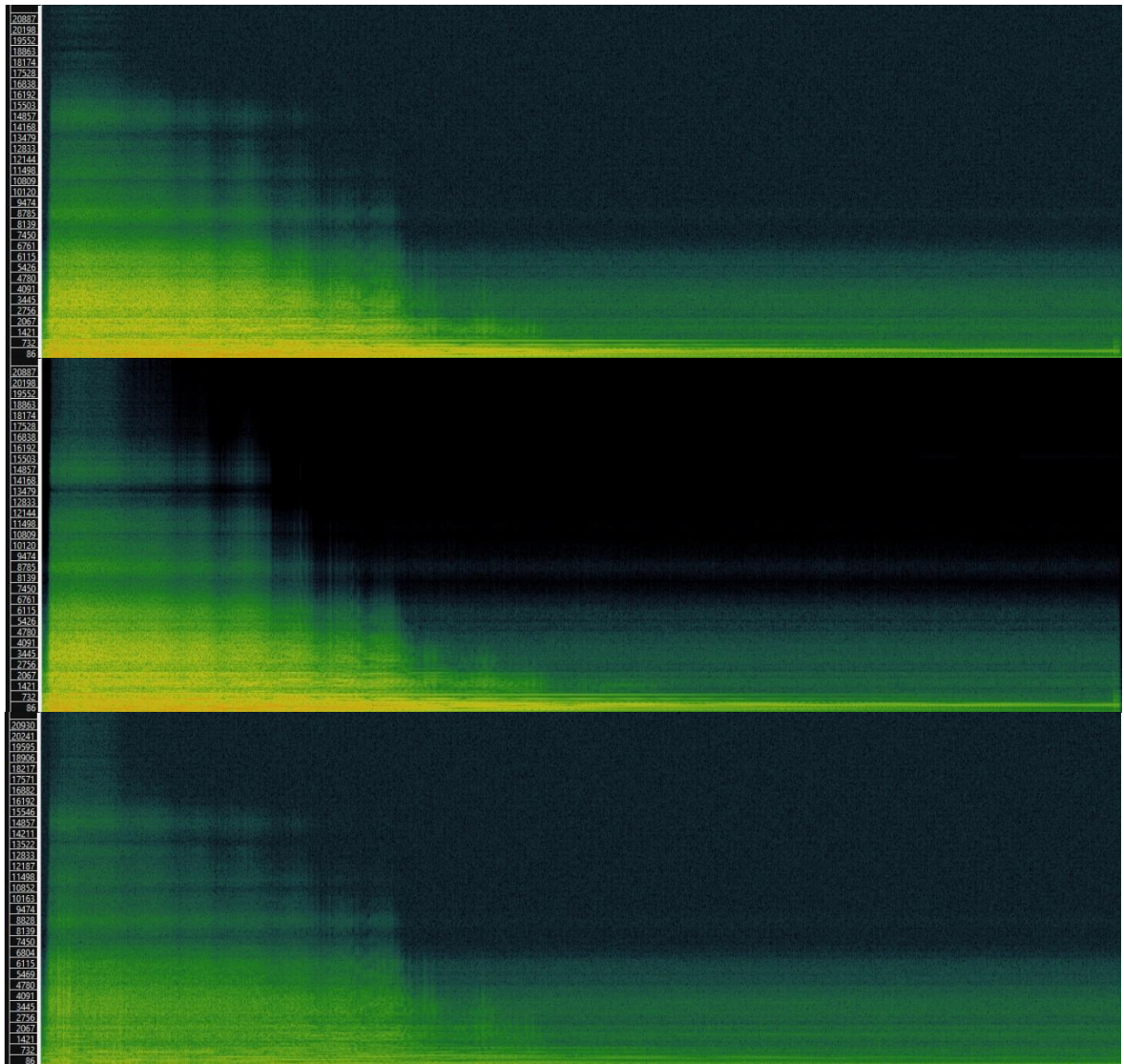


Zoom auf Ende des Ausschwingens der Saiten Revv Amp

Anhang 1.2: Spektrogramme

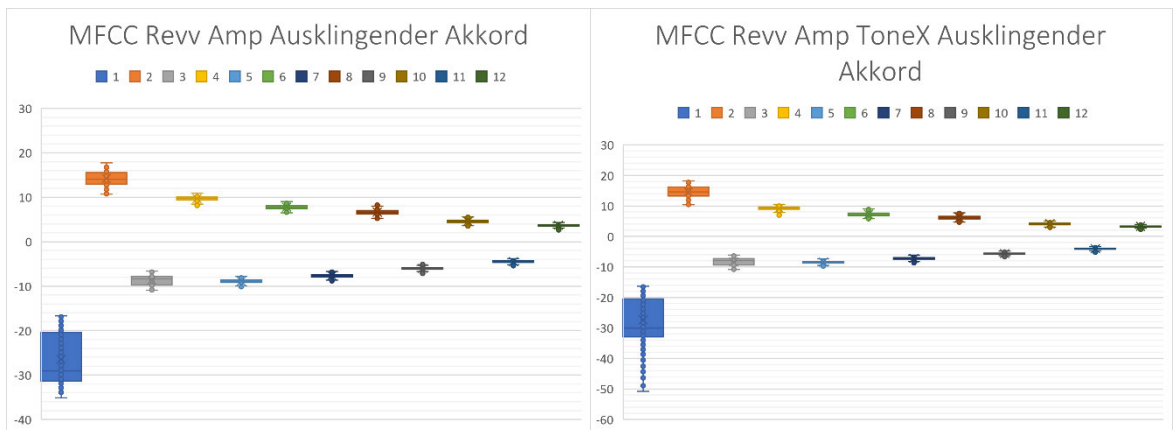
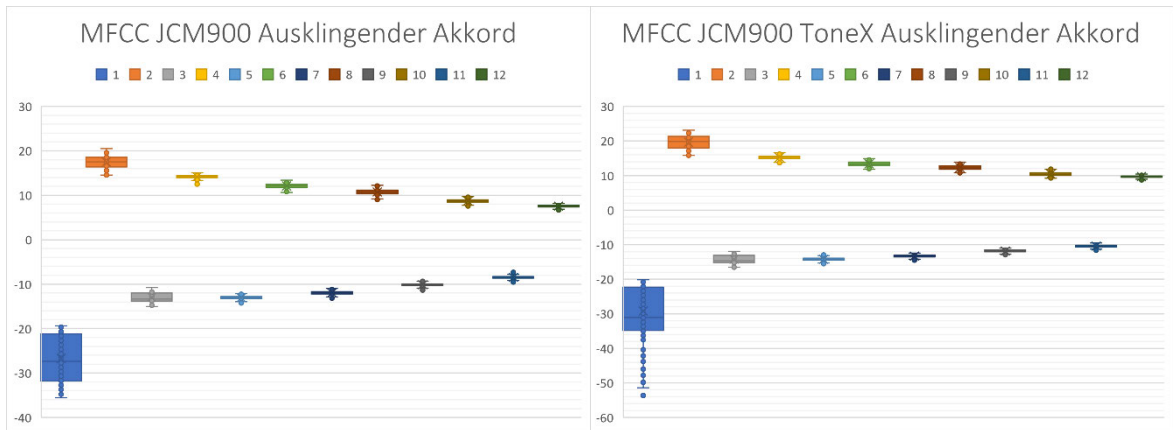


JCM900 (Von oben nach unten: Röhrenverstärker, ToneX, Differenzsignal)



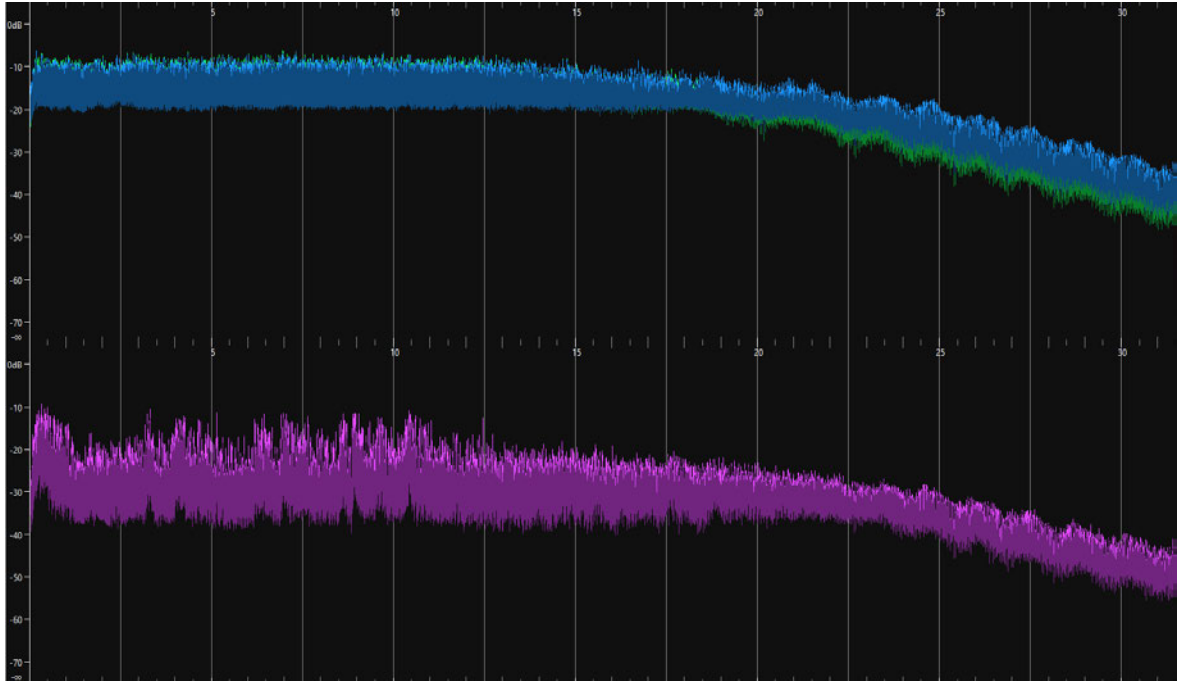
Revv Amp

Anhang 1.3: MFCC

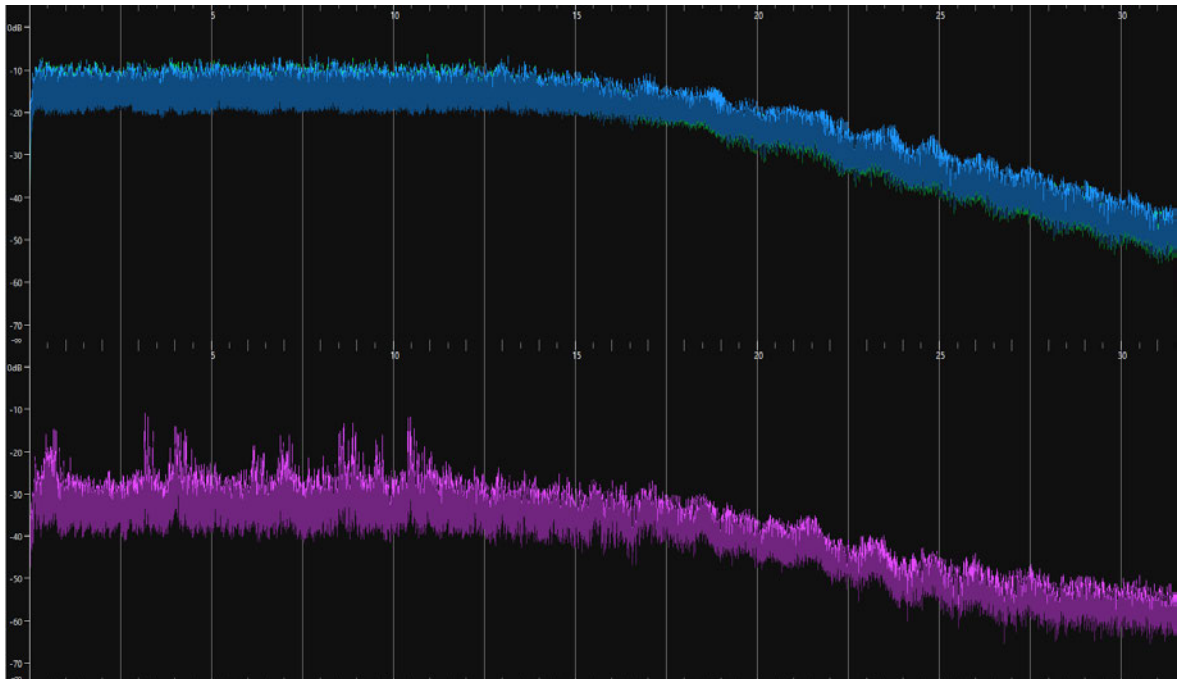


Anhang 2: Akkordwechsel

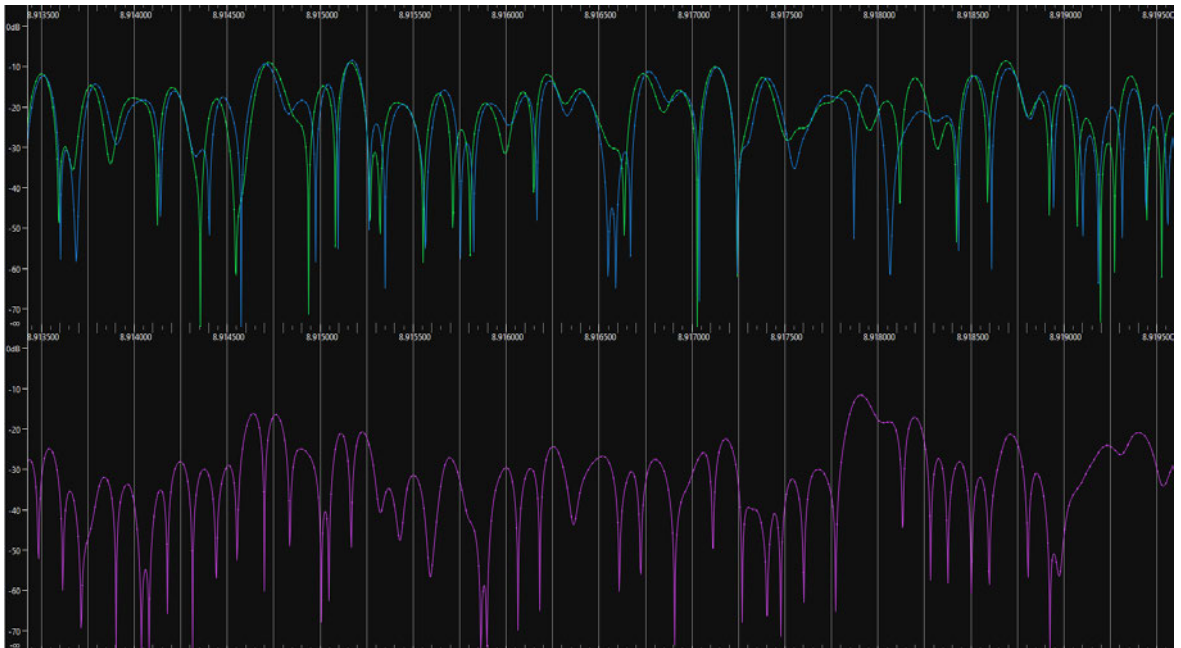
Anhang 2.1: Wellenformdiagramme



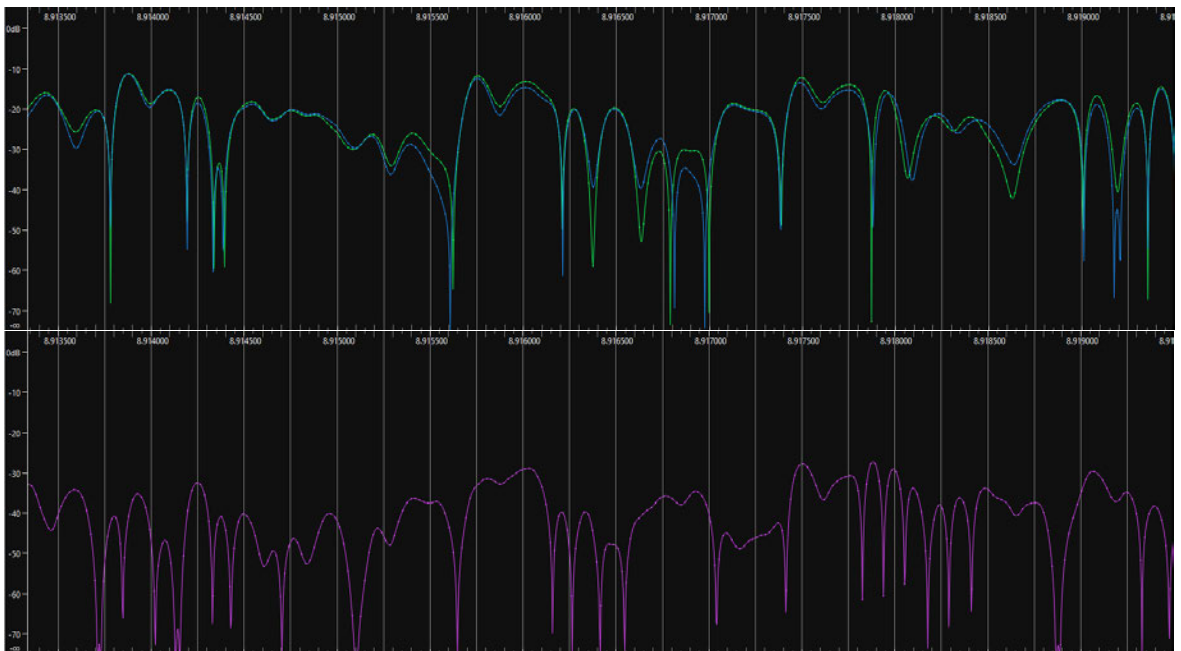
Gesamtsignal JCM900



Gesamtsignal Revv Amp

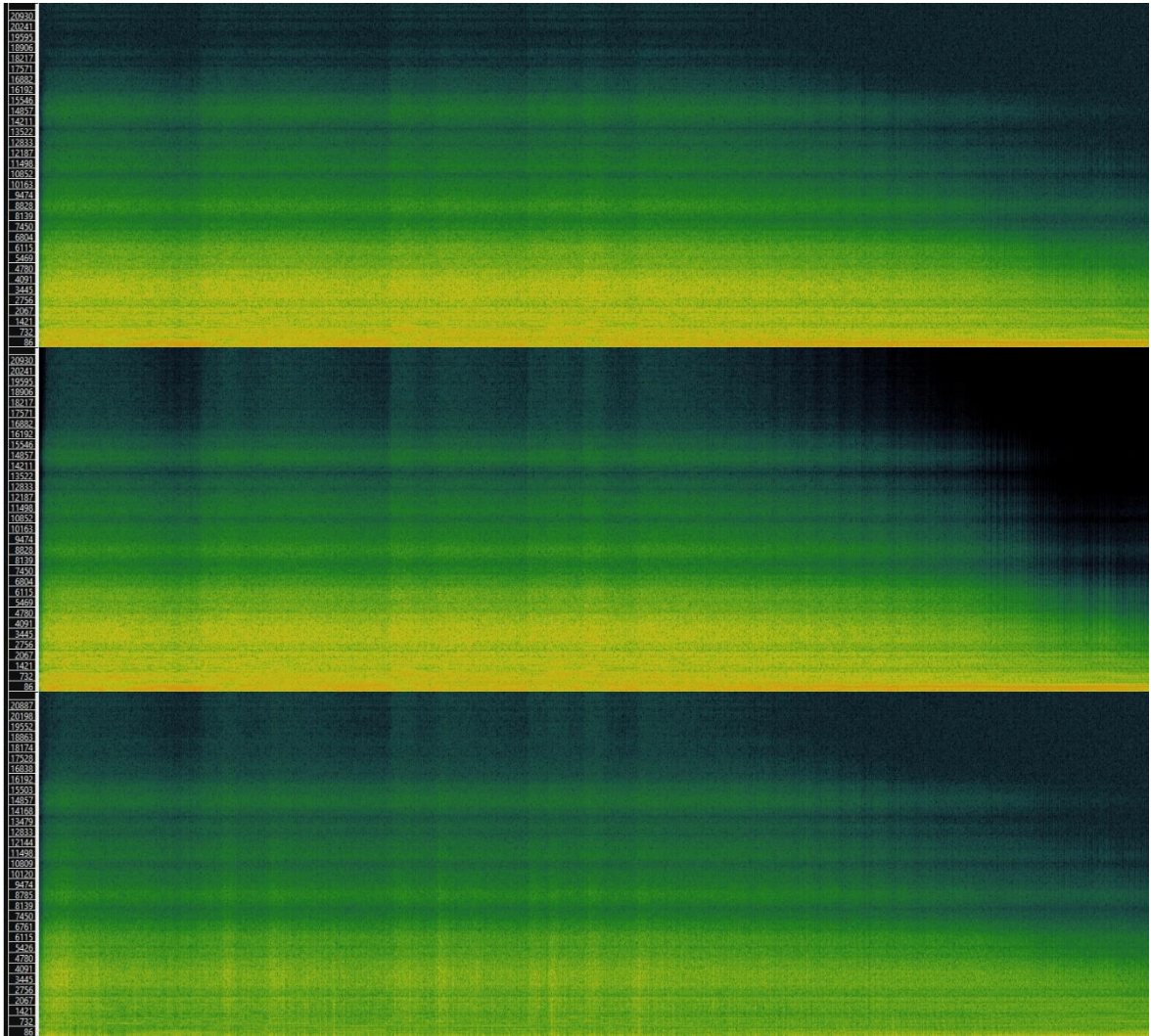


Zoom auf hohen Differenzpegel bei Decay eines Akkordanschlags JCM900

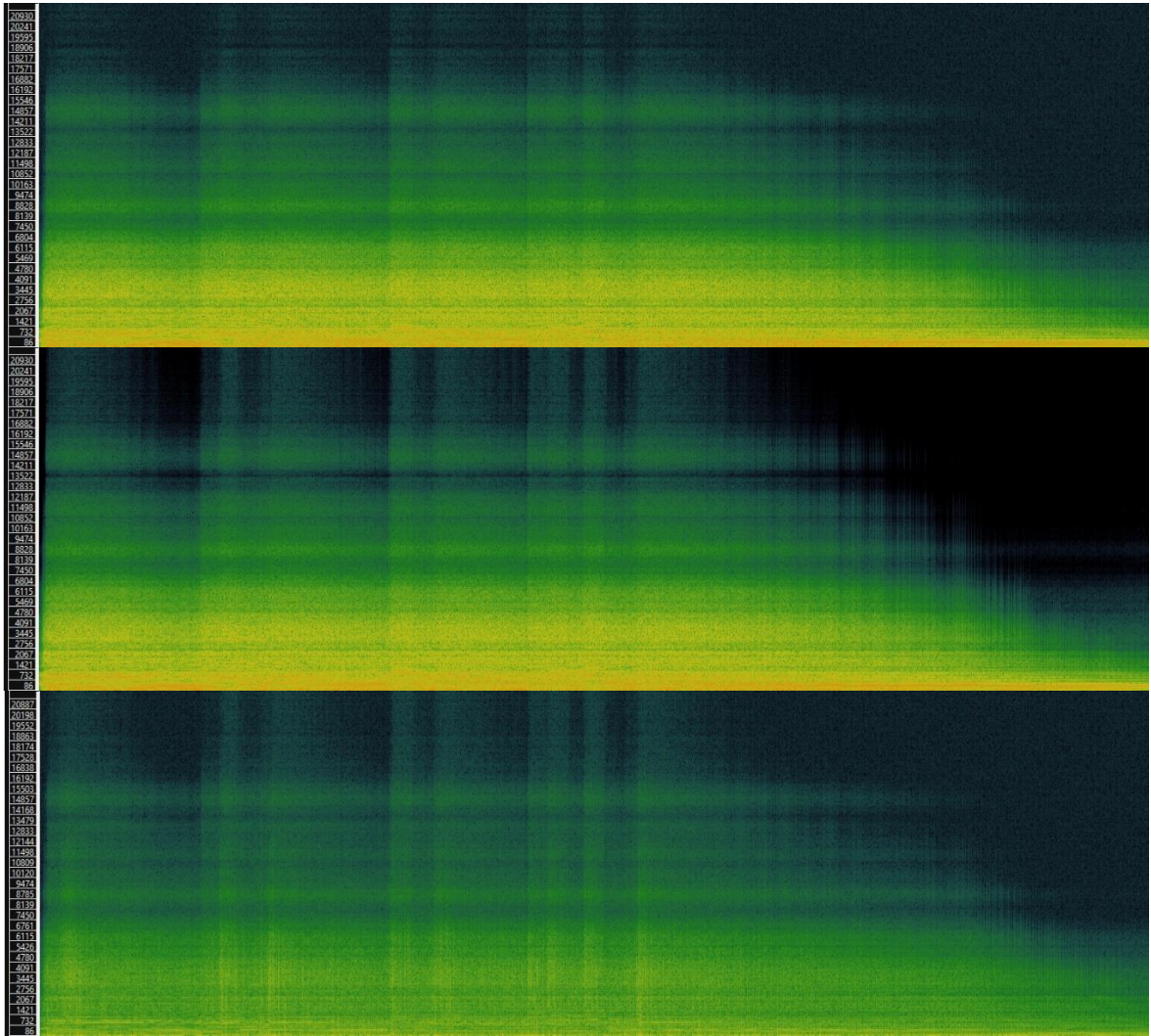


Zoom auf hohen Differenzpegel bei Decay eines Akkordanschlags Revv Amp

Anhang 2.2: Spektrogramme

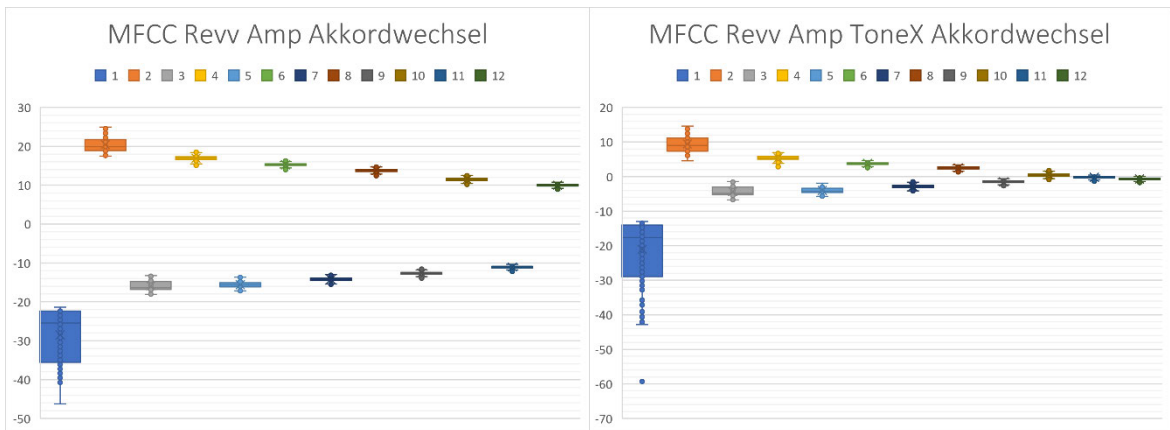
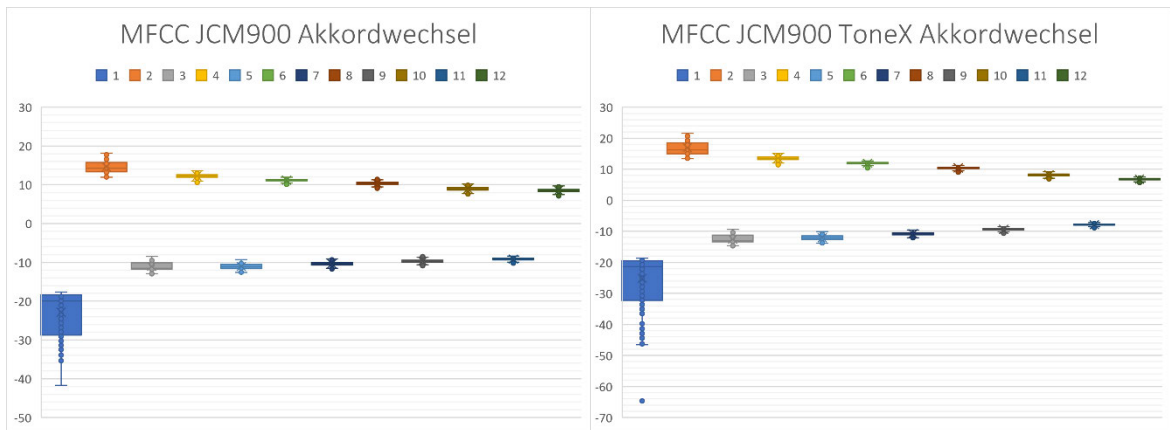


JCM900



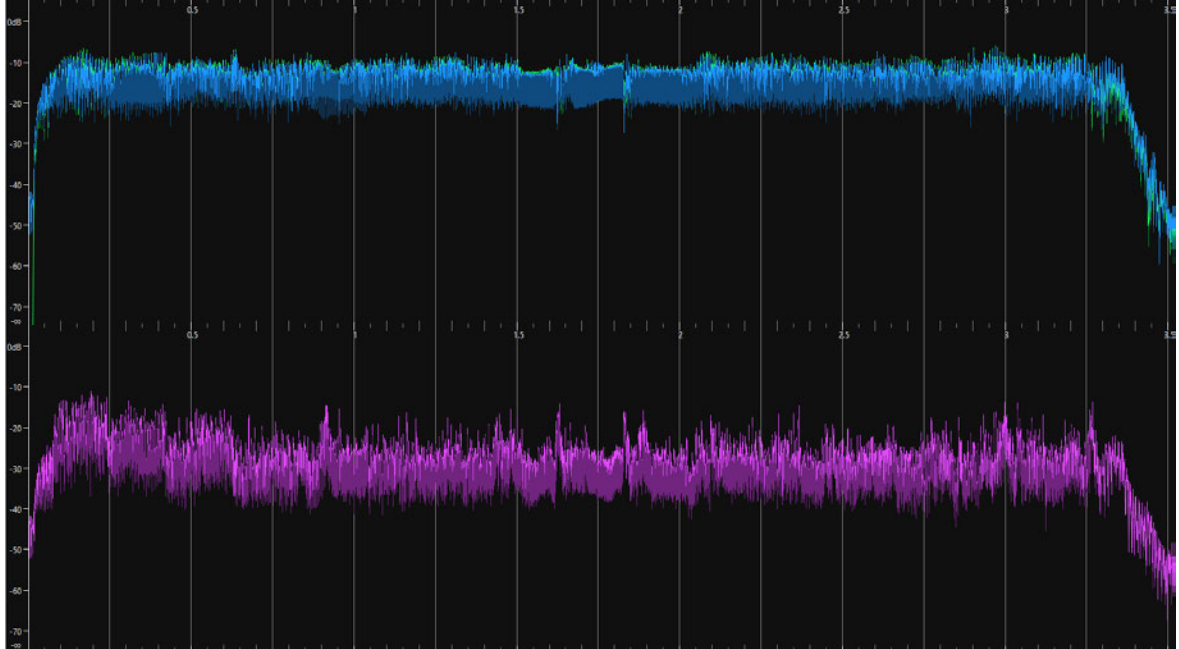
Revv Amp

Anhang 2.3: MFCC

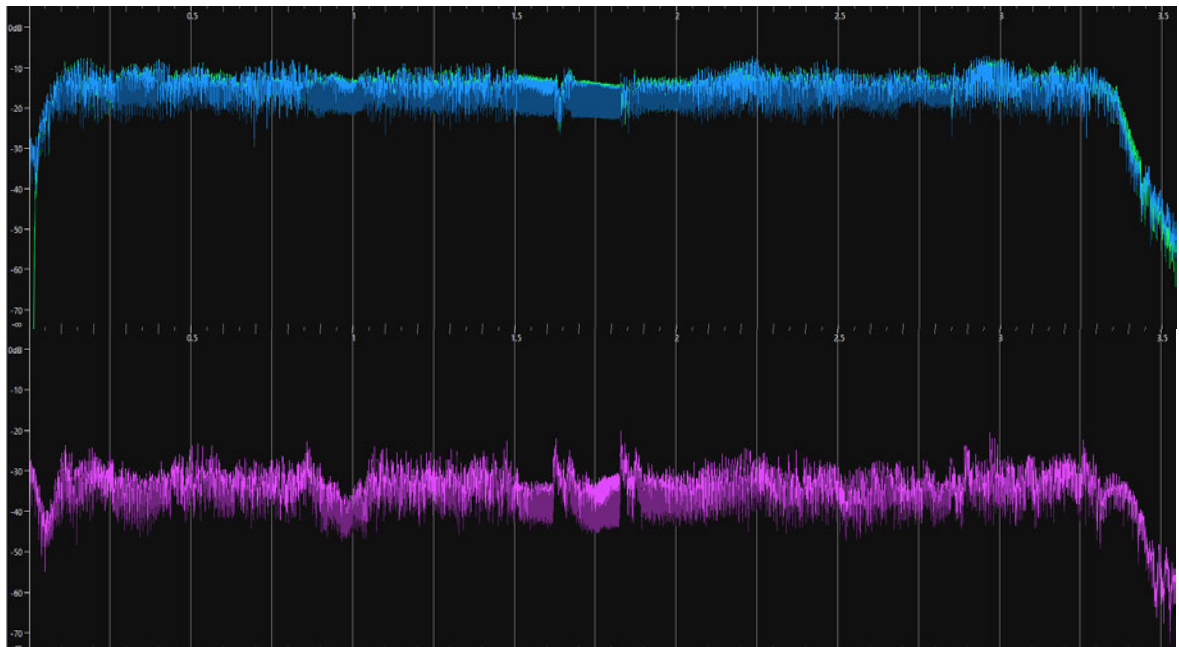


Anhang 3: Arpeggio

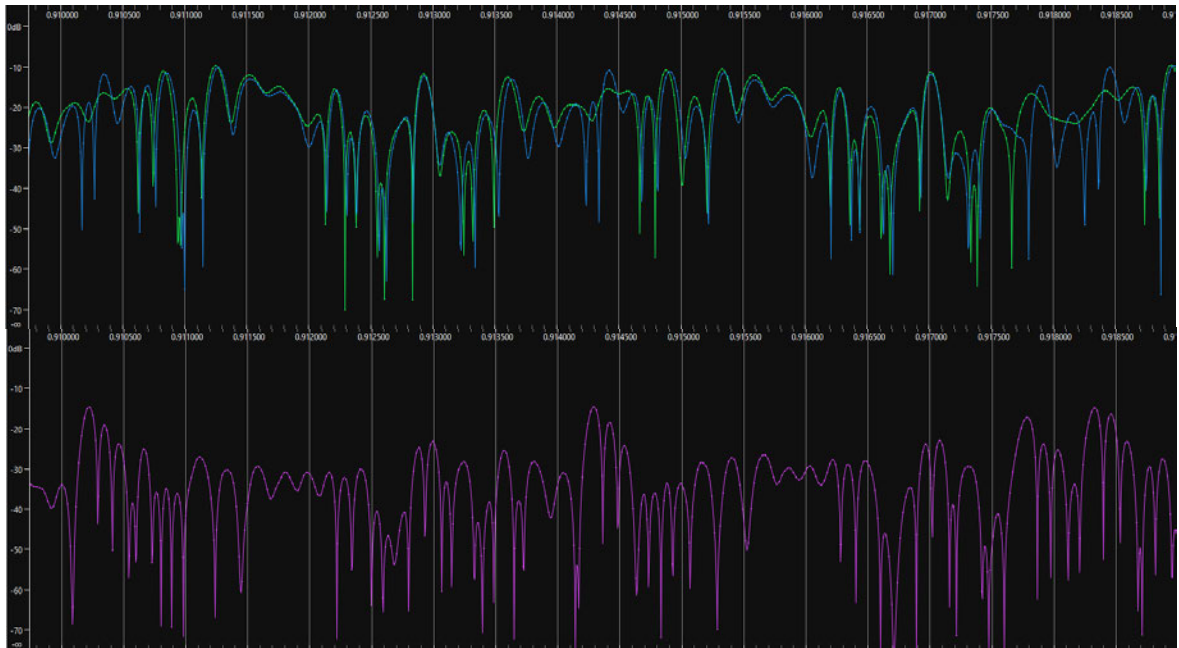
Anhang 3.1: Wellenformdiagramme



Gesamtsignal JCM900



Gesamtsignal Revv Amp

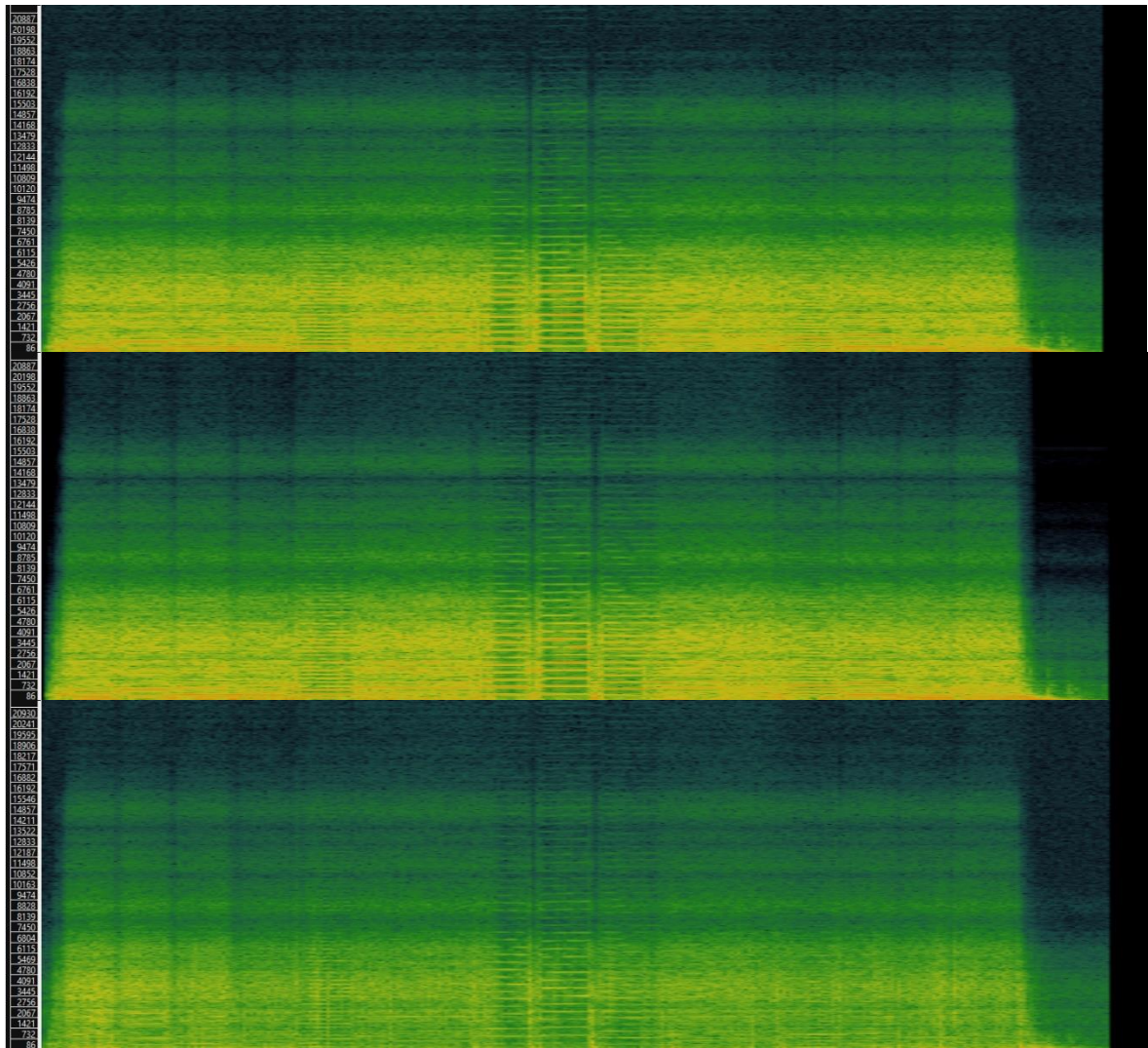


Zoom auf Decay beim Anschlagen einer tiefen Saite JCM900

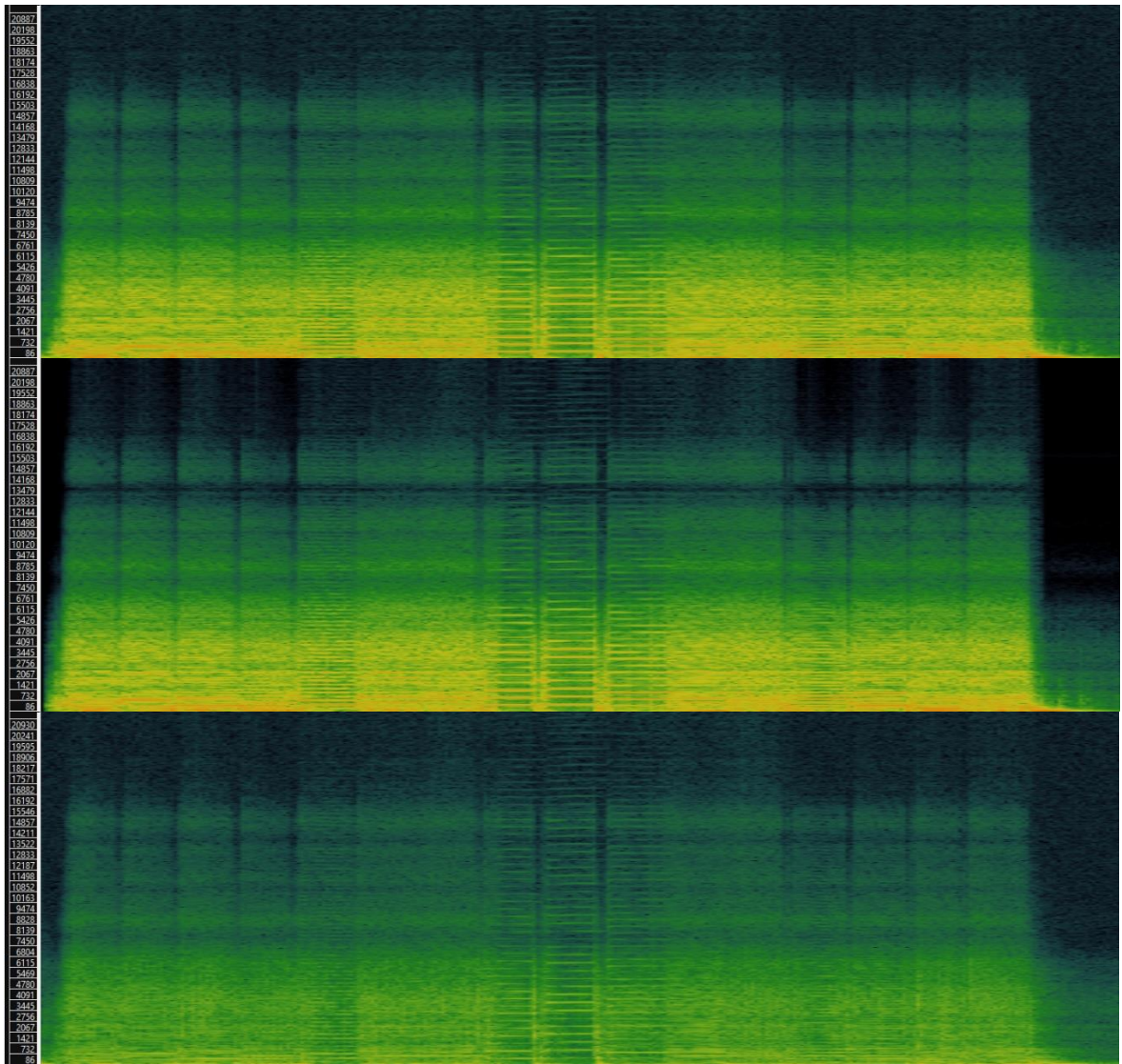


Zoom auf Decay beim Anschlagen einer tiefen Saite Revv Amp

Anhang 3.2: Spektrogramme

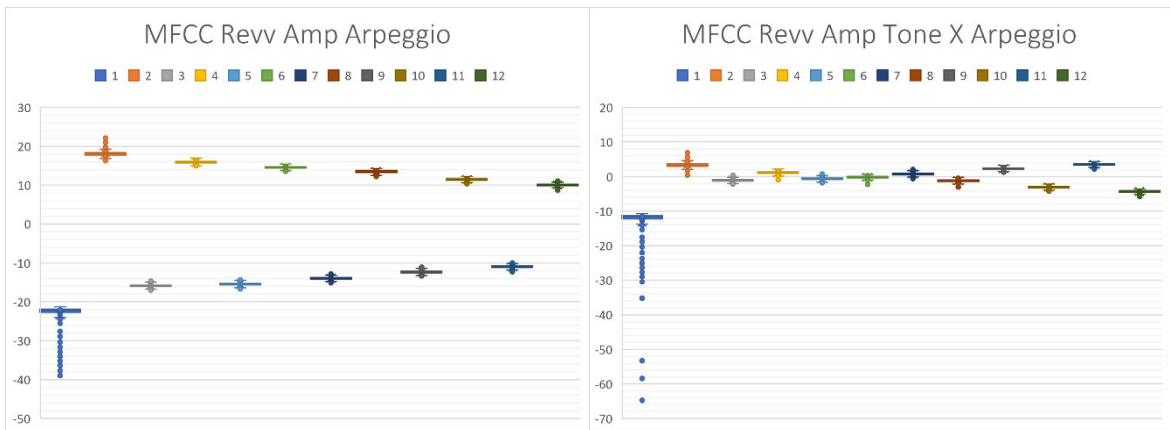
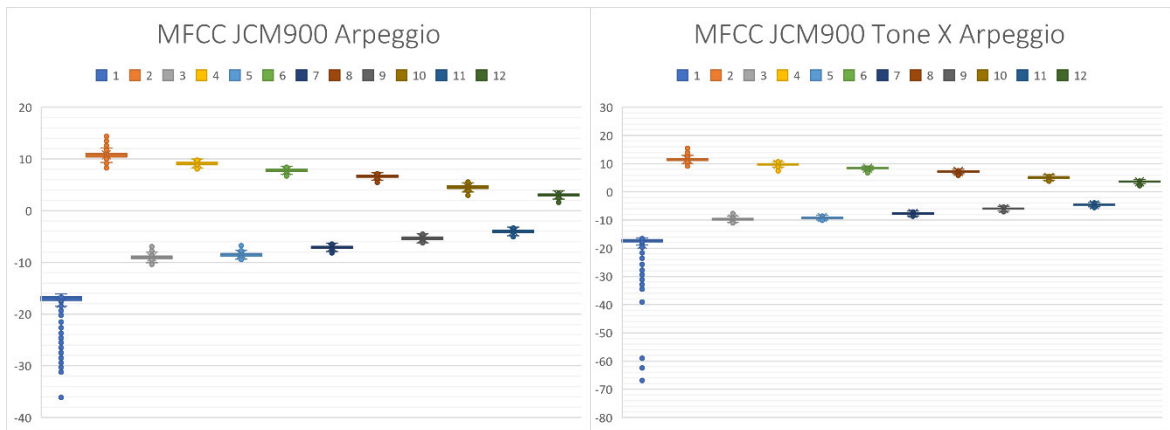


JCM900



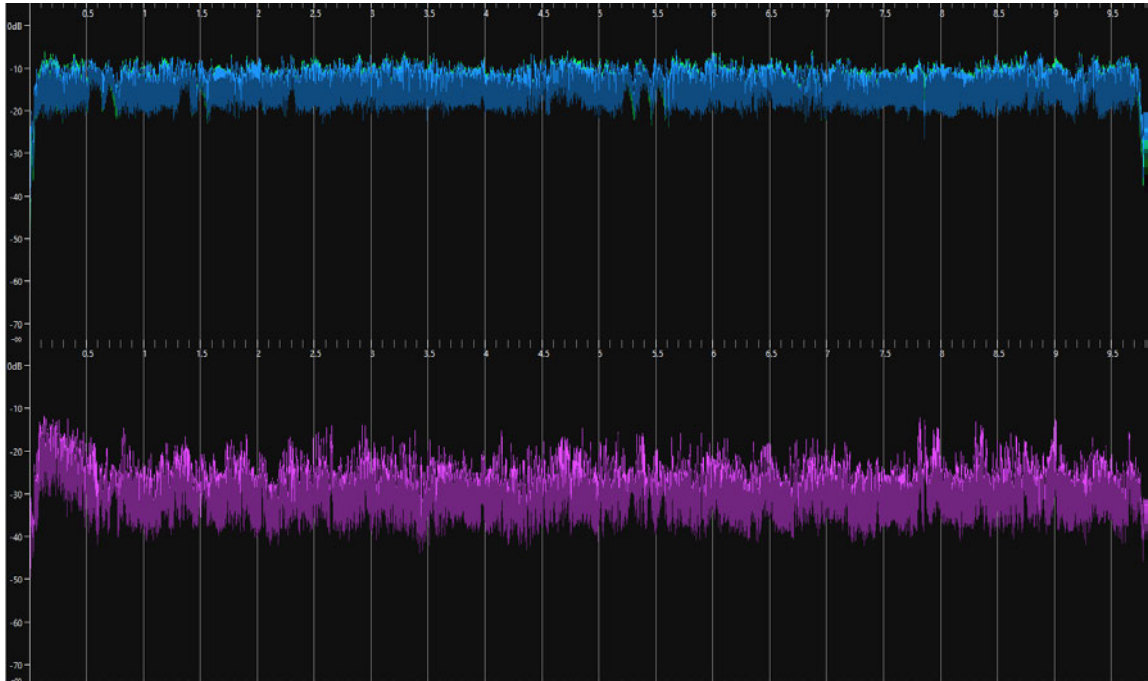
Rev Amp

Anhang 3.3: MFCC

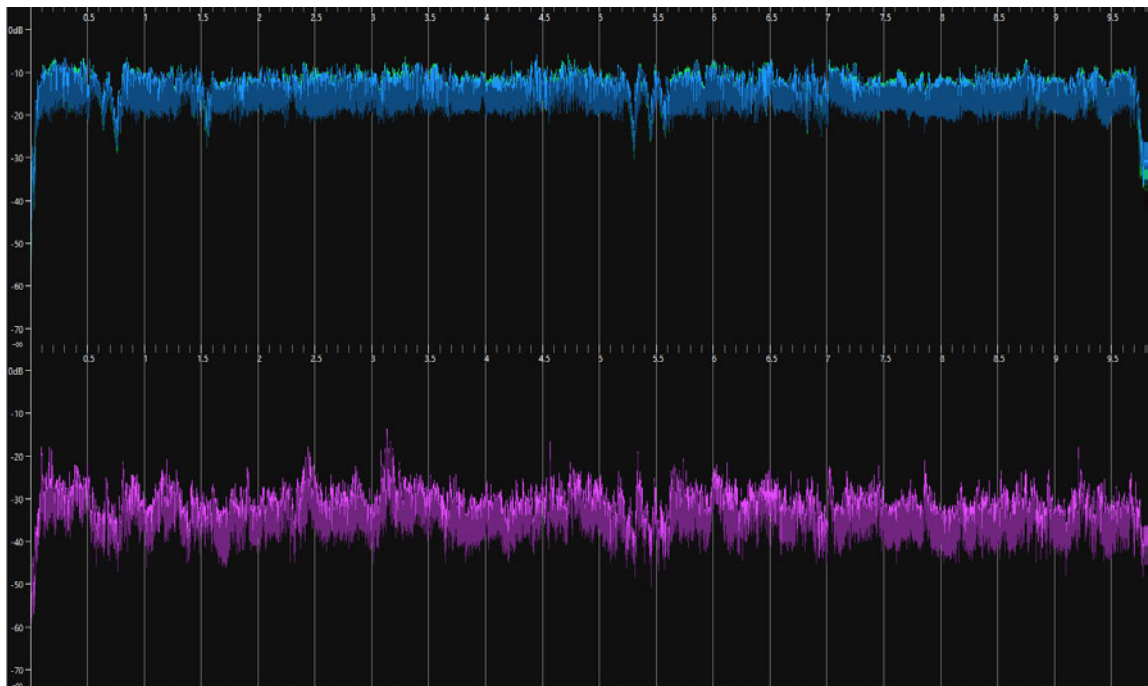


Anhang 4: Kurzes Riff

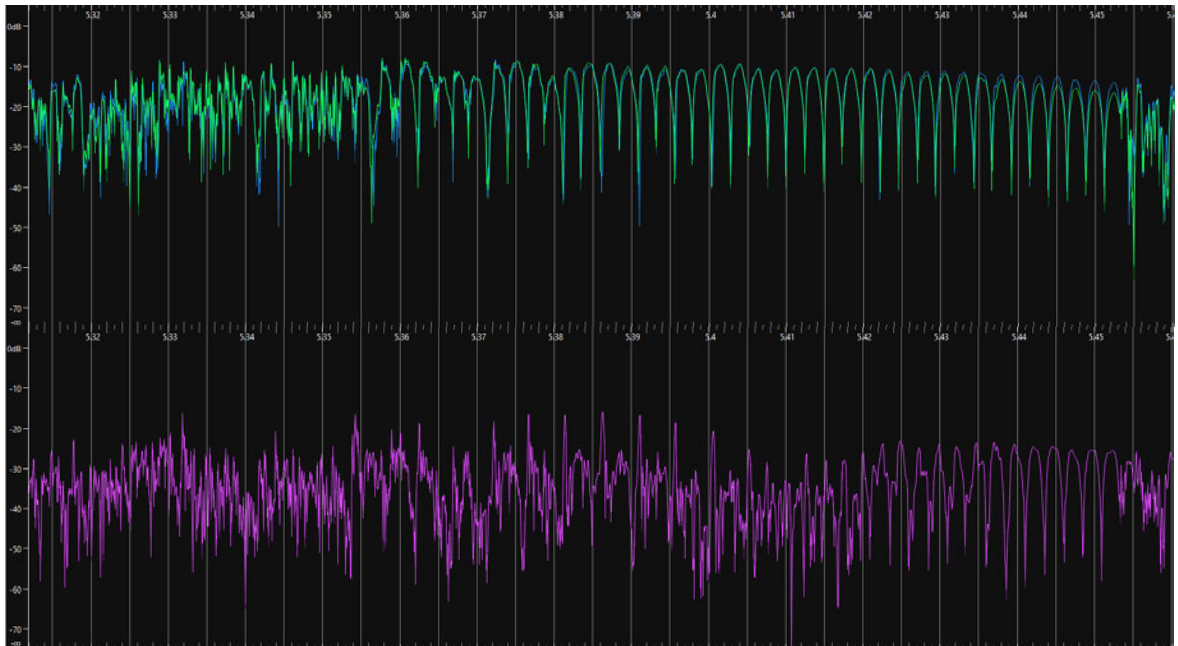
Anhang 4.1: Wellenformdiagramme



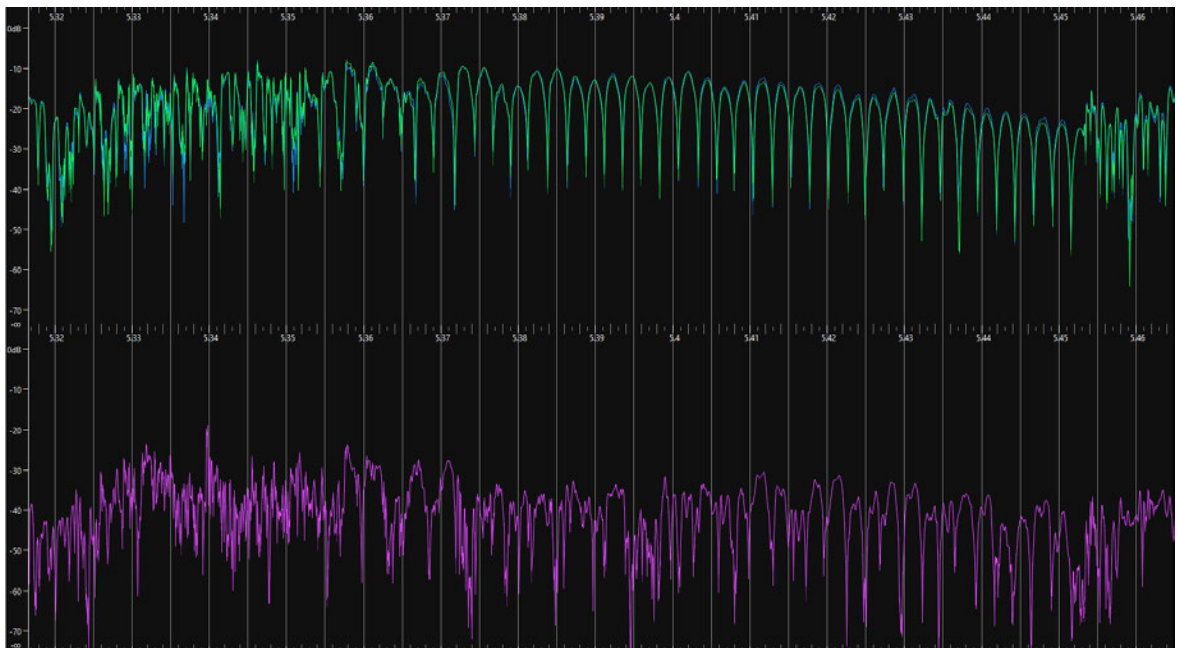
Gesamtsignal JCM900



Gesamtsignal Revv Amp

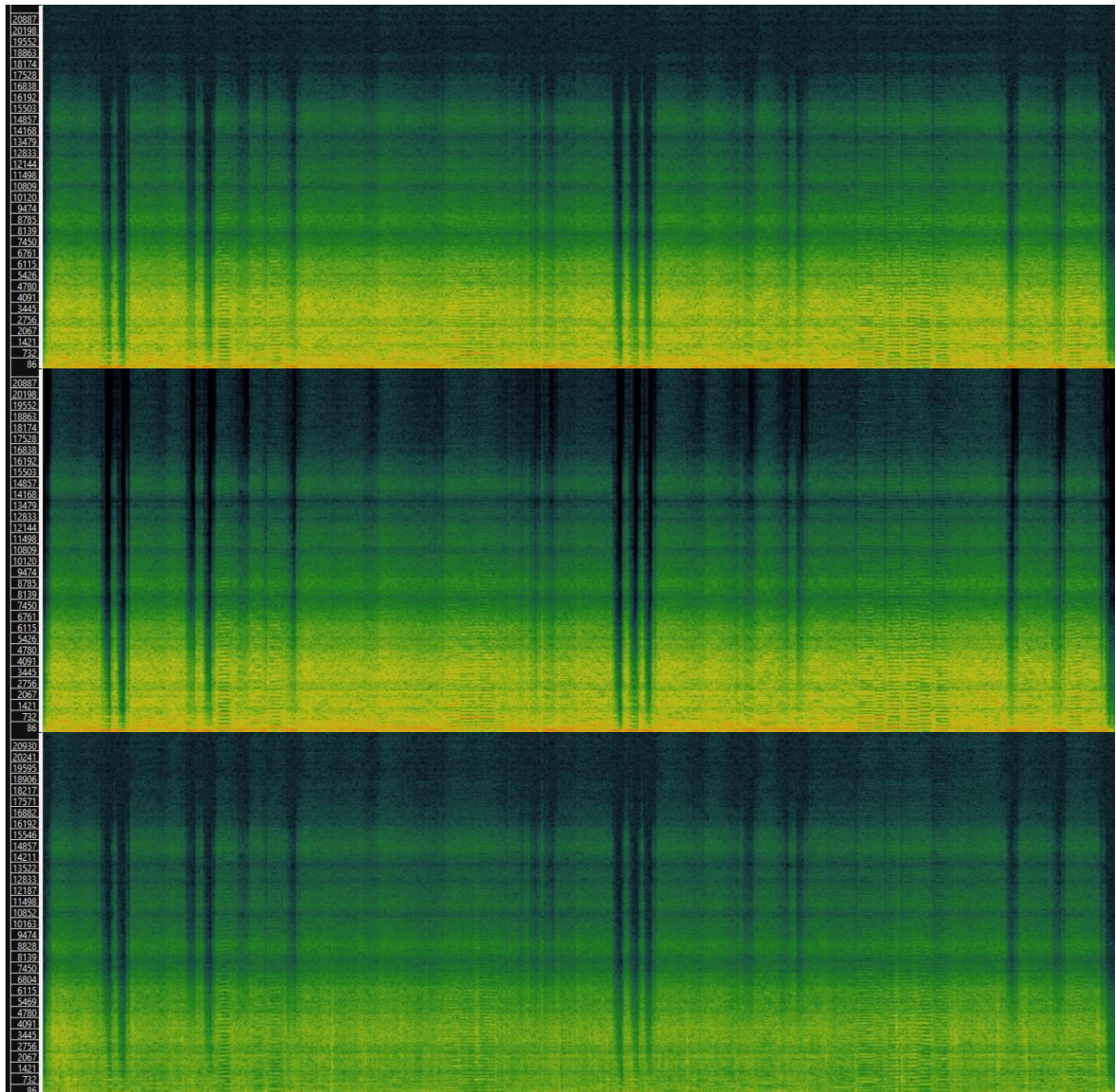


Zoom auf einzelnen Palm Mute JCM900

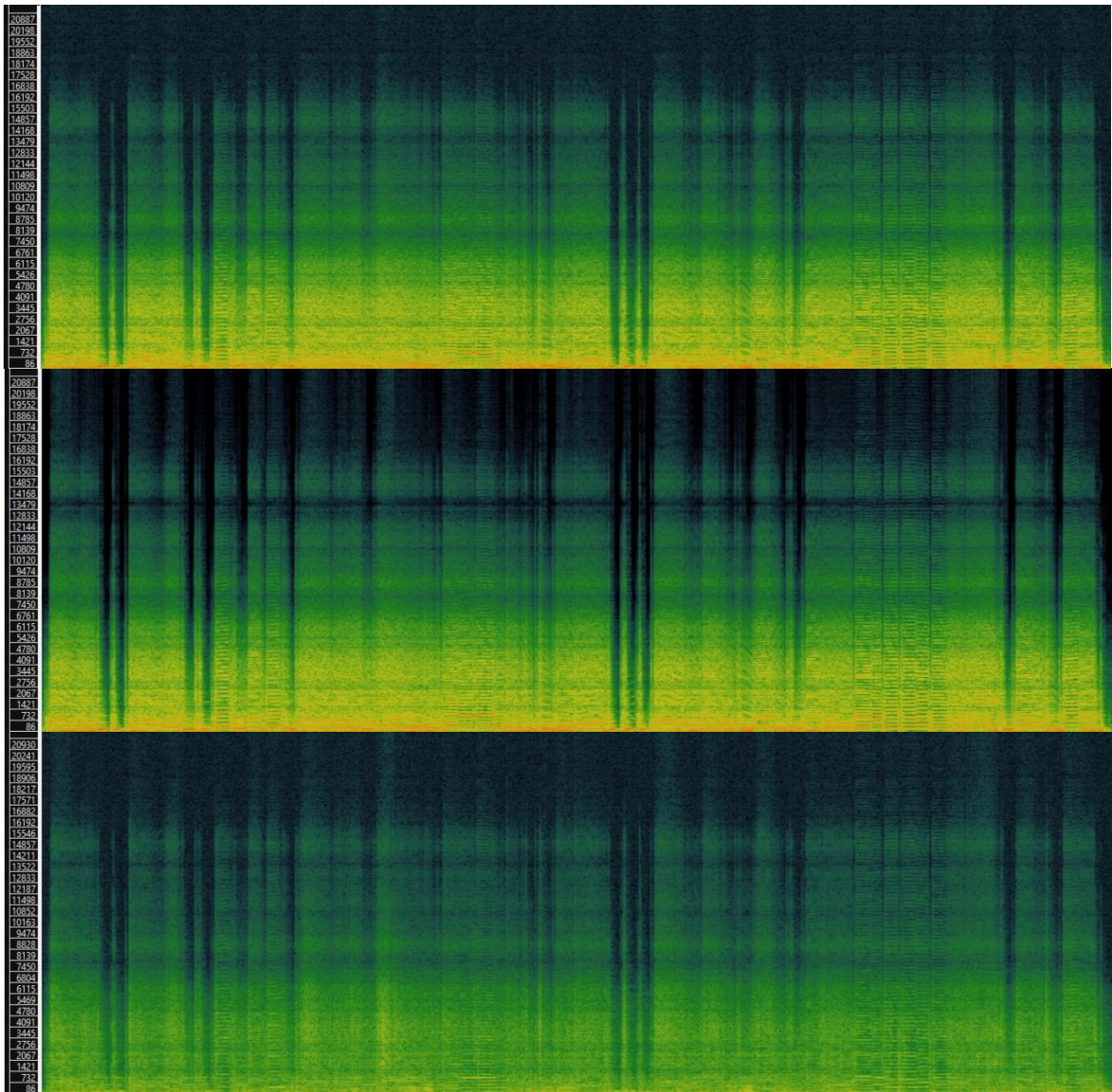


Zoom auf einzelnen Palm Mute Revv Amp

Anhang 4.2: Spektrogramme

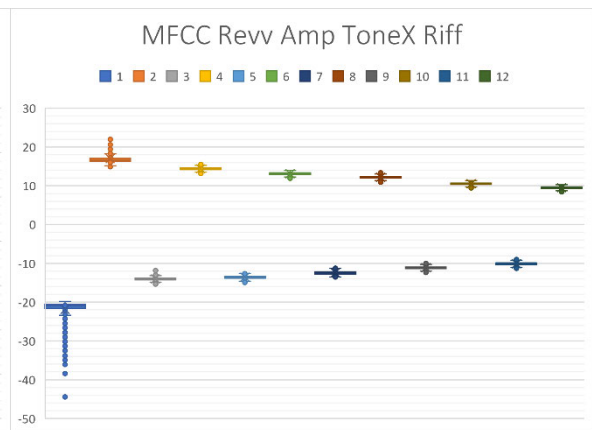
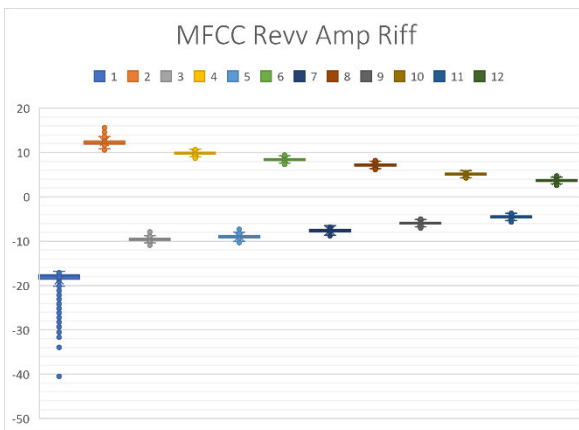
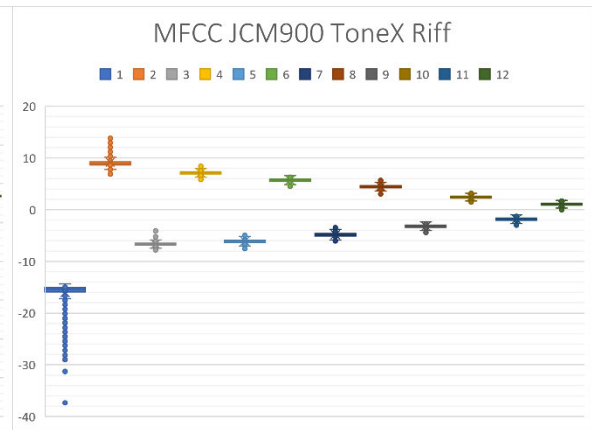
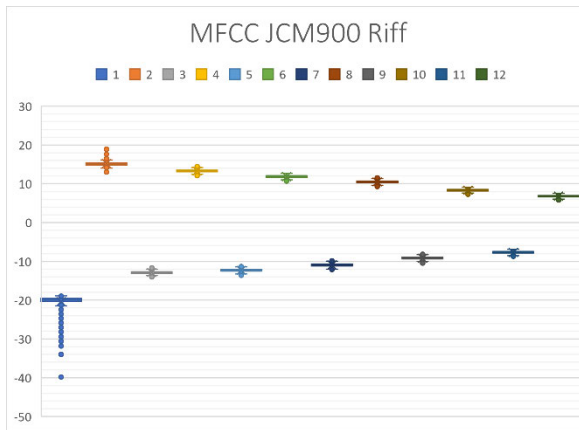


JCM900



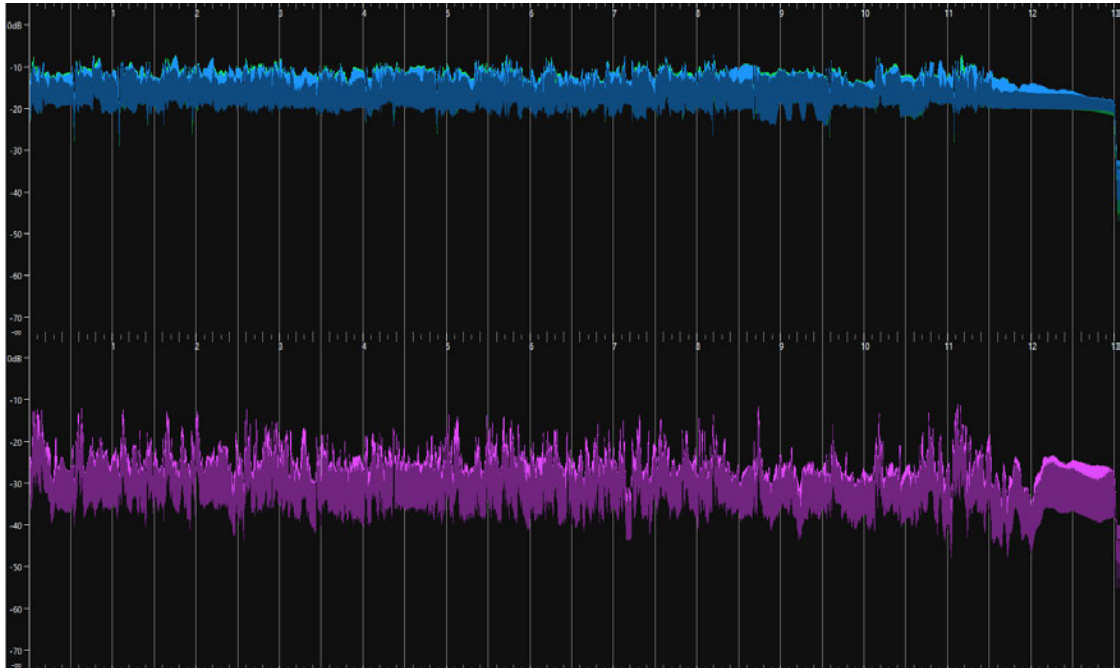
Revv Amp

Anhang 4.3: MFCC

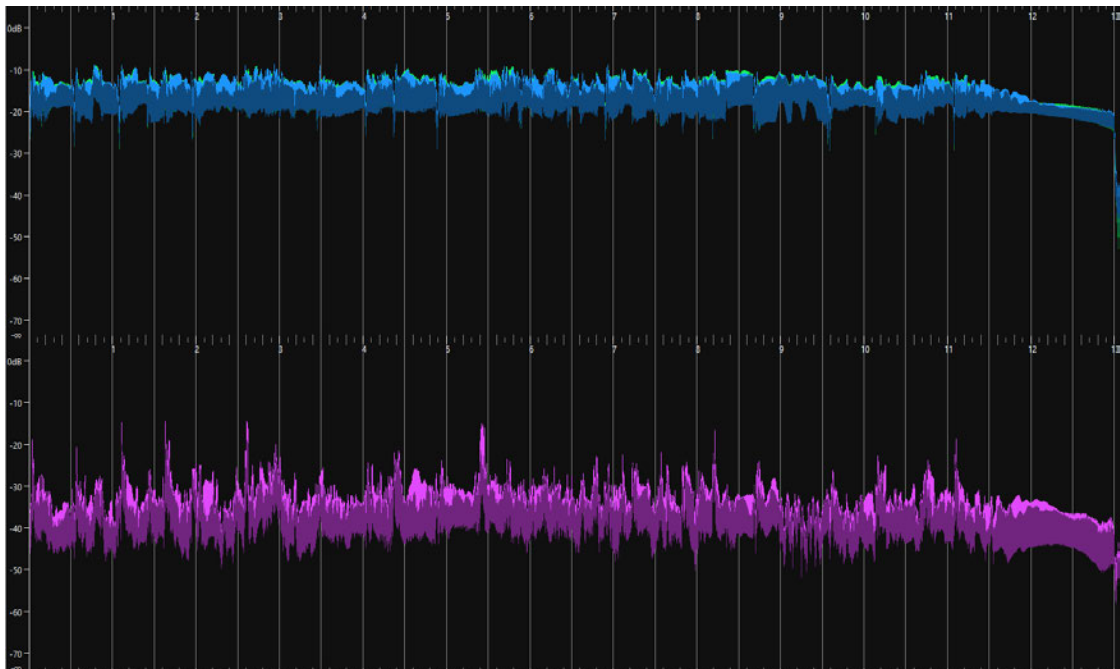


Anhang 5: Kurzes Solo

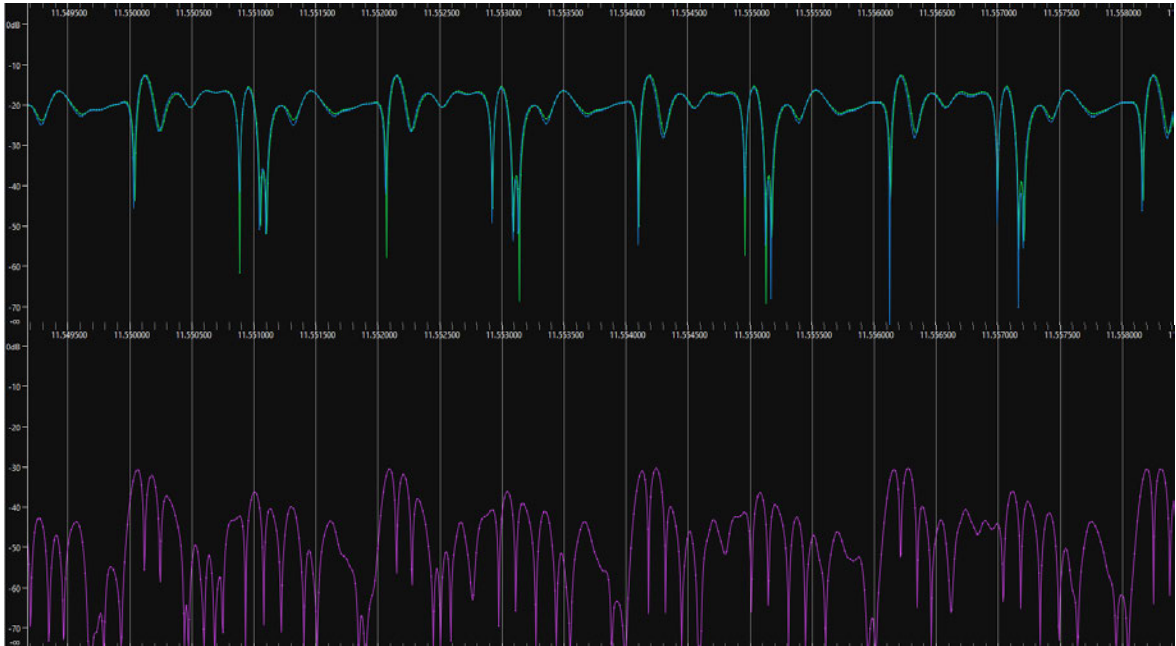
Anhang 5.1: Wellenformdiagramme



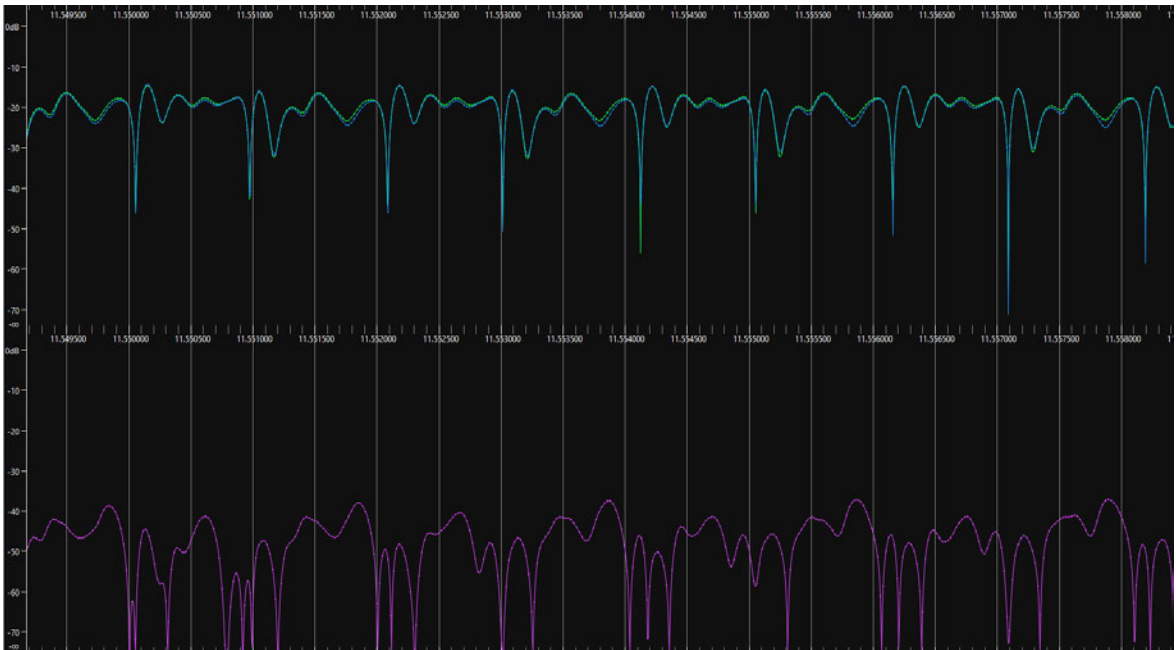
Gesamtsignal JCM900



Gesamtsignal Revv Amp

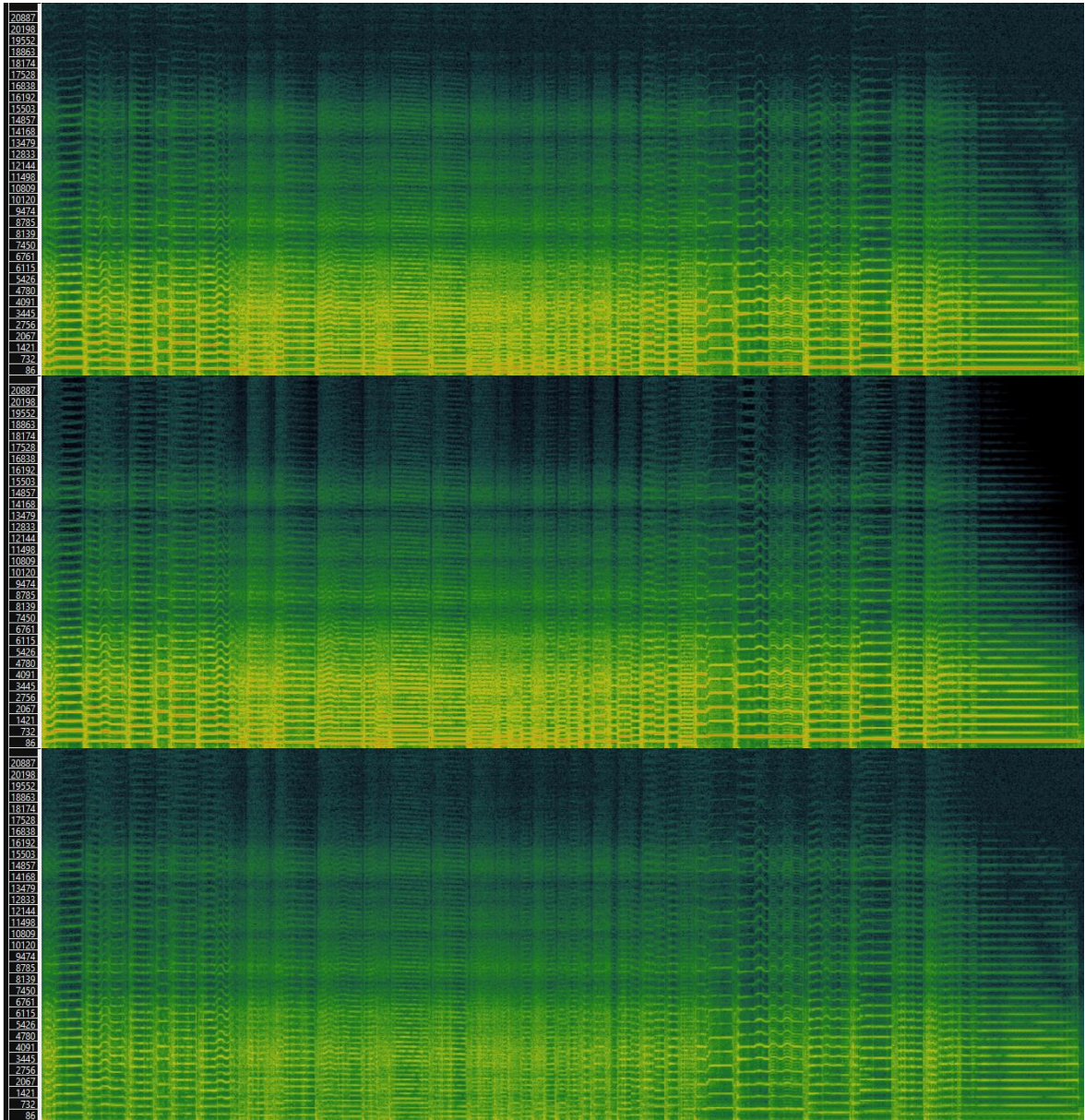


Zoom auf Sustain der letzten Note JCM900

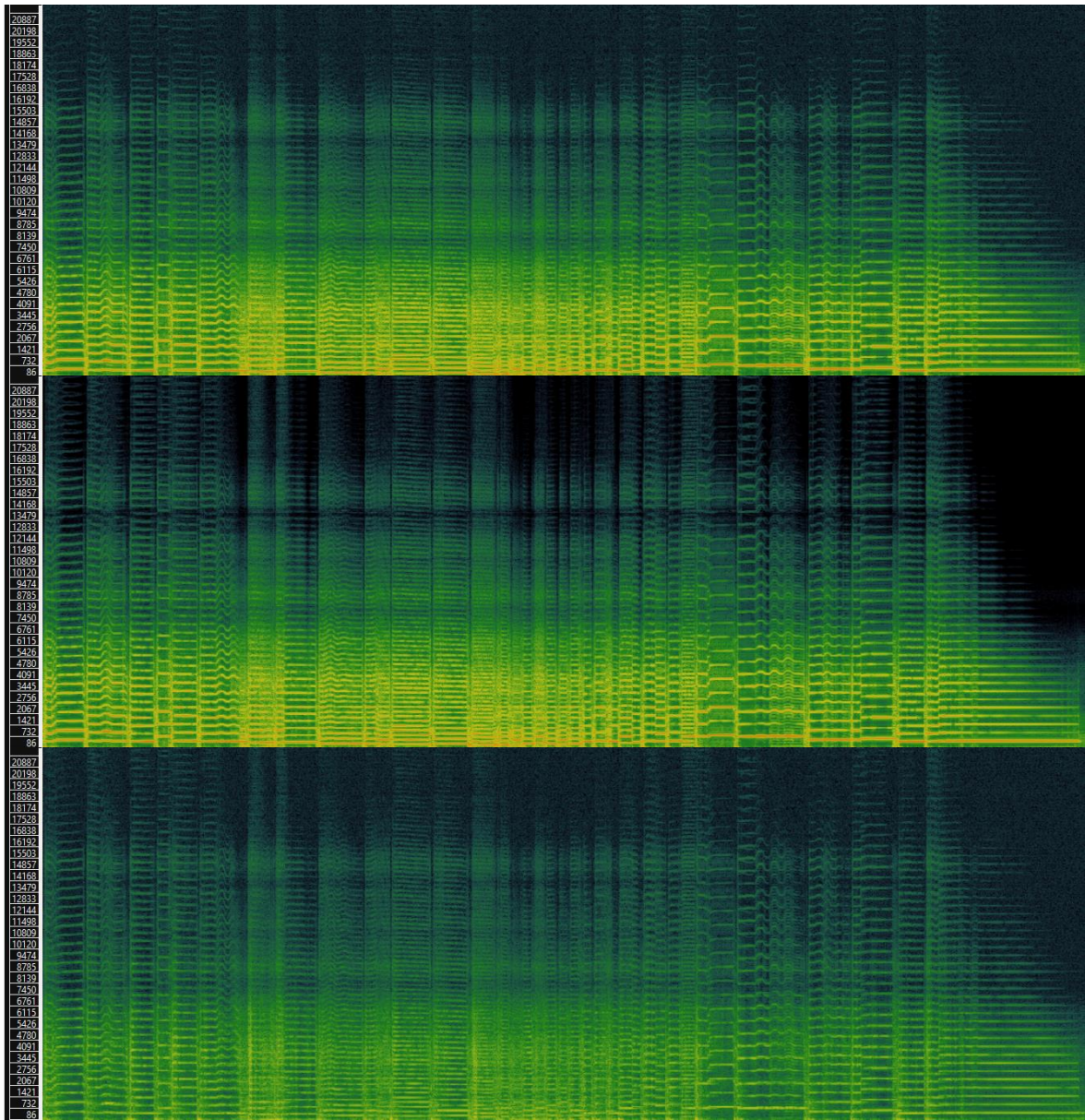


Zoom auf Sustain der letzten Note Revv Amp

Anhang 5.2: Spektrogramme

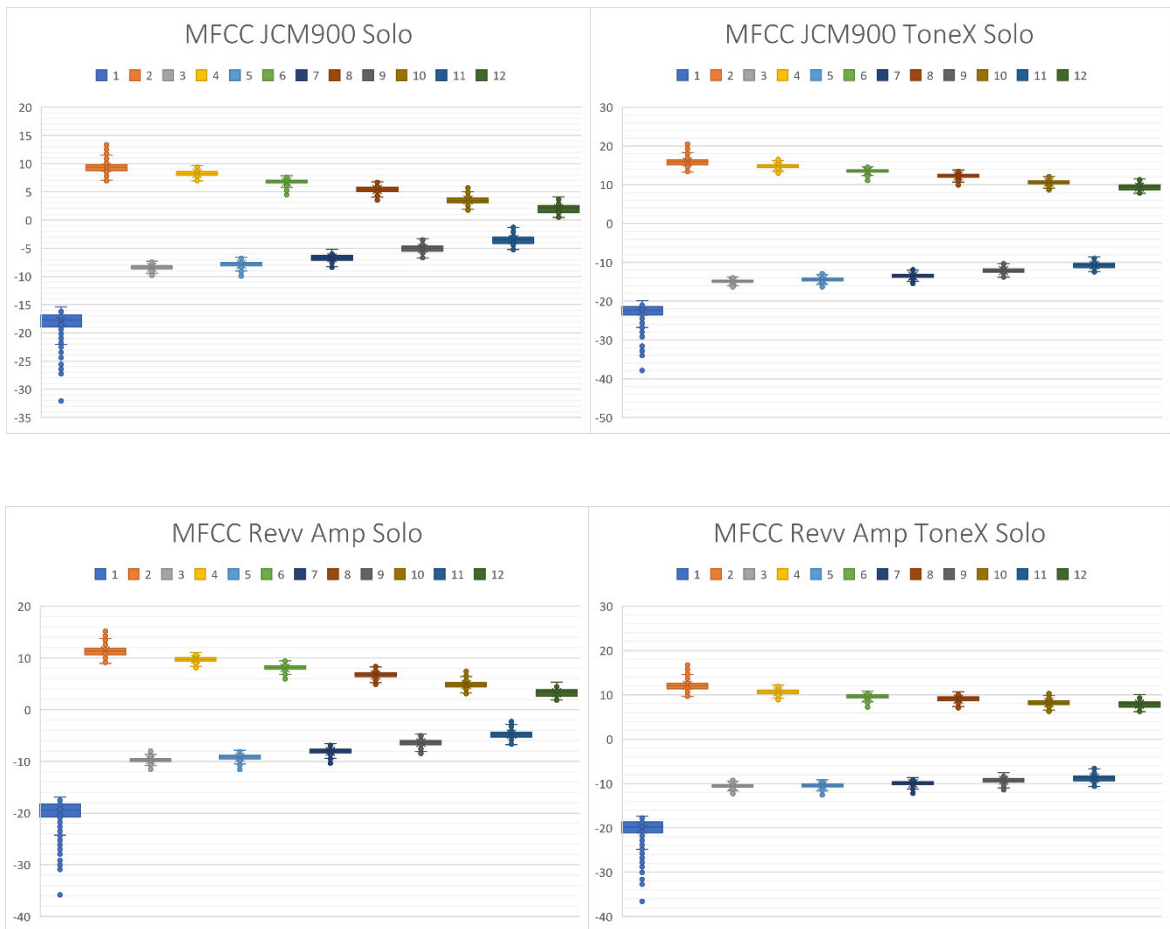


JCM900



Rev Amp

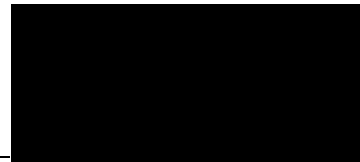
Anhang 5.3: MFCC



VI. Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Leipzig, 15.12.2023



Ort, Datum

Unterschrift