FH Salzburg

Stimmbearbeitung in der Popmusik

Die Klangästhetik und Technik synthetisch klingender Stimmen

Bachelorarbeit

Verfasser*in: Philipp Schinwald

Vorgelegt am Bachelorstudiengang MultiMediaArt, Fachhochschule Salzburg

Begutachtet durch: DI (FH) Martin Löcker

Salzburg, am 18.04.2025

Danksagung

An dieser Stelle möchte ich mich kurz bei den Menschen bedanken, die mich beim Schreibprozess dieser Bachelorarbeit unterstützt haben.
Mein besonderer Dank gilt meinem Betreuer DI (FH) Martin Löcker, der mir mit seinen wertvollen Anregungen und popkulturellen Beispielen sowie seiner fachlichen Expertise und seiner konstruktiven Kritik maßgeblich zu der vorliegenden Arbeit beigetragen hat.
Ebenfalls geht ein großes Dankeschön an meinen Onkel Reinhold Schinwald, der mich überhaupt auf die Thematik dieser Arbeit gebracht hat. Er hat mich auch in dem Prozess des Schreibens mit Denkanstößen, Literatur und spannenden Diskussionen begleitet. Ohne diese Unterstützung hätte diese Arbeit in ihrer jetzigen Form nicht entstehen können.
Des Weiteren bedanke ich mich bei meiner Familie und meinen Freunden, die immer ein offenes Ohr für mich hatten, wenn es um die Bachelorarbeit ging.

Salzburg, am 06.03.2025
Philipp Schinwald

Eidesstattliche Erklärung

Hiermit versichere ich, Philipp Schinwald, geboren am 09.12.2002 in Salzburg, dass ich die Grundsätze wissenschaftlichen Arbeitens nach bestem Wissen und Gewissen eingehalten habe und die vorliegende Bachelorarbeit von mir selbstständig verfasst wurde. Zur Erstellung wurden von mir keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

Ich versichere, dass ich die Bachelorarbeit weder im In- noch Ausland bisher in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der*den Betreuer*innen und Begutachter*innen vorgelegten Arbeit übereinstimmt.

Salzburg, am 18.04.2025

.....	.....
Philipp Schinwald	Personenkennzeichen

Kurzfassung

Vor- und Nachname:	Philipp Schinwald
Institution:	Fachhochschule Salzburg
Studiengang:	MultiMediaArt Bachelor
Titel der Bachelorarbeit:	Stimmbearbeitung in der Popmusik
Betreuer*in:	DI (FH) Martin Löcker
Schlagwörter:	
1. Schlagwort:	Popmusik
2. Schlagwort:	Stimmbearbeitung
3. Schlagwort:	Synthetik

Die vorliegende Bachelorarbeit beschäftigt sich mit dem Thema der Stimmbearbeitung zur Erzeugung synthetisch klingender Stimmen in der Popmusik mit der Forschungsfrage *Welche Techniken werden in der Popmusik eingesetzt, um Stimmen so zu bearbeiten, dass sie synthetisch klingen, und wie hat sich der ästhetische Ansatz entwickelt?* Die Faszination für synthetisch generierte Stimmen begann bereits früh und kam in den 1930er Jahren in der Popmusik als Stilmittel auf. Prägend waren dabei vor allem die Effekte Vocoder, Auto-Tune und Talkbox, auf die sich in der vorliegenden Arbeit fokussiert wird. Der Vocoder hat seinen Ursprung in der Telekommunikation, wurde im zweiten Weltkrieg zur Kommunikation zwischen den USA und Großbritannien eingesetzt und zog erst dann in die Musikwelt ein. Die Technik hinter Auto-Tune wurde vom Ölingenieur Harold Andy Hildebrand entwickelt, der durch die Technik, die er angewendet hat, um Öl zu lokalisieren, auf diese Idee kam. Diese Erfindung hat die Popmusik nachhaltig geprägt. Die Talkbox erschien 1939 erstmals in Form einer Puppe, in der Gestalt einer Gitarre, und wurde mit stetiger Weiterentwicklung ein Teil der Musikindustrie. Neben diesen drei Technologien lässt sich mit den meisten anderen Effekten ebenfalls ein *unnatürlicher* Gesangseffekt erzeugen, wenn diese intensivere Einstellungen aufweisen. Alle diese Effekte sind eine Interaktion zwischen Mensch und Maschine, wodurch auch Fragen der Echtheit und philosophische Gedanken des Posthumanismus aufkommen, auf die ebenfalls eingegangen wird. In der heutigen Musikproduktion ist die Bearbeitung von Gesang nach wie vor essenziell und wird häufig als Stilmittel eingesetzt. Im Rap beispielsweise ist Auto-Tune nicht mehr wegzudenken. Aufgrund dieser Wichtigkeit von Stimmbearbeitung, insbesondere dessen Verwendung als kreatives Stilmittel und der prägenden Rolle in der Popmusik, befasst sich die vorliegende Arbeit mit dieser Thematik, um den Ursprung und die Technik hinter den heutigen Trends zur Stimmbearbeitung und der Klangästhetik aufzuzeigen.

Abstract

This bachelor's thesis is about the topic of vocal processing for the creation of synthetic-sounding vocals in pop music, with the research question *What techniques are used in pop music to process vocals, so that they sound synthetic, and how has the aesthetic approach evolved?* To explore this subject, a combination of literature review and case analyses was used, to cover both, theoretical foundations and practical applications. The fascination for synthetic vocals begun as a stylistic tool in pop music in the 1930s. Important technologies in this development were the vocoder, Auto-Tune, and the talkbox, which are the focus of this bachelor's thesis. The vocoder has its origin in telecommunications, was later used in World War II and then got into the music industry. The technology behind Auto-Tune was developed by oil engineer Harold Andy Hildebrand, who came up with the idea for Auto-Tune through his work, which has had a lasting impact on pop music. The talkbox first appeared in 1939 in the form of a puppet that looks like a guitar and became an important part of the music industry. In addition to these three effects, many other vocal processing techniques can create an unnatural vocal effect when set to more intense levels. All these effects represent a fusion of human and machine, raising questions about authenticity and posthumanism, which are also discussed in this bachelor's thesis. In contemporary music production, vocal processing remains essential and is very often used as a stylistic tool. In rap, for example, Auto-Tune has become one of the main effects in the vocal chain. Given the significance of vocal processing, especially its use as a creative stylistic element and its defining role in pop music, this bachelor's thesis tries to analyse the origins and techniques behind modern trends in vocal processing and sound aesthetics.

Inhaltsverzeichnis

Abkürzungsverzeichnis 1

1. Einleitung 2

2. Begriffserklärungen und deren theoretischer Hintergrund 3

2.1. Popmusik 3

2.2. Gesang und Stimmbearbeitung in der Popmusik 4

2.3. Synthetisch klingende Stimmen 4

2.4. Ästhetik 5

3. Forschungsstand 6

4. Forschungsdesign 7

5. Historischer Überblick der Stimmbearbeitung für synthetisch klingende Stimmen 8

5.1. Die ersten Sprachsynthesen 8

5.2. Die Entwicklung des Vocoders 8

5.2.1. Anwendungen abseits der Musik 9

5.2.2. Einzug in die Popmusik 9

5.3. Die Entstehung von Auto-Tune 11

5.4. Die Geschichte der Talkbox 13

6. Techniken der Stimmbearbeitung und deren Funktionsweisen 14

6.1. Vocoder 15

6.2. Auto-Tune 17

6.3. Talkbox 19

7. Ausblick auf zukünftige Entwicklungen 20

8. Fazit 21

Literaturverzeichnis 23

Abbildungsverzeichnis 27

Glossar 28

Abkürzungsverzeichnis

Aufl. – Auflage

bzw. – beziehungsweise

etc. – et cetera

Hrsg. – Herausgeber

o.D. – ohne Datum

S. – Seite

usw. – und so weiter

1. Einleitung

In der heutigen Popmusik sind Effekte wie Hall, Delay oder auch solche, die eine Stimme unnatürlich und synthetisch klingen lassen, etablierte Mittel zur Bearbeitung des Gesangs. Unnatürlich meint dabei Klänge, die verfremdet werden und nicht der natürlichen Klangästhetik einer Stimme entsprechen und synthetische Klänge, die anmuten, als wären sie durch elektronische Mittel erzeugt worden.

Die Intention ist es zumeist eine außergewöhnliche Klangcharakteristik zu schaffen. Insbesondere in der heutigen Zeit, in der künstliche Intelligenz innerhalb der Musikindustrie immer relevanter wird und ganze Musikstücke mittlerweile anhand einer Texteingabe generiert werden können, ist diese Klangästhetik, die zwischen einem Menschen und einem Algorithmus entsteht für die heutige Zeit von Interesse und Relevanz. Prinzipiell lassen sich mit verschiedenen Bearbeitungsalgorithmen synthetisch klingende Stimmen erzeugen, insbesondere, wenn man Effekte überspitzt anwendet. In der vorliegenden Arbeit wird sich allerdings auf die prägendsten und bekanntesten Effekte zur Stimmbearbeitung für synthetisch klingende Stimmen beschränkt. Dabei handelt es sich um den Vocoder, Auto-Tune und die Talkbox, die durch ihre Anwendungen in Musikstücken die Popmusik nachhaltig geprägt haben. Diese Bachelorarbeit beschäftigt sich also mit der Bearbeitung von Gesang und nicht mit der Sprachsynthese.

Das Ziel der vorliegenden Arbeit ist es, die technischen Hintergründe dieser Effekte zu beleuchten und die Entwicklung dieser Klangästhetik genauer zu analysieren, um die Forschungsfrage *Welche Techniken werden in der Popmusik eingesetzt, um Stimmen so zu bearbeiten, dass sie synthetisch klingen, und wie hat sich der ästhetische Ansatz entwickelt?* beantworten zu können. Das beinhaltet auch zu reflektieren, welchen Einfluss diese Techniken auf die Popmusik mit sich brachten. Um dieses Ziel zu erreichen und die Forschungsfrage zu beantworten wurden eine Literaturrecherche, sowie kurze Analysen von relevanten Musikstücken durchgeführt. Dadurch, und durch detaillierte Erklärungen, wie etwaige Effekte anzuwenden und welche Parameter vorhanden sind, bietet diese Arbeit eine Brücke zwischen den theoretischen Grundlagen und der praktischen Anwendung in der Musikproduktion zu schaffen.

Die ersten Kapitel dieser Arbeit beschäftigen sich mit dem Forschungsstand und dem Forschungsdesign, worauf theoretische Grundlagen und begriffliche Erklärungen folgen. Anschließend wird auf die Geschichte von synthetisch klingenden Stimmen eingegangen, was kurz auch reine Sprachsynthesen einschließt. Nachdem sich diese Arbeit im Wesentlichen auf drei Effekte beschränkt, wird die Entstehung und Entwicklung von dem Vocoder, von Auto-

Tune und der Talkbox genauer betrachtet. Um in weiterer Folge die Funktionsweisen dieser Effekte verstehen zu können, wird im darauffolgenden Kapitel auf die Techniken eingegangen und verschafft zudem einen Überblick über gängige Effekte. Im letzten Teil der vorliegenden Arbeit bekommt man einen Ausblick auf zukünftige Entwicklungen, was primär künstliche Intelligenz thematisiert. Im abschließenden Fazit werden die gewonnenen Ergebnisse am Ende der Arbeit zusammengefasst. Zum besseren Verständnis werden wichtige Begriffe in einem Glossar erklärt, das sich auf den letzten Seiten befindet.
Die Motivation über dieses Thema zu schreiben, kommt aus der Faszination, die Stimme abseits vom Gesang im klassischen Sinne als Instrument zu verwenden. Der Musikproduzent Zedd hat in seinen beiden Musikstücken *Stay* und *The Middle* den Vocoder eingesetzt, um mit dem Gesang Akkorde zu bilden. Diese beiden Musikstücke entwickelten sich im Verlauf der Auseinandersetzung mit dem Thema zu zentralen Inspirationsquellen. Um zu verstehen, wie sich diese Klangästhetik entwickelt hat und welche Techniken dahinterstecken, wurde dies zum Thema dieser wissenschaftlichen Arbeit.

2. Begriffserklärungen und deren theoretischer Hintergrund

2.1. Popmusik

Popmusik wird in der Literatur oftmals auch als *Pop-Musik*, *populäre Musik*, *populärkulturelle Musik*, *Rock- und Popmusik* oder auch nur als *Pop* bezeichnet. Im Englischen wird überwiegend der Begriff *Popular Music* verwendet. Zur Vereinfachung ist in der vorliegenden Arbeit nur von *Popmusik* die Rede. Das Wort *Pop* ist dabei die Abkürzung von *populär*. Das Verständnis von Popmusik ist sehr unterschiedlich: Man liest in manchen Fällen von Musik, die wenig anspruchsvoll ist, und in anderen Fällen Musik, die weit verbreitet und beispielsweise im Radio zu hören ist (Seifert, 2018). Popmusik lässt sich also nicht auf einen bestimmten Musikstil beschränken. Popmusik kann sowohl Schlager oder Folklore sein, sowie nicht-westliche Musik. Sie entstand in den 1950er Jahren und geht über die Grenzen der Musik hinaus, es handelt sich vielmehr um ein sozial konstruiertes Phänomen (Diederichsen, 2014). Gemeint sind auch inhaltliche Aspekte, wie Aussagen, Texte oder Symbole, optische Wahrnehmungen, wie Kleidung, Videos, Bilder und Frisuren und körperliche Empfindungen, wie das Tanzen oder der Geruch bei Live-Auftritten. Erst also durch das Fantum bzw. die Konsument*innen entsteht Popmusik bzw. die damit einhergehende Popkultur. Sie kann sozialpolitische Themen aufgreifen, kann aber auch allein der Unterhaltung dienen, ohne einen tiefsinnigen Inhalt zu transportieren. Der Musiksoziologe und Philosoph Theodor W. Adorno schrieb der Popmusik eine niedrigere künstlerische Bedeutung als beispielsweise ernste Musik zu, sagte jedoch auch, dass diese anspruchsvoll sein kann. Dadurch, dass die Wünsche der Hörer*innen von den Musikproduzent*innen erfüllt werden wollen, entsteht laut Adorno

eine Standardisierung, wodurch der Aufbau der Musiktitel in vielen Fällen ähnlich ist. Man versuche laut Adorno, dass die Musik von vielen Menschen gehört und der erzielte Gewinn maximiert wird (Seifert, 2018).

2.2. Gesang und Stimmbearbeitung in der Popmusik

Eine Studie von The International Society for Music Information Retrieval (ISMIR) (Bittner et al., 2018) untersuchte die Musikpräferenzen von Nutzer*innen auf Spotify in den USA und Kanada. Den Hörer*innen wurden offene Fragen dabei gestellt, welche Faktoren für sie die Wichtigsten in der Musik sind. An der ersten Stelle der Umfrage befindet sich die Stimmung bzw. die Emotionen, die durch ein Musikstück ausgelöst werden. Direkt an der zweiten Stelle kommt die Stimme bzw. der Gesang, gefolgt vom Songtext. Alle weiteren Faktoren, wie Rhythmus, Struktur, Harmonien etc. befinden sich weiter hinten in den Ergebnissen. Der Gesang wird im Gehirn anders verarbeitet als Instrumentalmusik und sorgt ebenfalls dafür, dass gesungene Melodien einprägsamer sind als instrumental gespielte (Bittner et al., 2018). Nachdem alle gesunden Menschen eine Stimme besitzen und summen oder singen können, fällt es ihnen in der Regel einfach, sich mit einer Stimme zu identifizieren. Sie ermöglicht einen Zugang zu den Emotionen und ist ein Mittel des Ausdrucks (Pfleiderer, 2010). Folglich lässt sich also erschließen, dass der Gesang eine zentrale Rolle in der Musik spielt. Deshalb ist es wichtig, dass ihr in der Musikproduktion besondere Aufmerksamkeit gewidmet wird. Dies geschieht unter anderem durch die Bearbeitung einer Gesangsaufnahme. Das Ziel der Bearbeitung ist dabei, dass Interpretationsfehler oder Störgeräusche entfernt werden, eine ausbalancierte Dynamik sichergestellt wird, die Hörgewohnheiten und Erwartungen von Hörer*innen berücksichtigt werden und durch Effekte eine gewünschte Klangästhetik geschaffen wird. Auf Letzteres konzentriert sich die vorliegende Arbeit. Eine Bearbeitung kann mittels Hardware-Geräten, Software-Programmen, Plugins oder eine Kombination aus Hardware und Plugin erfolgen. Über eingebaute Algorithmen und die selbst eingestellten Parameter der jeweiligen Effekte lassen sich Audiosignale bearbeiten (Maempel, 2021).

2.3. Synthetisch klingende Stimmen

„Das älteste, echteste und schönste Organ der Musik, das Organ, dem unsere Musik allein ihr Dasein verdankt, ist die menschliche Stimme“ (Wagner, o.D., zitiert nach Schuster, 2020, S. 134). Die Stimme ist das meistgehörte Instrument des Menschen, ob es Gesang in der Musik ist oder gesprochene Sprache bei einer Unterhaltung. Dementsprechend schnell fällt es auf, wenn dieses menschliche Instrument bearbeitet wurde. In der Regel klingen Stimmen unnatürlich, wenn Töne über eine längere Zeit ohne zu schwingen gehalten werden (Crockett, 2016). Der Einzug von dieser unnatürlichen Klangästhetik brachte Fragen der Authentizität mit sich. Nachdem die Stimmbearbeitung zu einem etablierten Vorgehensprozess innerhalb der

Popmusik wurde, verschob sich somit auch die Bewertung von Authentizität. Diese wurde traditionell mit Live-Performances assoziiert, in der modernen Musikproduktion spielt vielmehr die Art und Weise, wie ein Musikstück produziert wird und wie Musik und Stimme miteinander harmonieren eine Rolle (Ahlers, 2018). Das zeigt auch die Studie von ISMIR (Bittner et al., 2018): Diese ergab, dass an zweiter Stelle der wichtigsten Faktoren für den Gesang steht, wie gut die Stimme zum restlichen musikalischen Gesamtbild passt. Erst an der vierten Stelle befindet sich die Authentizität bzw. Echtheit einer Stimme (Bittner et al., 2018). In der digitalen Musikproduktion kann Gesang technisch manipuliert oder gar komplett synthetisch sein, oftmals auch ohne, dass dies auffällt. Deshalb hat die Authentizität weniger mit der Natürlichkeit einer Stimme zu tun, sondern vielmehr mit den emotionalen Assoziationen, die mit diesem Klang verbunden werden (Daniel & Hillebrandt, 2019). Laut der von ISMIR durchgeführten Studie liegt *Wie natürlich die Stimme klingt* an neunter Stelle der wichtigsten Faktoren für den Gesang, während sich *Wie emotional die Stimme ist* an siebter Stelle befindet (Bittner et al., 2018). Die Verwendung von Auto-Tune und die damit einhergehende Korrektur aller abweichenden Tonhöhen kann zum Verlust eines fesselnden und spannenden Klangs führen. Kommt es bei einem Gesang zu Abweichungen, ist oft unklar, ob die*der Sänger*in dies intendiert hat oder etwa so emotional berührt war, dass die Stimmführung nicht mehr zur Gänze kontrolliert werden konnte. Wenn nun also alle Töne durch die Nachbearbeitung genau getroffen werden, können so Emotionen verloren gehen (Schuster, 2020, S. 137–142). Denn wie Schuster (2020) es beschreibt: „Die Ausdrucksweise, wie man singt, spricht oder rappt drückt viel über Emotionen und Empfindungen des Künstlers aus. Da die Gesangsstimme dem Menschen und seiner natürlichen Ausdrucksweise am nächsten kommt, hat sie schon durch minimale Veränderungen einen großen Wirkungsspielraum“ (S. 139). Je nach Einsatz kann aber auch das Gegenteil der Fall sein: Verleiht man der menschlichen Stimme einen synthetischen und unnatürlichen Klang, entsteht eine neue Stimmcharakteristik, die außergewöhnlich oder gar extrem wirken und somit interessant für die Zuhörer*innen sein kann. Effekte wie Auto-Tune können eine Stimme an die gesanglichen Grenzen bringen, die üblicherweise für eine gefühlsbetonte Haltung stehen, wodurch Emotionen bei dem Hören entstehen können. Insgesamt lässt sich also schlussfolgern, dass es fallabhängig ist, ob durch die Nachbearbeitungen Emotionen dazugewonnen werden oder verloren gehen (Schuster, 2020).

2.4. Ästhetik

Der Begriff *Ästhetik* kommt aus dem Griechischen (*aísthēsis*) und heißt übersetzt *sinnliche Wahrnehmung*, es ist im übertragenen, wortwörtlichen Sinne also die Theorie von sinnlicher Wahrnehmung und Erkenntnis. Es gibt allerdings keine eindeutige Definition dieses Begriffes. So liest man beispielsweise, dass die Ästhetik die Theorie der sinnlichen Erkenntnis, der Kunst

oder des Schönen sei. All diesen Definitionen kann man allerdings widersprechen: Nachdem Träume oder Vorstellungen ästhetisch sein können, ist die Eingrenzung auf sinnliche Erkenntnisse ungeeignet. Die Theorie der Kunst ist ebenfalls deplatziert, da auch die Natur ästhetisch sein kann, genauso wie ein Gewitter, wodurch auch die Theorie der Schönheit unpassend ist (Kaiser, 2017). Maria E. Reicher (2005) beschreibt die Ästhetik als Theorie der ästhetischen Eigenschaften, Erfahrungen und Gegenstände, wodurch sie subjektive und objektive Ästhetik miteinschließt (S. 16). In der Popmusik bezieht sich die Ästhetik nicht ausschließlich auf das Werk, sondern vielmehr auf die körperliche und sinnliche Erfahrung, die während dem Hören der Musik entsteht (Daniel & Hillebrandt, 2019).

3. Forschungsstand

Zur Ausarbeitung dieser Arbeit wurde eine systematische Literaturrecherche durchgeführt. Quellen aus wissenschaftlichen Datenbanken, wie Google Scholar oder die SpringerLink-Datenbank wurden bevorzugt, um relevante und wissenschaftliche Quellen identifizieren zu können. Abseits davon wurden auch einige Online-Artikel von Musikmagazinen und relevanten Unternehmen herangezogen, die oftmals einen praxisnahen Zugang zur Thematik liefern konnten. Neben Internetquellen wurden auch Fachbücher zur Beantwortung der Forschungsfrage herangezogen.

Um für das Thema relevante Literatur zu finden, wurden gezielte Fachbegriffe angewendet, nach denen gesucht wurde. Beispiele hierfür wären *Stimmbearbeitung*, *synthetische Stimmen* und *Popmusik*. Es wurde Literatur herangezogen, die sowohl den historischen Hintergrund beschreibt als auch die technischen Aspekte erklärt. Bei den technischen Aspekten wurde nach Quellen gesucht, die direkt von den Erfinder*innen des jeweiligen Effekts stammen, wie das Patent für Auto-Tune oder einen Artikel zur Funktionsweise des Vocoders von dessen Erfinder.

Bei Literatur, die nicht von den Erfinder*innen stammt wurde stets darauf geachtet, dass eine Nennung der Autor*innen, sowie ein Impressum bei Internetquellen vorhanden ist. Ausschließlich bei relevanten Unternehmen oder Institutionen wurden auch Quellen ohne die Nennung der Autor*innen gewählt.

Die historische Entwicklung von synthetisch klingenden Stimmen in der Popmusik und die genaueren technischen Hintergründe mit den Auswirkungen auf die Popmusik und die zukünftige Entwicklung wurden in der Literatur bisher noch nicht gemeinsam in einen Kontext gesetzt. Die vorliegende Arbeit soll daher eine Brücke bilden, die eben nicht nur die historischen und technischen Aspekte beleuchtet, sondern auch die nachhaltige Wirkung auf das Publikum und die Popmusik, was auch das Betrachten von spezifischen Beispielen inkludiert.

4. Forschungsdesign

Stimmbearbeitung und künstlich klingender Gesang werden in der wissenschaftlichen Forschung aus unterschiedlichen Blickwinkeln betrachtet. Die Thematik kann sowohl technisch und historisch als auch kulturell und philosophisch eingeordnet werden. Dabei kommen neben der Forschungsfrage *Welche Techniken werden in der Popmusik eingesetzt, um Stimmen so zu bearbeiten, dass sie synthetisch klingen, und wie hat sich der ästhetische Ansatz entwickelt?* auch Fragen auf, wie *Was waren die musikalischen Meilensteine in der Popmusik, die dazu beigetragen haben, synthetisch klingende Stimmen zu etablieren?* oder *Wie wirken sich diese Effekte auf die Authentizität aus?* Die Hypothese der vorliegenden Arbeit ist, dass synthetisch klingende Stimmen durch ihr erstmaliges Auftreten in der Popmusik für eine gewisse Faszination bei den Hörer*innen gesorgt haben. Diese Klangästhetik könnte schnell zu einem kreativen Stilmittel geworden sein, das aber auch zu einer gewissen emotionalen Distanz zu der*dem Sänger*in führen könnte. Ebenfalls wird angenommen, dass künstliche Intelligenz in der zukünftigen Entwicklung vor allem hinsichtlich der Sprachsynthese von großer Bedeutung sein wird, wodurch es auch zu unnatürlich klingenden Stimmen kommen kann.

Um das Thema wissenschaftlich zu beleuchten, wird die Methodik der systematischen Literaturrecherche angewendet, die vorwiegend theoretische Grundlagen liefert. Dabei sollen Fachliteratur, Studien sowie Patente herangezogen werden. Dieser theoretische Rahmen soll mit praktischen und relevanten Beispielen und kurzen Fallanalysen aus der Popmusik gepaart werden, um eine Kombination aus Theorie und Praxis zu schaffen. Die Datenbasis besteht aus Studien, wie der Studie von ISMIR, die aussagt, dass der Gesang ein zentrales Element in der Musik spielt (Bittner et al., 2018). Weiters werden technische Dokumentationen, wie die Beschreibung der Funktionsweise des Vocoders vom Erfinder des Vocoders Homer Dudley in dem *Bell System Technical Journal* von 1940 herangezogen. Darüber hinaus befinden sich Patente in der Datenbasis, wie das 1999 veröffentlichte Patent von Auto-Tune, das von dem Erfinder Harold Andy Hildebrand konzipiert wurde und die Funktionsweise von diesem Effekt beschreibt, sowie Musikbeispiele, wie *Believe* von Cher. Zunächst werden theoretische Grundlagen und zentrale Begriffe erläutert, bevor im Anschluss die historische Entwicklung synthetisch erzeugter Stimmen thematisiert wird. In diesem Abschnitt werden bereits einige Beispiele aus der Popmusik genannt, um das Thema greifbar zu machen. Danach wird auf die technischen Hintergründe und Funktionsweisen eingegangen, sowie die Erklärungen einiger Parameter für den praktischen Bezug. Zum Abschluss soll ein Ausblick zeigen, wie die Zukunft dieser Klangästhetik aussehen wird.

5. Historischer Überblick der Stimmbearbeitung für synthetisch klingende Stimmen

5.1. Die ersten Sprachsynthesen

Anfang des 11. Jahrhunderts soll Gerbert von Aurillac ein Gerät aus Bronze gebaut haben, welches er *Sprechenden Kopf* taufte, und die Wörter *Ja* und *Nein* synthetisieren konnte. Aufgrund mangelnder Aufzeichnungen ist es schwierig, die Funktionsweise dieses Geräts genauer zu erläutern bzw. ist es fraglich, ob es sich bei diesem Gerät nicht um eine Legende handelt. Rund 800 Jahre später entstand der *Sprechapparat des Wolfgang von Kempelen*, mit dem es möglich war, mehrere Laute zu erzeugen (Deutsches Museum, o.D.). Dieser Apparat versuchte die menschliche Anatomie, die es zum Sprechen braucht, nachzuahmen: Ein Blasebalg soll die Lunge simulieren an dem eine Schnur mit Gewicht befestigt ist, das, ähnlich dem menschlichen Zwerchfell, den Balg aufzieht und wieder mit Luft befüllt. Die Luft strömt anschließend durch eine speziell angefertigte Zungenpfeife, die als Stimmlippen fungieren soll. Ein Gummitrichter und eine Holzscheibe übernehmen die Aufgaben von Mund und Nase und ein Rohrblatt aus Elfenbein die der Stimmbänder. Da Mechanismen für Zunge und Zähne auch nach gescheiterten Versuchen fehlten, können mit dieser Apparatur nur bedingt Sprachlaute erzeugt werden (Brackhane, 2011).
In den 1930er Jahren entwickelte Homer Dudley in den Bell Laboratories in den USA den Voder, dessen Name sich aus *Voice operated demonstrator* zusammensetzt und einen wichtigen Meilenstein in der Geschichte der Sprachsynthese ausmacht (Gold, 1990). Das Gerät, das Dudley erstmals 1939 auf der Weltausstellung in New York City präsentierte, basiert darauf, dass man zuerst zwischen einem Summton (*Buzz*) und einem Rauschsignal (*Hiss*) wählt (Campbell & Dean, 1993). Mit dem Summton können Vokale und Nasallaute erzeugt werden, mit dem Rauschsignal sind Frikative und Affrikaten möglich. Diese Klänge werden anschließend in einen von zehn Bandpassfiltern geleitet, den man über eine spezielle Klaviatur auswählt. Schlussendlich wird das Signal auf einen Lautsprecher übertragen und somit hörbar. Das Prinzip basiert auf Dudleys vorherige Arbeit am Vocoder, auf dessen Funktionsweise im Kapitel 6.1 genauer eingegangen wird (Gold, 1990).

5.2. Die Entwicklung des Vocoders

Der Channel Vocoder leitet sich ab von *Voice Coder* wurde ebenfalls von Homer Dudley entwickelt. Es war das erste System, das eine externe Spracheingabe analysieren und verarbeiten konnte. Daher handelt es sich ebenfalls um eine Form der Sprachsynthese. Das

Prinzip ähnelt dem des Voders, mit dem Unterschied, dass es eine externe Eingabe anstelle der beiden integrierten Oszillatoren gibt (Gold, 1990).

5.2.1. Anwendungen abseits der Musik

Der Vocoder hat seinen Ursprung in der Telekommunikation. Homer Dudleys Ziel war es, eine Maschine zu bauen, die es durch die Reduktion der Frequenzbandbreite schaffte, Telefongespräche auf eine größere Entfernung zu übertragen. Das Eingabesignal – also die Stimme – wurde dabei in mehrere Frequenzspektren aufgeteilt und nur die Frequenzbänder, die es für Sprachverständlichkeit brauchte, wurden übertragen. Im Gegensatz zu den Übertragungen mit voller Bandbreite, war mit dieser Methode weniger Kupferkabel notwendig, da weniger Daten übertragen werden mussten.

Das Militär entdeckte diese Technologie als Methode für verschlüsselte Unterhaltungen auf eine längere Entfernung. So wurde das Prinzip im Zweiten Weltkrieg zwischen Winston Churchill und Franklin Delano Roosevelt verwendet, um Konversationen zwischen Großbritannien und den USA führen und geheim halten zu können (Izotope, 2018). Das Kommunikationssystem zwischen den Alliierten nannte sich SIGSALY und löste das zuvor vorhandene System *A-3 Screambler* 1943 ab, das von deutschen Behörden abgehört werden konnte. Das über 50 Tonnen schwere Gerät wurde von A. B. Clark und Alan Turing ebenfalls in den Bell Laboratories konzipiert und an zwölf Orten weltweit installiert, wodurch über 3000 Gespräche während dem Krieg sicher geführt werden konnten (Weadon, o.D.). Das Gerät, das einen ganzen Raum gefüllt hat, ist die größte bekannte Maschine, die zur Verschlüsselung verwendet wurde und hatte eine Übertragungsrate von etwa 1700 Bit pro Sekunde. 1965 folgten weitere, kompaktere Geräte, wie die KY-3, die nur noch die Größe eines Schrankes hatte. Weitere Entwicklungen hatten nur noch die Größe eines gewöhnlichen Telefons (Schmeh, 2022).

5.2.2. Einzug in die Popmusik

Das erste Musikstück, das den Vocoder verwendete, stammt von den Bell Laboratories selbst und ist eine hörbare Demonstration dieser Technologie am Beispiel des irischen Volkslieds *Love's Old Sweet Song*. 1968 kam es zur ersten kommerziellen Verwendung des Vocoders: Bruce Haack baute einen eigenen Vocoder und verwendete ihn für sein Album *The Electronic Record for Children*. Inspiriert von Dudleys Erfindung baute Bob Moog 1971 ebenfalls einen eigenen Vocoder, der erstmals im Soundtrack von Wendy Carlos zum Film *Clockwork Orange* zu hören war und dadurch ein größeres Publikum erreichte (Gale, 2018).

Abbildung 1: Vocoder von Bob Moog. Quelle: Izotope (2018).

1974 hat die deutsche Band Kraftwerk mit der Verwendung des Vocoders im Album *Autobahn* dieses Tool in den Mainstream gebracht. Durch dessen Erfolg breitete sich diese Technologie in unterschiedlichen Musikrichtungen aus: Herbie Hancock brachte sie mit seinem Song *I Thought It Was You* von 1978 und *Better Bet Your Love* von 1979 auch in den Jazz-Funk-Bereich (Gale, 2018). 1981 zeigte Laurie Anderson wie weit man mit dem Vocoder gehen konnte: Ihr Song *O Superman* besteht, bis auf wenige Elemente, ausschließlich aus ihrer Stimme mit dem Vocoder-Effekt (Schwarz, 2019). In die Hip-Hop-Bewegung gelang der Vocoder durch Afrika Bambaataa und The Soulsonic Force und deren 1982 veröffentlichten Song *Planet Rock*. Der Erfolg des Songs sorgte dafür, dass sich immer mehr Künstler*innen im Hip-Hop mit Stimmbearbeitung und Vocoding auseinandersetzten (Gale, 2018). Ab diesem Zeitpunkt war die Verbreitung des Vocoders in der Popmusik nicht mehr aufzuhalten. Daft Punk beispielsweise sorgte dafür, dass der Vocoder in die Disco Musik Einzug hielt: Sie nutzten für ihre Stimmbearbeitung überwiegend den Variphrase VP-9000, den sie von Roland bereits vor der offiziellen Veröffentlichung des Geräts erhalten haben (Eilers, 2024). Einer ihrer bekanntesten Musikstücke ist *Harder, Better, Faster, Stronger*. Es handelt sich um einen Electro House Song, der 2001 auf dem Album *Discovery* erschienen ist und sich durch Gesang auszeichnet, der mit einem Vocoder bearbeitet wurde. Das Instrumental basiert auf einem Sample aus dem 1979 veröffentlichten Song *Cola Bottle Baby* von Edwin Birdsong, allerdings

in einer beschleunigten Variante. Der Song thematisiert den schnellen technologischen Fortschritt und die einhergehende Produktivität, wodurch die Brücke zu synthetisch klingenden Stimmen geschlossen wird. In der zweiten Hälfte des Songs wird die Stimme als eine Art Gitarrensolo verwendet, sowie in einzelne Abschnitte aufgeteilt und rhythmisch neu zusammengesetzt, wie es bei Vocal Chops der Fall ist. Der durch den Vocoder verfremdete Gesang zeigt, wie die Grenzen zwischen Stimme und Instrument verschwimmen, vorrangig bei sehr hohen und sehr tiefen Noten, was einen wichtigen klangästhetischen Aspekt des Musikstücks darstellt. Kanye West verwendete einen Ausschnitt aus *Harder, Better, Faster, Stronger* als Sample für seinen 2007 erschienenen Song *Stronger*, der zu einem seiner meistgestreamten Musikstücke wurde (Smith, 2021).
Seither hat sich der Vocoder auch in Form von Software weiterentwickelt und ist somit für eine breitere Masse an Musikproduzent*innen verfügbar (Eilers, 2024). Ein modernes Beispiel, in dem der Vocoder zu einem essenziellen Element eines Songs wurde, ist *The Middle* von Zedd, Maren Morris und Grey, der 2018 erschienen ist. Bei diesem Song wird zwischen Abschnitten mit Vocoder im Hintergrund und Gesang ganz ohne Vocoder gewechselt. Auch Halsey und Cashmere Cat verwenden in ihrem Song *Hopeless* den Vocoder. In diesem Musikstück gibt es allerdings auch Passagen, in denen ausschließlich der Gesang mit dem Vocoder-Effekt zu hören ist. *Gateway Car* von Taylor Swift wird mit dem Vocoder eröffnet, genauso wie *Migraine* von Twenty One Pilots. Daraus lässt sich schließen, dass dieser Effekt auch in der zeitgenössischen Popmusik nach wie vor eingesetzt wird.

5.3. Die Entstehung von Auto-Tune

In den 80er Jahren fand der Wechsel von analoger zu digitaler Studiotechnik statt, der eine erhebliche Veränderung in der Art und Weise darstellte, wie Musik produziert wurde. Diese technische Revolution hat auch den Vocoder weiterentwickelt. Daraus entstand Auto-Tune – ein Tool, das die einzelnen Töne einer Gesangslinie so korrigieren kann, dass dieser durchgängig in der gewählten Tonlage bleibt (Schwarz, 2019). Der Erfinder von Auto-Tune und Gründer von Antares Audio Technologies, Harold Andy Hildebrand, war ursprünglich Ölingenieur und hat auf der Suche nach Öl mit Autokorrelation gearbeitet. Diese Methode, die er später für Auto-Tune angewendet hat, ermöglichte es, die Tonhöhe eines Gesangs erkennen und anschließend korrigieren zu können (Crockett, 2016). In der Ölindustrie verwendete Hildebrand Autokorrelation, um Ölvorkommen lokalisieren zu können. Dabei handelt es sich um eine Technologie von Texas Instruments und des Massachusetts Institute of Technology. Reflektionen von seismischen Wellen zu unterschiedlichen Zeitpunkten werden dabei analysiert, um Muster erkennen und in weiterer Folge Öl finden zu können. In der Audiotechnologie hat Hildebrand diese Idee dazu verwendet, um Muster bzw. Periodizitäten in einem Audiosignal zu erkennen, was schlussendlich Erkenntnisse hinsichtlich Tonhöhe

schafft (Marshall, 2024). Die Funktionsweise von Auto-Tune wird im Kapitel 6.2 genauer beschrieben.

Vor der Erfindung von Auto-Tune war die Korrektur einzelner Töne eines Gesangs äußerst zeitaufwendig: Zuerst haben Sänger*innen die einzelnen Passagen eines Musikstücks mehrfach aufgenommen. Anschließend sorgten Produzent*innen dafür, die am wenigsten verstimmten Teile jeder Aufnahme herauszuschneiden und zu einer stimmigen Gesangsspur zusammenzusetzen. Diesem eher mühsamen und unpräzisen Verfahren nahm sich Harold Andy Hildebrand an und arbeitete an einer Lösung: Die Vision von Hildebrand, der Auto-Tune 1997 veröffentlichte, war es, ein Werkzeug für die präzise Korrektur von Gesang zu schaffen, die einfach und schnell anwendbar ist. Die ersten drei Jahre sollte dieses neue Werkzeug vor der Öffentlichkeit geheim gehalten werden, weil das Publikum kritisch gegenüber „gekünstelter" Musik stand. Der Fall Milli Vanilli, bei dem 1989 bekannt wurde, dass die beiden Künstler lediglich zu den Gesangsspuren anderer Playback aufführten, trug wesentlich zur verstärkten kritischen Auseinandersetzung mit Authentizität in der Musikproduktion bei. Aus diesem Grund wurde Auto-Tune anfangs sehr subtil eingesetzt, um einen natürlichen Klang beizubehalten. Neben der subtilen Anwendung eröffnete das Tool jedoch auch neue kreative Möglichkeiten: Cher brachte 1998 mit ihrem Song *Believe* einen neuartigen Klang in die Charts, der durch den gezielten überhöhten Einsatz von Auto-Tune geprägt war und die Musikindustrie verändert hat. Es entstand der Begriff des sogenannten *Cher-Effekts*, der eine unnatürlich starke Tonhöhenkorrektur in der Stimme bezeichnet (Crockett, 2016). Mark Taylor und Brian Rawling, die Musikproduzenten von Chers *Believe*, verwendeten dazu die schnellstmögliche *Retune-Geschwindigkeit* und hielten es anfänglich geheim, auf welchem Weg sie diesen Stimmklang erzeugt haben. Auf der einen Seite wollten die beiden diese neue Technologie für sich behalten, auf der anderen Seite bestand die Angst, dass das Image von Cher darunter leidet, was auch die Plattenfirma von Cher befürchtete. Aus diesem Grund kommunizierte man der Öffentlichkeit, es hätte sich um einen bestimmten Vocoder-Effekt gehandelt, was durch das Aufkommen von erfolgreichen Produktionen, die Auto-Tune verwendeten, aufgeflogen ist (Schwarzer, 2023). Besonders in den Versen ist dieser Effekt hörbar: in *I can't break through* und *It's so sad that you're leaving* (Cher, 1998) sowie im dritten Refrain. Cher wechselt also zwischen einem natürlichen und einem unnatürlichen, maschinell transformierten Klang. Bei der Stimme in den Versen wird darüber hinaus ein Equalizer mit einem High- und Lowcut verwendet, der zu einem dezenten Radioeffekt führt.

Anfang der 2000er Jahre fand Auto-Tune besonders im Rap eine neue Heimat und entwickelte sich dort zu einem zentralen Element der Stimmbearbeitung. 2005 wendete der Rapper T-Pain in seinem Song *I'm Sprung* Auto-Tune zum ersten Mal auf Rap-Gesang an. Die Meinungen über Auto-Tune waren allerdings gespalten und es gab Gegenwind seitens einiger Musiker*innen. 2009 beispielsweise ging die Band Death Cab for Cutie bei der 51. Grammy-

Verleihung mit blauen Schleifen auf die Bühne, um so gegen diese neue Technik zu protestieren. Dennoch verwendeten Rapper*innen und Sänger*innen wie Kanye West und Jennifer Lopez Auto-Tune in ihren Musikstücken, weshalb es heutzutage nur mehr schwer aus der Stimmbearbeitung in der Musik wegzudenken ist und in den meisten Musikproduktionen, wenn auch oftmals nur subtil, verwendet wird (Schwarzer, 2023). Ein zeitgenössisches Beispiel, in dem Auto-Tune hörbar eingesetzt wird, ist *WHEN I WAS OLDER* von Billie Eilish. In diesem Song wird neben Auto-Tune allerdings auch der Vocoder eingesetzt. Im Rap verwendet beispielsweise Juice WRLD Auto-Tune in einem starken Ausmaß, unter anderem in seinem Song *Big*.

5.4. Die Geschichte der Talkbox

Bei der Talkbox handelt es sich um einen Effekt, bei dem eine beliebige Klangquelle, beispielsweise eine Gitarre oder ein Synthesizer, mithilfe des Mundes moduliert wird. Im Kapitel 6.3 wird die Funktionsweise dieses Effekts genauer beschrieben. Die erste Version eines derartigen Systems entstand 1939 von Alvino Rey, der die erste E-Gitarre und den ersten Tonabnehmer von Gibson miterfunden hatte. Er schuf eine Puppe in Form einer Gitarre, die er *Stringy, the Talking Guitar* nannte, und mit der er auf Bühnen aufgetreten ist. Rey spielte Gitarre, während seine Frau Luise King den Gitarrenklang mithilfe eines Kehlkopfmikrofons modulierte. Durch Bewegungen des Mundes der Puppe wurde *Stringy, the Talking Guitar* eine Stimme verliehen. Noch im selben Jahr erfand Gilbert Wright das *Sonovox*, der dasselbe Prinzip anwendete, jedoch mit einem Wandler anstelle des Kehlkopfmikrofons. Das Gerät und dessen Klang wurde häufig für Jingles im Radio verwendet und fand auch in der Filmmusik seinen Einzug. 1964 baute Pete Drake eine eigene Version: Der Klang seiner Steel-Gitarre wurde an einen Lautsprecher geschickt, an dem ein Trichter befestigt war. Der Trichter war an eine Röhre angeschlossen, die den Klang der Gitarre direkt in den Mund von Pete Drake leitete. Ähnlich wie bei Rey und Wright wurde auch hier der Mund zum Modulieren des Klangs genutzt. Massentauglich wurde dieser Effekt allerdings erst durch Doug Forbes, der die Idee von Drake in eine kleinere Version komprimierte. Er verwendete einen kleinen 30 Watt Verstärker und einen Plastikschlauch, gab alles in eine Umhängetasche und nannte das Produkt *The Bag*. 1973 baute Bob Heil diese Idee aus und schuf eine Talkbox, die keinen Verstärker benötigte und nannte diese *Heil Talk Box*. Durch Peter Frampton's Song *Do You Feel Like We Do* aus 1973 gewann die *Heil Talk Box* an Bekanntheit und prägte den Klang in den 1970er Jahren mit. *Do you Feel Like We Do* war Teil des Live Albums *Frampton Comes Alive!*, welches laut Rolling Stone (2015) eines der meist verkauften Live Alben aller Zeiten ist. Der Effekt wird in diesem Musikstück nur für eine Passage verwendet, die allerdings etwa vier Minuten des 14-minütigen Songs ausmacht und in Form eines Solos vorkommt. Frampton verwendet eine E-Gitarre als Eingangssignal, das er anschließend mit seinem Mund moduliert.

Er wechselt zwischen rein melodischen Passagen und wörtlichen Phrasen, wie *Do You Feel Like We Do*, wodurch eine Fusion aus Stimme und Instrument entsteht und sein Gitarrensolo um eine gesanglich-melodische Ebene erweitert wird (Mandler, 2023). Nachdem es sich bei der Talkbox, anders als beim Vocoder und Auto-Tune, um ein elektroakustisches System und nicht ausschließlich einer elektronischen Signalverarbeitung handelt, entsteht ein organischer Klang mit menschlichen Eigenschaften (Steinwachs, 2021).
Seither hat sich diese Technik stetig weiterentwickelt und wurde in bekannten Musikstücken, wie *Livin' on a Prayer* aus 1986 und *It's My Life* aus 2000 von Bon Jovi verwendet. Auch in der zeitgenössischen Popmusik wird dieser Effekt immer wieder angewendet: Byron Chambers spielte die Talkbox im Intro und am Ende des 2016 erschienenen Songs *24K Magic* von Bruno Mars (Mandler, 2023).

6. Techniken der Stimmbearbeitung und deren Funktionsweisen

Digitale Technologien sorgen dafür, dass die menschliche Stimme nicht mehr ausschließlich an einen Menschen gebunden ist, sondern manipuliert oder synthetisiert werden kann und Musik somit in einem hybriden Prozess entsteht. Mensch und Maschine gehen immer mehr ineinander über und die Grenzen verschwimmen. David Bowie spielte bereits in den 1970er Jahren bei seinem Song *Oh! You Pretty Things* mit der Idee, dass sich der Mensch in ein übermenschliches Wesen verwandelt, den *Homo Superior*. Grimes beispielsweise spricht in ihrem Song *We Appreciate Power* eine Herrschaft von künstlicher Intelligenz an und verwendet selbst die Ästhetik der starken Stimmbearbeitung, um so das Thema auch klanglich zu repräsentieren. Es handelt sich hierbei nur um zwei von vielen Beispielen, die zeigen, wie eng die Popmusik mit den Konzepten des Posthumanen arbeitet. Dies ist womöglich einer der Gründe, weshalb roboterhafte Stimmen in diesem Metier des Öfteren aufgegriffen wurden und werden (Berlich, 2022). Brian Eno argumentiert, dass man ein erfolgreiches Musikstück innerhalb weniger Millisekunden erkennen würde, allerdings nicht durch die Melodie oder den Akkordaufbau, sondern vielmehr durch den Klangcharakter. Er verdeutlicht damit, dass die Klangästhetik eine bedeutende Rolle in der Popmusik spielt, wodurch auch die Verwendung von synthetisch klingenden Stimmen zu erklären wäre (Wicke, 2008). Üblicherweise wurde die Musik als Ausdruck von einem Subjekt verstanden (Diederichsen, 2016). In der Popmusik verschwindet allerdings oftmals das Subjekt, beispielsweise durch digitale Manipulationen, die neue Möglichkeiten der Selbstinszenierung schaffen (Berlich, 2022). Diederichsen (2016) kritisiert diesen Gedanken allerdings, denn die Eliminierung des Subjekts führe nicht zu einer Entmenschlichung, sondern vielmehr zu einer Kontrolle der Klänge. Selbst wenn versucht wird, eine künstliche Manipulation vorzunehmen, so stehe noch immer der Mensch mit seinen Konzepten und Entscheidungen hinter diesen klanglichen Manipulationen, was wiederrum zu einer Übermenschlichung und gesteigerten Subjektivierung führe (S. 101).

In der Popmusik werden unterschiedliche Effekte für die Bearbeitung des Gesangs verwendet. Die Intention einiger klassischer Effekte ist es nicht, die Stimme zu verfremden, sondern zu verbessern. Grundsätzlich lässt sich allerdings mit der übertriebenen Verwendung von klassischen Effekten in vielen Fällen ein unnatürlicher Klang erzeugen. Übliche Effekte sind beispielsweise Hall, Delay, Equalizer, Chorus, Phaser, Flanger, Verzerrung und Pitch Shifter. Jeder dieser Effekte wirkt sich auf das Audiosignal in unterschiedlicher Weise aus, sei es durch die zeitliche Verschiebung desselben Signals oder die Manipulation von Frequenzen (Maempel, 2021). All diese Effekte werden im Glossar genauer beschrieben. Um die Funktionsweisen vom Vocoder, von Auto-Tune und der Talkbox zu verstehen, wird im Folgenden auf diese Effekte eingegangen.

6.1. Vocoder

Ein Vocoder schafft es mithilfe bestimmter Algorithmen die menschliche Stimme oder auch andere Audiosignale zu analysieren und anschließend wieder zu synthetisieren. Er besteht aus zwei Eingängen und einem Ausgang: Die Eingänge setzen sich aus einem Modulator und einem Träger zusammen. Als Modulator kann jedes Audiosignal dienen, typischerweise eine menschliche Stimme oder ein rhythmisches Signal. Der Träger ist zumeist ein harmonisch reiches Signal, wie ein Synthesizer-Klang, der die Tonhöhe und Klangfarbe bestimmt. Die Tonhöhe kann zumeist über eine Klaviatur oder über MIDI bestimmt werden.
Zuerst wird das Modulatorsignal in Filterbänken in Hinsicht auf seine Frequenzeigenschaften und Dynamik analysiert, durch verschiedene Bandpassfilter geleitet und anschließend in mehrere Frequenzbänder aufgeteilt. Die jeweiligen Amplituden der einzelnen Frequenzbänder werden mithilfe eines Hüllkurven-Verfolgers analysiert und auf das Trägersignal übertragen. Ist die Amplitude bei dem Modulatorsignal beispielsweise bei 100 Hz sehr niedrig, aber bei 200 Hz sehr hoch, so wird die Lautstärke bei dem Trägersignal bei 100 Hz abgeschwächt und bei 200 Hz verstärkt. Nachdem das für alle Frequenzbänder passiert, sorgt somit eine größere Anzahl von Frequenzbändern für eine detaillierte Übertragung der Klangcharakteristika des Modulators auf den Träger und somit zu einer höheren klanglichen Ähnlichkeit zum Modulatorsignal. Das modulierte Signal wird schließlich durch einen DCA (*Digitally Controlled Amplifier*) bei digitalen Geräten, oder einem VCA (*Voltage Controlled Amplifier*) bei analogen Geräten verstärkt und anschließend bearbeitet oder über einen Lautsprecher hörbar gemacht (Caudet, 2023).

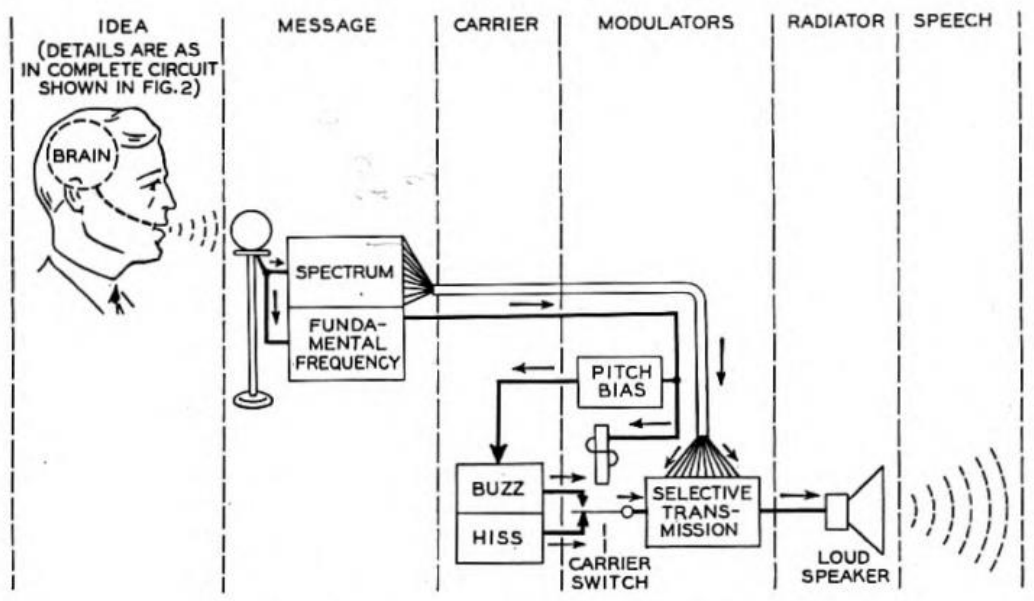

Abbildung 2: Schematische Darstellung des Vocoders. Quelle: Dudley (1940, S. 508).

Homer Dudley, der Erfinder des Vocoders, hat diese Idee 1940 in The Bell System Technical Journal verschriftlicht. Zur besseren Verständlichkeit hat er eine Skizze des grundlegenden Signalflusses eines Vocoders angefertigt, wie er von ihm entwickelt wurde (siehe Abbildung 2). Der Prozess kann in sechs Abschnitte unterteilt werden: Am Beginn steht die Idee (*Idea*), also das, was gesagt oder gesungen werden möchte. Diese Idee wird anschließend über das Sprechen in ein hörbares, akustisches Audiosignal umgewandelt. Diese Sprachnachricht (*Message*), die von einem Menschen erzeugt wurde, wird von einem Mikrofon aufgenommen und in ein Spektrum (*Spectrum*) und eine Grundfrequenz (*Fundamental Frequency*) aufgeteilt. Das Spektrum beschreibt dabei die frequenzabhängige Energieverteilung der Sprache, also wie stark gewisse Frequenzen vorhanden sind. Die Grundfrequenz beschreibt die tiefste Frequenz, die für die Tonhöhe verantwortlich ist. Die gewonnenen Daten bezüglich der spektralen Zusammensetzung des Eingangssignals werden nun auf ein Trägersignal (*Carrier*) übertragen. Dabei wird zwischen einem Summton (*Buzz*), also einem periodischen Signal für stimmhafte Laute, wie Vokale, und einem Rauschsignal (*Hiss*), also einem nicht-periodischen Signal für stimmlose Klänge, wie Frikative und Affrikaten, unterschieden. Durch einen Trägerschalter (*Carrier Switch*) wird anschließend entschieden, ob der Summton oder das Rauschsignal verwendet wird, abhängig von den analysierten Sprechmerkmalen. Die Tonhöhensteuerung (*Pitch Bias*) verändert die Grundfrequenz des Trägersignals. Die spektralen Informationen des Modulators, also des Eingangssignals, werden dann durch eine selektive Übertragung (*Selective Transmission*) mit mehreren Filtern auf das Trägersignal übertragen. Anschließend werden die einzeln modulierten Frequenzbänder zusammengeführt, im Radiator an den Ausgangsverstärker weitergeleitet und über einen

Lautsprecher (*Loudspeaker*) ausgegeben. Zu hören ist die über den Vocoder veränderte Sprache (*Speech*) (Dudley, 1940).

In der zeitgenössischen Musikproduktion wird als Trägersignal zumeist eine Sägezahnschwingung verwendet, da diese viele harmonische Schwingungen aufweist. Verwendet man die Stimme als Modulatorsignal, ist es ratsam, die einzelnen Wörter des Gesangs stark und möglichst klar zu betonen, um so nach dem Vocoder-Algorithmus noch möglichst viel Sprachverständlichkeit beizubehalten. Die korrekte Tonhöhe beim Singen spielt dabei eine untergeordnete Rolle, da diese ohnehin von dem Vocoder bzw. dem Träger übernommen wird.
Die häufigsten und essenziellsten Parameter, die sich bei einem modernen Software-Vocoder einstellen lassen, sind *Bands*, *Frequency Range*, *Formant* und *Unvoiced*, neben den klassischen Parametern wie *Attack*, *Release* und *Dry/Wet*. Bei der *Bands*-Einstellung lässt sich die Anzahl der Frequenzbänder einstellen. Eine niedrige Anzahl sorgt für eine geringere Auflösung, was aber auch für einen „altmodischeren“ Klang sorgt. Mit der *Frequency Range* lässt sich das Frequenzspektrum einstellen, das auf den Träger übertragen werden kann. Der Parameter *Formant* oder manchmal auch *Shift* bezeichnet, erhöht oder verringert die Frequenzen der Filter, was den Klang höher oder tiefer macht, ähnlich einer weiblichen oder männlichen Stimme. Mit *Unvoiced* lassen sich auch Plosive, Frikative und Affrikaten nachbilden, indem diese nicht vom Träger, sondern einem zusätzlichen Rauschgenerator erkannt und erzeugt werden (Rewak, 2022).

6.2. Auto-Tune

Auto-Tune ist ein Effekt, der die Frequenz eines Audiosignals messen und an die nächstgelegene Tonhöhe einer bestimmten Tonleiter anpassen kann. Diese Korrektur geschieht in Echtzeit, wodurch der Einsatz im Tonstudio, sowie im Livebetrieb möglich ist. Harold Andy Hildebrand (1999) hat den Algorithmus hinter diesem Effekt in seinem Patent US 5,973,252 niedergeschrieben und erklärt wie Auto-Tune in einer analogen Form aufgebaut ist: Ein Line-Level-Eingang oder ein Eingang von einem Mikrofon wird zuerst über einen Analog-Digital-Wandler in ein digitales Format umgewandelt, laut Hildebrand mit einer Samplerate von 44,1 kHz und einer Bittiefe von 16-Bit. Anschließend wird die aktuelle Tonhöhe des Signals in einem Mikrochip mithilfe von Autokorrelation bestimmt, die Hildebrand bereits bei seiner Arbeit als Ölingenieur angewendet hat. Jedes Audiosignal ist eine Schwingung, die sich in einer gewissen Periode wiederholt. Mit Autokorrelation lässt sich diese Periode bestimmen, wodurch sich schließlich die Frequenz und somit auch die Tonhöhe bestimmen lassen. Die aktuelle Periode bzw. Tonhöhe des Eingangssignals wird danach mit einer vorgegebenen Skala oder einem MIDI-Eingang verglichen und bei einer Abweichung korrigiert. Diese Korrektur erfolgt

durch Resampling. Dabei handelt es sich um eine Technik, die das Signal beschleunigt, um so eine höhere Frequenz und somit einen höheren Ton zu erzielen, oder verlangsamt, um die Frequenz und somit die Tonhöhe zu senken. Die Geschwindigkeit des Übergangs zwischen den einzelnen Tönen lässt sich dabei einstellen, was sich auf die Natürlichkeit des Klangs auswirkt. Auto-Tune bietet auch die Möglichkeit eines künstlichen Vibratos, also das gleichmäßige Verändern der Frequenz nach oben und unten um ein Tonzentrum. Schlussendlich wird das mit Auto-Tune korrigierte digitale Signal mit einem Digital-Analog-Wandler wieder
in ein analoges Signal umgewandelt und kann anschließend bearbeitet oder über ein Audiointerface ausgegeben werden (Hildebrand, 1999).

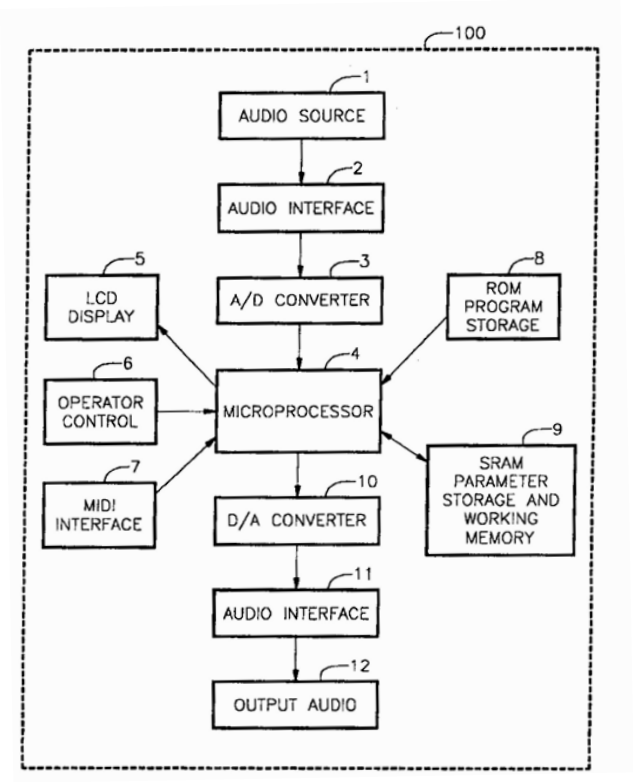

Abbildung 3: Blockdiagramm der Tonhöhenbestimmung von Auto-Tune. Quelle: Hildebrand (1999, S. 1).

Anhand des Blockdiagramms (siehe Abbildung 3) in seinem Patent veranschaulicht Hildebrand den groben Aufbau von Auto-Tune. Das Audiosignal (*Audio Source*) wird über ein Audiointerface in das System eingespeist. Dort wird das analoge Audiosignal mit einem Analog-Digital-Wandler (*A/D Converter*) in ein digitales Signal umgewandelt und im Mikroprozessor (*Microprocessor*) weiterverarbeitet. Im Mikroprozessor findet die Tonhöhenerkennung mittels Autokorrelation und die Tonhöhenkorrektur durch Resampling statt. Um die Statusinformationen sehen zu können, ist ein LCD Display vorhanden. Über ein Bedienfeld (*Operator Control*) können Parameter wie die Geschwindigkeit des Übergangs zwischen zwei Tönen eingestellt werden. Wenn nicht eine der vorgegebenen Tonleitern verwendet werden soll, besteht außerdem die Möglichkeit der Eingabe der Tonhöhen mittels MIDI (*MIDI Interface*). Firmware und Algorithmen befinden sich im *ROM Programm Storage*, vorübergehend gespeicherte Parameter werden in *SRAM Parameter Storage and Working Memory* zwischengespeichert. Über den Digital-Analog-Wandler (*D/A Converter*) wird das digitale Signal wieder in ein analoges Signal umgewandelt, wodurch es anschließend wieder durch ein Audiointerface beispielsweise über einen Lautsprecher ausgegeben werden kann (Hildebrand, 1999).

Wenngleich es sich bei der Erklärung von Hildebrand um eine analoge und von externer Technik unabhängige Version handelt, war die erste Version von Auto-Tune digital. Die Software funktioniert grundsätzlich allerdings ähnlich: Jeder Ton des Eingangssignals wird in Echtzeit hinsichtlich seiner Grundfrequenz analysiert. Dies geschieht mithilfe der Fast-Fourier-Transformation, die eine Schallwelle in ihre einzelnen Sinuswellen und somit Frequenzen aufteilt. Nachdem die Grundfrequenz erkannt wurde, wird der jeweilige Ton in die Richtung der nächstgelegenen Note von der eingestellten Tonleiter nach oben oder unter verschoben. Der Rhythmus und das Timing ändern sich dabei nicht. Einstellbare Parameter sind *Retune Speed*, *Humanisieren*, *Eingangstyp* und *Benutzerdefinierte Skalen*. *Retune Speed* steht dabei für die Geschwindigkeit der Tonhöhenkorrektur. Eine schnelle Geschwindigkeit sorgt für den *Cher-Effekt*, also einen roboterhaften Klang. Diese Einstellung ist vor allem in elektronischer Musik, Pop und im Hip-Hop zu hören. Eine subtile Korrektur erfolgt durch eine niedrige Retune-Geschwindigkeit, da natürliche Schwankungen beibehalten werden und wird insbesondere in Rock, Soul und R&B Musik angewendet. Mit *Humanisieren* klingen gehaltene Töne natürlicher, indem auf diese eine geringere Retune-Geschwindigkeit angewendet wird. Bei dem *Eingabetyp* wählt man die Tonlage, die voraussichtlich verwendet wird, damit eine präzisere Tonhöhenkorrektur vorgenommen werden kann. Um die Zieltöne der Korrektur festzulegen, kann über die *benutzerdefinierten Skalen* eine vorgegebene Skala ausgewählt oder eine eigene Skala definiert werden (Antares, 2023).
Als Gegenstück zur automatischen Korrektur von Tonhöhen, gibt es auch die Möglichkeit einer händischen Berichtigung mit einem Plugin namens *Melodyne*. Dabei handelt es sich um ein Werkzeug von Celemony Software GmbH, mit dem eine detaillierte Manipulation der einzelnen Noten innerhalb eines Audiosignals möglich ist. Das Audiosignal wird hinsichtlich Tonhöhe, Timing, Lautstärke und Formanten analysiert, was auch bei polyphonen, also mehrstimmigen Signalen, möglich ist. Diese Parameter sind anschließend bearbeitbar und ein Vibrato kann auf einzelne Töne hinzugefügt werden. Durch diese detaillierte und manuelle Bearbeitung kann so einerseits ein natürlicher Klang für die Stimme oder aber auch andere Instrumente beibehalten werden. Andererseits bietet dieses Werkzeug auch die Möglichkeit neue Klänge ausgehend vom Eingangssignal zu kreieren (Celemony, o.D.).

6.3. Talkbox

Um die Aufgabe einer Talkbox besser verstehen zu können, ist es hilfreich, kurz die Entstehung des menschlichen Stimmklangs zu betrachten: Die Luft strömt aus der Lunge über die Luftröhre in den Kehlkopf. Dort versetzt die Luft die elastischen Stimmlippen in Schwingungen, wodurch ein Ton entsteht. Für die Stimmfarbe sorgen die Resonanzräume, wie Luftröhre und Lungenraum, Stirn-, Nasen-, und Mundhöhle, sowie der Schädel. Die Mundhöhle mit der Zungenstellung spielt dabei die wichtigste Rolle bei der Bildung von

Vokalen (Michels, 2008, S. 23). Überträgt man diesen Mechanismus auf Instrumente, so wird beispielsweise bei einer Gitarre die Aufgabe der menschlichen Lunge übersprungen. Töne entstehen durch das Zupfen der Saiten und dessen Vibrationen. Nachdem eine Gitarre keinen Mundraum besitzt, ist die Bildung von Vokalen nur bedingt möglich. Eine Talkbox hilft dann, diese Töne in den Mund einer Person zu befördern. Es handelt sich dabei also um ein Effektgerät, mit dem typischerweise der Klang einer Gitarre oder eines Synthesizers über den menschlichen Mundraum moduliert werden kann (Mandler, 2023).
An die Heil Talkbox, die von Peter Frampton bekannt gemacht wurde, wird als Eingang üblicherweise ein Gitarrensignal verwendet, das in der Talkbox verstärkt und über einen Hornkompressionstreiber, also einen Lautsprecher, ausgegeben wird. Das vom Hornkompressionstreiber ausgegebene Signal wird anschließend durch ein langes Rohr geleitet, das in den Mund der*des Künstlerin*Künstlers führt. Dort kann anschließend die*der Sänger*in durch das Singen oder Sprechen den Gitarrenklang so verändern, dass der Gitarre metaphorisch eine Stimme verliehen wird, da sich mit unterschiedlichen Stellungen von Zunge und Lippen der Resonanzraum verändert. Über ein zusätzliches Gesangsmikrofon am Mund des singenden oder sprechenden Menschen kann das Resultat aufgenommen und über einen Lautsprecher wiedergeben werden. Der Unterschied zum Vocoder liegt also darin, dass bei einer Talkbox ein akustisches Audiosignal in den Mund geleitet und somit ein elektroakustisches System angewendet wird. Bei einem Vocoder wiederrum wird die über ein Mikrofon aufgenommene Sprache ausschließlich elektronisch verarbeitet (Steinwachs, 2021). „Ein Vocoder lässt die Stimme wie ein Instrument klingen und eine Talkbox lässt ein Instrument wie eine menschliche Stimme klingen“ (Mandler, 2023, Abschnitt Talkbox vs. Vocoder).

7. Ausblick auf zukünftige Entwicklungen

„Die Stimme ist das zentrale Element, das wird sich nicht so schnell verändern – sondern nur ihr Klang“ (Schwarz, 2019, Abschnitt Harmonien im Prisma). Es ist davon auszugehen, dass neue Effekte entstehen, die den Klang einer Stimme verändern können. Bereits bestehende Effekte, wie Auto-Tune und Vocoder werden sich vor allem in der digitalen Domäne stetig weiterentwickeln und durch erweiterte Algorithmen immer präziser werden, was gleichzeitig neue Möglichkeiten für Sounddesign bedeutet. Ebenfalls wird es zu einer Kombination aus mehreren Effekten kommen, wie etwa einem Vocoder mit Auto-Tune-Funktionen. Die Klangcharakteristik, beispielsweise von Vocodern aus den 1970er und 1980er Jahren wird für einen Vintage-Klang erhalten bleiben, allerdings mit neuen Einstellungsebenen, wie die Anpassung von Formanten und Filterbänken (Jones, 2019). Besonders in Hinblick auf künstliche Intelligenz kann sich die Art und Weise wie Musik produziert und bearbeitet wird sehr verändern. Auto-Tune könnte mit KI-Technologien noch präziser werden, vor allem bei komplexen Audiosignalen (Miranda, 2024). Insbesondere in dem Bereich der Sprachsynthese

kommt es durch künstliche Intelligenz zu neuen Möglichkeiten: Mithilfe von Werkzeugen wie ACE Studio ist es möglich, Stimmen per MIDI-Eingabe oder mit Songtexten vollständig generieren zu lassen (ACE Studio, o.D.). Über Suno AI beispielsweise lassen sich ganze Musikstücke inklusive des Gesangs durch Texteingaben und Beschreibungen des gewünschten Musikstils und der gewünschten Eigenschaften generieren. Je nach Werk können so realistische, aber auch unnatürliche Stimmen entstehen (King, 2023). Daraus lässt sich schließen, dass es einige neue Werkzeuge zur Stimmsynthese in der Zukunft geben wird, was neue Chancen zur kreativen Bearbeitung von Gesang mit sich ziehen kann.
Laut einer Studie der Confédération Internationale des Sociétés d'Auteurs et Compositeurs (CISAC, 2024) werden bis zum Jahr 2028 20 % der Streaming-Einnahmen und 60 % der Einnahmen aus Musikbibliotheken KI-generierte Musikstücke ausmachen, was einen Verlust von jährlich etwa 4 Milliarden Euro für Musikschaffende bedeutet. Es ist also davon auszugehen, dass die Verwendung von KI-basierten Tools einen hohen Stellenwert in der Popmusik bilden wird.

8. Fazit

Die menschliche Stimme ist wohl das älteste Musikinstrument der Menschheit und ein zentrales Element in der Popmusik. Dementsprechend schnell fällt es auf, wenn diese unnatürlich und synthetisch klingt, was in der Geschichte der Popmusik allerdings oft als Stilmittel verwendet wurde und weiterhin wird.
Um das Thema der Klangästhetik und Technik synthetisch klingender Stimmen durch Stimmbearbeitung genauer zu beleuchten, wurden zu Beginn der Arbeit wichtige Begrifflichkeiten geklärt. Einen Großteil der Arbeit macht die Untersuchung diverser Effekte aus, insbesondere dem Vocoder, Auto-Tune und der Talkbox, um den historischen, als auch den technischen Hintergrund zu erklären. Es wurden ebenfalls Themen wie die Ästhetik von synthetisch klingenden Stimmen behandelt, wobei auch die Philosophie des Posthumanismus aufgegriffen wurde.
Die einhergehende Fragestellung *Welche Techniken werden in der Popmusik eingesetzt, um Stimmen so zu bearbeiten, dass sie synthetisch klingen, und wie hat sich der ästhetische Ansatz entwickelt?* wurde in dieser Arbeit mithilfe der hinzugezogenen Literatur beantwortet. Die Erkenntnisse sind, dass es einige Techniken bzw. Effekte gibt, die eine Stimme synthetisch klingen lassen, auch, wenn es in einigen Fällen nicht die ursprüngliche Intention war. Die Art und Weise wie ein Effekt verwendet wird und vor allem in welcher Intensität ist dabei ausschlaggebend. Drei Bearbeitungstechnologien haben die Popwelt allerdings durch ihren künstlichen Klang besonders geprägt: Der Vocoder, Auto-Tune und die Talkbox. Der Vocoder analysiert die Stimme, zerlegt sie in ihre Einzelteile und überträgt diese Informationen auf ein anderes Audiosignal. Auto-Tune analysiert ebenfalls die Stimme und passt die Tonhöhe der

einzelnen Töne einer Aufnahme so an, dass sich diese automatisch in einer zuvor ausgewählten Tonlage befindet. Dieser Effekt wird zum Großteil subtil eingesetzt und hat die Popmusik durch intensive Verwendung nachhaltig geprägt. Bei der Talkbox wird der Klang eines Instruments durch einen Schlauch in den Mund eines Menschen geführt und dort moduliert. Bei all diesen Effekten handelt es sich um eine Kombination aus menschlichen Eingaben und technischer Bearbeitung, wodurch man bei der Frage der Authentizität und Ästhetik landet. Nachdem es sich bei der Stimme um das älteste Instrument der Menschheitsgeschichte handelt, fallen bereits minimale Veränderungen auf. Diese Veränderungen können sowohl dafür sorgen, dass Emotionen verloren gehen, es können aber auch Emotionen entstehen. Laut Diederichsen kommt es bei einer Manipulation der Stimme nicht etwa zu einer Entmenschlichung, sondern zu einer Kontrolle der Klänge, da eine Manipulation in der Regel mit einem klanglichen Konzept verbunden und vom Menschen beabsichtigt ist.

Die Faszination für synthetisch klingende Stimmen begann bereits Anfang des zweiten Jahrtausends mit dem *Sprechenden Kopf* von Gerbert von Aurillac. Nachdem im Laufe der Zeit immer mehr Geräte entstanden, die eine Stimme synthetisieren konnte, war es nur eine Frage der Zeit, bis sich dieser Effekt auch in der Musik verbreitete. 1939 gab es mit *Stringy, the Talking Guitar* von Alvino Rey die erste Talkbox, die in einem Musikstück verwendet wurde. Nachdem die Talkbox stetig weiterentwickelt wurde, wurde sie vor allem in der Rockmusik beliebt und wurde von Künstler*innen wie Peter Frampton genutzt. Der Vocoder, der 1968 bei einem Musikstück der Bell Laboratories erstmals zu hören war, ursprünglich aus der Telekommunikation stammt und später primär in der elektronischen Tanzmusik von Musikproduzent*innen wie Kraftwerk und Daft Punk verwendet wurde, hatte, wie auch die Talkbox, das Ziel, synthetisch zu klingen. Beide Effekte wurden als futuristische Effekte, auch in Filmen, eingesetzt. Cher verwendete Auto-Tune erstmals in ihrem Song *Believe,* wodurch sie die Popkultur geprägt hat. Heutzutage ist dieser Effekte vor allem im Rap anzutreffen.

Zukünftig ist davon auszugehen, dass künstliche Intelligenz eine wichtige Rolle in der Popmusik spielen wird, vorwiegend im Bereich der Sprachsynthese. Generell wäre es für eine weitere Forschungsarbeit spannend, den Bereich der reinen Sprachsynthese und dessen Rolle und Ästhetik in der Popmusik genauer zu beleuchten. Ebenfalls wäre eine genauere Analyse der zukünftigen Entwicklung durch künstliche Intelligenz und die verschiedenen Funktionsweisen von derartigen Algorithmen für nachfolgende Arbeiten von Interesse.

Literaturverzeichnis

ACE Studio. (o. D.). *AI singing voice generator*. Acestudio.ai. Abgerufen am 28. Februar 2025, von https://acestudio.ai/

Ahlers, M. (2018). Komposition und Produktion von populärer Musik. In H. Schramm (Hrsg.), *Handbuch Musik und Medien* (S. 1–29). Springer VS. https://doi.org/10.1007/978-3-658-21943-7_17-1

Antares. (2023, 09. August). *Die Wissenschaft hinter Auto-Tune: Wie es funktioniert*. Antarestech.com. Abgerufen am 05. Februar 2025, von https://www.antarestech.com/community/the-science-behind-auto-tune?srsltid=AfmBOooMrYqBooP9MLjk2Euwrw0b4v_PimXmx7HPithvZ50NB9k2BsCT

Berlich, S. (2022). Posthumanismus und Pop-Musik. In *Where Are We Now? - Orientierungen nach der Postmoderne* (S. 273–284). transcript Verlag. https://doi.org/10.1515/9783839462560-022

Bittner, R., Demetriou, A., Jansson, A., & Kumar, A. (2018). *Vocals in music matter: The relevance of vocals in the minds of listeners*. International Society for Music Information Retrieval. Abgerufen am 03. März 2025, von http://ismir2018.ircam.fr/doc/pdfs/98_Paper.pdf

Brackhane, F. (2011). *Die Sprechmaschine Wolfgang von Kempelens – Von den Originalen bis zu den Nachbauten* [Masterarbeit, Universität des Saarlandes]. Leibnitz-Institut für Deutsche Sprache. Abgerufen am, 19. Jänner 2025 von https://ids-pub.bsz-bw.de/files/3400/Brackhane_Die_Sprechmaschine_Wolfgang_von_Kempelens_2011.pdf

Buck, J. R., Oppenheim, A. V., & Schafer, R. W. (2004). *Zeitdiskrete Signalverarbeitung* (2. Aufl.). Pearson Studium.

Campbell, J. J. P., & Dean, R. A. (1993). A history of voice coding. *Digital Signal Processing*, *3*(3), 220–224. https://doi.org/10.1006/dspr.1993.1027

Caudet, E. (2023, 31. Juli). *Vocoder: Funktionswiese, Geschichte und beste Modelle*. Wood and Fire. Abgerufen am 04. Februar 2025, von https://woodandfirestudio.com/vocoder/

Celemony. (o. D.). *Was ist Melodyne?* Celemony.com. Abgerufen am 10. Februar 2025, von https://www.celemony.com/de/melodyne/what-is-melodyne

Chen, H.-W., Lee, J.-H., Lin, B.-Y., Chen, S., & Wu, S.-T. (2017, 01. Dezember). Liquid crystal display and organic light-emitting diode display: present status and future perspectives. *Light, Science & Applications*, *7*(3). https://doi.org/10.1038/lsa.2017.168

Cher. (1998). *Believe*. Warner Bros Records.

Confédération Internationale des Sociétés d'Auteurs et Compositeurs. (2024, November). *Study on the economic impact of Generative AI in the Music and Audiovisual industries*. CISAC. Abgerufen am 28. Februar 2025, von https://members.cisac.org/CisacPortal/cisacDownloadFileSearch.do?docId=47661&lang=en

Crockett, Z. (2016, 26. September). *The mathematical genius of auto-Tune*. Priceonomics. Abgerufen am 17. Februar 2025, von https://priceonomics.com/the-inventor-of-auto-tune/

Daniel, A., & Hillebrandt, F. (2019). Die multiple Formation der Popmusik: Eine praxissoziologische Perspektive. In *Die Praxis der Popmusik: Soziologische Perspektiven* (1. Aufl.). Springer VS. https://doi.org/10.1007/978-3-658-22714-2

Deutsches Museum. (o. D.). *Der Vocoder: Die beeindruckende Geschichte einer großen Erfindung*. Google Arts & Culture. Abgerufen am 19. Jänner 2025, von https://artsandculture.google.com/story/vQUBYdBg1jWLJg?hl=de

Diederichsen, D. (2014). *Über Pop-Musik* (1. Aufl.). Kiepenheuer & Witsch.

Diederichsen, D. (2016). Sound Art Und Die Romantik Des Posthumanen. *POP*, *5*(1), 99–102. https://doi.org/10.14361/pop-2016-0116

Dresen, C., & Klindt, M. (2022). *Eine kleine Handreichung zur Digitalisierung von Audio- und Videomaterial*. digiS - Forschungs- und Kompetenzzentrum Digitalisierung Berlin.

Dudenredaktion & Wöllstein, A. (Hrsg.). (2016). *Die Grammatik: Unentbehrlich für richtiges Deutsch* (9. Aufl.). Dudenverlag. Abgerufen am 19. März 2025, von https://www.ciando.com/img/books/extract/3411911778_lp.pdf

Dudley, H. (1940, Oktober). The Carrier Nature of Speech. *The Bell System Technical Journal*, *19*(4), 495–651. Abgerufen am 01. Februar 2025, von https://www.worldradiohistory.com/Archive-Bell-System-Technical-Journal/40s/Bell-1940d.o.pdf

Eilers, L. (2024, 18. März). *Daft Punk Gear: Die wichtigsten Synths, Sounds und Effekte*. Gearnews.de. Abgerufen am 22. Jänner 2025, von https://www.gearnews.de/daft-punk-equipment-synths-sounds-effekte/

Gale, D. (2018, Juli 18). *The vocoder: History and deployment*. MusicTech. Abgerufen am 20. Jänner 2025, von https://musictech.com/features/the-history-of-the-vocoder/

Gold, B. (1990). A History of Vocoder Research at Lincoln Laboratory. *The Lincoln Laboratory Journal*, *3*(2), 163–202. Abgerufen am 19. Jänner 2025, von https://archive.ll.mit.edu/publications/journal/pdf/vol03_no2/3.2.1.vocoder.pdf

Hildebrand, H. A. (1999). *Pitch detection and intonation correction apparatus and method* (US 5,973,252). United States Patent. https://patents.google.com/patent/US5973252

Hubert, F. (2023). *Piezoelektrische EMV-Filter*. Springer Vieweg. https://doi.org/10.1007/978-3-658-41858-8

Izotope. (2018, 18. Mai). *A Brief History of the Vocoder*. Izotope.com. Abgerufen am 20. Jänner 2025, von https://www.izotope.com/en/learn/a-brief-history-of-the-vocoder.html?srsltid=AfmBOoruau65TqIGaKqF034139WEOLrSD-jnjjIBa_Qk4Yd7diAO4wFN

Jones, S. (2019, 01. Mai). *Auto-Tune & Vocoder — Wir erklären den Unterschied zwischen diesen beiden beliebten Effekten*. Antarestech.com. Abgerufen am 28. Februar 2025, von https://www.antarestech.com/community/auto-tune-vocoder-we-explain-the-difference?srsltid=AfmBOooc_bgpZrrVryCV8UeUV1MjSfkr-40hTl08xJu6jUvCd5pUZ1R4

Kaiser, J. (2017). *Singen in Gemeinschaft als ästhetische Kommunikation: Eine Ethnographische Studie* (1. Aufl.). Springer VS.

King, H. (2023, 20. Dezember). *Exclusive: Gen AI music app Suno comes out of stealth*. Axios. Abgerufen am 28. Februar 2025, von https://www.axios.com/2023/12/20/suno-gen-ai-music-microsoft

Kody, A. (2023, 14. Juli). *Vocal chops 101: everything you need to know to make your own (with free vocal samples)*. Native Instruments. Abgerufen am 18. März 2025, von https://blog.native-instruments.com/vocal-chops

Lent, K. (1989). An efficient method for pitch shifting digitally sampled sounds. *Computer music journal*, *13*(4), 65–71. https://doi.org/10.2307/3679554

Maempel, H.-J. (2021). Audiobearbeitung. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (S. 1–30). Springer VS. https://doi.org/10.1007/978-3-662-60357-4_29-1

Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE.* Institute of Electrical and Electronics Engineers, *63*(4), 561–580. https://doi.org/10.1109/proc.1975.9792

Mandler, J. (2023, 19. März). *Talkbox: In einfachen Worten erklärt*. delamar.de. Abgerufen am 02. Februar 2025, von https://www.delamar.de/gitarre/gitarren-effektgeraete/talkbox-68983/

Marshall, O. (2024). The oleaginous voice: Auto-Tune, linear predictive coding, and the security-petroleum complex. *History and Technology*, *40*(3), 276–296. https://doi.org/10.1080/07341512.2024.2402580

Messitte, N. (2023, 05. Jänner). *Ultimate Guide to Audio Effects*. Izotope.com. Abgerufen am 09. Februar 2025, von https://www.izotope.com/en/learn/guide-to-audio-effects.html?srsltid=AfmBOorztS-RCXzzNC51C1h88x2Ole2ooIiniKzgIvtWtCqWneev2UQq#reverb

Messitte, N. (2024, 21. März). *What is distortion in music? When and how to use it*. Izotope.com. Abgerufen am 10. Februar 2025, von https://www.izotope.com/en/learn/what-is-distortion-in-music-when-and-how-to-use-it.html?srsltid=AfmBOooTauhLTHdmUCdIdB78r3lRrYfEjX7og6obH1FBmAsSmLmiQ6w-

Michels, U. (2008). *dtv-Atlas Musik: Systematischer Teil. Musikgeschichte von den Anfängen bis zur Gegenwart*. dtv Verlagsgesellschaft.

Miranda, C. (2024, 14. Oktober). *Was ist Auto-Tune? Das bemerkenswerte Tool, das die moderne Musik neu definierte.* Moises.ai. Abgerufen am 08. Februar 2025, von https://moises.ai/de/blog/inspiration-de/was-ist-auto-tune/

Pfleiderer, M. (2010). *Vocal pop pleasures. Theoretical, analytical and empirical approaches to voice and singing in popular music*. Abgerufen am 25. März 2025, von https://doi.org/10.5429/2079-3871(2010)v1i1.7en

Reicher, M. E. (2005). *Einführung in die Philosophische Ästhetik* (1. Aufl.). Wissenschaftliche Buchgesellschaft.

Rewak, M. (2022, 26. April). *What is a vocoder? How to use vocoders in your music*. Splice. Abgerufen am 04. Februar 2025, von https://splice.com/blog/how-vocoder-works/

Rothstein, J. (1995). *MIDI: A Comprehensive Introduction* (2. Aufl.). Atlantic Books.

Schmeh, K. (2022). *Codeknacker gegen Codemacher: Die faszinierende Geschichte der Verschlüsselung* (4. Aufl.). Springer VS. https://doi.org/10.1007/978-3-658-34189-3

Schuster, S. (2020). *Wirkungs-Elemente der Musik: Inspirationen & Erkenntnisse für Musiker & Musikliebhaber*. Books on Demand.

Schwarz, M. (2019, 5. August). *Maschinenstimmen: Vom Vocoder zum Prisma*. Redbull.com. Abgerufen am 20. Jänner 2025, von https://www.redbull.com/at-de/mugm-vocoder-von-moroder-bis-flut-prismizer-stimme

Schwarzer, M. (2023, 19. Oktober). *Auto-Tune: Wie ein Cher-Song die Popmusik für immer veränderte*. RedaktionsNetzwerk Deutschland. Abgerufen am 05. Februar 2025, von https://www.rnd.de/kultur/auto-tune-wie-ein-cher-song-die-popmusik-fuer-immer-veraenderte-DYZYSGODE5BJVO4E52HQQZWJFA.html

Seebode, J., & Sendlmeier, W. F. (2006). *Formantkarten des deutschen Vokalsystems*. TU Berlin, Institut für Sprache und Kommunikation. Abgerufen am 19. März 2025, von http://www.static.tu.berlin/fileadmin/www/10002019/Forschung/Formantkarten_des_deutschen_Vokalsystems_01.pdf

Segundo, C. S. (2022, 23. Februar). *Das große Equalizer 1x1*. delamar.de. Abgerufen am 10. Februar 2025, von https://www.delamar.de/musikproduktion/eq-tutorial-1-2095/

Seifert, R. (2018). *Popmusik in Zeiten der Digitalisierung: Veränderte Aneignung - veränderte Wertigkeit*. transcript Verlag. https://doi.org/10.1515/9783839444825

Smith, P. (2021, 13. Oktober). *Harder Better Faster Stronger: Behind Daft Punk's indestructible classic*. Dig!. Abgerufen am 26. Februar 2025, von https://www.thisisdig.com/feature/harder-better-faster-stronger-daft-punk-song-story/

Spotify. (o. D.). *Was ist Spotify?* Spotify.com. Abgerufen 19. März 2025, von https://support.spotify.com/at/article/what-is-spotify/

Steinwachs, M. (2021, 24. März). *Was ist eine Talkbox? Funktionsweise und Geschichte: Warum eine Talkbox kein Vocoder ist*. Amazona. Abgerufen am 02. Februar 2025, von https://www.amazona.de/was-ist-eine-talkbox-funktionsweise-und-geschichte/

Rolling Stone (2015, 29. April). *50 greatest live albums of all time*. Rolling Stone. Abgerufen am 27. Februar 2025, von https://www.rollingstone.com/music/music-lists/50-greatest-live-albums-of-all-time-173246/the-replacements-the-shit-hits-the-fans-1985-154247/

Weadon, P. D. (o. D.). *Sigsaly Story*. United States Department of Defense. Abgerufen 20. Jänner 2025, von https://media.defense.gov/2021/Jul/13/2002761629/-1/-1/0/SIGSALY%20STORY.PDF

Weinzierl, S. (2023). Audiosignale und -systeme. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (S. 1–22). Springer VS. https://doi.org/10.1007/978-3-662-60357-4_1-1

Werwein, M. (2021). A/D- und D/A-Wandler für Audiosignale. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (S. 1–15). Springer VS. https://doi.org/10.1007/978-3-662-60357-4_32-1

Wicke, P. (2008). *Popscriptum: Das Sonische in der Musik*. Forschungszentrum Populäre Musik der Humboldt-Universität zu Berlin. https://doi.org/10.18452/20288

Zölzer, U., & Lerch, A. (2025). Digitale Audio-Effekte. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (S. 1–22). Springer VS. https://doi.org/10.1007/978-3-662-60357-4_28-1

Abbildungsverzeichnis

Abbildung 1: Vocoder von Bob Moog. Quelle: Izotope (2018). 10

Abbildung 2: Schematische Darstellung des Vocoders. Quelle: Dudley (1940, S. 508). 16

Abbildung 3: Blockdiagramm der Tonhöhenbestimmung von Auto-Tune. Quelle: Hildebrand (1999, S. 1). 18

Glossar

Predelay

Bei einem Predelay handelt es sich um einen Parameter, der üblicherweise bei Halleffekten aufzufinden ist und die Verzögerung von dem Nachhall beeinflusst (Maempel, 2021).

Fast-Fourier-Transformation (FFT)

Mithilfe des Algorithmus der FFT lässt sich die diskrete Fourier Transformation in Echtzeit berechnen. Mit ihr ist es möglich, die Energieverteilung von Frequenzen aus einem Audiosignal zu analysieren (Weinzierl, 2023).

Vocal Chops

Bei Vocal Chops handelt es sich um die Technik, eine Gesangsspur in einzelne Ausschnitte aufzuteilen und diese Ausschnitte so anzuordnen, dass ein neuer Rhythmus oder eine neue Melodie entsteht (Kody, 2023).

Spotify

Spotify ist eine digitale Plattform, auf der Musik, Podcasts und Videos auf Smartphones, dem Computer oder anderen Geräten gestreamt werden können (Spotify, o.D.).

Frikative

Diese Artikulationsart wird auch Reibelaute genannt und entsteht, wenn der Luftstrom beim Sprechen durch eine Engstelle im Mundraum gepresst wird. Beispiele hierfür sind das alveolare [s] oder das dentale [f] (Dudenredaktion & Wöllstein, 2016).

Affrikate

Wenn nach einem Frikativ ein Plosiv, wie ein [t] oder [p], am selben Artikulationsort folgt, lassen sich sogenannte Affrikate bilden. Affrikate wären beispielsweise ein [ts] wie in Zopf und [pf] wie in Pfanne (Dudenredaktion & Wöllstein, 2016).

Bandpassfilter

Ein Bandpassfilter ist ein Filter, der nur ein gewisses Frequenzband durchlässt. Alle Frequenzen, die sich außerhalb des eingestellten Frequenzbands befinden, werden entweder nicht oder in einer abgeschwächten Stärke durchgelassen, was sich in der Regel einstellen lässt (Lerch & Zölzer, 2025).

Oszillator

Oszillatoren erzeugen hörbare, periodische Audiosignale, die bei Synthesizern in der Regel als Ausgangssignal verwendet werden (Maempel, 2021).

Bandbreite

Eine Bandbreite definiert die Größe des Frequenzbereichs, die verarbeitet oder übertragen wird (Weinzierl, 2023).

Allpassfilter

Bei dem Allpassfilter handelt es sich um einen Filter, der zwar alle Frequenzen durchlässt, aber Phasenverschiebungen abhängig von der Frequenz verursacht (Lerch & Zölzer, 2025).

Hall

Wenn sich Schallwellen in einem Raum bewegen, werden sie durch physische Hindernisse teilweise absorbiert oder wieder reflektiert. Dieses physikalische Phänomen kann mit einem Hall-Effekt digital oder auch über ein analoges Gerät nachgebaut werden (Messitte, 2023). Vor der digitalen Studiotechnik wurde Hall mit echten Räumen erzeugt, indem ein Audiosignal über Lautsprecher in einem Raum ausgegeben und anschließend mit Mikrofonen wieder aufgenommen wurde, wodurch die Aufnahmen einen natürlichen Hall erhielten. In den 1950er Jahren kamen mechanische Methoden wie die Hallplatte auf, die durch einen elektrodynamischen Wandler in Schwingungen versetzt und mit Hilfe eines Piezo-Tonabnehmers abgenommen wurde. Zu Beginn der 1960er Jahre wurde die erste Idee eines digitalen Halls konzipiert. In Echtzeit und auf handelsüblichen Computern anwendbar wurden diese Algorithmen allerdings erst in den 1970er Jahren, was den Beginn der Verwendung in der Musikproduktion darstellte. Die Größe des Raumes, sowie Eigenschaften, wie die Nachhallzeit und ein Predelay sind in der Regel bei klassischen Hallgeräten einstellbar. Weitaus weniger Einstellungsmöglichkeiten bieten Faltungsalgorithmen, die es allerdings schaffen können, einen Raum mit hoher klanglicher Präzision zu simulieren. Aber auch das Gegenteil ist möglich, indem mit synthetischen Samples gearbeitet wird. Das Ergebnis sind unnatürliche Klänge bzw. Hallräume, die so in der Natur nicht vorkommen (Maempel, 2021). Bezogen auf den Gesang, kann mithilfe von Hall einer Stimme Tiefe und Natürlichkeit verschafft werden, es kann aber auch als kreatives Stilmittel eingesetzt werden, beispielsweise durch einen „verwaschenen“ Klang bei einem erhöhten Einsatz (Messitte, 2023).

Delay

Ein Delay-Plugin nimmt das Eingangssignal auf und gibt es nach einer einstellbaren Zeit wieder aus. Das Signal wird also beliebig vervielfacht und zeitlich verschoben. Wie sich dieses

neue Signal verhält, lässt sich über ein Bedienfeld bestimmen. In der Akustik entsteht dieser Effekt bei weit entfernten reflektierenden Oberflächen, sodass das Originalsignal und die Reflexion getrennt voneinander zu hören sind. In der Stimmbearbeitung wird dieser Effekt in erster Linie dafür genutzt, um Tiefe zu schaffen oder gewisse Wörter zu betonen (Messitte, 2023). Eine hörbare Verzögerung entsteht, wenn zwischen den beiden Signalen ein Zeitabstand von mindestens 40 Millisekunden liegt. Eine derart geringe Verzögerung wird zumeist zur räumlichen Erweiterung eines Audiosignals genutzt, ähnlich dem Hall, da keine ausgeprägten Echos entstehen. Ein deutliches Echo wird erst ab 100 Millisekunden hörbar (Maempel, 2021).

Equalizer

Mit einem Equalizer lässt sich das Frequenzspektrum eines Audiosignals verändern, indem bestimmte Frequenzen verstärkt, abgeschwächt oder ganz entfernt werden können. Störfrequenzen können entfernt und das Klangbild hinsichtlich seiner Frequenzverteilung geformt werden. Es sind aber auch kreative Effekte, wie der einer Telefon- oder Radiostimme möglich, indem hohe und tiefe Frequenzanteile entfernt werden (Segundo, 2022). Dynamische Equalizer agieren automatisch und ändern ihre Einstellungen abhängig vom Pegel der zu bearbeitenden Frequenzanteile (Maempel, 2021).

Kompressor

Kompression sorgt dafür, dass die Dynamik eines Audiosignals reduziert wird, indem laute Passagen eines Audiosignals leiser und leise Passagen lauter gemacht werden können. In der Musikproduktion wird dieser Effekt oftmals in Form von Sidechain Kompression angewendet: Ein Signal beeinflusst die Dynamik eines anderen. So kann beispielsweise das Instrumental zu dem Zeitpunkt, an dem Gesang zu hören ist, reduziert werden, damit die Stimme klarer zu hören ist (Messitte, 2023). Übliche Parameter sind *Threshold*, also der Pegel, ab dem der Kompressor arbeitet, *Ratio*, das für das Verhältnis der Lautstärke zwischen dem eingehenden und dem ausgehenden Signal oberhalb des Thresholds steht, *Attack*, bei dem es sich um die Zeit handelt, die der Kompressor benötigt, um zu reagieren, *Release*, das die Zeit definiert, bis der Kompressor nach dem Unterschreiten des Thresholds mit dem Eingreifen aussetzt und *Make-Up Gain*, das zur Verstärkung des Ausgangssignal verwendet wird (Maempel, 2021).

Chorus

Der Chorus wird oft als Dopplungseffekt angewendet und lässt ein Audiosignal voller und breiter wirken. Dies geschieht, indem dasselbe Audiosignal mit einer leichten zeitlichen Verschiebung und einer minimalen Tonhöhendifferenz abgespielt wird, um eine mehrstimmige

Wirkung zu erzielen (Messitte, 2023). Die Verzögerungszeiten zwischen dem Eingangssignal und dem gedoppelten Signal liegen zwischen 15 und 30 Millisekunden (Maempel, 2021).

Phaser

Bei einem Phaser handelt es sich um einen Modulationseffekt, der das Eingangssignal dupliziert und die Kopie durch mehrere Allpassfilter schickt, die Phasenauslöschungen und Kammfiltereffekte erzeugen. Um Dynamik zu erzeugen, verändert sich die Phasenlage periodisch. Dieses Signal wird anschließend wieder mit dem Originalsignal zusammengemischt, wodurch ein wellenartiges Klangbild entsteht. Die Modulationsgeschwindigkeit, die Modulationstiefe, der Modulationsmittelpunkt sowie die Rückkopplung sind in der Regel einstellbar. Eine häufige Verwendung des Effekts findet bei Synthesizern und Gitarren statt und ist daher häufig in Form von Bodeneffektgeräten aufzufinden (Maempel, 2021; Messitte 2023).

Flanger

Ein Flanger funktioniert ähnlich wie ein Chorus, allerdings mit Verzögerungszeiten von ein bis zehn Millisekunden. Durch die minimale Verschiebung und erneutem Zusammenmischen der zwei Signale entsteht ein Kammfiltereffekt, der sehr metallisch klingt (Maempel, 2021).

Verzerrer

Durch die Übersteuerung eines Audiosignals entsteht eine Verzerrung, wodurch zusätzliche Obertöne entstehen. Nachdem eine leichte Verzerrung bei analogen Geräten häufig auftaucht, wird dieser Effekte in einer abgeschwächten Form oftmals zur Simulation von Hardwaregeräten verwendet (Messitte, 2024). Über Parameter, wie *Gain*, *Boost* oder *Drive* kann die Stärke der Verzerrung eingestellt werden, wodurch auch bestimmt wird, wie stark ein Audiosignal abgeschnitten wird, was sich *Clipping* nennt. Häufig wird dieser Effekt bei E-Gitarren eingesetzt (Maempel, 2021).

Pitch Shifter

Ein Pitch Shifter kann die Tonhöhe eines Audiosignals mithilfe von Resampling verändern und so eine Stimme beliebig hoch oder tief klingen lassen: Die Samplerate wird erhöht oder gesenkt, was nicht nur Änderungen der Tonhöhe, sondern auch der Länge und der Formanten bewirkt. Um der Längenänderung entgegenzuwirken wird das Signal mit der neuen Abtastrate anschließend komprimiert oder verlängert. Dieser Vorgang beinhaltet oftmals Störgeräusche, wie Clicks, Pops oder andere Modulationseffekte. Durch die Änderung der Formanten kann eine Aufnahme bereits nach dem Pitchshift einer Oktave unerkennbar werden. Mit dem Einsatz von linear predictive filters und der Fast-Fourier-Transformation ist es möglich, die

Formanten beizubehalten, was allerdings eine erhöhte Rechenleistung erfordert (Lent, 1989, S. 65).

Linear predictive filter

Mithilfe eines Linear predictive filters lassen sich Audiosignale aufgrund vergangener Werte vorhersagen (Makhoul, 1975).

MIDI

MIDI steht für *Musical Instrument Digital Interface* und ist ein Protokoll, das Daten wie Tonhöhe und die Lautstärke auf eine Software übertragen kann. Es beschreibt dabei nicht, wie etwas klingen soll, es dient er als Eingabegerät für Daten (Rothstein, 1995).

Filterbank

Bei einer Filterbank handelt es sich um eine Reihe an Filtern, die ein Audiosignal hinsichtlich ihrer Frequenzbereiche aufteilen bzw. analysieren können (Weinzierl, 2023).

A/D-Wandler

Ein Analog-Digital-Wandler wandelt analoge Signale, wie Schallwellen, in digitale Signale, wie Zahlenwerte (Werwein, 2021).

D/A-Wandler

Ein Digital-Analog-Wandler ist das Gegenstück zum A/D-Wandler und wandelt digitale Signale, wie Zahlenwerte, in analoge Signale, wie Schallwellen (Werwein, 2021).

Bittiefe

Die Bittiefe gibt an, wie viele Bits bei einer Quantisierung eines Audiosignals genutzt werden und beschreibt somit die Auflösung des Signales. Übliche Werte sind 8 Bit, 16 Bit, 24 Bit und 32 Bit (Dresen & Klindt, 2022).

Abtastrate

Die Abtastrate, auch Samplerate bezeichnet, beschreibt wie oft in der Sekunde ein analoges in ein digitales Audiosignal umgewandelt wird. Gängige Werte sind 44,1 kHz, 48 kHz und 96 kHz (Buck et al., 2004).

LCD Display

Ein LCD Display, ausgeschrieben Liquid Crystal Display, stellt Bilder mithilfe von flüssigen Kristallen und darauf einstrahlendem Licht dar (Chen et al., 2018).

Piezo-Tonabnehmer

Piezo-Tonabnehmer befinden sich direkt auf der Oberfläche eines Festkörpers und können mechanische Energie, die durch die Vibrationen des Körpers entsteht, in elektronische Energie umwandeln, wodurch kein Mikrofon notwendig ist (Hubert, 2023).

Kammfiltereffekt

Wird ein Signal ident dupliziert und das duplizierte Signal leicht verzögert mit dem Ausgangssignal erneut zusammengeführt, werden gewisse Frequenzen angehoben, andere wiederrum gesenkt, was als Kammfiltereffekt bezeichnet wird. Diese Anhebung bzw. Senkung ist dabei partialtonähnlich (Maempel, 2021).

Formanten

Bei Formanten handelt es sich um Frequenzbänder, die resonieren, also durch einen Raum besonders verstärkt werden. Diese entstehen beim Menschen im Vokaltrakt und sind individuell. Besonders betroffen sind dabei Vokale, wie *a*, *e*, *i*, *o* und *u* (Seebode & Sendlmeier, 2006).