

Hochschule der Medien Stuttgart
Fakultät Electronic Media
Audiovisual Media Creation and Technology (AV3)



Masterarbeit zum Thema:
Einfluss der Bearbeitungszeit von Vocals in der Popmusikproduktion
auf ihre wahrgenommene Qualität im Kontext der Mischung.

Zur Erlangung des akademischen Grades Master of Science

Vorgelegt von:
Daniel Knüttel
Kandelweg 7
79258 Hartheim
E-mail: dk123@hdm-stuttgart.de

Matrikelnummer: 40223
Fachsemester: 4
Studiengang: Audiovisual Media Creation and Technology (AV3)
Bearbeitungsende: 19.08.2025

Erstprüfer: Prof. Oliver Curdt
Zweitprüfer: Johannes Wohlleben

Ehrenwörtliche Erklärung

Hiermit versichere ich, Daniel Knüttel, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel: „Einfluss der Bearbeitungszeit von Vocals in der Popmusikproduktion auf ihre wahrgenommene Qualität im Kontext der Mischung“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden. Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Datum: 18.08.2025

Unterschrift: 

Vorwort

Die Leidenschaft für Musikproduktion begleitet mich nun seit mehr als 12 Jahren. Besonders die Popmusik hat mich schon immer fasziniert. Eine Kultur, die sich durch unfassbare Vielfalt auszeichnet und Raum für Kreativität lässt, sich immer wieder neu zu erfinden.

Während meines Bachelorstudiums in Audiovisuellen Medien mit dem Schwerpunkt Ton an der Hochschule der Medien in Stuttgart, konnte ich meine technischen Fähigkeiten vertiefen und mich künstlerisch im Bereich der Musikproduktion weiterentwickeln. Mit dieser Masterarbeit darf ich nun meine musikalische Reise fortsetzen.

Die Vocal-Produktion zählt für mich zu den spannendsten und facettenreichsten Bereichen der Popmusik. Vocals sollen für den Zuhörer leicht greifbar sein, sind in ihrer Produktion jedoch keineswegs simpel. Mit ihnen steht und fällt ein Song. Sie geben demselben einen emotionalen Leitfaden, ein Gesicht und eine Seele. Vor allem im Alltag als Musikproduzent spielt aus ökonomischer Sicht heutzutage die effiziente Vocal-Bearbeitung eine wichtige Rolle. Die Musikszene befindet sich in einem stetigen Wandel und ist schnelllebiger geworden. In der Produktion und besonders bei der Bearbeitung von Vocals erachte ich es daher als sinnvoll, ein gutes Mittel zwischen Wirtschaftlichkeit und Kreativität anzustreben. Diese Arbeit ist Ausdruck meiner Faszination und meiner Motivation, die Kunst der Vocal-Bearbeitung wissenschaftlich zu erkunden.

Mein besonderer Dank gilt Professor Oliver Curdt, der mich und meine Musik stets unterstützt und mir Raum gegeben hat, meine Ideen zu verwirklichen. Ebenso danke ich meinem Zweitkorrektor Johannes Wohlleben für seine wertvolle Begleitung. Ein herzliches Dankeschön gilt auch allen, die mich auf diesem Weg moralisch unterstützt haben.

Kurzfassung

Diese Masterarbeit untersucht den Einfluss unterschiedlicher Bearbeitungsumfänge von Gesangsspuren (Vocals) auf deren wahrgenommene Qualität im Gesamtmix moderner Popmusik. Im Fokus steht die Frage, ob eine detaillierte und damit zeitintensive Nachbearbeitung von Hörenden als qualitativer Mehrwert empfunden wird. Es wurden qualitative Experteninterviews mit professionellen Produzenten sowie ein Hörtest mit drei Testgruppen (Laien, audiophile Hörer und professionelle Produzenten) durchgeführt. Alle Testgruppen konnten die unbearbeitete Version identifizieren. Ein signifikanter Unterschied zwischen der kurz und der ausführlich bearbeiteten Version wurde hingegen kaum festgestellt. Die Untersuchung legt nahe, dass sich das Verhältnis zwischen Investition und Wahrnehmung der Vocal-Bearbeitung im Sinne des Pareto-Prinzips beschreiben lassen.

Abstract

This master's thesis examines the impact of varying degrees of vocal editing on the perceived quality of vocals within the overall mix of contemporary pop music. The central question is whether detailed and time-intensive post-processing is perceived as a qualitative improvement. Qualitative expert interviews with professional music producers and a listening test with three groups (lay listeners, audiophiles, and professional producers) were conducted. All groups were able to identify the unprocessed version; however, significant differences between the briefly and extensively edited versions were rarely observed. The study suggests that the relationship between editing investment and perceived vocal quality can be interpreted in terms of the Pareto principle.

Inhaltsverzeichnis

| | |
|--|-----------|
| EHRENWÖRTLICHE ERKLÄRUNG | 2 |
| VORWORT | 3 |
| KURZFASSUNG | 4 |
| ABSTRACT | 4 |
| INHALTSVERZEICHNIS | 5 |
| ABBILDUNGSVERZEICHNIS | 8 |
| TABELLENVERZEICHNIS | 8 |
| ABKÜRZUNGSVERZEICHNIS | 9 |
| HINWEISE ZUR ARBEIT | 10 |
| 1 EINLEITUNG | 11 |
| 1.1 RELEVANZ UND KONTEXT VON VOCALS IN DER POPMUSIKPRODUKTION | 11 |
| 1.2 ZIELSETZUNG UND FORSCHUNGSFRAGEN | 12 |
| 1.3 AUFBAU DER ARBEIT | 13 |
| 2 GRUNDLAGEN | 15 |
| 2.1 BEDEUTUNG UND ENTWICKLUNG VON VOCALS IN DER POPMUSIK IN DEN LETZTEN ZWEI JAHRZEHNEN | 15 |
| 2.2 ANFORDERUNGEN BEI DER PRODUKTION VON MODERNER POPMUSIK UND DEREN VOCALS | 17 |
| 2.3 DER RECORDING-PROZESS | 20 |
| 2.3.1 <i>Technische Komponenten</i> | 20 |
| 2.3.2 <i>Psychologische Komponenten</i> | 24 |
| 2.3.3 <i>Weitere entscheidende Faktoren</i> | 25 |
| 2.3.3.1 Raumakustik | 25 |
| 2.3.3.2 Monitoring beim Recording | 25 |
| 2.3.3.3 Gain-Staging | 26 |
| 2.4 TOOLS UND TECHNIKEN ZUR VOCAL-BEARBEITUNG | 27 |
| 2.4.1 <i>Das Editieren</i> | 27 |
| 2.4.2 <i>Vocal-Comping</i> | 28 |
| 2.4.3 <i>Vocal-Layering</i> | 30 |
| 2.4.4 <i>Intonationskorrektur</i> | 31 |
| 2.4.5 <i>Equalizer</i> | 32 |
| 2.4.6 <i>Kompressor</i> | 37 |
| 2.4.7 <i>Limiter</i> | 39 |

| | | |
|----------|--|-----------|
| 2.4.8 | <i>De-Esser</i> | 40 |
| 2.4.9 | <i>Saturation Tools</i> | 41 |
| 2.5 | DAS PARETO-PRINZIP | 43 |
| 2.5.1 | <i>Der Bezug zur Vocal-Bearbeitung</i> | 44 |
| 3 | METHODIK | 45 |
| 3.1 | FORSCHUNGSDESIGN: QUALITATIVE DATENERHEBUNG DURCH EXPERTENINTERVIEWS | 45 |
| 3.1.1 | <i>Aufbau der Experteninterviews: Evaluation des Mindestmaßes an Bearbeitung</i> | 46 |
| 3.2 | UNTERSCHIEDLICHE BEARBEITUNGSSTUFEN VON VOCALS IM HÖRVERGLEICH | 48 |
| 3.2.1 | <i>Demomaterial</i> | 49 |
| 3.2.1.1 | Unbearbeitete Version | 49 |
| 3.2.1.2 | Mindest-Bearbeitung (30 min.) | 50 |
| 3.2.1.3 | Detaillierte Bearbeitung (120 min.) | 53 |
| 3.3 | TECHNISCHER VERSUCHSAUFBAU | 56 |
| 3.4 | VERSUCHSDURCHFÜHRUNG | 56 |
| 3.5 | AUSWAHL DER PROBANDEN | 57 |
| 4 | ERGEBNISSE | 60 |
| 4.1 | ERGEBNISSE DER EXPERTENINTERVIEWS | 60 |
| 4.1.1 | <i>Konsens und Unterschiede zur Mindestbearbeitung</i> | 60 |
| 4.1.2 | <i>Perspektiven zur Detaillierten Vocal-Bearbeitung</i> | 63 |
| 4.1.3 | <i>Weitere Erkenntnisse aus den Experteninterviews</i> | 64 |
| 4.2 | ERGEBNISSE DES HÖRTESTS: WAHRNEHMUNGSUNTERSCHIEDE ZWISCHEN BEARBEITUNGSSTUFEN | 66 |
| 4.2.1 | <i>Allgemeine Beobachtungen</i> | 67 |
| 4.2.2 | <i>Vergleich unbearbeitet vs. bearbeitet</i> | 68 |
| 4.2.3 | <i>Kurz Bearbeitet vs. Detailliert bearbeitet</i> | 69 |
| 4.2.4 | <i>Unterschiede zwischen den Testgruppen</i> | 70 |
| 4.3 | ABGLEICH DER ERKENNTNISSE: EXPERTENMEINUNGEN, PRAXISANALYSEN UND HÖRTESTERGEBNISSE | 72 |
| 5 | DISKUSSION | 73 |
| 5.1 | INTERPRETATION DER ERGEBNISSE IM KONTEXT DER MUSIKPRODUKTION | 73 |
| 5.2 | BEDEUTUNG DES MINDESTMAßES BEZÜGLICH EFFIZIENZ UND PRODUKTIONSQUALITÄT | 75 |
| 5.3 | RELEVANZ DES PARETO-PRINZIPS FÜR DIE VOCAL-BEARBEITUNG | 77 |
| 5.4 | METHODISCHE REFLEXION: EINSCHRÄNKUNGEN UND VERZERRUNGEN DER UNTERSUCHUNG | 78 |
| 6 | FAZIT | 82 |
| 6.1 | ZUSAMMENFASSUNG DER WICHTIGSTEN ERKENNTNISSE | 82 |
| 6.2 | BEANTWORTUNG DER FORSCHUNGSFRAGEN | 83 |

| | | |
|----------|---|-----------|
| 6.3 | AUSBLICK AUF ZUKÜNFTIGE FORSCHUNGSMÖGLICHKEITEN | 85 |
| 7 | QUELLENVERZEICHNIS | 89 |
| | DIGITALER ANHANG | 94 |

Abbildungsverzeichnis

| | |
|---|----|
| ABBILDUNG 1: SCHEMATISCHE DARSTELLUNG KONDENSATORMIKROFON (LINKS) UND TAUCHSPULENMIKROFON (RECHTS). (EIGENE DARSTELLUNG) | 21 |
| ABBILDUNG 2: BENUTZEROBERFLÄCHE MIT TAKE-LANES IN ABLETON LIVE | 29 |
| ABBILDUNG 3: FREQUENZAUFTEILUNG DER MENSCHLICHEN STIMME IN FÜR DIE MUSIKPRODUKTION RELEVANTE BEREICHE (SCHEMATISCH, EIGENE DARSTELLUNG) | 35 |
| ABBILDUNG 4: BENUTZEROBERFLÄCHE FABFILTER PROQ3 | 51 |
| ABBILDUNG 5: BENUTZEROBERFLÄCHE ANTARES AUTOTUNE | 51 |
| ABBILDUNG 6: BENUTZEROBERFLÄCHE UAD 1176LN KOMPRESSOR | 52 |
| ABBILDUNG 7: BENUTZEROBERFLÄCHE IZOTOPE RX DE-ESSER..... | 53 |
| ABBILDUNG 8: AUDIOSPUR NACH KORREKTUR DES TIMINGS UND COMPING IN ABLETON LIVE | 53 |
| ABBILDUNG 9: BENUTZEROBERFLÄCHE CELEMONY MELODYNE | 54 |
| ABBILDUNG 10: BENUTZEROBERFLÄCHE VOCALIGN PROJECT5 | 55 |

Tabellenverzeichnis

| | |
|---|----|
| TABELLE 1: ÜBERSICHT DER INTERVIEWPARTNER (PRODUZENTEN) | 47 |
|---|----|

Abkürzungsverzeichnis

| | |
|-------------|---------------------------------------|
| A/D-Wandler | Analog/Digital-Wandler |
| AI | Artificial Intelligence |
| BPM | Beats per minute |
| dB | Dezibel |
| dBFS | Dezibel relative to Full Scale |
| DAW | Digital Audio Workstation |
| EQ | Equalizer |
| FET | Field Effect Transistor |
| Hz | Hertz |
| KI | Künstliche Intelligenz |
| LUFS | Loudness Units relative to Full Scale |
| PCM | Pulse Code Modulation |
| VCA | Voltage Controlled Amplifier |
| Vari-Mu | Variable-Mu-Kompressor |

Hinweise zur Arbeit

Die in dieser Masterarbeit verwendeten Personenbezeichnungen beziehen sich immer gleichermaßen auf weibliche, männliche und generell alle in LGBTQAI+ einbegriffenen Personen, sowie alle dazwischen und außerhalb. Auf eine Doppelnennung und gegenderte Bezeichnungen wird zugunsten einer besseren Lesbarkeit verzichtet.

Für sprachliche Optimierungen in Bezug auf Grammatik wurde punktuell das KI-gestützte Textfeedback-Tool „ChatGPT“ (OpenAI, 2025) eingesetzt. Inhaltliche Konzeption, Analyse, Auswertung und wissenschaftliche Argumentation wurden davon nicht beeinflusst und sind vollständig eigenständig erarbeitet.

1 Einleitung

1.1 Relevanz und Kontext von Vocals in der Popmusikproduktion

Die Produktion von Popmusik unterliegt in den letzten Jahrzehnten einem ständigen Wandel. Technologische Entwicklungen, neue Vertriebsformen und sich ändernde Hörgewohnheiten haben nicht nur die Art verändert, wie Musik klingt und wahrgenommen wird. Vielmehr haben sich aufgrund dessen auch Prozesse innerhalb der Musikindustrie wie z.B Recording, Bearbeitung und Veröffentlichung angepasst.

Dabei nimmt die Stimme eine zentrale Rolle ein. Die Vocals sind in der Regel der prominenteste Bestandteil eines Popsongs – sie transportieren nicht nur den Text, sondern geben der Produktion auch die emotionale Identität. Gerade in der Popmusik sind sie eng mit Wiedererkennungswert und Markenbildung rund um die Künstler verknüpft. Entsprechend hoch sind die Anforderungen an Faktoren wie z.B Klangqualität, Sprachverständlichkeit, stilistische Konsistenz und emotionale Wirkung [1], [2], [3],[4].

Im Rahmen dieser Arbeit wird der Begriff „Vocals“ verwendet, um die Gesangsspuren in einer Musikproduktion zu bezeichnen. Der englische Terminus ist im Produktionskontext weit verbreitet und schließt sowohl Haupt- als auch Begleitstimmen, Harmonien und Layer-Spuren mit ein. Synonym wird im Text gelegentlich von „Gesangsspuren“ oder „Stimme“ gesprochen, wobei stets dasselbe gemeint ist.

Vocal-Bearbeitung ist längst nicht mehr nur ein isolierter Nachbearbeitungsschritt, sondern ein integraler Bestandteil der Popmusikproduktion. Tools wie Auto-Tune, Equalizer, Kompressoren, Sättigung oder Hallräume werden gezielt eingesetzt, um einen Sound zu erzeugen, der als zeitgemäß, professionell und „releasetauglich“ gilt.

Dabei verschwimmen zunehmend die Grenzen zwischen technischer Korrektur, kreativer Gestaltung und stilistischer Notwendigkeit.

Im Spannungsfeld zwischen ästhetischem Anspruch und Produktionsökonomie stellt sich zunehmend die Frage, welche Bearbeitungsschritte in Bezug auf Vocals für eine moderne Popproduktion tatsächlich notwendig sind – und welche Eingriffe klanglich oder wahrnehmungspsychologisch kaum ins Gewicht fallen. Diese Entwicklung macht die gezielte Untersuchung der Vocal-Bearbeitung innerhalb aktueller Popproduktionen nicht nur aus technischer Perspektive, sondern auch aus ästhetischer und ökonomischer Sicht relevant [1], [3],[4].

1.2 Zielsetzung und Forschungsfragen

Ziel dieser Arbeit ist es, den Einfluss des Bearbeitungsumfangs von Vocals in Popmusik Produktionen auf deren wahrgenommene Qualität im finalen Mix zu untersuchen. Das Wort „Mix“ bezieht sich hier auf die fertig gemischte Summe einer Musikproduktion. Synonym werden in dieser Arbeit je nach Kontext Begriffe wie „Musikmix“, „Mischung“, „Gesamtmischung“ oder „Summe“ verwendet, welche jedoch dieselbe Bedeutung haben.

Es wird insbesondere der Vergleich zwischen unbearbeiteten, schnell bearbeiteten und detailliert bearbeiteten Gesangsspuren fokussiert. Genauere Erläuterungen dazu finden sich ab Kapitel 3. Die Untersuchung erfolgt im Rahmen von Experteninterviews mit erfahrenen Musikproduzenten, sowie eines qualitativen Hörversuchs mit drei Testgruppen: Laien, audiophile Hörer und professionelle Muskschaffende (im folgenden „professionals“).

Im Zentrum dieser größtenteils empirischen Untersuchungen steht die Rezeption von Konsumenten, woraus sich folgende Forschungsfrage ergibt:

Welchen Einfluss hat der Umfang der Vocal-Bearbeitung in der modernen Popmusikproduktion auf die wahrgenommene Qualität der Vocals im Gesamtmix?

Untergeordnete Forschungsfragen:

- Wie deutlich wird Vocal-Bearbeitung von Hörenden als qualitätssteigernd wahrgenommen?
- Inwiefern unterscheiden sich die Wahrnehmungen zwischen unterschiedlichen Hörergruppen?
- Wo könnte ein „Sweet Spot“ zwischen Aufwand und hörbarem Mehrwert liegen?
- Inwieweit lässt sich der Zusammenhang zwischen Bearbeitungsaufwand und Wirkung im Sinne des Pareto-Prinzips interpretieren?

1.3 Aufbau der Arbeit

Kapitel 2 behandelt die technischen und gestalterischen Grundlagen der Vocal-Produktion. Hier wird der historische Hintergrund von Vocals in der Popmusik beleuchtet. Außerdem geht es um aktuelle Anforderungen bei der Produktion sowie Tools und Techniken.

Kapitel 3 beschreibt die methodische Vorgehensweise, insbesondere die Entwicklung des Demomaterials, den Ablauf des qualitativen Hörversuchs sowie der Experteninterviews.

Kapitel 4 präsentiert die Ergebnisse aus Interviews und Hörtests, die anschließend in Kapitel 5 diskutiert und im Hinblick auf Effizienz, Wahrnehmung und das Pareto-Prinzip eingeordnet werden.

Kapitel 6 fasst die zentralen Erkenntnisse zusammen und bietet einen Ausblick auf weiterführende Forschungsperspektiven.

2 Grundlagen

2.1 Bedeutung und Entwicklung von Vocals in der Popmusik in den letzten zwei Jahrzehnten

Seit dem Aufkommen der Popmusik in den 1950er Jahren ist die Stimme ein zentrales Element. In den letzten zwei Jahrzehnten hat sich vor allem die Stimmästhetik und somit auch die technische Bearbeitung der Vocals stark verändert. Die Erwartungen an Bearbeitungsqualität sind gestiegen – nicht zuletzt durch digitale Produktionsstandards und veränderte Hörgewohnheiten im Streamingzeitalter. Das wirkt sich nicht nur auf die kompositorische und inhaltliche Gestaltung aus, sondern auch auf die gesamte Nachbearbeitung. In der heutigen Popmusik sind Vocals weit mehr als nur ein transportierendes Element für Text und Melodie: Sie fungieren als Trägermedium, das maßgeblich über die Wiedererkennbarkeit, Emotionalität und Marktgängigkeit eines Songs entscheidet [1], [3], [4], [5],[6].

Eine aktuelle Langzeitstudie von Gerdes und Siedenburg (2023) bestätigt die zentrale Rolle von Lead-Vocals in der Popmusikproduktion ebenfalls aus mixingtechnischer Perspektive. Die Analyse von rund 6.000 Songs aus dem Zeitraum 1946 bis 2020 zeigt, dass sich die Lautstärkeverhältnisse zwischen Vocals und Begleitmaterial über die Jahrzehnte stark verändert haben. Während in frühen Produktionen die Vocals noch deutlich über dem Mix lagen, pendelte sich der durchschnittliche Lautstärkeunterschied ab den 1970er-Jahren bei etwa 1–2 dB ein. Jedoch sind in „vokalzentrierten“ Stilrichtungen wie kommerziellem Pop, Country oder Rap, die Vocals heute deutlich prominenter gemischt als etwa in Rock- oder Metalproduktionen. Ein weiterer interessanter Aspekt ist, dass Solo-Künstler tendenziell höhere Vocal-Level aufweisen als Bandproduktionen, was die kommerzielle Fokussierung auf die Stimme als identitätsmerkmal unterstreicht.

Die Studie macht deutlich, dass die klangliche Gestaltung von Vocals über rein technische Aspekte hinausgeht und auch ästhetisch-strategische Überlegungen umfasst [6].

Mit dem Aufkommen digitaler Produktionsumgebungen und leistungsfähiger Plug-ins (Software-Erweiterungen wie z.B Effekt- und Bearbeitungsmodule) haben sich die Möglichkeiten zur Produktion von Vocals erheblich erweitert – sowohl im technischen als auch im kreativen Sinne. Wo früher analoges Recording und damalige zur Verfügung stehende Technik die natürlichen Grenzen des Produktionsstandards setzten, eröffnen Tools und Techniken wie Autotune, Melodyne oder Vocal-Comping heutzutage nahezu grenzenlose Bearbeitungsmöglichkeiten. Die stetig wachsende Rechenleistung und damit verbundene geringere Beschränkung bei der Nutzung zahlreicher Plugins in DAW's (Digital Audio Workstation), begünstigen diesen Umstand ebenfalls. Das Klangideal moderner Popmusik hat sich, unter anderem dadurch, in den letzten Jahrzehnten stark verändert. Während in den frühen Jahren der Popmusik meist die natürliche Stimme und eine gute Aufnahme im Vordergrund standen, ist heute ein stark korrigierter und bearbeiteter, „Larger than life“ Vocal-Sound häufig Standard – insbesondere im Mainstream-Pop [2], [4], [7].

Die Hörgewohnheiten des Publikums haben sich also verändert. Popmusikkonsumierende sind heute an eine hochgradig bearbeitete, oft stilisierte Stimmästhetik gewöhnt und verlangen regelrecht danach. Für Muskschaffende sind die Grenzen zwischen technischer Optimierung und künstlerischer Gestaltung von Vocals dabei zunehmend fließend geworden. Gleichzeitig hat sich mit dem Siegeszug von Streamingdiensten und variablen Hörumgebungen auch die Art und Weise verändert, wie Songs rezipiert werden. Dies hat ebenfalls direkte Auswirkungen auf die Anforderungen an Muskschaffende [3].

Die Bedeutung von Vocals als zentrales Gestaltungselement hat somit in jeglicher Hinsicht deutlich zugenommen. Dies macht eine gezielte Auseinandersetzung mit der Bearbeitungstiefe, dem Produktionsaufwand und der wahrgenommenen Qualität von Gesang in der Popmusik zunehmend relevant. Insbesondere vor dem Hintergrund aktueller Produktionsrealitäten, ökonomischer Zwänge und gängiger stilistischer Konventionen [1], [2], [8].

2.2 Anforderungen bei der Produktion von moderner Popmusik und deren Vocals

Die Erwartungen an moderne Popmusikproduktionen befinden sich wie Popkultur selbst, in einem ständigen Wandel – sowohl aus technischer, als auch aus ästhetischer Sicht. Dennoch gibt es einige Faktoren, von denen die Stimmbearbeitung grundsätzlich abhängt. Doch vorab gilt es zu klären, warum diese überhaupt notwendig ist [1], [2].

Die menschliche Stimme ist ein klanglich sehr individuelles und zugleich wandelbares Instrument in der Musikproduktion. Anders als synthetische Sounds oder elektronische Instrumente unterliegt sie natürlichen Schwankungen in ihrer Tonhöhe, Dynamik, Artikulation und damit natürlich ebenfalls in ihrer Frequenzverteilung. Pisanski et al. (2020) zeigen, dass sich die Grundfrequenz der Stimme (F_0) bei ein und derselben Person – etwa beim Übergang von Sprechen zu Schreien – um ein Vielfaches verändern kann, während die individuelle Charakteristik dennoch erhalten bleibt [9]. Diese Eigenschaften machen die Stimme als Instrument einerseits sehr einzigartig und organisch, andererseits aber auch technisch anspruchsvoll im Kontext einer professionellen Musikproduktion [10], [11], [12].

Nebst der physiologischen Komponente gibt es weitere Einflussfaktoren bei der Produktion wie z.B die Aufnahmekette (siehe Kapitel 2.3), sowie die Raumakustik (Kapitel 2.3.3.1).

Moderne Vocal-Bearbeitung hat in erster Linie das Ziel, Tonale Stabilität, Konsistenz in Dynamik und Klang, und dadurch Sprachverständlichkeit herzustellen, ohne die Charakteristik der Stimme zu verlieren. Um sich im Mix durchzusetzen, muss die Stimme dynamisch kontrolliert und im Frequenzbild aufgeräumt sein. Letzteres spielt vor allem im Kontext der Mischung insgesamt eine wichtige Rolle, da die Stimme frequenztechnisch ihren Platz im Mix hat, weshalb es üblich ist, störende Anteile zu entfernen. Auch eventuelle störende Nebengeräusche, Zischlaute, Resonanzen oder Atemgeräusche können die Verständlichkeit beeinträchtigen. Neben diesen funktionalen Korrekturen spielt auch die stilistische Ausrichtung eine Rolle. Details zur Bearbeitung mit dem Equalizer finden sich in Kapitel 2.4.5. Trotz der Individualität jeder Stimme lassen sich einige allgemeingültige Konventionen identifizieren [11], [12], [13].

Die Bearbeitung der Vocals zielt oft darauf ab, einen oder mehrere der o.g Bereiche zu optimieren. Sie erfolgt dabei typischerweise in mehreren Schritten. Dazu mehr im Kapitel „Grundlagen“.

Neben der Anpassung der Signale selbst, gibt es weitere übergeordnete Produktionsanforderungen. Der gängige qualitative Standard der Bearbeiteten Vocals sollte unabhängig von Wiedergabemedium, Abhörumgebung oder Endgerät gesichert sein, was heutzutage, bei einer großen Vielfalt an elektronischen Wiedergabegeräten und einer ortsunabhängigen Verfügbarkeit von Musikinhalten herausfordernd sein kann. Wie schon Weinzierl (2008) erkannte, muss deshalb oftmals von zunehmend ungünstigen Abhörbedingungen ausgegangen werden [13].

Die Durchsetzungsfähigkeit der Vocals im Mix, sowie der Mischung selbst gegenüber anderen Songs, sollte besonders im Musikstreamingkontext sowie im Kontext sozialer Medien (Spotify, TikTok, YouTube usw.) beachtet werden. Die Bearbeitung muss jedoch nicht ausschließlich technischen Konventionen genügen. In vielen aktuellen poprelevanten Genres wie z.B. Trap, Hyperpop oder Dance-Pop, ist eine stark bearbeitete Stimme heute nicht nur akzeptiert, sondern gehört zur „stilistischen DNA“ derselben [1], [5],[3].

Hinzu kommen ökonomische und strukturelle Anforderungen: Der Produktionsprozess ist zunehmend fragmentiert, demokratisiert und digitalisiert. Vocals werden häufig dezentral aufgenommen – in Home- und Projektstudios, auf Tour oder mit mobilen Setups – und die Postproduktion übernimmt die Aufgabe, diese Aufnahmen auf einen marktkonformen Qualitätsstandard zu bringen.

Gleichzeitig steigt der Output-Druck in einer stark von Algorithmen dominierten Logik, in Bezug auf Produktion und Veröffentlichung. Kurze Aufmerksamkeitsspannen, Playlistenplatzierung und virale Rezeption sind heutzutage wesentliche Faktoren. Auch Produktionszeiträume haben sich im Zuge der Streaming-Ökonomie erheblich verkürzt: Anstelle langfristiger Albumproduktionen orientieren sich viele Künstler heute am Rhythmus algorithmischer Sichtbarkeit und veröffentlichen regelmäßig neue Singles – teils im Abstand weniger Wochen, um kontinuierlich im Feed der Plattformen präsent zu bleiben [14],[8].

Das bedeutet: Es bleibt oft wenig Zeit für aufwendige Detailarbeit, und dennoch ist ein „larger-than-life“, professionell klingender Vocal-Sound Pflicht. Hohe Qualität bei quantitativ hohem Output haben sich zum neuen Standard entwickelt [3], [15], [16].

In diesem Spannungsfeld zwischen ästhetischem Ideal, technischer Möglichkeit und ökonomischer Realität stellt sich die Frage, in welchem

Umfang die Vocal-Bearbeitung für ein von Konsumenten Akzeptiertes Produkt tatsächlich notwendig ist – und ab wann zusätzlicher Aufwand keinen wahrnehmbaren Mehrwert für Musikkonsumenten mehr bringt. Genau an dieser Stelle setzt die vorliegende Arbeit an: Sie untersucht, ob Vocal-Bearbeitung generell als qualitativer Mehrwert in der Mischung wahrgenommen wird, ob ein funktionales Mindestmaß an Bearbeitung ausreicht, um den heutigen Anforderungen gerecht zu werden, oder ob sich eine differenzierte Nachbearbeitung hörbar auszahlt.

2.3 Der Recording-Prozess

Vor jeder Bearbeitung von Vocals steht der Recording-Prozess. Technisch qualitative Aufnahmen sind Grundvoraussetzung, um überhaupt ein bestmögliches Ergebnis in der Nachbearbeitung erzielen zu können.

2.3.1 Technische Komponenten

Neben Know-how und Erfahrung ist eine hochwertige Signalkette für eine qualitative Aufnahme unerlässlich [11], [12], [13]. Zu dieser Kette zählen folgende Komponenten:

MIKROFON

Die Auswahl eines passenden Mikrofones steht beim Aufnahmeprozess an erster Stelle. Mittlerweile bietet der Markt hier eine sehr breite Auswahl, wobei sich verschiedene Bautypen und Hersteller klanglich voneinander unterscheiden [12]. Je nach Genre und Stimme sind somit unterschiedliche Mikrofone geeignet. Auch die Platzierung des Mikrofons sowie des Sängers im Raum beeinflusst die Aufnahme maßgeblich. Dies ist z.B in Bezug auf den Nahbesprechungseffekt der Fall. Mehr zum Thema Raumakustik findet sich in Kapitel 2.3.3.1.

Bei gängigen Mikrofonen im Tonstudio wird grundsätzlich zwischen drei Wandler Prinzipien unterschieden:

- **Elektrostatisch (Kondensatormikrofon):** (Abb.1) arbeitet nach dem Prinzip eines elektrostatischen Wandlers. Dabei fungieren eine dünne, leitfähige Membran und eine feststehende Rückplatte als Kondensator. Zwischen diesen beiden Elementen liegt eine elektrische Spannung an. Trifft Schall auf die Membran, wird sie in Schwingung versetzt, wodurch sich der Abstand zur Rückplatte verändert. Diese Veränderung bewirkt eine Kapazitätsänderung des Kondensators, aus der ein elektrisches Signal erzeugt wird, das dem ursprünglichen Schallverlauf entspricht. Die meisten Kondensatormikrofone erzeugen selbst keine Spannung und benötigen daher eine Phantomspeisung (meist +48 Volt), um betrieben zu werden. Aufgrund ihrer hohen Empfindlichkeit und dem weitestgehend linearen Frequenzgang, bieten diese ein hohes Maß an Klangqualität und sind am besten für ruhige Studioumgebungen geeignet [12].

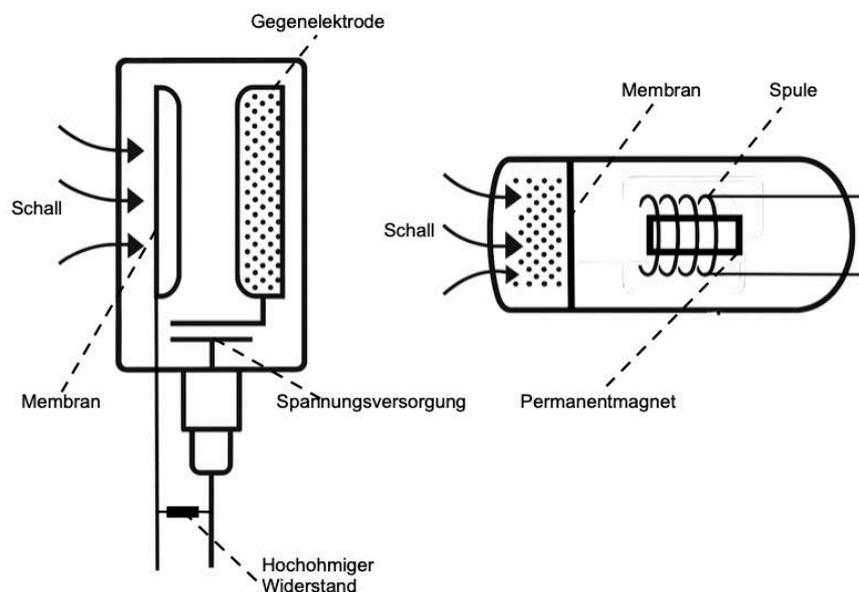


Abbildung 1: schematische Darstellung Kondensatormikrofon (links) und Tauchspulenmikrofon (rechts). (eigene Darstellung)

- **Dynamisch (Tauchspulenmikrofon):** (Abb. 1) arbeitet nach dem elektromagnetischen Prinzip. Im Inneren befindet sich eine Membran, die mit einer kleinen Spule verbunden ist. Diese Spule ist in einem festen Magnetfeld positioniert. Wenn Schallwellen auf die Membran treffen, versetzt der Schalldruck die Membran und damit die Spule in Bewegung. Durch diese Bewegung der Spule im Magnetfeld wird eine elektrische Spannung induziert, die dem Schallverlauf entspricht. Dynamische Mikrofone benötigen keine externe Spannungsversorgung. Sie gelten im Gegensatz zu Elektrostaten als robust, preiswert und eher unempfindlich gegenüber hoher Lautstärke. Neben Studioaufnahmen sind sie besonders geeignet für den Live-Bereich oder die Abnahme lauter Schallquellen wie Schlagzeug oder Gitarrenverstärker [12].
- **Dynamisch: (Bändchenmikrofon):** gehört zur Gruppe der dynamischen Mikrofone, verfolgt jedoch ein anderes Wandlerprinzip. Anstelle einer Tauchspule befindet sich zwischen den Polen eines Magneten ein hauchdünnes, elektrisch leitfähiges Metallbändchen. Dieses Bändchen fungiert gleichzeitig als Membran und als beweglicher Leiter. Durch Eintreffen von Schall auf das Bändchen, beginnt es im Magnetfeld zu schwingen und es wird, wie oben beschreiben, Spannung induziert. Sie zeichnen sich durch einen nahezu linearen Frequenzgang und ein gutes Impulsverhalten aus [12].

VORVERSTÄRKER

Das nächste Glied in der Aufnahmekette stellt der Mikrofon-Vorverstärker dar. Dieser dient dazu das elektrische Ausgangssignal eines Mikrofons auf ein nutzbares Pegelniveau anzuheben. Neben dieser Verstärkung findet auch eine Impedanzwandlung von einem hochohmigen Input zu einem niederohmigen Output statt [13]. Die notwendige Vorverstärkung (Gain)

lässt sich dabei meist individuell einstellen. Der Vorverstärker bringt als aktives Glied in der Signalverarbeitungskette seine eigene Klangcharakteristik mit sich und prägt somit den Klang des Signals wesentlich. Hochwertige Vorverstärker zeichnen sich vor allem durch einen geringen Klirrfaktor und wenig Eigenrauschen aus. Aus technischer Sicht ist eine möglichst hohe Linearität in Bezug auf die Signalverarbeitung wünschenswert. In der Produktionspraxis, insbesondere in der (Pop) Musikproduktion, ist eine gewisse Klangfärbung oftmals wünschenswert [17]. Vorverstärker gibt es in Form eigenständiger Geräte, aber auch in Eingangskanalzügen von Mischpulten. Bei kleineren Setups sind Vorverstärker als all-in-one Lösung heutzutage oftmals in Kombination mit A/D-Wandlern in Audio-Interfaces verbaut.

ANALOG/DIGITAL WANDLER

Das letzte Glied in der Kette stellen Analog/Digital Wandler (A/D-Wandler) dar. Dieser wandelt analoge Audiosignale, wie sie beispielsweise von einem Mikrofon erzeugt werden, in digitale Daten um. Dabei wird das kontinuierliche analoge Signal in regelmäßigen Abständen abgetastet (Sampling) und in diskrete Werte umgerechnet (Quantisierung). Die Qualität der Wandlung hängt dabei unter anderem von der Abtastrate und der Bittiefe ab. A/D-Wandler bestimmen maßgeblich die Auflösung und somit ebenfalls die Klangqualität des aufgenommenen Signals [12].

Mikrofon, Vorverstärker und A/D-Wandler bilden gemeinsam die Grundlage einer hochwertigen Aufnahme. Ihre technischen Eigenschaften und ihr Zusammenspiel entscheiden bereits im frühen Stadium über die klangliche Qualität und den Charakter des aufgenommenen Signals. Das eingesetzte Mikrofon und der verwendete Vorverstärker sind ausschlaggebende Komponenten wenn es darum geht, wie stark eine Vocal-Aufnahme nachbearbeitet werden muss – darauf weist auch Produzent 2 in seinem Interview hin [18].

Neben diesen Komponenten ist zusätzlich diverses Zubehör wie Kabel, Popschutz etc. von Nöten, um ein hochwertiges Ergebnis bei der Aufnahme zu erzielen.

2.3.2 Psychologische Komponenten

Neben den technischen Voraussetzungen ist besonders die gesangliche Performance der Person vor dem Mikrofon ausschlaggebend für das Aufnahmeergebnis. Der erforderliche Zeitaufwand für die nachträgliche Bearbeitung der Vocals hängt maßgeblich von der Qualität dieser Darbietung ab – ein Punkt, den alle befragten Experten im Interview übereinstimmend bestätigen [17], [18], [19], [20], [21], [22]. Daher spielt auch die psychologische Komponente beim Vocal-Recording eine entscheidende Rolle. Die emotionale Verfassung der Sängerin oder des Sängers beeinflusst unmittelbar Intonation, Artikulation und klangliche Ausdrucksstärke. Eine gezielt gestaltete Aufnahmesituation, etwa durch Vertrauen in die mitarbeitenden Personen, eine ruhige Umgebung und eine angenehme Atmosphäre, steigert nachweislich die Qualität der Performance. Bereits kleinste psychologische Eingriffe, wie etwa die Wahl eines vertrauten bzw. optisch passenden Mikrofons oder die Möglichkeit zur freien Körperhaltung, können sich hörbar auf das Ergebnis auswirken. Körperliche Faktoren wie Atmung, Bewegungsfreiheit und Haltung haben ebenfalls (im Sinne der „embodied cognition“) eine Wirkung auf das Singen: Sie beeinflussen nicht nur den Klang, sondern auch das emotionale Erleben der Performance – und somit deren Authentizität und Ausdrucksstärke im finalen Mix [23], [24].

2.3.3 Weitere entscheidende Faktoren

Neben den bereits genannten Faktoren wie Technik und Psychologie gibt es weitere, die das Recording beeinflussen. Aufgrund des Umfangs dieser Themen, können nicht alle im Detail behandelt werden. Jedoch sollen die Folgenden aufgrund ihrer Wichtigkeit hier nicht unerwähnt bleiben.

2.3.3.1 *Raumakustik*

Die Raumakustik spielt eine wesentliche Rolle beim Recording von Vocals. Reflexionen, stehende Wellen oder zu lange Nachhallzeiten im Aufnahmeraum können den Klang der Stimme verfälschen und somit gezielte Platzierung und Durchsetzung im Mix erschweren. In Räumen, die nicht professionell akustisch optimiert sind – z.B. in Home- oder Projektstudios – kann der Einsatz von Absorbern oder mobilen Vocal-Booths helfen. So können störende Signalanteile minimiert, oder zu lange Nachhallzeiten kontrolliert werden. Anders als z.B. bei Klassik- oder Jazzproduktionen wird in der Popmusik oft bewusst so trocken wie möglich aufgenommen, um maximale Kontrolle über Nachhall und Rauminformation in der Postproduktion zu gewährleisten. Eine trockene, direkt klingende Aufnahme bildet die bestmögliche Grundlage für flexible Klanggestaltung in der Nachbearbeitung [12].

2.3.3.2 *Monitoring beim Recording*

Ein präzises und latenzfreies Monitoring ist entscheidend für die stimmliche Performance beim Recording. Die Art, wie Sängerinnen und Sänger sich selbst über Kopfhörer hören, beeinflusst unmittelbar Intonation, Timing und Ausdruck. Ein gut ausbalancierter Monitoring-Mix – häufig mit leicht erhöhter Vocal-Lautstärke – kann das Sicherheitsgefühl beim Singen deutlich steigern, während ein unzureichendes Monitoring oft zu Intonationsproblemen oder Unsicherheiten führt.

Besonders wichtig ist dabei, dass sich die Sänger latenzfrei hören können. An zweiter Stelle steht die Bearbeitung des Monitorsignals mit Effekten wie Reverb, Kompression oder Auto-Tune, um ein Gefühl von „Einbettung“ in den Song und eine klanglich bereits bearbeitete Stimme zu vermitteln. Viele Künstler empfinden dies als deutlich angenehmer als das Hören ihres trockenen Signals.

Technisch lässt sich dies entweder über klassische Hardware-Monitorwege (z. B. Mischpulte) oder über Low-Latency-Monitoring innerhalb der DAW realisieren. Zusätzlich bieten viele moderne Audio-Interfaces die Möglichkeit, über die Funktion „Direct Monitoring“ individuelle Monitormischungen zu erstellen. Bei einigen hochwertigen Audio-Interfaces ist auch die Einbindung von o.g. Effekten zur Vocal-Bearbeitung möglich. Dabei wird das Eingangssignal direkt zum Kopfhörerausgang geführt, ohne dass wahrnehmbare Latenz entsteht. So kann sich die singende Person optimal auf ihre Performance konzentrieren, ohne klangliche Irritationen [12], [13], [25], [26], [27].

2.3.3.3 Gain-Staging

Ein korrektes Gain Staging ist essenziell für saubere Vocalaufnahmen. Ziel ist es, das Eingangssignal so einzustellen, dass es einen ausreichend hohen Pegel aufweist, um ein gutes Signal-to-Noise-Verhältnis zu erreichen, dabei jedoch Verzerrungen oder Clipping vermeidet. In der Praxis liegt der optimale Aufnahmepegel typischerweise zwischen -12 dBFS und -6 dBFS in der DAW, um ausreichend Headroom für Pegelspitzen zu lassen. Bereits auf der Ebene des Preamps erfolgt die erste kritische Pegelanpassung. Ist das Signal zu leise, muss es später stark verstärkt werden, was zu erhöhtem Rauschen führen kann. Ein zu hoher Eingangspegel hingegen kann zu digitaler Übersteuerung führen und das Signal unbrauchbar machen. Ein durchdachter Gain-Staging-Prozess schafft daher die technische Grundlage für alle weiteren Bearbeitungsschritte [28], [29].

2.4 Tools und Techniken zur Vocal-Bearbeitung

Die gezielte Nachbearbeitung von Gesangsaufnahmen gilt als Standard in der modernen Popmusikproduktion und ist neben der eigentlichen Aufnahme entscheidend für die klangliche Qualität der fertigen Vocals im Mix [29]. Dabei sollen die Vocals nicht nur für sich betrachtet aufgewertet werden. Vielmehr sollen sie sich auf den Kontext der gesamten Mischung positiv auswirken und in diesem gut funktionieren. Dabei kommen verschiedene Werkzeuge und Techniken zum Einsatz, um Faktoren wie z.B. Timing, Intonation, Dynamik und Klangcharakter technisch zu optimieren oder stilistisch zu gestalten. Hierbei spielt die Bearbeitung sowohl im Kontext der Mischung als auch des Genres eine wichtige Rolle [11], [17],[29].

Welche Tools in der Praxis tatsächlich als besonders relevant gelten, wurde auch im Rahmen der Experteninterviews deutlich: Alle befragten Musikproduzenten bestätigten die Wichtigkeit bestimmter Bearbeitungsschritte im Umgang mit Vocals. Zu den am häufigsten genannten Werkzeugen und Techniken zählen die Intonationskorrektur (Autotune & Melodyne), Editing, und Vocal-Comping. Außerdem wurden Equalizer, Kompressor und Limiter, De-Esser sowie Saturation genannt [17], [18], [19], [20], [21], [22]. Im Folgenden werden daher die aus Sicht der Praxis grundlegenden Tools und Techniken der Vocal-Bearbeitung näher vorgestellt und eingeordnet.

2.4.1 Das Editieren

Der erste Schritt der Vocal-Bearbeitung besteht meist im grundlegenden Editieren des aufgenommenen Audiomaterials. Ziel dieses Arbeitsschritts ist es, die Aufnahme von störenden oder ungewollten Bestandteilen zu befreien und so eine saubere Arbeitsgrundlage für alle weiteren Bearbeitungsschritte zu schaffen [25].

Dazu gehört zunächst das Entfernen unerwünschter Nebengeräusche, etwa Klicken, Knackser oder Hintergrundgeräuschen in den Gesangspausen. Auch Atmer werden in modernen Popsongs häufig ganz oder teilweise entfernt, vor allem dann, wenn diese durch spätere (oft starke) Kompression zu dominant wirken oder im Mix störend auffallen [29]. In eher puristischen Songs werden sie oftmals bewusst nicht entfernt oder nur leicht in der Lautstärke reduziert, um die Natürlichkeit der Aufnahme zu erhalten.

Ein weiterer optionaler Bestandteil des Editierens ist das manuelle Anpassen des Timings. Hierbei werden einzelne Vocal-Clips - in Form von Worten oder Silben - im Raster verschoben, um rhythmische Ungenauigkeiten zwischen Vocals und Instrumental zu korrigieren. Dies erfolgt in der Regel noch vor einer möglichen automatisierten Quantisierung oder Time-Stretching-Prozessen und trägt zur rhythmischen Präzision der Vocalspur bei [11], [19].

2.4.2 Vocal-Comping

Beim Comping werden mehrere aufgezeichnete Vocal-Takes einer Gesangsperformance zu einem finalen, stimmigen Take zusammengesetzt. Diese Technik ist besonders in der Popmusik weit verbreitet und ermöglicht es, aus verschiedenen Durchläufen die jeweils besten Phrasen, Silben oder sogar einzelnen Wörter auszuwählen. Somit kann ein möglichst rhythmisch präziser, ausdrucksstarker und musikalisch überzeugender Take erzeugt werden, ohne dass der Künstler diesen an einem Stück performen muss.

Comping setzt voraus, dass die Sängerin denselben Abschnitt mehrmals einsingt – oftmals mit dem Ziel, verschiedene Varianten in Ausdruck, Timing oder Intonation zur Auswahl zu haben. In der Praxis werden diese Takes in sogenannten Takes-Lanes oder Playlists untereinander organisiert und anschließend systematisch durchgehört. Dabei wird für jede Stelle entschieden, welche Version performativ am besten passt.

Abbildung 2 zeigt Take-Lanes einer Gesangsaufnahme in der DAW Ableton Live. Hier wurden für dieselbe Stelle im Song fünf Takes aufgezeichnet, aus welchen sich nun der „ideale Take“ zusammensetzen lässt.

Vocal-Comping wird selbst bei internationalen Chartproduktionen und bekannten Künstlern standardmäßig eingesetzt. Auch wenn diese über ein hohes gesangliches Niveau verfügen, ist Comping keineswegs überflüssig – vielmehr trägt es dazu bei, die bestmögliche Performance aus mehreren starken Durchläufen herauszuarbeiten. Der Produzent Ken Lewis, der unter anderem für Usher und Lenny Kravitz gearbeitet hat, betont in einem Interview, dass Vocal-Comping ein essenzieller Bestandteil professioneller Popproduktionen ist, da es selbst herausragende Takes noch optimieren kann [30].

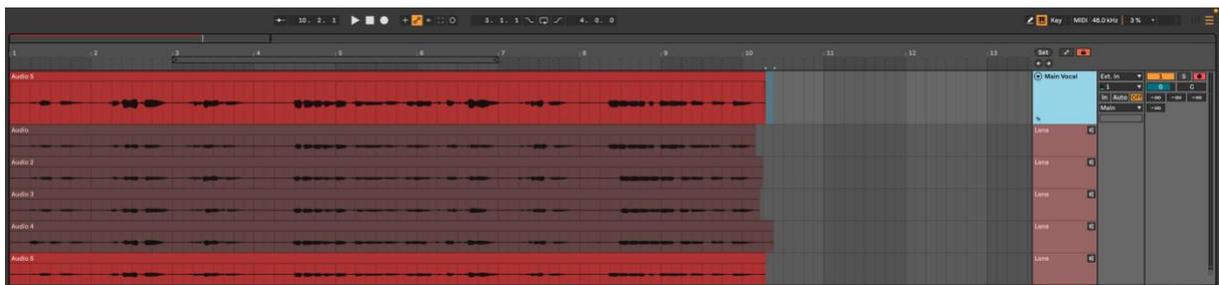


Abbildung 2: Benutzeroberfläche mit Take-Lanes in Ableton Live

Ein gutes Vocal-Comping zeichnet sich nicht nur durch technische Sauberkeit aus, sondern ebenfalls durch musikalisches Feingefühl: Die Übergänge zwischen einzelnen Takes sollten möglichst unauffällig und flüssig wirken. Oft sind zusätzliche kleine Schnittkorrekturen, Crossfades oder Lautstärkeanpassungen notwendig, um eine natürlich und durchgehend wirkende Performance zu erzeugen. Aus diesen Gründen geht Comping oft mit dem Editieren der Vocals einher oder wird bereits davor erledigt [11], [31].

2.4.3 Vocal-Layering

Eine weitere gestalterische Technik in der Vocal-Produktion ist das gezielte Layering von Gesangsspuren. Dabei werden mehrere Gesangsstimmen übereinandergeschichtet, um den Klang voller, breiter oder ausdrucksstärker zu gestalten. Diese können entweder identisch eingesungen oder bewusst variiert werden. Das Layering von Vocals ist insbesondere in der modernen Popproduktion weit verbreitet und kann sowohl funktionalen Zwecken (z.B. mehr Durchsetzungsfähigkeit der Stimme in bestimmten Parts wie dem Refrain, Anreicherung von „dünnen“ Stimmen) als auch ästhetischen Zielen (z.B. Verbreiterung der Stimme um diese Mächtiger wirken zu lassen, Harmoniestimmen als Chor) dienen [11], [29].

Zu den trivialsten Formen des Vocal Layerings gehören Unisono-Dopplungen, bei denen die Hauptstimme ein zweites oder drittes Mal exakt gleich eingesungen wird, um sie im Mix zu verstärken und räumlich zu verbreitern [11]. Dabei bleibt der gesungene Inhalt identisch, jedoch unterscheiden sich die Signale technisch durch minimale Abweichungen in Timing, Intonation und Dynamik, wodurch es zu leichten Schwebungen kommt, die den Klang dichter, lebendiger und Breiter wirken lassen können. Das wird von Hörenden oft als angenehm empfunden [32]. Beim Prinzip der Harmoniestimmen handelt es sich um zusätzliche Gesangsspuren, die in festgelegten Intervallen (z.B. Terzen, Quinten oder Oktaven) zur Hauptstimme verlaufen und sich tonal in das harmonische Gerüst des Songs einfügen. Einfach umzusetzen sind z.B. Oktav-Layer, bei denen eine zusätzliche Stimme eine Oktave höher oder tiefer eingesungen wird. Darüber hinaus kommen in bestimmten Stilrichtungen geflüsterte Stimmen (Whisper-Tracks), gesprochene Layer oder bewusst raue Alternativtakes zum Einsatz, um Akzente zu setzen [11], [29].

Technisch sauberes Layering erfordert Präzision im Timing und in der Intonation, um Phasenprobleme, Unsauberkeiten oder ungewollte Doppelschläge zu vermeiden. Die einzelnen Layer werden häufig separat bearbeitet, beispielsweise durch gezieltes Panning, unterschiedliche EQ-Einstellungen oder dezente Sättigung, um eine klangliche Variation zu schaffen. Im Hinblick auf Sprachverständlichkeit und Ästhetik sollten die Layer-Vocals stets im Gesamtkontext betrachtet werden, da zu viele überlagerte Spuren zu verwaschenen Vocals oder einem überladenen Mix führen können.

Layering ist nicht nur eine kreative Entscheidung, sondern bewusster Teil des technischen Bearbeitungsprozesses, der eng mit Editing, Comping und Mixing verbunden ist [33].

2.4.4 Intonationskorrektur

Die Korrektur der Tonhöhe ist bei der Vocal-Bearbeitung in der modernen Popmusikproduktion heutzutage nicht mehr wegzudenken. Selbst bei technisch sauberen Gesangsaufnahmen kommt es oft zu kleineren Intonationsabweichungen, die im Kontext einer hochwertigen Produktion in der Mischung auffallen. Besonders da sich das musikalische Hören der Menschen über die letzten Jahrzehnte stark an die Ästhetik dieser Korrekturen gewöhnt hat, kann ein Fehlen derselben bei modernen Produktionen stark auffallen [34]. Um die Intonation gezielt zu korrigieren, stehen verschiedene Werkzeuge zur Verfügung – von vollautomatischen bis hin zu manuellen Lösungen [35], [25], [29].

Zu den gängigsten Tools für die automatisierte Tonhöhenkorrektur zählt AutoTune des Marktführers Antares. Hierbei wird die Stimme automatisch auf die nächstliegende Tonhöhe innerhalb einer vorgegebenen Tonart gezogen. Die Stärke der Korrektur kann über Parameter wie „Retune Speed“ oder „Humanize“ beeinflusst werden.

In vielen modernen Popproduktionen wird AutoTune nicht nur zur unauffälligen Intonationskorrektur eingesetzt, sondern oft auch bewusst als stark hörbares Stilmittel [11]. Bekannt wurde diese Technik insbesondere durch den Song „Belive“ der Künstlerin Cher aus dem Jahre 1998 [36].

Für eine detaillierte und flexiblere Bearbeitung kommt hingegen Software wie z.B. Celemony Melodyne zum Einsatz. Melodyne analysiert das Audiomaterial und stellt die Tonhöheninformationen grafisch dar, sodass einzelne Töne, Phrasen oder sogar Vibrato-Verläufe exakt händisch angepasst werden können. Diese manuelle Korrektur erlaubt nicht nur eine besonders natürlich wirkende Beeinflussung der Tonhöhe, sondern auch die gezielte Entscheidung, was korrigiert wird und was bewusst erhalten bleibt [10].

In der Praxis werden beide Ansätze häufig kombiniert [21]. Eine schnelle automatische Grundkorrektur kann durch gezielte manuelle Eingriffe ergänzt werden. Hierbei spielt nicht zuletzt auch der Faktor Zeit eine wichtige Rolle. Während Autotune nur weniger Handgriffe bedarf, kann manuelle Intonationskorrektur sehr zeitaufwendig sein, wenn sie gewissenhaft durchgeführt wird [36], [37]. Dies wurde auch beim Erstellen des in Kapitel 3.2.1 erwähnten Demomaterials berücksichtigt. In der kurzen Bearbeitungszeit von 30 min. kam lediglich eine schnelle, automatische Intonationskorrektur zum Einsatz, während in der längeren Bearbeitungsphase ausreichend Zeit für eine präzisere manuelle Korrektur mit Melodyne zur Verfügung stand.

2.4.5 Equalizer

Der Equalizer (EQ) ist eines der zentralen Werkzeuge in der Vocal-Bearbeitung und gehört zu den Standardprozessoren auf jeder Spur im Mix. Mit ihm wird das Frequenzspektrum der Stimme gezielt geformt, um eventuelle frequenztechnische Probleme zu beheben, die

Sprachverständlichkeit zu erhöhen und die Stimme klanglich in den Gesamtkontext des Songs einzubetten.

Die Anwendung eines oder mehrerer EQs erfolgt in der Regel in zwei aufeinanderfolgenden Schritten. Die technische Korrektur und die klangästhetische Gestaltung [38].

Technische Korrektur

Zunächst dient der EQ dazu, unerwünschte Frequenzanteile zu entfernen oder zu reduzieren. Dazu zählt etwa das Absenken tiefer Frequenzen unterhalb von 80–120 Hz, welche meist keine musikalisch relevante Information der Stimme enthalten. Diese tiefen Frequenzen werden in der Regel mit einem High-Pass-Filter entfernt [39]. Hierbei sollte darauf geachtet werden, die Flankensteilheit nicht zu extrem zu wählen, da sonst ungewünschte Artefakte im Zeitbereich (ringing) entstehen können, was die Transientenwiedergabe erheblich beeinträchtigen kann. Aufgrund dessen empfiehlt sich, eher mit moderaten Flankensteilheiten von z.B. -12, -18 dB/Okt zu arbeiten [40].

Außerdem können störende Resonanzen, ungewollte nasale Mittenfrequenzen oder zu harsche Höhen mit einem schmalbandigen Bell-Filter gezielt abgesenkt werden. Eine genauere Betrachtung der Frequenzbereiche folgt [41], [13], [29].

Klangästhetische Anpassung

Nach der technischen Korrektur folgt in der Regel die gestalterische Klangformung. Dies kann in der Praxis optional mit einem zweiten Equalizer geschehen, welcher in Serie zum ersten geschaltet wird. Hier wird die Stimme so angepasst, dass sie genügend Durchsetzungskraft im Mix bekommt, jedoch ohne unnatürlich oder zu scharf zu wirken.

Eine moderate Anhebung kann helfen, die Präsenz und Sprachverständlichkeit zu verbessern, oder der Stimme Brillanz zu verleihen, ohne sie zwingend lauter erscheinen zu lassen. Es ist auch möglich, diese Effekte in übertriebenem Ausmaß einzusetzen und abschließend mit Werkzeugen wie dem De-Esser dagegen zu steuern. Somit kann eine möglichst brillant klingende Vocal erzeugt werden, welche jedoch keine stark störenden Höhenanteile aufweist. Technisch gesehen erscheint das wenig sinnvoll, in der Praxis führt es jedoch oftmals zum gewünschten Ergebnis. Etablierte Tools wie z.B. der in der Musikproduktion weit verbreitete Pultec EQP-1A (sowie dessen Emulationen und Nachbauten) arbeiten nach genau diesem Prinzip.

Die Eingriffe, welche mit einem EQ vorgenommen werden, hängen stark vom Klangcharakter der Originalaufnahme, vom Genre und vom gewünschten Klang des Endergebnisses ab. In der modernen Popmusikproduktion ist ein eher vordergründiger, brillanter und durchsetzungsfähiger Vocal-Sound üblich [17], [38], [11], [13], [29].

Ein detaillierter Überblick über die Frequenzaufteilung einer menschlichen Stimme bzw. die in der Musikproduktion relevanten Bereiche zeigt Abb. 3. Die markierten Frequenzbereiche sind als Richtwerte zu sehen und können je nach Signal variieren. Die in der Abbildung verwendeten Bezeichnungen für die Bereiche unterscheiden sich in der Praxis ebenfalls stark. Es handelt sich dabei um eine schematische Darstellung, um das grundlegende Prinzip zu veranschaulichen. Im unteren Bereich von ca. 20-100 Hz befinden sich die „Rumble“ Frequenzen. Sie sind tief, oft störend und enthalten für den Anwendungsfall der Popmusikproduktion meist keine relevanten Signalanteile. Darüber befindet sich die „Body Region“, zwischen ca. 150-300 Hz. Hier befindet sich der „Körper“ bzw. die „Wärme“ der Stimme, weshalb starke Absenkungen in diesem Bereich schnell zu einer sehr dünnen Stimme und damit Zerstörung des originalen Character führen kann.

In der „Boxy Region“ lässt sich die Durchsetzung in den tiefen Mitten steuern. Zu viel lässt die Stimme „Boxy“ wirken, bei zu wenig Signalanteilen in diesem Bereich wirkt die Stimme oft hohl. In der Midrange von 1 bis 5 kHz, vor allem im eher oberen Bereich, liegt meist die Präsenz der Stimme. Eine Anhebung kann hier zu besserer Durchsetzungsfähigkeit im Mix, sowie zu einer Verbesserung der Sprachverständlichkeit beitragen. Besonders dieser Bereich sollte mit anderen Instrumenten (z.B Gitarren, Synthesizer usw.) abgestimmt werden, da diese Frequenztechnisch hier oft miteinander konkurrieren. In der „Sibilance Region“ von ca. 5,5 bis 9 kHz befinden sich die Sibilanten, also S-,Z- und Zischlaute. Diese werden von Hörenden oft als „scharf“ und unangenehm empfunden, wenn dieser Frequenzbereich im Signal eine zu hohe Amplitude aufweist [42]. Hier kann mit einem Equalizer oder aber auch mit einem De-Esser (Frequenzabhängiger Kompressor) gegengesteuert werden.

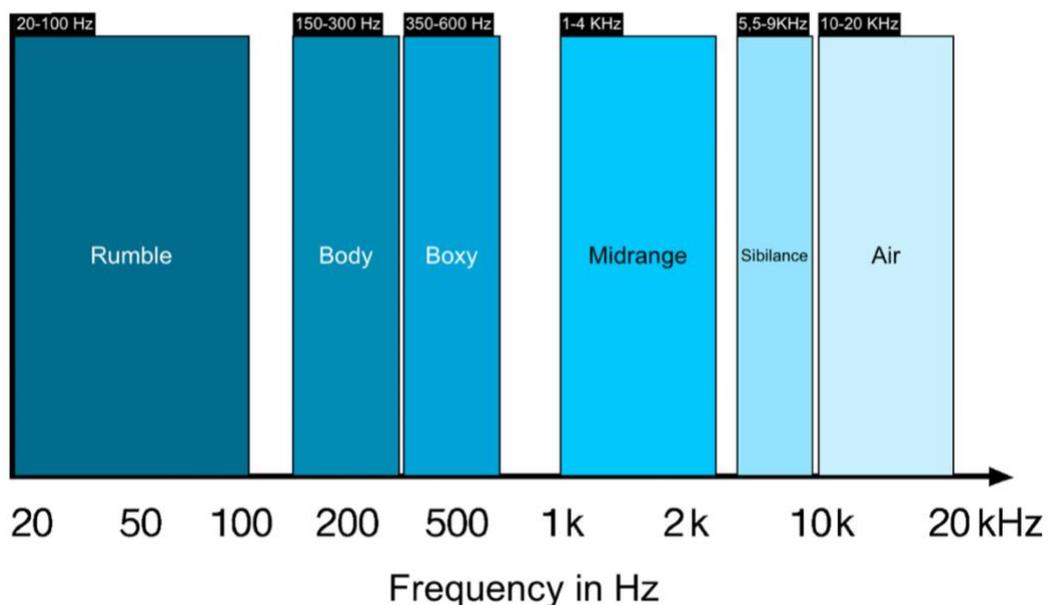


Abbildung 3: Frequenzaufteilung der Menschlichen Stimme in für die Musikproduktion relevante Bereiche (schematisch, eigene Darstellung)

Weitere Erläuterungen zu allen Tools und Bearbeitungstechniken finden sich weiter unten in diesem Kapitel. Im obersten Frequenzbereich von ca. 10 bis 20 kHz befindet sich die „Air Region“. Eine Anhebung dieser Frequenzen kann der Stimme „Luftigkeit“ oder „Brillanz“ verleihen [43].

Wichtig bei der Anwendung des EQs und auch anderer Tools, ist die Beurteilung des bearbeiteten Signals im Kontext der gesamten Mischung. Eine Vocalspur, die solo gut klingt, kann im Zusammenspiel mit Instrumenten andere Eingriffe wie z.B. stärkere Absenkungen in überlappenden Frequenzbereichen oder gezielte Anhebungen für mehr Durchsetzungskraft benötigen. Das Arbeiten mit dem EQ erfordert daher oftmals wiederholtes Hören und Anpassen im Produktionsverlauf [11], [12], [29].

Verschiedene Typen von Equalizern

In der Vocal-Bearbeitung können je nach Anwendungsfall verschiedene Typen von Equalizern zum Einsatz kommen, die sich in ihrer Funktionsweise unterscheiden. Die Wahl des richtigen Equalizer-Typs ist vom Eingangssignal und den klanglichen Anforderungen des Ergebnisses abhängig.

Statisch-parametrische Equalizer sind der Standard in der Produktion. Sie arbeiten mit eingestellten Frequenzbändern, die unabhängig vom Eingangssignal je nach Einstellung dauerhaft an- oder abgesenkt werden. Sie eignen sich besonders für grundlegende Korrekturen [12].

Ein weiteres hilfreiches Tool stellen *dynamische Equalizer* dar. Sie reagieren auf das Eingangssignal und passen die Verstärkung eines Frequenzbands nur dann an, wenn ein bestimmter eingestellter Schwellenwert – der sogenannte Threshold - über- oder unterschritten wird. Sie verbinden die Funktionen eines Equalizers mit der Wirkweise eines Kompressors und sind besonders nützlich bei problematischen Frequenzen, die nur zeitweise auftreten [44].

Aufgrund der Vollständigkeit soll der *Linear-Phase-Equalizer* nicht unerwähnt bleiben. Dieser arbeitet mit einer Phasenkorrektur, bei der alle Frequenzen zeitlich gleichbehandelt werden. Dadurch entstehen keine Phasenverschiebungen im bearbeiteten Signal, was besonders bei parallel bearbeiteten Spuren oder beim präzisen Mastering hilfreich sein kann. Allerdings verursachen Linear-Phase-EQs aufgrund ihrer Lookahead-Zeit eine Latenz und auch Pre-Ringing, weshalb sie für die Vocal-Bearbeitung weniger geeignet sind [40].

2.4.6 Kompressor

Die dynamische Bearbeitung durch Kompressoren zählt neben dem einsetzen des EQs ebenfalls zu den grundlegenden Arbeitsschritten in der Vocalproduktion. Ziel ist es, Pegelschwankungen innerhalb der

Gesangsaufnahme zu kontrollieren und damit die Stimme im Gesamtmix verständlich und präsent zu positionieren. Darüber hinaus kann Kompression auch gestalterisch eingesetzt werden, etwa zur Erzeugung von mehr Energie oder klanglicher Dichte [29].

Ein Kompressor reduziert den Pegel eines Signals, sobald dieser einen festgelegten Schwellenwert überschreitet. Die Stärke der Reduktion wird über die Ratio definiert, während Attack- und Release-Zeiten regeln, wie schnell der Kompressor auf Pegelveränderungen reagiert. Gerade bei Vocals ist die Abstimmung dieser Parameter entscheidend, da sie maßgeblich beeinflussen, wie natürlich das bearbeitete Ergebnis wahrgenommen wird. So kann eine zu kurze Attack-Zeit Transienten abschwächen, während eine zu lange Release-Zeit zu unerwünschtem Pumpen führen kann [12].

Gängige Bautypen von Kompressoren sind:

- **VCA-Kompressoren** (Voltage Controlled Amplifier) sind präzise, schnell und vielseitig einsetzbar. Sie eignen sich besonders gut für kontrollierte, transparente Kompression bei Pop-Vocals und Sprachaufnahmen. Beispiel: UA dbx 160, SSL Bus Compressor.
- **FET-Kompressoren** (Field Effect Transistor) reagieren sehr schnell und erzeugen eine leichte klangliche Färbung. Sie sind gut geeignet, um schnelle Pegelspitzen abzufangen und werden daher auch in Kombination mit trägeren Typen wie z.B. Opto-Kompressoren verwendet. Beispiel: Universal Audio 1176.
- **Opto-Kompressoren** (Optisch gesteuerte Kompression) arbeiten langsamer und weicher. Sie reagieren eher musikalisch auf Pegelschwankungen und erzeugen oft eine angenehme, natürliche Kompression, die sich besonders für expressive Pop-Vocals eignet. Beispiel: Teletronix LA-2A.
- **Vari-Mu-Kompressoren** (Röhrenkompressoren mit variabler Verstärkung) zeichnen sich durch einen eher weichen regelverlauf und

eine dadurch musikalisch angenehme Kompression aus. Sie sind eher subtil und daher gut einsetzbar für Balladen oder Genres, bei denen natürliche Klangfarbe im Vordergrund steht. Beispiel: Fairchild 670, Manley Vari-Mu [45].

Neben der klassischen Insert-Kompression wird in der Vocal-Bearbeitung auch parallele Kompression verwendet. Dabei wird eine stark komprimierte Version der Stimme dem Originalsignal dezent beigemischt. Auf diese Weise kann der Gesamteindruck an Lautheit und Dichte gewinnen, ohne die Natürlichkeit der Main Vocal zu verlieren. Ebenso gängig ist, wie o.g der serielle Einsatz mehrerer Kompressoren mit unterschiedlichen Charakteristiken, etwa zur Kombination schneller Peak-Kontrolle mit nachfolgender trägerer Pegelglättung [46].

Ähnlich wie beim Equalizer ist die Wahl eines Kompressors stets im Kontext des Genres, der aufgenommenen Stimme sowie des Produktionsziels zu betrachten. Kompression ist dabei längst nicht mehr nur als technische Maßnahme zu verstehen, sondern als integrativer Bestandteil der klangästhetischen Gestaltung des Vocal-Mix [47].

2.4.7 Limiter

Als Spezialform des Kompressors dient der Limiter in der Vocal-Bearbeitung in erster Linie der harten Begrenzung von Pegelspitzen. Im Unterschied zum klassischen Kompressor arbeitet ein Limiter mit einer sehr hohen Ratio, in der Regel $\infty:1$. Er greift sofort ein, sobald das Signal einen festgelegten Schwellenwert überschreitet. Damit verhindert er zuverlässig Übersteuerungen und schützt nachfolgende Signalverarbeitungsschritte oder den Summenbus vor Clipping [13].

In der Praxis werden Limiter häufig am Ende der Vocal-Signalkette eingesetzt, um letzte Pegelspitzen abzufangen, die trotz vorheriger Kompression noch auftreten können. Besonders in dicht arrangierten

Produktionen können Limiter helfen, die Stimme konstant präsent zu halten, ohne dass einzelne Silben oder Worte im Mix untergehen. Der Limiter greift dabei meist nur sehr kurzzeitig und in geringem Maß ein, um die Natürlichkeit der Stimme nicht zu beeinträchtigen [48].

Während Kompressoren oft auch gestalterisch eingesetzt werden, erfüllt der Limiter überwiegend eine eher technisch-funktionale Aufgabe im Rahmen der Dynamikbearbeitung. Dennoch kann auch hier eine klangliche Färbung auftreten, weshalb man ihn je nach Anwendungsfall bewusst einsetzen sollte [49].

2.4.8 De-Esser

Mit einem De-Esser lassen sich gezielt Sibilanten (Zischlaute) in Gesangsaufnahmen, insbesondere im Bereich der Laute „S“, „Z“ und „Sch“ reduzieren. Diese liegen meist im Frequenzbereich zwischen 4 kHz und 10 kHz. Sie können überbetont auftreten und als störend empfunden werden. Wie stark diese auftreten bzw. wie sehr sie reduziert werden sollten, ist abhängig von Mikrofon, Aufnahmekette, Eingangssignal und der Bearbeitungskette.

Ein De-Esser arbeitet ähnlich wie ein Kompressor, welcher nur in einem bestimmten Frequenzbereich agiert. Sobald der Pegel im definierten Frequenzbereich einen bestimmten Schwellenwert überschreitet, wird das Signal kurzzeitig abgesenkt. Die Wirkung beschränkt sich dabei auf die Sibilanten, während der übrige Frequenzbereich unbeeinflusst bleibt. Ziel ist es, scharfe Höhenanteile im Signal zu kontrollieren, ohne die Brillanz oder Präsenz der Stimme insgesamt zu verlieren.

Je nach Software oder Hardware können verschiedene Einstellungen vorgenommen werden, darunter Frequenzbereich, Schwellenwert, Stärke der Absenkung und Art des Filters (Shelf oder Bandpass) [13].

Ein korrekt eingesetzter De-Esser wirkt subtil und unauffällig. Eine übermäßige Anwendung kann schnell zu einem lispelnden und dumpfen Klang der Stimme führen, weshalb das Tool mit Bedacht eingesetzt werden sollte. Auch die Kombination mehrerer eher dezenter De-Essing-Schritte, beispielsweise einmal gezielt auf 5 kHz und ein weiteres Mal auf 8 kHz, kann zum gewünschten Ergebnis führen [50].

2.4.9 Saturation Tools

Saturation-Tools dienen der gezielten Anreicherung eines Audiosignals mit harmonischen Obertönen und gehören zu den klanggestalterischen Werkzeugen in der Vocal-Bearbeitung [12]. Im Gegensatz zu rein technischen Werkzeugen wie z.B Kompression zielt Sättigung primär auf die klangliche Verdichtung, Wärme und Durchsetzungskraft ab, die in modernen Pop-Produktionen häufig erwünscht ist. Technisch basiert Sättigung auf dem Prinzip der nichtlinearen Verzerrung, wie sie ursprünglich z.B in analogen Röhrenverstärkern oder Tonbandgeräten auftritt. Diese analoge Charakteristik wird in digitalen Plugins heute gezielt nachgebildet. Je nach Modell und vorgenommener Einstellung kann die Sättigung subtil oder deutlich hörbar und aggressiv ausfallen.

In der Vocal-Bearbeitung wird Sättigung häufig eingesetzt, um die Stimme voller und präsenter wirken zu lassen, ohne sie ausschließlich durch Lautstärkeanhebung oder Equalizing zu betonen. Besonders da heutzutage aufgrund der fortgeschrittenen Technologie die Aufnahmen oft sehr sauber sind, kann eine moderate Sättigung dazu beitragen, der Stimme Charakter zu verleihen.

Darüber hinaus kann Sättigung auch dabei helfen Vocals besser in eine Mischung zu integrieren und sie mit dem Instrumental verschmelzen zu lassen [29].

In der Praxis sind je nach Tool verschiedenste Arten von Sättigung umsetzbar. Dezentere Bandsättigung, röhrenähnliche Sättigung bis hin zu Multiband-Saturation, bei der nur bestimmte Teile des Frequenzspektrums gezielt angesprochen werden. Sättigung wird ebenfalls nicht nur als Insert verwendet, sondern auch subtil parallel beigemischt. Wie bei allen klangverändernden Prozessen gilt auch hier: Der Einsatz von Saturation erfordert Feingefühl. Eine zu starke oder unkontrollierte Anwendung führt schnell zu einem überladenen oder verzerrten Klangbild, das die Verständlichkeit und Natürlichkeit der Stimme beeinträchtigt [51].

2.5 Das Pareto-Prinzip

Im Zusammenhang mit der Frage, wie viel Aufwand in Relation zum Ergebnis sinnvoll ist, wird in dieser Arbeit das Pareto-Prinzip als Grundlage herangezogen. Es beschreibt ein grundsätzliches Ungleichgewicht zwischen Input und Output und liefert damit einen hilfreichen theoretischen Rahmen, um Produktionsprozesse hinsichtlich ihrer Effizienz zu bewerten.

Das Pareto-Prinzip, auch bekannt als 80/20-Regel, geht auf den italienischen Ökonomen Vilfredo Pareto zurück. Er stellte Ende des 19. Jahrhunderts bei einer Untersuchung von Vermögensverteilungen in Italien fest, dass ein Großteil des Besitzes von Grund und Boden in Italien auf einen vergleichsweise kleinen Teil der Bevölkerung entfiel [52]. Diese Beobachtung wurde später von Joseph M. Juran aufgegriffen. Dieser erkannte, wie universell sich das Prinzip auf wirtschaftliche und organisatorische Prozesse übertragen ließ. Konkret beschreibt er mit dem Prinzip, dass häufig rund 80 Prozent eines Ergebnisses mit lediglich 20 Prozent des Gesamtaufwands erzielt werden [53].

Die Bedeutung in der Praxis: Ein kleiner Anteil an eingesetzten Ressourcen, Maßnahmen oder Ursachen ist für den größten Teil des Resultats verantwortlich. Umgekehrt ist der verbleibende Anteil des Ergebnisses – also die letzten Prozente hin zur „Perfektion“ – meist nur durch einen überproportional hohen Mehraufwand erreichbar. Das Zahlenverhältnis 80/20 ist dabei symbolisch zu verstehen und kann je nach Kontext variieren. Entscheidend ist die zugrundeliegende Idee eines Effizienzverhältnisses, das in vielen Bereichen beobachtet werden kann. So zum Beispiel im Projektmanagement, der Qualitätskontrolle, aber auch in kreativen Prozessen wie der Musikproduktion. Gerade in kreativen und produktionsorientierten Arbeitsfeldern stellt sich die Frage, bis zu welchem Punkt zusätzlicher Aufwand in der Produktion tatsächlich zu einem wahrnehmbaren Mehrwert für den Endkonsumenten führt.

Ein Ziel der Arbeit ist es, den Sweet-Spot zu identifizieren, an dem die wahrgenommene Qualität im Gesamtmix durch weitere Bearbeitung nur noch geringfügig steigt. Hier bietet das Pareto-Prinzip ein hilfreiches Modell zur Bewertung von Effizienz, Arbeitsaufwand und resultierender Wirkung [52], [54].

2.5.1 Der Bezug zur Vocal-Bearbeitung

Auch im Bereich der Musikproduktion lässt sich das Pareto-Prinzip auf konkrete Produktionsabläufe anwenden – hier im Hinblick auf die Bearbeitung von Gesangsspuren. In vielen Fällen lassen sich durch wenige, zielgerichtete Eingriffe, wie z.B. schnelles Comping, Tuning und einfache EQ und Dynamikbearbeitung, bereits stark wahrnehmbare klangliche Veränderungen im Kontext der Gesamtmischung erzielen. Es stellt sich die Frage ob der zusätzliche Zeitaufwand für feine manuelle Korrekturen wie etwa detaillierte Intonationskorrektur, kleinteiliges Editieren, Automationen oder Timing-Korrekturen und Layering zwangsläufig zu einer proportional wahrnehmbaren Veränderung bzw. Qualitätssteigerung führen.

Im Rahmen des durchgeführten Hörvergleichs wurde untersucht wie sich Gesangsaufnahmen, die in verschiedenen zeitlichen Rahmen bearbeitet wurden, in Bezug auf die subjektive Wahrnehmung von diversen Probanden unterscheiden. Damit soll geklärt werden, ob der Großteil der subjektiven klanglichen Qualität mit einem vergleichsweise geringen Bearbeitungsaufwand erreichbar ist. Im späteren Verlauf der Arbeit wird erörtert, ob sich diese mit dem Pareto-Prinzip in Verbindung bringen lassen. Gerade in Produktionskontexten mit begrenztem Zeitbudget kann diese Erkenntnis eine Orientierung bieten. Nicht jeder Bearbeitungsschritt rechtfertigt zwangsläufig seinen Aufwand, wenn er in Relation zur wahrgenommenen Qualität betrachtet wird. Die Anwendung des Pareto-Prinzips in der Vocal-Produktion eröffnet somit ein Reflexionsfeld, das sowohl wirtschaftliche als auch gestalterische Relevanz besitzt.

3 Methodik

Die vorliegende Arbeit verfolgt einen qualitativen Forschungsansatz. Bei Experteninterviews mit Produzenten steht die Erhebung subjektiver Einschätzungen im Mittelpunkt. Im anschließend durchgeführten Hörvergleich steht die auditive Wahrnehmung im Zentrum der Untersuchung. Qualitative Forschung eignet sich insbesondere für Fragestellungen, bei denen es weniger um quantifizierbare Ergebnisse, sondern vielmehr um individuelle Meinungen, Erfahrungen und Kontexte geht. Durch diese qualitative Herangehensweise können sowohl praktisch erprobte Produktionsstandards als auch auditive Wahrnehmungen verschiedener Rezipienten aus unterschiedlichen Blickwinkeln betrachtet und ausgewertet werden [55].

3.1 Forschungsdesign: Qualitative Datenerhebung durch Experteninterviews

Um aktuell übliche Vorgehensweisen in Bezug auf Vocal-Produktion in der Popmusik-Branche zu evaluieren, wurden im Vorfeld der Arbeit Experteninterviews durchgeführt. Der qualitative Ansatz ermöglicht es, nicht nur technische Aspekte (z. B. verwendete Effekte, Signalverarbeitung), sondern auch subjektive Einschätzungen und individuelle Herangehensweisen zu erfassen, die sich nicht durch quantitative Methoden abbilden lassen.

In der Analyse wird darauf geachtet, ob es gemeinsame Muster gibt oder ob die Vorgehensweisen variieren. Die Methodik orientiert sich an einem halbstrukturierten Interview-Ansatz, bei dem die Experten zwar durch Leitfragen gelenkt werden, aber auch die Möglichkeit haben, frei über ihre Arbeitsweise zu sprechen. Dies erlaubt eine detaillierte und praxisnahe Erhebung der relevanten Informationen.

3.1.1 Aufbau der Experteninterviews: Evaluation des Mindestmaßes an Bearbeitung

Die halbstrukturierten Experteninterviews wurden mit dem übergeordneten Ziel durchgeführt, vorab die Frage zu klären, wie sich ein Mindestmaß in der Vocal-Bearbeitung darstellt. Untergeordnetes Ziel war es, unterschiedliche Herangehensweisen und persönliche Einschätzungen zu erfassen. Dadurch konnten tiefgehende Einblicke in den Produktionsprozess von Vocals sowie in die bevorzugten Workflows der Produzenten gewonnen werden.

Die Auswahl der Interviewpartner erfolgte nach gezielten Kriterien, um möglichst aussagekräftige und praxisnahe Daten zu generieren. Berücksichtigt wurden ausschließlich Musikproduzenten, die aktiv in der Branche tätig sind und über nachweisbare regelmäßige Veröffentlichungen und Erfolge verfügen. Zudem lag der Fokus auf Personen, die sich regelmäßig mit der Bearbeitung von Vocals befassen und entsprechende Expertise in diesem Bereich aufweisen. Um eine breite Perspektive auf die Thematik zu erhalten, wurden Produzenten aus unterschiedlichen Musikgenres wie etwa Rock, Indiepop, Austropop und Deutschpop mit einbezogen. Die Produktionen der befragten Personen lassen sich trotz des Einflusses verschiedener Genres alle im Bereich der modernen Popmusik einordnen. Zur besseren Übersicht sind alle Interviewten Produzenten im Folgenden tabellarisch aufgelistet.

| Kürzel | Rolle | Genre | Erfahrung | Relevanz |
|---------------|----------------------------------|--------------------|------------------|--|
| Produzent 1 | Produzent, DJ | Dance, Hyperpop | Ca. 8 Jahre | Schwerpunkt Vocals |
| Produzent 2 | Produzent, Mixing Engineer | Pop, Rock | >30 Jahre | Viel Vocal Mixing, Gold, Platin, Grammy |

| | | | | |
|-------------|----------------------------------|-----------|-----------|--|
| Produzent 3 | Produzent | Pop | >10 Jahre | Schwerpunkt Vocals, Gold Platin |
| Produzent 4 | Produzent, Mixing Engineer | Pop | <15 Jahre | Viel Vocals, Gold, Platin |
| Produzent 5 | Produzent | Pop | <10 Jahre | Writer, Producer Schwerpunkt Vocals |
| Produzent 6 | Produzent, Mixing Engineer | Pop, Rock | <20 Jahre | Schwerpunkt Vocals |

Tabelle 1: Übersicht der Interviewpartner (Produzenten)

Die Durchführung der Interviews erfolgte entweder schriftlich, per E-Mail oder telefonisch. Die schriftlichen Interviews wurden direkt dokumentiert, die telefonisch geführten Interviews aufgezeichnet und anschließend transkribiert. Die Audioaufnahmen sowie Transkripte sind der Arbeit als *Digitaler Anhang* beigefügt. Als Erhebungsmethode diente das o.g. leitfadengestützte Interview. Es erlaubte den Experten ihre persönliche Einschätzung darzulegen. Die Fragen zielten darauf ab, einen möglichst umfassenden Eindruck aus deren Produktionspraxis zu erhalten. Bevor die Leitfragen gestellt wurden, wurde den Produzenten folgendes Szenario dargestellt:

„Kontext: Du produzierst einen Popsong (länge ca. zwei min.). Die für den Song aufgenommenen Vocals liegen Dir völlig unbearbeitet vor. Das Material wurde im Studio-Kontext mit professionellem Equipment aufgenommen. Alle folgenden Fragen beziehen sich ausschließlich auf die Vocal-Produktion.“

Bei den Telefongesprächen wurde nach der Einführung die erste Frage gestellt. Danach verlief das Gespräch offen, orientierte sich jedoch an den Leitfragen. Diese lauteten in abgekürzter Form:

- Was hältst du für das absolute Mindestmaß, um Vocals in deiner Produktion professionell klingen zu lassen?
- Gibt es Bearbeitungsschritte, die überbewertet werden?
- Welche Schritte zählen für dich zu einer detaillierten Vocal-Bearbeitung?
- Wie stark hängt die Vocal-Bearbeitung vom Sänger ab?
- Hast du eine persönliche Faustregel für effiziente Vocal-Bearbeitung?

Die vollständigen Fragen sind aufgrund der Ausführlichkeit als „*Digitaler Anhang 1*“ beigefügt.

Im Hinblick auf ethische und datenschutzrechtliche Aspekte wurden alle Interviewpartner vorab umfassend über den Zweck der Studie informiert und gaben ihre Einwilligung zur Teilnahme. Zudem wurden sämtliche der erhobenen Daten anonymisiert, sodass keine Rückschlüsse auf einzelne Personen möglich sind. Somit sollen zusätzliche Verzerrungen der Aussagen vermieden werden. Damit wurde den gängigen ethischen Standards der qualitativen Sozialforschung Rechnung getragen und die Vertraulichkeit der bereitgestellten Informationen sichergestellt. Die Konformität nach DSGVO wurde beachtet. Die Produzenten werden im Folgenden als Produzent 1 bis 6 bezeichnet.

3.2 Unterschiedliche Bearbeitungsstufen von Vocals im Hörvergleich

Um der Leitfrage dieser Arbeit auf den Grund zu gehen, wurde ein Hörvergleich mit verschiedenen Testgruppen durchgeführt. Mit diesen wurde evaluiert, ob eine detailliertere und damit zeitaufwendigere Bearbeitung von Vocals im Gegensatz zu einem Mindestmaß an Bearbeitung

einen Einfluss auf ihre wahrgenommene Qualität im Kontext der Gesamtmischung hat. Der Hörversuch diene außerdem zur Klärung der sekundären Fragestellungen, wie stark Vocal-Bearbeitung überhaupt wahrgenommen wird, ob sich zwischen den Gruppen grundsätzlich unterschiedliche Wahrnehmungen abzeichnen, sowie der Identifizierung eines „Effizienz-Sweetspots“.

3.2.1 Demomaterial

Für den durchgeführten Hörvergleich wurden zwei Hörbeispiele bzw. Demos produziert. Es handelt sich um Deutschsprachige Popsongs mit einer Länge von 2,5 min., aus denen ein Ausschnitt von ca. 20 Sekunden ausgewählt wurde. Es handelt sich dabei bewusst um einen Ausschnitt, bei dem die Vocals klar im Vordergrund des Songs stehen.

Die Hörbeispiele wurden in jeweils drei Vocal-Bearbeitungsstufen erstellt: unbearbeitet, 30 min. Bearbeitung und 120 min. Bearbeitung. Daraus ergeben sich in Summe sechs Stimuli. Die produzierten Stimuli (Demo 1, Demo 2) in ihren drei unterschiedlichen Bearbeitungsstufen (A,B,C) befinden sich im *Digitalen Anhang 4* dieser Arbeit. Aufgrund des subjektiven Lautstärke Eindrucks wurden alle Demos einheitlich mit -8,5 LUFS gerendert, um ihre Vergleichbarkeit sicherzustellen – unabhängig von der individuellen Abhörlautstärke der Testpersonen. Der gewählte Wert von -8,5 LUFS spiegelt einen in der aktuellen Popmusikproduktion üblichen Mittelwert der Lautheit fertig gemasterter Titel realistisch wider [56], [57]. Alle Stimuli wurden als PCM-WAV in 24-Bit / 48kHz exportiert. Die konkrete Produktion der unterschiedlichen Versionen wird im Folgenden erläutert.

3.2.1.1 Unbearbeitete Version

Die unbearbeitete Version wurde keinerlei nachträglicher Vocal Bearbeitung unterzogen. Die Vocals wurden Aufgenommen, aber nicht bearbeitet.

Dennoch sollen die Fakten zur Aufnahme hier nicht unerwähnt bleiben, da diese ebenfalls das Ausgangsmaterial für die beiden Bearbeiteten Versionen darstellte. Beide Demos wurden - mit einem „UAD Apollo twin x“ Audio-Interface sowie einem „Neumann TLM 102“ Mikrofon (Druckgradientenempfänger, Niere) - in einer ruhigen Studioumgebung aufgenommen. Diese Komponenten bieten eine optimale technische und praxisnahe Grundlage. Aufgenommen wurden zwei unterschiedliche Deutschsprachige (Männliche) Pop-Künstler. Für Demo 1 (A,B,C) Sänger 1 und für Demo 2 (A,B,C) Sänger 2.

3.2.1.2 Mindest-Bearbeitung (30 min.)

Die Bearbeitung der 30-min.-Demos orientierte sich an den Aussagen der befragten Produzenten im Rahmen der Experteninterviews. Ziel war es, eine praxisnahe Umsetzung eines als „Mindestmaß“ verstandenen Bearbeitungsumfangs zu realisieren. Die Auswahl und Gewichtung der Bearbeitungsschritte orientierte sich dabei an den in Kapitel 4.1 dargestellten Expertenaussagen.

Vorab wurde ein grundlegendes Editing durchgeführt, bei dem sowohl Anfang als auch Ende jedes Takes sauber geschnitten und gefadet wurde.

Anschließend kam eine standardisierte Vocal Chain zum Einsatz, welche in Bezug auf die in den Interviews genannten Tools gestaltet wurde. Diese bestand aus einem Equalizer (FabFilter Pro-Q3), welcher als Hochpassfilter, zur Korrektur störender Frequenzbereiche sowie zu einer Anhebung der Höhen für mehr Brillanz und Sprachverständlichkeit genutzt wurde. (Abb. 4.)



Abbildung 4: Benutzeroberfläche fabfilter ProQ3

Als zweites Plugin der Kette kam Antares Auto-Tune zur automatisierten Intonationskorrektur zum Einsatz. In Abbildung 5 sind die final eingestellten Parameter zu sehen. Die Einstellungen erfolgten auf experimenteller Basis, wobei jene Parameterkombination gewählt wurde, die im Hörvergleich die beste Anpassung an das jeweilige Eingangssignal ermöglichte. Es folgten ein UAD Teletronix 1176 als Kompressor mit schneller Regelzeit (Abb. 6)



Abbildung 5: Benutzeroberfläche Antares Autotune

zum Abfangen der ersten Pegelspitzen, sowie ein Teletronix LA-2A als leicht klangfärbender und eher trägerer Opto-Kompressor. Zum Schluss folgte ein De-Esser zur Reduktion der

im Signal vorhandenen Sibilanten (Abb. 7) Die Effekte wurden unter Zeitvorgabe von 30 min. mit praxisüblichen Voreinstellungen auf den Kanal geladen und auf das betreffende Signal angepasst. Diese Bearbeitungsschritte wurden ebenfalls auf Layer-Vocals wie Backings und Harmonies angewendet.



Abbildung 6: Benutzeroberfläche UAD 1176LN Kompressor

Der festgelegte Zeitrahmen von 30 min. bezieht sich auf eine gesamte Version des Songs, mit einer Länge von ca. zwei min. Somit bildeten diese Demos die „schnell bearbeitete“ Vergleichsbedingung im späteren Hörversuch. Ziel war es, innerhalb dieses begrenzten Rahmens und auf Basis der Expertenaussagen ein möglichst qualitativ hochwertiges Ergebnis zu erzielen. Dabei wurde bewusst darauf verzichtet, sich in Details zu verlieren. Stattdessen lag der Fokus darauf, mit den verfügbaren Mitteln und Standardwerkzeugen das bestmögliche klangliche Resultat zu erreichen. Um klangliche Vergleichbarkeit zwischen den 30 min. und 120 min. Bearbeitungsversionen zu gewährleisten, wurden die Grundeinstellungen der Reverb- und Delay-Effekte in beiden Versionen identisch eingesetzt. Zum Einsatz kamen zwei statische Reverbs aus dem Plug-in Valhalla VintageVerb: eine kurze Plate-Emulation mit einer Nachhallzeit von 0,21 Sekunden sowie ein längerer Room-Hall mit 2,7 Sekunden Decay. Beide Reverbs wurden mit Lo-Cut- und Hi-Cut-Filtern versehen, um unerwünschte tiefe bzw. scharfe hohe Frequenzanteile zu reduzieren und die Einbettung der Stimme im Mix zu optimieren. Als Delay wurde das Standard-Delay Plugin von Ableton Live mit einem $\frac{1}{4}$ Delay verwendet.



Abbildung 7: Benutzeroberfläche Izotope RX De-esser

3.2.1.3 Detaillierte Bearbeitung (120 min.)

Die 120-min.-Versionen wurden mit dem Ziel erstellt, eine möglichst differenzierte und professionell produzierte Gesangsbearbeitung zu realisieren. Im Unterschied zur 30 min. bearbeiteten Variante lag der Fokus hier auf präziser Korrektur und gestalterischer Feinabstimmung aller relevanten Parameter. Die Vocal-Chain der 30 min. Version wurde dabei in der 120 min. Version unverändert als Basis genutzt.

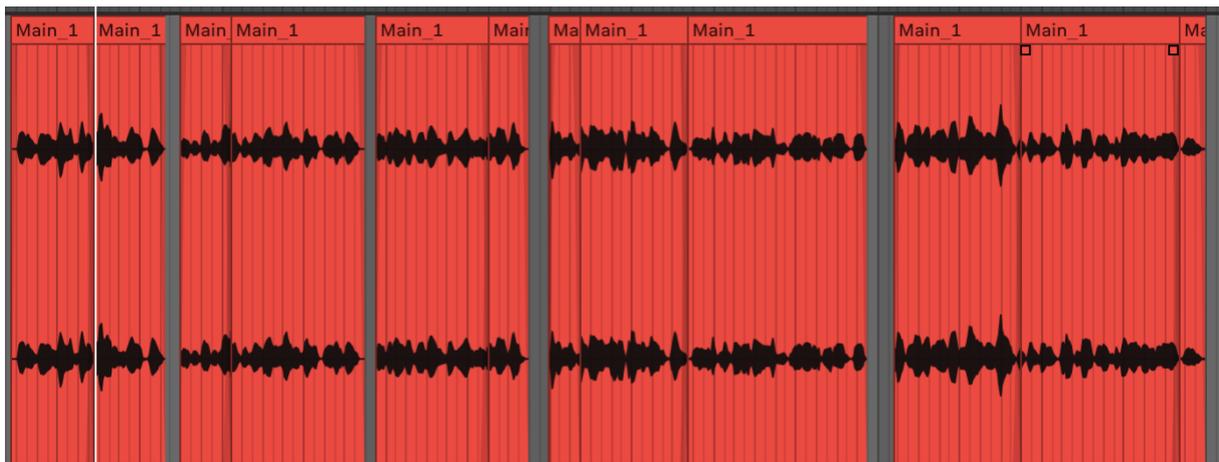


Abbildung 8: Audiospur nach Korrektur des Timings und Comping in Ableton Live

Ein wesentlicher Unterschied bestand in der Editierung sämtlicher vorhandener Vocalspuren. Neben der Entfernung störender

Nebengeräusche und übermäßiger Atmer wurde auch das Timing einzelner Phrasen und Worte manuell korrigiert, um rhythmische Ungenauigkeiten zu beheben. Darüber hinaus stand ausreichend Zeit zur Verfügung, um durch detailliertes Comping aus mehreren Takes einen „optimalen Take“ zusammenzustellen. (Abb. 8) Zusätzlich zu den Kompressoren der verwendeten Vocal-Kette, wurde an einigen Stellen mit Lautstärke Automationen gearbeitet.

Ein weiteres zentrales Merkmal der 120 min. bearbeiteten Versionen war die manuelle Intonationskorrektur mit Melodyne. Hierbei wurden gezielt einzelne Wörter und Phrasen tonal und rhythmisch angepasst, bevor anschließend (aus stilistischen Gründen) die weitere Bearbeitung mit Auto-Tune erfolgte. Diese Kombination ermöglichte eine besser kontrollierbare Intonation, als dies allein mit Antares Autotune der Fall gewesen wäre. (Abb. 9)

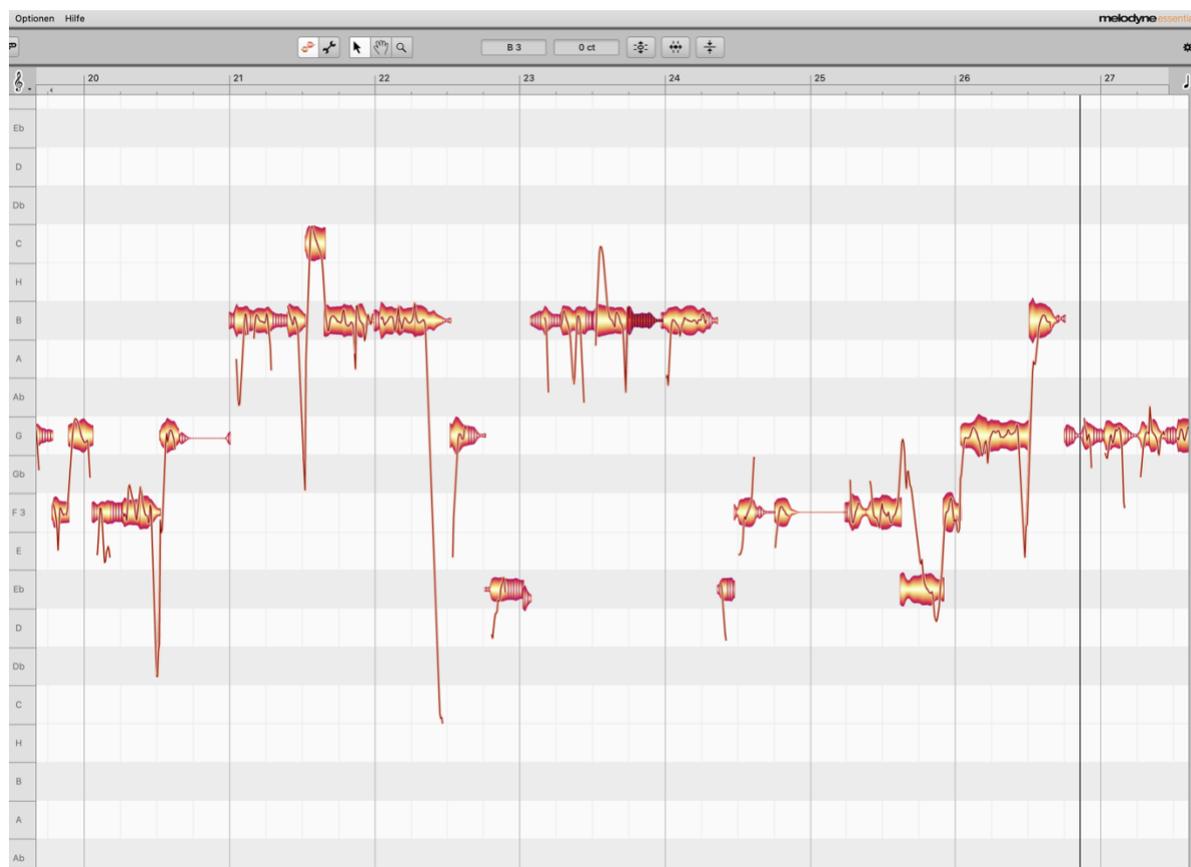


Abbildung 9: Benutzeroberfläche Celemony Melodyne

Ergänzend wurden alle weiteren Vocal-Layer (Harmonien, Dopplungen und Backings) im Timing exakt an die Lead-Vocal angepasst, um eine klare rhythmische Kohärenz im Gesamtbild zu gewährleisten. Die zeitlichen Anpassungen wurden sowohl händisch als auch mit Hilfe des Plugins „Vocalign Project“ umgesetzt. Dieser zusätzliche Schritt diente nicht nur der zeitlichen Präzision, sondern unterstützte auch die Durchsetzungskraft der Vocals im Mix. In Abb. 10 ist das Plugin Vocalign zu sehen. Als Guide-Track (Sidechain) wurde hier die Main Vocal genutzt. Das Plugin wurde in diesem Fall auf der Layerspur „Harmonie Hook“ eingesetzt. Rechts oben lässt sich die Match-Time einstellen welche angibt, wie stark (in ms) das anzupassende Signal zeitlich vom Guide Signal differiert. Um den Effekt des Alignments in den Hörbeispielen sehr deutlich zu machen, wurde diese Zeit auf 0 ms eingestellt.



Abbildung 10: Benutzeroberfläche VocAlign Project5

3.3 Technischer Versuchsaufbau

Zum Hören der Stimuli wurde ein Setup in ruhiger Umgebung aufgebaut. Zum Abhören wurden Kopfhörer des Typs Beyerdynamic DT 770Pro genutzt. Um den potenziellen Störfaktor der Raumakustik auszuschließen, fiel die Wahl bewusst auf Kopfhörer anstatt auf Studiomonitore. Als Vorverstärker bzw. D/A Wandler diente das Audiointerface „Apollo Twin X Quad“ von Universal Audio. Ableton Live wurde sowohl als DAW zur Wiedergabe der Stimuli benutzt als auch um zwischen diesen verzögerungsfrei zu wechseln. Um das Umschalten zwischen den Hörproben (A,B,C) möglichst einfach zu gestalten, wurde als Steuerelement ein kleines MIDI-Keyboard verwendet. Mithilfe hochwertiger Komponenten wurde so eine möglichst präzise und unverfälschte Wiedergabe sichergestellt.

3.4 Versuchsdurchführung

Der Hörversuch wurde überwiegend im Rahmen persönlicher Interviews durchgeführt. Die Teilnehmenden hörten die Stimuli zuerst in strukturierter Reihenfolge. Sie wussten dabei nicht, welcher Stimulus welcher Bearbeitungsversion (unbearbeitet, kurz oder lang bearbeitet) entsprach. Um Voreingenommenheit zu vermeiden, wurden die drei Versionen pro Demo anonymisiert und jeweils mit den Buchstaben A, B und C gekennzeichnet. Die Zuordnung dieser Buchstaben erfolgte zufällig und variierte je nach Demo. In einem begleitenden Gespräch erfolgte dann die Beantwortung der Leitfragen. Da das Gespräch im Rahmen eines semi-strukturierten Interviews stattfand, wurde den Testpersonen neben der strukturierten Beantwortung der Leitfragen Raum für freie Kommentare und persönliche Eindrücke gelassen, um auch spontane Wahrnehmungen zu erfassen. Die Testpersonen hörten jeweils alle drei Versionen eines Songs im direkten Vergleich untereinander. Dadurch fanden alle Bewertungen innerhalb derselben Abhörsituation statt, wodurch interne Vergleichbarkeit gewährleistet wurde. Den Probanden wurde ausdrücklich

gestattet, die Stimuli mehrfach anzuhören, um eine fundierte Einschätzung vornehmen zu können, ohne unter Zeitdruck zu geraten. Über ein MIDI-Steurelement war es den Teilnehmenden möglich verzögerungsfrei und nahtlos zwischen den Stimuli zu wechseln. Außerdem wurde bei jeder Testperson die Abhörlautstärke am Anfang des Hörversuchs festgelegt und während des Hörblocks nicht verändert.

In einigen Fällen war ein persönliches Treffen nicht möglich. Hier wurde der Versuch von den Teilnehmenden selbstständig durchgeführt, jedoch mit telefonischer Begleitung bzw. Vor- und Nachbesprechung. So sollte eine vergleichbare Durchführung des Versuchs in Bezug auf Aufgabenstellung und Rückmeldemöglichkeit sichergestellt werden. Die Durchführung erfolgte in ruhiger Umgebung mit selbst gewähltem Abhörsystem (professionelle Studio-Kopfhörer). Trotz fehlender vollständiger Kontrolle über das technische Setup wird aufgrund der oben genannten Maßnahmen und dem Umstand, dass die auf diese Weise Interviewten Personen alle als Audiotechnisch versiert oder Audio-Professionals einzustufen sind, hier von einer ausreichenden Vergleichbarkeit der Höreindrücke ausgegangen. Dies wurde aufgrund der gewählten Methodik als ausreichend erachtet, da der Fokus bei dieser qualitativen Untersuchung darauf liegt, subjektive Wahrnehmung und Unterschiede zwischen Bearbeitungsgraden zu erfassen. Die technische Präzision der Abhörumgebung spielt bei diesem Vergleich eine untergeordnete Rolle. Durch diese flexible Handhabung konnte die Teilnahme auch unter eingeschränkten Bedingungen gewährleistet werden, ohne die Vergleichbarkeit der Aussagen wesentlich zu beeinträchtigen.

3.5 Auswahl der Probanden

Um möglichst vielfältige Perspektiven auf die wahrgenommene Qualität zu erhalten, wurde der Hörversuch mit drei unterschiedlichen Testgruppen durchgeführt: Laien, audiophile Personen und Professionelle aus dem

Musikbereich (Produzenten). Diese Differenzierung ermöglicht einen Vergleich der Wahrnehmung abhängig vom individuellen Erfahrungsgrad.

Definition der Testgruppen:

Laien: Musikhörende Personen ohne jegliche Erfahrung im Bereich Audio- oder Musikproduktion. Diese Gruppe bildet die üblichen Konsumenten ab. Im Folgenden anonymisiert genannt als Laien 1 bis 5.

Audiophile Hörende: Menschen mit genereller Erfahrung im Audibereich. Diese kann auf Beruflicher Erfahrung und/oder Schulbildung basieren. Diese Probanden haben jedoch keine relevante Kenntnis im Bereich Popmusik-Produktion und somit auch nicht im Bereich Vocal-Produktion, sondern sind in anderen Bereichen der Audiobranche tätig. Dazu gehören z.B Instrumentalisten, Techniker für Beschallung, Live-Ingenieure und Toningenieure. Die Rezipienten kennen sich mit den Grundlagen der Audiotechnik aus, haben ein geschultes Gehör und können ihre Eindrücke aus fachlicher Sicht schildern. Sie verfügen jedoch über kein praktisches Wissen, welche Bearbeitungsschritte und Techniken in einer Popmusik-Produktion gängig sind. Im Folgenden anonymisiert genannt als Audiophile Person 1 bis 5.

Professionals: Musikproduzenten, vorwiegend im Bereich Popmusik, die häufig mit Vocal-Bearbeitung zu tun haben und sich in diesem Bereich auskennen. Sie wissen mit welchen Tools sie zum gewünschten Ergebnis kommen und sind in der Lage ihre Eindrücke noch differenzierter zu benennen als die Gruppe der audiophilen Hörenden.

Während bei Laien zu erwarten ist, dass sie die Stimuli tendenziell intuitiv und emotional bewerten, verfügen audiophile Probanden über ein besonders sensibles Gehör für Details, jedoch ohne Kenntnis über den Produktionstechnischen Hintergrund.

Professionelle hingegen können eine technische und praxisnahe Perspektive mit einbringen, die sich an aktuellen Produktionsstandards der Popmusikindustrie orientiert. Im Folgenden werden diese anonymisiert genannt als Profi 1 bis 5.

Durch diese bewusste Auswahl verschiedener Hörergruppen lassen sich eventuelle Unterschiede in der Wahrnehmung von Bearbeitungsgraden identifizieren und Rückschlüsse auf die zielgruppenspezifische Relevanz bestimmter Produktionsentscheidungen ziehen. Vor allem im Hinblick auf die Fragestellung, ob ein hoher Bearbeitungsaufwand im Mix subjektiv als Qualitätsgewinn wahrgenommen wird, liefert dieser Gruppenvergleich wertvolle Erkenntnisse. Es wurden insgesamt 5 Probanden pro Gruppe interviewt, um einerseits eine ausreichende Vielfalt an Höreindrücken zu erfassen und gleichzeitig eine vertiefte qualitative Analyse zu ermöglichen. Einige der befragten Laien stammten aus dem persönlichen (nicht Familiären) Umfeld der forschenden Person. Eine bewusste Reflexion möglicher Beeinflussungen durch die persönliche Nähe wurde im Auswertungsprozess berücksichtigt.

4 Ergebnisse

4.1 Ergebnisse der Experteninterviews

Die Auswertung aller durchgeführten Interviews erfolgte in Anlehnung an die qualitative Inhaltsanalyse nach Mayring [58]. Dabei wurde bewusst auf ein formales Kategoriensystem verzichtet. Stattdessen wurden die transkribierten Aussagen thematisch strukturiert, um zentrale Motive, Wahrnehmungen und Bewertungen der Interviewpartner herauszuarbeiten. Die Leitfragen des Interviews dienten dabei als grundlegender Orientierungsrahmen. Ziel war es, wiederkehrende Muster ebenso wie individuelle Perspektiven der Interviewees sichtbar zu machen. Die Transkription der geführten Interviews erfolgte mithilfe eines KI-gestützten Transkriptionstools. Die automatisch erzeugten Transkripte wurden anschließend manuell überprüft und anonymisiert. Die inhaltliche Interpretation basiert ausschließlich auf dem überprüften Material.

4.1.1 Konsens und Unterschiede zur Mindestbearbeitung

Im Rahmen dieser Masterarbeit wurden sechs erfahrene Musikproduzenten zur Einschätzung eines notwendigen Mindestmaßes an Bearbeitungsschritten bei Vocals befragt. Ziel war es, herauszufinden, welche technischen Maßnahmen für erfahrene Produzenten als unverzichtbar gelten, wenn nur begrenzte Bearbeitungszeit zur Verfügung steht, und inwiefern sich daraus eine praxisnahe Mindestbearbeitungszeit ableiten lässt. Die Befragung erfolgte schriftlich oder mündlich anhand eines standardisierten Fragenkatalogs.

Ein deutlicher Konsens zeigte sich hinsichtlich grundlegender Bearbeitungsschritte, die selbst unter Zeitdruck als unerlässlich gelten. Nahezu alle befragten Produzenten nannten Equalizing, Kompression sowie

eine Form von Intonationskorrektur (meist automatisch via Auto-Tune) als Mindestanforderung für professionell klingende Vocals. Auch das Entfernen störender Nebengeräusche (z. B. Atmer, Mouth Clicks) und ein dezentes De-Essing wurden wiederholt erwähnt [17], [18], [19], [20], [21], [22].

Die Einschätzungen deuten darauf hin, dass sich mit diesen Maßnahmen, abhängig von Workflow und Erfahrung, innerhalb von etwa 30 min. eine grundlegend überzeugende Vocal-Spur erzeugen lässt. Laut einiger Experten werden dabei Tools wie vorgefertigte Effektketten (Vocal-Chains) oder Templates gezielt genutzt, um Prozesse zu beschleunigen [17], [21]. Diese Templates bieten eine gute Ausgangslage, von welcher aus sich anschließend gezielt weiterarbeiten lässt. Der Fokus liegt dabei auf einem technisch sauberen und im Mix funktionierenden Ergebnis, nicht auf ästhetischer Perfektion. Dies ist besonders im Hinblick auf die Forschungsfrage dieser Arbeit relevant, bei der die Vocals im Kontext der Mischung betrachtet werden.

Trotz einiger inhaltlicher Überschneidungen variiert die Bewertung der Bearbeitungstiefe. Während z.B. Produzent 2 betont, dass es für ihn kein starres Mindestmaß gibt, sondern ausschließlich das klangliche Resultat zählt, heben andere hervor, dass gewisse Schritte aus Effizienzgründen weggelassen werden können, ohne dass dies die hörbare Qualität wesentlich beeinträchtigt [18], [20], [21].

Insbesondere bei aufwendigen Maßnahmen wie manueller Tonhöhenbearbeitung via Melodyne, Comping oder Layering waren die Einschätzungen differenzierter. Einige Produzenten merkten an, dass diese Schritte im finalen Mix kaum wahrgenommen würden, obwohl sie viel Zeit beanspruchen [19], [21].

Auch die starke Abhängigkeit der Bearbeitung von der Gesangsperformance wurde mehrfach betont. Dazu mehr in Kapitel 4.2.1, sowie Kapitel 2.3.

Auf Grundlage der Expertenaussagen deren Abgleich mit gängiger Fachliteratur und Masterclasses anderer relevanter Popmusik-Produzenten lässt sich ein praxisnahes Mindestmaß an Bearbeitungsschritten formulieren. Daraus lässt sich eine Mindestbearbeitungszeit ableiten, die für erfahrene Produzenten als praxisorientierter Richtwert gelten kann. Die von den Experten als unverzichtbar genannten Maßnahmen (EQ, Kompression, Auto-Tune, De-Essing) wurden daher in der „Kurzbearbeitung“ der Stimuli berücksichtigt. Zur zügigen Anwendung der genannten Bearbeitungsschritte wurde eine Mindestbearbeitungszeit von 30 min. praktisch evaluiert [13], [59], [60].

4.1.2 Perspektiven zur Detaillierten Vocal-Bearbeitung

Neben der Einschätzung eines praxisnahen Mindestmaßes an Bearbeitungsschritten gaben die befragten Produzenten auch Einblicke in ihren Umgang mit detaillierteren Formen der Vocal-Bearbeitung. Dabei wurde deutlich, dass die Entscheidung für oder gegen eine aufwendige Nachbearbeitung stets im Spannungsfeld zwischen ästhetischem Anspruch, Produktionskontext und verfügbaren Ressourcen steht.

Als häufig genannte Maßnahmen im Rahmen einer erweiterten Bearbeitung wurden die manuelle Intonationskorrektur mit Melodyne, genaueres Vocal-Comping, Volume-Automatiken, multibandbasierte Dynamikbearbeitung sowie die räumliche Feinabstimmung mittels Reverb und Delay-Design genannt. Einige Produzenten betonten zudem die Bedeutung subtiler Gestaltungsmittel wie Formant-Pitching, Parallelsättigung oder gezielter Verzerrung einzelner Layer, um klangliche Individualität einzelner Spuren zu erzeugen [17], [21].

Auffällig war, dass eine detaillierte Bearbeitung seltener als rein technische Optimierung, sondern vielmehr als gestalterisches Mittel beschrieben wurde. Die Grenze zwischen technischer Notwendigkeit und kreativem Eingriff wurde dabei fließend beschrieben. Gleichzeitig betonten mehrere Befragte, dass eine detaillierte Bearbeitung nicht automatisch zu einem besseren Ergebnis führe. In manchen Fällen könne sie sogar kontraproduktiv wirken – etwa dann, wenn der natürliche Ausdruck der Stimme dadurch verloren gehe oder eine überbearbeitete Künstlichkeit entstehe [18], [19]. Produzent 4 warnte in diesem Zusammenhang auch vor dem Einsatz übermäßig vieler Plugins und überladener Vocal-Ketten, da dies schnell zum Verlust von Druck und Volumen der Vocals führen kann [20].

Insgesamt verdeutlichen die Aussagen, dass detaillierte Bearbeitungsmaßnahmen zwar sinnvoll sein können, um professionelle

Ergebnisse zu erzielen, jedoch immer im Kontext der musikalischen Zielsetzung, der Qualität der Ausgangssignale und der Genrespezifischen Konventionen zu bewerten sind.

Für die Demos zur detaillierten Bearbeitung wurde aufgrund Praktischer Erkenntnisse die vierfache Zeit von 120 min. veranschlagt. Um eine noch detailliertere Bearbeitung zu erreichen, wäre eine Bearbeitung über die 120 min hinaus durchaus möglich, doch für die Zwecke dieser Untersuchung reichen diese aus, um den qualitativen Unterschied sichtbar zu machen. Dies erwies sich als ausreichend, um alle über das Mindestmaß hinausgehenden Tools und Techniken anzuwenden. Dazu zählen alle in Kapitel 3.2.1 genannten Bearbeitungsschritte. Auch wenn die erstellten Demos nur Songausschnitte mit ca. 20 Sekunden Spielzeit darstellen, beziehen sich die veranschlagten Bearbeitungszeiten auf einen ganzen Popsong mit ca. 2:20 min. Spielzeit.

Mit den gewonnenen Eindrücken lässt sich im weiteren Verlauf analysieren, inwiefern das Mindestmaß im Vergleich zu einer detaillierten Bearbeitung klanglich unterscheidbar ist und ob sich die Erkenntnisse im Sinne des Pareto-Prinzips interpretieren lassen. Zur Erstellung des Demomaterials unter Berücksichtigung der gewonnenen Erkenntnisse siehe Kapitel 3.2.1.

4.1.3 Weitere Erkenntnisse aus den Experteninterviews

Neben konkreten Einschätzungen zur Mindest- und Detailbearbeitung ergaben die Experteninterviews auch weiterführende Beobachtungen, die andere Aspekte des Vocal-Produktionsprozesses betreffen und kontextualisierend herangezogen werden können.

Ein durchgängig genanntes Thema war die Relevanz der Gesangsperformance für den gesamten Bearbeitungsaufwand. Mehrere Produzenten betonten, dass die Qualität und Ausdrucksstärke der Rohaufnahme maßgeblich bestimmt, wie viele nachträgliche Korrekturen

notwendig sind. Aufnahmen mit sauberer Intonation, stabiler Rhythmik und guter emotionaler Überzeugung erfordern deutlich weniger Eingriffe im Post-Processing [17], [18]. In diesem Zusammenhang wurde auch angemerkt, dass der Produzent bei Studiosessions idealerweise, gezielt in Rücksprache mit dem Künstler vorausschauend an den Vocals arbeitet. Dies kann z.B durch einzelne Phrasierungsanweisungen oder das gezielte Recording von Takes für späteres Comping geschehen [20], [22].

Ebenfalls erwähnt wurde auch die Wichtigkeit von schnellem und gezieltem Arbeiten in Bezug auf Studiosessions mit Künstlern. Diese erwarten meistens sehr schnell ein „fertig-klingendes“ Ergebnis, was einen schnellen Workflow von Session-Produzenten voraussetzt. Das ist für alle in der Session arbeitenden Personen von großer Wichtigkeit, da auf dieser Grundlage weitere Entscheidungen für die Fertigstellung eines Songs getroffen werden. Um gute Stimmung, ein angenehmes Arbeitsumfeld und damit auch bestmögliche Ergebnisse zu gewährleisten, sollten daher von Produzentenseite lange Wartezeiten für die Mitarbeitenden unbedingt vermieden werden [21].

Ein weiterer Aspekt betrifft die Subjektivität von Qualität in der Vocal-Produktion. Während manche Produzenten technische Perfektion anstreben, steht für andere Produzenten der puristische Gedanke im Vordergrund. Sie legen eher Wert auf stimmlichen Ausdruck und Authentizität der Performance. Demnach ist eine stark bearbeitete Stimme nicht per se „besser“, sondern muss im jeweiligen musikalischen Kontext bewertet werden. Die Aussagen der Interviews im Abgleich mit literarischen Quellen zeigen: Insbesondere in der modernen Popmusik gelten technisch saubere, aber stark bearbeitete Vocals als Standard, während in anderen Stilen (z. B. Singer-Songwriter oder Indie-Pop) bewusst auf ein gewisses Maß an Rohheit oder Natürlichkeit gesetzt wird [61].

Auch die ökonomische Dimension wurde vereinzelt zum Thema: Arbeitszeit ist in professionellen Studioumgebungen eng mit Kosten verbunden. Einige Befragte wiesen darauf hin, dass effizientes Arbeiten nicht nur klanglich, sondern auch wirtschaftlich notwendig sei. Vor allem gelte das bei Projekten mit begrenztem Budget oder engem Zeitrahmen [19], [21].

Zusammenfassend lässt sich feststellen, dass Vocal-Bearbeitung stets als ein Zusammenspiel aus technischen Möglichkeiten, künstlerischem Anspruch und pragmatischen Rahmenbedingungen verstanden wird. Die Interviews unterstreichen, dass es hier keine allgemeingültige Norm gibt. Vielmehr ist es ein kontinuierliches Abwägen zwischen Aufwand, gewünschter Ästhetik und Zweckmäßigkeit.

4.2 Ergebnisse des Hörtests: Wahrnehmungsunterschiede zwischen Bearbeitungsstufen

Ziel des durchgeführten Hörtests war es zu untersuchen, ob und in welchem Umfang sich unterschiedliche Bearbeitungsstufen von Gesangsspuren auf deren wahrgenommene Qualität im Kontext eines Popmusikmixes auswirken. Im Mittelpunkt stand dabei die Frage, ob eine detaillierte und zeitintensive Bearbeitung gegenüber einem praxisorientierten Mindestmaß an Bearbeitung tatsächlich als qualitativer Mehrwert wahrgenommen wird. Der genaue Aufbau des Tests, die verwendeten Stimuli sowie die Gruppeneinteilung der Teilnehmenden wurden im Methodik-Abschnitt (Kapitel 3.2.1 und 3.5) ausführlich beschrieben.

Zur besseren Übersicht wurden Aussagen der Interviewpartner stichpunktartig und in Tabellenform dieser Arbeit als *Digitaler Anhang 5* beigelegt. Diese Darstellung dient der Strukturierung qualitativer Inhalte und erhebt keinen Anspruch auf statistische Repräsentativität.

4.2.1 Allgemeine Beobachtungen

Die Auswertung der Aussagen aller drei Testgruppen (Laien, Audiophile, Professionals) zeigt das nahezu allen Teilnehmenden Personen diverse Unterschiede zwischen den drei vorgespielten Versionen wahrnehmen konnten. Dabei war die Wahrnehmung der drei unterschiedlichen Versionen bei Demo 1 und Demo 2 weitestgehend gleich. Besonders auffällig war die Bewertung der unbearbeiteten Versionen, die unabhängig vom Erfahrungsniveau mehrheitlich als „roh“, „unbearbeitet“, „flach“, „trocken“ oder „noch nicht releasetauglich“ beschrieben wurde [62], [63], [64], [65], [66], [67].

Die beiden bearbeiteten Versionen (30-min.- und 120-min.-Bearbeitung) wurden hingegen überwiegend als „fertig“, „professionell“ oder „releasefähig“ eingeschätzt. Gleichzeitig zeigte sich, dass viele Teilnehmende – insbesondere in der Laien- und audiophilen Gruppe – Schwierigkeiten hatte, zwischen diesen beiden bearbeiteten Varianten eindeutige Unterschiede auszumachen [68], [69], [70]. Häufig wurden Aussagen wie „klingen fast gleich“, oder „sehr ähnlich“ getroffen. Unterschiede wurden oft nur vermutet oder emotional begründet, etwa mit Aussagen wie „B klingt für mich irgendwie voller“ [67].

Dass der Unterschied von der Unbearbeiteten Version zu den beiden anderen als sehr stark empfunden wurde, sowie das zwischen den beiden bearbeiteten Versionen nur selten ein Unterschied ausgemacht werden konnte, ist ein wiederkehrendes Muster, welches über alle Testgruppen hinweg besteht. Einige Hörer betonten sogar explizit, dass sie ohne direkten Vergleich keine Unterschiede zwischen 120 min.-Version und 30 min.-Version bemerkt hätten [67], [69].

Ein Großteil der Teilnehmenden gaben an, dass sie die Möglichkeit des direkten Vergleichs (A/B/C-Schaltung) als hilfreich empfanden, um überhaupt auf feine Unterschiede aufmerksam zu werden.

Dieses Feedback deutet darauf hin, dass differenzierte Bearbeitungen unter realen Hörbedingungen – etwa im Streaming- oder Radio-Kontext – deutlich weniger ins Gewicht fallen könnten als im kontrollierten Testumfeld [62], [66], [67], [71], [72].

4.2.2 Vergleich unbearbeitet vs. bearbeitet

Der deutlichste Wahrnehmungsunterschied im Hörtest zeigte sich zwischen der unbearbeiteten Version und den beiden bearbeiteten Versionen. Unabhängig von Hörerfahrung oder Testgruppe wurde die unbearbeitete Fassung nahezu einstimmig als qualitativ deutlich schwächer eingestuft. Sie wurde meist mit den o.g. Attributen beschrieben.

Zentraler Kritikpunkt war dabei die mangelnde Einbettung der Stimme in den Gesamtmix. Viele Teilnehmende beschrieben die Vocals der unbearbeiteten Version als herausstechend und „über den Instrumenten“. Es wurde oftmals, ein fehlender Hall angemerkt. Besonders auffällig war, dass diese Einschätzungen nicht nur von professionellen Hörern, sondern auch von Laien sehr eindeutig geäußert wurden, was auf eine allgemein geteilte Hörwahrnehmung hinweist. Auch die Intonation wurde vereinzelt kritisiert – allerdings eher aus den Gruppen der erfahreneren Hörer.

Ein weiterer Aspekt, der häufig genannt wurde, war die fehlende „Fertigkeit“ der unbearbeiteten Version. Viele Versuchspersonen gaben an, dass sich die unbearbeitete Fassung „wie ein Demo“, oder „nicht produziert“ anhört und nicht wie ein finales Produktionsergebnis. Auch audiophile Hörer, die grundsätzlich für den Erhalt einer gewissen Natürlichkeit in der Bearbeitung sensibilisiert sind, stuften die unbearbeitete Version als unausgereift ein und bewerteten sie insgesamt klar am schwächsten [62], [64], [68], [71], [72], [73], [74], [75].

Insgesamt ergibt sich aus den Rückmeldungen ein klares Bild: Die Unterscheidung zwischen unbearbeitet und bearbeitet wurde von allen Gruppen deutlich vorgenommen, und die unbearbeitete Version wurde durchgehend als unterlegen eingeschätzt.

4.2.3 Kurz Bearbeitet vs. Detailliert bearbeitet

Im Vergleich zwischen der 30-min.- und der 120-min.-Version äußerten viele Teilnehmende, dass sie keine oder nur sehr geringe Unterschiede zwischen den beiden Bearbeitungsstufen wahrgenommen hätten.

Laie 5 etwa gab an: „A und B haben sich für mich genau gleich angehört“, während Audiophile Person 2 formulierte, dass C (unbearbeitet) eher sehr roh klingt, während: „30 und 120 eigentlich identisch klingen“ [64], [66]. Auch in der Gruppe der Professionals wurde diese Einschätzung geäußert. Profi 2 beschrieb die beiden bearbeiteten Versionen als „relativ ähnlich, sogar fast gleich“ [68].

Einzelne Teilnehmende unter den erfahreneren Testpersonen konnten vereinzelt Unterschiede benennen. So wurden z.B Schmatzer in der kurz bearbeiteten Version gegenüber der 120 min. Version festgestellt. Dies kann auf das detailliertere Editieren der Audioclips zurückzuführen sein [70].

Des Weiteren gab es auch einige Aussagen bei denen die 30 min. Version als ausgereifteste und am besten produzierte Version interpretiert wurde, welche der 120 min. Version qualitativ überlegen sei [62], [67], [68].

Jedoch wurden auch einige Bearbeitungsschritte genannt, die nicht durchgeführt wurden. So z.B merkte eine Testperson aus der Professionals Gruppe an das die 30 min.-Version mehr Dynamik hätte, während die 120 min.-Version stärker komprimiert sei, während eine weitere Person der lang bearbeiteten Version mehr Höhenanteile attestierte [62] [63].

Einige Stimmen wiesen zudem darauf hin, dass Unterschiede nur im direkten A/B-Vergleich bemerkbar seien. So wurde in mehreren Interviews die Möglichkeit zur gezielten Umschaltung zwischen den Versionen als notwendig beschrieben, um überhaupt eine differenzierte Wahrnehmung vornehmen zu können. Produzent 3 äußerte, dass der Unterschied von 30 auf 120 min. ohne direkten Vergleich komplett untergehe und betonte, dass dieser aus seiner Sicht bei alltäglichem Konsum nicht hörbar sei [69].

Bei Nachfrage nach einer Einordnung im Hinblick auf die Professionalität der Stimme (von „Stufe 1-Unbearbeitet“ bis „Stufe 4-Sehr Professionell“), wurde die Unbearbeitete Version von allen Personen auf Stufe 1 bis 2 eingeordnet. Die beiden Bearbeiteten Versionen wurden nahezu gleichermaßen auf Stufe 3 und 4 eingeordnet (Gut produziert bis sehr professionell).

Mit der Bitte um Nennung einer favorisierten Version wurde die 30-min.-Version am häufigsten genannt und insgesamt einmal mehr bevorzugt als die 120-min.-Version, während die unbearbeitete Version von keiner der Testpersonen als Favorit gewählt wurde.

Trotz vereinzelter Aussagen über minimale Unterschiede blieb eine klare Zuordnung oder Bevorzugung zwischen 30- und 120-min.-Version in der Mehrheit der Fälle aus. Die Wahrnehmungen waren oft nicht konsistent und klar benennbar. Meistens basierten diese Einschätzungen eher auf Intuition als auf konkreten auditiven Merkmalen. Beide Versionen bewegten sich für alle Teilnehmenden auf einem vergleichbar hohen Qualitätsniveau.

4.2.4 Unterschiede zwischen den Testgruppen

Beim Vergleich der Rückmeldungen aus den drei Testgruppen zeigten sich weniger Unterschiede in der grundsätzlichen Bewertung der Stimuli als vielmehr in der Art und Weise, wie Unterschiede wahrgenommen und verbalisiert wurden. Die Laiengruppe beschrieb Unterschiede vorrangig

emotional oder intuitiv – etwa durch Aussagen wie „irgendwie klarer“ oder „fühlt sich runder an“. Konkrete technische Begriffe wurden kaum verwendet, stattdessen dominierte ein Eindrucksbild, was eher durch allgemeine Attribute beschrieben wurde [65], [71], [73].

Die audiophilen Hörer zeigten insgesamt ein differenzierteres Klangbewusstsein. Unterschiede wurden hier vor allem im Bereich von Räumlichkeit, Transparenz der Vocals im Mix oder EQ-Bearbeitung benannt [74], [75].

Teilnehmende aus der Gruppe der Professionals beschrieben Unterschiede am detailliertesten und mit gezieltem Bezug auf technische Parameter wie Kompression, Intonationskorrektur, Stereoverbreiterung oder Transientenbearbeitung. Jedoch wurden auch in dieser Gruppe einige Aussagen getroffen, die nicht der tatsächlichen Bearbeitung entsprachen. Gleichzeitig äußerten die Professionals häufiger eine gewisse Skepsis gegenüber den praktischen minimal relevanten Unterschieden – besonders im Hinblick auf alltägliche Hörumgebungen von Musikkonsumenten und wirtschaftlich begrenzte Produktionszeiten [67], [68], [69].

Auffällig war zudem, dass sich die Professionals häufiger explizit auf ihre Produktionsroutinen oder reale Anwendungsszenarien bezogen, während Laien und Audiophile primär aus einer Hörkonsumentenperspektive argumentierten. Trotz dieser Unterschiede in der Begriffswahl und der Tiefe der Analysen lassen sich keine grundlegend entgegengesetzten Bewertungen feststellen. Insbesondere in Bezug auf die Abgrenzung zwischen unbearbeitetem und bearbeitetem Material war der Konsens in allen Gruppen hoch.

4.3 Abgleich der Erkenntnisse: Expertenmeinungen, Praxisanalysen und Hörtestergebnisse

Die Aussagen der befragten Experten (Produzenten) in den Interviews decken sich in zentralen Punkten mit den Ergebnissen des Hörtests. Sowohl in den Experteninterviews als auch in der Rezeption durch die Testpersonen wurde die unbearbeitete Version eindeutig als qualitativ unterlegen eingestuft. Die von Produzenten angesprochenen Merkmale wie Hallraum zur Einbettung der Stimme und die Verwendung von genannten Effekten spiegeln sich nahezu identisch in den Rückmeldungen aller Hörenden wider.

Deutliche Unterschiede zeigen sich hingegen bei der Einschätzung der Bearbeitungsdauer. Während die Produzenten tendenziell einen hohen zeitlichen Aufwand durch Bearbeitungsschritte (z.B. Timingkorrekturen oder manueller Intonationskorrektur mit Melodyne) mit einem besseren Ergebnis verknüpfen, konnten die Testpersonen kaum hörbare Unterschiede zwischen der 30- und der 120-minütigen Bearbeitung feststellen. Die Annahme, dass mehr Bearbeitungszeit automatisch zu einer besseren Klangwirkung führt, wurde durch den Hörtest nicht gestützt.

Insgesamt lässt sich festhalten: Die grundlegende Bedeutung von Vocal-Bearbeitung zur Erzeugung eines professionellen Höreindrucks wurde von beiden Seiten bestätigt. Die Expertenmeinung zur Relevanz von Detailarbeit und zeitintensiver Optimierung wurde dagegen aus Sicht der Testpersonen nur eingeschränkt bestätigt. Dies weist auf eine Wirkungssättigung der Bearbeitung aus Sicht der Rezipienten hin.

5 Diskussion

5.1 Interpretation der Ergebnisse im Kontext der Musikproduktion

Die im Rahmen des Hörtests gewonnenen Ergebnisse lassen sich in mehrfacher Hinsicht in den Kontext moderner Popmusikproduktion einordnen. Zunächst bestätigt sich die grundlegende Annahme, dass eine gewisse Bearbeitungstiefe bzw. ein bestimmtes Maß an Bearbeitung in der Vocal-Produktion notwendig ist, damit die Mischung von Hörenden als professionell und „releasefähig“ wahrgenommen zu wird. Die durchgängige Abwertung der unbearbeiteten Version über alle Testgruppen hinweg legt nahe, dass Bearbeitungsschritte wie Kompression, Intonationskorrektur, Anpassung durch Equalizer und räumliche Einbettung von Vocals in eine Mischung heute als konstitutive Bestandteile einer zeitgemäßen Popmusikproduktion gelten. Dies gilt unabhängig vom individuellen Hörvermögen oder technischen Vorwissen der Rezipienten. Die Identifikation dieser Unterschiede von nahezu allen Testpersonen, zeigt dass diese über ein grundsätzlich geeignetes Urteilsvermögen verfügten und in der Lage waren, relevante klangliche Unterschiede bewusst wahrzunehmen.

Bemerkenswert ist jedoch, dass zwischen der schnell bearbeiteten (30 min.) und der detailliert bearbeiteten (120 min.) Version keine signifikanten Unterschiede in der Wahrnehmung festgestellt wurden. Dies war weder in Bezug auf den Qualitätseindruck noch auf die Präferenz der Fall. Diese Beobachtung zieht sich durch alle untersuchten Testgruppen und deutet darauf hin, dass ab einem gewissen funktionalen Mindestmaß an Bearbeitung ein professioneller Höreindruck im Gesamtkontext erreicht werden kann, der durch weiteren Aufwand und zusätzliche Zeitinvestition nicht zwangsläufig verbessert wird.

Im professionellen Produktionskontext könnte dies bedeuten, dass eine effizientere Herangehensweise an die Vocal-Bearbeitung unter bestimmten Bedingungen nicht nur wirtschaftlich sinnvoll, sondern auch ästhetisch ausreichend ist. Gerade in Hinblick auf Produktionsbudgets, Deadlines und gezielter Einteilung von Ressourcen bietet dieses Ergebnis eine praxisnahe Relevanz: Nicht jeder zusätzliche Bearbeitungsschritt scheint sich unmittelbar auf die wahrgenommene Qualität auszuwirken.

Gleichzeitig zeigen die Rückmeldungen der professionellen Teilnehmenden, dass die Entscheidung über den Grad der Bearbeitung stets in einem Spannungsfeld zwischen funktionaler Korrektur und ästhetischer Gestaltung zu betrachten ist. Während technische Bearbeitung zum Erreichen eines gewissen Qualitätsstandards als Ziel vorhanden bleibt, wurde mehrfach betont, dass der Aufwand in Relation zur tatsächlichen Wahrnehmbarkeit stehen muss. Insofern verweist das Ergebnis auch auf eine mögliche Differenz zwischen technischer Möglichkeit und hörpraktischer Notwendigkeit.

Die Tatsache, dass viele Unterschiede nur im direkten Vergleich auffielen, unterstreicht zudem die starke Relevanz der Hörumgebung und der Wahrnehmungssituation, welcher Rezipienten ausgesetzt sind. Unter alltäglichen Abhörbedingungen wie etwa via Streaming mit In-Ear-Kopfhörern, im Autoradio oder auf anderen mobilen Geräten wie z.B. Bluetooth Lautsprechern dürften sich minimale Bearbeitungsunterschiede weit weniger stark auswirken als im gezielten A/B-Vergleich unter kontrollierten Bedingungen mit Studiokopfhörern.

Die Ergebnisse unterstreichen auch, dass die bloße Verfügbarkeit professioneller Tools nicht automatisch zu einem professionellen Ergebnis führt. Zwar sind alle genannten Werkzeuge heute leicht zugänglich und standardisiert, doch ihre wirkungsvolle Anwendung hängt maßgeblich von der Hörerfahrung und Urteilskraft des Anwenders ab. In den

Experteninterviews wurde mehrfach betont, dass es nicht nur auf das „Was“, sondern vor allem auf das „Wie“ und „Wieviel“ ankommt – also auf die Fähigkeit, zu erkennen, wann eine Maßnahme wirklich notwendig ist und wie sie sich im Gesamtklang auswirkt.

Diese Erfahrung ist nur bedingt durch Software oder Tutorials vermittelbar. Sie basiert auf kontinuierlicher Auseinandersetzung mit Klang, Ästhetik verschiedener Genres und Hörgewohnheiten und damit auf einem Erfahrungswissen, das in der modernen Popproduktion mindestens genauso wichtig ist wie technisch korrektes Arbeiten. Die besten Tools können die klangliche Wirkung einer Entscheidung nicht selbst bewerten, sie liefern lediglich das Potenzial. Insofern bleibt trotz aller Automatisierung und Standardisierung die subjektive, erfahrungsbasierte Einschätzung des Sounds der Vocals im Mix ein entscheidender Faktor für die resultierende Qualität einer Produktion.

5.2 Bedeutung des Mindestmaßes bezüglich Effizienz und Produktionsqualität

Die Ergebnisse des Hörtests legen nahe, dass ein technisch und ästhetisch überzeugender Vocal-Mix bereits mit einem reduzierten Bearbeitungsaufwand erreicht werden kann – vorausgesetzt, es wird ein gewisses funktionales Mindestmaß an Bearbeitung eingehalten. Dieses Mindestmaß umfasst laut Aussagen der befragten Produzenten grundlegende Maßnahmen wie Klanganpassung durch EQ, dynamische Kontrolle über das Signal via Kompression, eine automatische Intonationskorrektur, De-Essing sowie eine klanglich stimmige Raumsimulation durch Reverb. Diese Bearbeitungen lassen sich – wie die 30-min.-Version im Hörversuch gezeigt hat – mit begrenztem Zeitaufwand umsetzen und führen dennoch zu einem Ergebnis, das in den meisten Fällen professionell wahrgenommen wurde.

Diese Beobachtung gewinnt im Produktionsalltag zunehmend Bedeutung, da Musikproduktionen heute häufig unter hohem Zeitdruck und mit begrenzten Budgets entstehen. Die Erkenntnis, dass ein hoher Produktionsstandard nicht zwangsläufig mit maximalem Bearbeitungsaufwand einhergehen muss, sondern bereits durch eine gezielte Auswahl wirksamer Maßnahmen erreichbar ist, hat direkte Relevanz für effizientes Arbeiten im Studio. Auch in der professionellen Testgruppe wurde mehrfach betont, dass Vocals durch eine gute Performance und eine standardisierte, auf das Signal angepasste Bearbeitungskette in vielen Fällen ausreichen, um konkurrenzfähige Ergebnisse zu erzielen und dass weitere Verfeinerungen nicht zwangsläufig zu einer hörbaren Verbesserung führen.

Das Mindestmaß dient in diesem Zusammenhang nicht als künstlerische Einschränkung, sondern vielmehr als Rahmen zur Orientierung für ökonomische Vocalproduktion ohne Qualitätseinbußen. Besonders in Arbeitsumgebungen mit hohem Output (z. B. Songwriting-Camps, Content-Produktion, Serienformate) stellt es eine praxistaugliche Referenz dar, um zeitlich planbare Workflows mit einem reproduzierbar professionellen Ergebnis zu verbinden.

Ein gezielt festgelegtes Mindestmaß an Bearbeitung kann in der modernen Popmusikproduktion also ausreichend sein. Selbst mit reduziertem Aufwand lassen sich demnach Ergebnisse erzielen, welche dem aktuellen Standard des Musikmarktes entsprechen.

Da sich zwischen den verschiedenen Hörergruppen keine systematisch unterschiedlichen Bewertungen der Bearbeitungsstufen zeigten, wurde auf eine weiterführende Diskussion dieses Aspekts verzichtet.

5.3 Relevanz des Pareto-Prinzips für die Vocal-Bearbeitung

Die Ergebnisse des Hörtests lassen sich in weiten Teilen im Sinne des Pareto-Prinzips interpretieren, welches auch als 80/20-Regel bekannt ist. Dieses besagt, dass in vielen Systemen 80 % des Ergebnisses durch 20 % des Gesamtaufwands erzielt werden. Weitere Erläuterungen hierzu finden sich in Kapitel 2.5.

Übertragen auf die Vocal-Bearbeitung im Kontext moderner Popmusikproduktion bedeutet dies, dass bereits durch gezielte Kernmaßnahmen, wie z.B jene in der 30 min. bearbeiteten Version ein Großteil des klanglichen Eindrucks geprägt wird.

Diese Interpretation wird durch die Beobachtung gestützt, dass zwischen der 30-minütigen und der 120-minütigen Version keine signifikanten Qualitätsunterschiede wahrgenommen wurden. Auch in den Experteninterviews wurde deutlich, dass viele Produzenten nicht jedes Detail der Bearbeitung als hörentscheidend betrachten, sondern gezielt nach dem „größten Hebel“ suchen. Das Pareto-Prinzip bietet damit einen soliden theoretischen Rahmen für die praktische Frage, wie viel Bearbeitung tatsächlich nötig ist und wo zusätzlicher Aufwand nur noch marginale Wirkung entfaltet.

Dennoch muss betont werden, dass das Prinzip nicht als starre Formel missverstanden werden darf. Die Relation von Aufwand und Wirkung kann je nach Genre, Zielsetzung der Produktion und Anspruch variieren. Besonders bei stark stilisierten Produktionen oder bei sehr exponierten Stimmen können Details bewusst hörbar gemacht oder stilistisch überhöht werden. Insofern ist das Pareto-Prinzip eher als Denkmuster zur Priorisierung zu verstehen, nicht als dogmatische Regel.

Für die Produktionspraxis kann dieser Ansatz für Muskschaffende und Produzenten dennoch hilfreich sein. In zeitsensiblen Projekten, im Bildungsbereich oder bei Künstlern, die sich selbst produzieren und

Vermarkten, liefert er eine Orientierung dafür, welche Bearbeitungsschritte zuerst fokussiert werden sollten. Er unterstützt eine differenzierte Auseinandersetzung mit Produktionsentscheidungen und der Balance zwischen Input und Output.

5.4 Methodische Reflexion: Einschränkungen und Verzerrungen der Untersuchung

Trotz der klar definierten Versuchsanordnung und einer konsequenten Auswertung, lassen sich einige methodische Einschränkungen und mögliche Verzerrungen identifizieren, die bei der Einordnung und Interpretation der Ergebnisse berücksichtigt werden sollten.

Es muss beachtet werden, dass der Hörversuch im qualitativen Rahmen mit 15 Menschen durchgeführt wurde. Die Ergebnisse lassen sich aufgrund der geringen Anzahl an Personen, so wie der Methodik selbst, nicht bedingungslos generalisieren.

Ein weiterer wesentlicher Aspekt betrifft die Standardisierung der Abhörsituation. Während ein Großteil der Hörtests – insbesondere mit der Laiengruppe – unter Aufsicht und unter kontrollierten Bedingungen durchgeführt wurde, fanden einzelne Interviews mit professionellen Teilnehmenden remote statt. Diese Remote-Hörtests erfolgten jedoch ausschließlich mit professionellem Equipment, wie hochwertigen Audiointerfaces, Studiokopfhörern und in ruhiger Umgebung. Die verwendete Technik wurde vorab abgefragt bzw. während des Gesprächs spezifiziert. Zwar stand somit eine grundsätzlich hochwertige Abhörumgebung zur Verfügung, dennoch war das eingesetzte Equipment nicht in allen Fällen identisch. Dies hätte zu leichten Unterschieden in der klanglichen Wahrnehmung führen können. Da es im Hörtest jedoch nicht um eine absolute Klangbewertung ging, sondern primär um den relativen Vergleich der Stimuli untereinander, ist davon

auszugehen, dass sich diese Unterschiede nur geringfügig auf die Ergebnisse ausgewirkt haben.

Zudem handelt es sich bei der Versuchssituation um einen künstlich geschaffenen Vergleichsrahmen. Die Teilnehmenden hörten die Stimuli im direkten Vergleich zwischen den Versionen, mit der Möglichkeit des wiederholten Umschaltens. Dies schärft das Bewusstsein für kleinste Abweichungen, entspricht aber nicht dem natürlichen Musikkonsum im Alltag, bei dem solche direkten Vergleiche in der Regel nicht stattfinden. In realen Rezeptionskontexten könnten die wahrgenommenen Unterschiede somit anders gewichtet oder gar nicht bewusst registriert werden.

Ein weiterer zentraler Punkt betrifft die Festlegung der Bearbeitungszeiten. Die Wahl von 30 und 120 min. als fixe Zeitrahmen diene der klaren Operationalisierung des Untersuchungsgegenstands. In der Praxis ist jedoch zu berücksichtigen, dass Produzenten sehr unterschiedlich arbeiten sowohl in Bezug auf Workflow als auch auf Arbeitsgeschwindigkeit. Was ein erfahrener Produzent in 30 min. umsetzt, könnte für einen anderen mehrere Stunden bedeuten, oder umgekehrt. Die gewählten Zeitspannen sind daher unbedingt nicht absolut zu verstehen, sondern als grobe Referenzpunkte, um den Unterschied zwischen einem reduzierten, eher funktional orientierten Bearbeitungsansatz und einer maximal detaillierten Ausarbeitung von Vocals in einer Popmusikproduktion zu erfassen. Eine Generalisierung dieser Zeiten auf andere Produktionssituationen ist daher nur eingeschränkt möglich.

Auch auf Seiten der Stimuli gibt es Einschränkungen. Alle Versionen der jeweiligen Demos basieren auf derselben Vocal-Aufnahme und die Bearbeitung erfolgte ausschließlich innerhalb der definierten Zeitspanne. Jedoch könnte die Auswahl anderer Genres bzw. anderer Songs als Stimuli, sowie die Wahl anderen Recording Equipments subtile Effekte auf die Gesamtwirkung haben. Vor allem der Bezug der Vocals zur restlichen

Instrumentierung eines Songs (Dichte des Arrangements zusätzlich zu den Vocals) kann hier eine Rolle spielen. Auch die psychologische Komponente im Sinne des Persönlichen Musikgeschmacks der Probanden könnte eventuell eine Auswirkung haben. Hinzu kommt, dass die Stimuli zugunsten der Praktikabilität des Hörvergleichs nur auf einen kurzen Ausschnitt des gesamten bearbeiteten Songs beschränkt waren. Auch wenn dieser gezielt gewählt wurde, kann ein kurzer Ausschnitt nicht die vollständige Dynamik und Vielfältigkeit der Vocals eines vollständigen Songs abbilden. Dadurch besteht die Möglichkeit, dass die Wirkung der Bearbeitung in ihrer Rezeption verzerrt bewertet wurde.

Die Tatsache, dass die Stimuli vom Versuchsleiter selbst produziert wurden, birgt darüber hinaus das Risiko einer Bestätigungs- oder Erwartungsverzerrung. Dies kann der Fall sein, wenn etwa unbeabsichtigt Annahmen über „gute“ Bearbeitung in die Produktion eingeflossen sind.

Die Gruppeneinteilung in Laien, Audiophile und Professionals basierte auf funktionalen Kriterien (z. B. Beruf, Ausbildung, Selbstbeschreibung). Diese können jedoch nicht scharf voneinander getrennt werden. Zwischen den Gruppen gab es Überschneidungen in der Hörerfahrung, und innerhalb jeder Gruppe streuten die individuellen Kompetenzen leicht. Dies erschwert eindeutige Aussagen über systematische Gruppenunterschiede.

Die Frage nach dem „Effizienz-Sweetspot“ lässt sich ebenfalls nur schwer pauschal beantworten, da die Bearbeitung oftmals sehr vom jeweiligen Anwendungsfall abhängig ist.

Zuletzt ist die qualitative Analyse der Rückmeldungen interpretationsabhängig. Zwar wurde bei der Auswertung auf thematische Kohärenz geachtet, dennoch ist eine gewisse Subjektivität in der Auswahl und Gewichtung der Aussagen unvermeidlich.

Trotz dieser Einschränkungen bietet die Untersuchung brauchbare Hinweise auf die Wirkung unterschiedlicher Bearbeitungsstufen in der

Popmusikproduktion. Die methodischen Grenzen betreffen vor allem die Übertragbarkeit und Generalisierbarkeit der Ergebnisse – nicht jedoch deren Aussagekraft im Rahmen des gewählten Forschungsdesigns.

6 Fazit

6.1 Zusammenfassung der wichtigsten Erkenntnisse

Diese Arbeit beschäftigte sich mit der Frage, inwieweit der Umfang der Vocal-Bearbeitung Einfluss auf die wahrgenommene Qualität im Mix moderner Popmusik hat. Ausgangspunkt war die Beobachtung, dass in der Musikproduktion häufig großer Aufwand in die Optimierung von Vocals investiert wird, obwohl unklar ist, wie stark sich dieser Mehraufwand tatsächlich auf die Rezeption durch Hörende auswirkt. Um dies zu untersuchen, wurden qualitative Interviews mit erfahrenen Produzenten durchgeführt und ein Hörversuch konzipiert, der unterschiedliche Bearbeitungsgrade systematisch vergleichbar machte.

Die Ergebnisse zeigen, dass bereits eine reduzierte Vocal-Bearbeitung – bestehend aus gezielten Eingriffen wie EQ, Kompression, De-Essing, automatischer Intonationskorrektur und räumlicher Einbettung – ausreicht, um einen professionellen Gesamteindruck zu erzeugen. Die Unterschiede zwischen unbearbeiteten und bearbeiteten Versionen waren für alle Hörergruppen deutlich wahrnehmbar, während die Unterscheidung zwischen der 30-minütigen und der 120-minütigen Bearbeitung häufig schwerfiel oder nicht eindeutig war.

Die Einschätzungen der Produzenten bestätigten diese Beobachtung: Viele nannten gezielte Kernmaßnahmen welche ihrer Erfahrung nach die größten klanglichen Auswirkungen haben und betrachteten viele Detailbearbeitungen eher als stilistische oder workflowbezogene Entscheidungen, denn als hörbare Qualitätsfaktoren. Besonders deutlich wurde dies beim Thema manuelle Intonationskorrektur und präzises Editing, das aus professioneller Sicht sinnvoll, aus Hörerperspektive aber kaum messbar ist.

Insgesamt ergibt sich daraus ein nicht-lineares Verhältnis zwischen Aufwand und Ergebnis. Der Großteil der wahrgenommenen Qualität kann offenbar mit einem relativ geringen Zeit- und Arbeitseinsatz erreicht werden, während zusätzlicher Aufwand zunehmend abnehmende Wirkung entfaltet. Dieses Muster lässt sich gut durch das Pareto-Prinzip beschreiben, das als theoretischer Bezugsrahmen zur Bewertung von Effizienz in Produktionsprozessen herangezogen werden kann.

6.2 Beantwortung der Forschungsfragen

Die zentrale Forschungsfrage lautete:

Welchen Einfluss hat der Umfang der Vocal-Bearbeitung in der modernen Popmusikproduktion auf die wahrgenommene Qualität der Vocals im Gesamtmix?

Antwort:

Ein gewisses Maß an Bearbeitung führte im Rahmen der Untersuchung zu einem signifikanten Anstieg der wahrgenommenen Qualität im Vergleich zu unbearbeiteten Vocals. Ein größerer Umfang der Vocal-Bearbeitung zeigte hingegen ab einem bestimmten Punkt keine signifikante Auswirkung mehr auf die wahrgenommene Qualität im Mix. Alle Testgruppen konnten den Unterschied zwischen unbearbeitetem und bearbeitetem Material eindeutig erkennen. Die Unterscheidung zwischen „kurz“ und „detailliert“ bearbeiteten Vocals blieb jedoch oft unscharf oder subjektiv.

Der Umfang der Bearbeitung hat somit einen Einfluss auf die wahrgenommene Qualität – dieses Verhältnis scheint jedoch nicht linear zu verlaufen.

Erste Teilfrage

Wie deutlich wird Vocal-Bearbeitung von Hörenden als qualitätssteigernd wahrgenommen?

Generell wird Vocal-Bearbeitung von allen Untersuchten Hörergruppen sehr deutlich wahrgenommen. Vor allem grundlegende Bearbeitungen wie EQ, Kompression, Intonationskorrektur und Reverb scheinen hier eine Rolle zu spielen. Alle befragten des Hörversuchs bewerteten einstimmig die unbearbeitete Version als qualitativ am schlechtesten. Die grundsätzliche Vocal-Bearbeitung scheint aufgrund des signifikanten Qualitätsunterschiedes als notwendig.

Zweite Teilfrage

Inwiefern unterscheiden sich die Wahrnehmungen zwischen unterschiedlichen Hörergruppen?

Es zeigten sich keine systematisch gravierenden Unterschiede zwischen Laien, audiophilen Hörenden und Professionals. Alle Gruppen konnten grobe Unterschiede klar erkennen, die Bewertungen der beiden bearbeiteten Versionen lagen jedoch weitgehend auf vergleichbarem Niveau.

Dritte Teilfrage

Wo liegt möglicherweise ein „Sweet Spot“ zwischen Aufwand und hörbarem Mehrwert?

Die Ergebnisse deuten darauf hin, dass bereits mit einem gezielten, zeitlich begrenzten Bearbeitungsansatz ein professionell wahrgenommener Vocal-Sound erzeugt werden kann. Darüberhinausgehende Maßnahmen verbessern zwar möglicherweise den technischen Standard der Produktion, sowie das Gewissen für „sauberes Arbeiten“, führen aber nicht zwingend zu einer wahrnehmbar höheren Qualität. Der „Sweet Spot“ liegt demnach dort,

wo Aufwand und Wirkung in einem funktional günstigen Verhältnis stehen. So zum Beispiel bei der Umsetzung weniger, aber wirkungsvoller Schritte, welche hier durch das Mindestmaßrepräsentiert werden.

Vierte Teilfrage

Inwieweit lässt sich der Zusammenhang zwischen Bearbeitungsaufwand und Wirkung im Sinne des Pareto-Prinzips interpretieren?

Aufgrund des offensichtlich nicht-linearen Zusammenhanges zwischen Input und Output, lässt sich eine deutliche parallele zum Pareto-Prinzip ableiten. Wie sich dieses in der Praxis anwenden lässt, ist jedoch stark von der individuellen Situation abhängig. Eine ausführliche Erörterung findet sich in Kapitel 5.3.

6.3 Ausblick auf zukünftige Forschungsmöglichkeiten.

Die vorliegende Arbeit liefert erste Hinweise darauf, wie sich unterschiedliche Bearbeitungsstufen von Vocals auf deren Wahrnehmung in einer Mischung auswirken. Dabei wurde deutlich, dass ein funktionales Mindestmaß an Bearbeitung in vielen Fällen bereits ausreicht, um einen professionellen Höreindruck zu erzeugen. Gleichzeitig ergeben sich aus der Untersuchung eine Reihe von offenen Fragen und Anknüpfungspunkten, die im Rahmen zukünftiger Forschung weiterverfolgt werden könnten.

Ein zentraler Ansatzpunkt betrifft die Erweiterung und Diversifizierung der Stimuli. Die vorliegenden Hörtests basierten auf zwei Songs und jeweils einem daraus ausgewählten Ausschnitt. In zukünftigen Studien könnten unterschiedliche Songabschnitte (z. B. Strophen, Refrains, Bridges) oder verschiedene Genres gezielt miteinander verglichen werden. Auch Unterschiede in der stimmlichen Charakteristik könnten eine wichtige Rolle spielen und zu einer Generalisierbarkeit der Ergebnisse beitragen.

Zum Beispiel könnten Unterschiede zwischen männlichen und weiblichen Stimmen, verschiedenen Gesangsstilen oder Ausdrucksformen (z. B. Popgesang vs. Rap, akustisch vs. elektronisch) untersucht werden.

Darüber hinaus wäre es sinnvoll, die Ansätze der Bewertungen und Analysen zu erweitern. Während in dieser Arbeit eine qualitative Analyse auf Basis subjektiver Einschätzungen im Vordergrund stand, könnten zukünftige Studien quantitative Methoden einsetzen. Geeignet wären etwa, Doppelblindtests mit Auswertung über standardisierte Bewertungsskalen. Auch neurophysiologische oder kognitive Messansätze wie etwa Blickverlaufsmessung, Reaktionszeittests könnten ergänzende Einsichten liefern, insbesondere im Hinblick auf unbewusste Wahrnehmungseffekte.

Zu Beginn dieser Arbeit stand die untergeordnete Forschungsfrage „*Welche Bearbeitungsschritte werden von Hörenden als qualitätssteigernd wahrgenommen?*“ im Raum. Diese wurde aufgrund des begrenzten Umfangs der Arbeit gestrichen. Daher wäre ein weiterer Forschungsimpuls, die Wirkung einzelner Bearbeitungsschritte gezielt zu isolieren. Während in dieser Studie komplette vordefinierte Bearbeitungsketten verglichen wurden, wäre es denkbar, spezifische Parameter wie ausschließlich Intonationskorrektur, Time-Alignment oder Kompression in kontrollierter Variation zu untersuchen. Dadurch ließen sich Rückschlüsse auf die Relevanz einzelner Werkzeuge in Bezug auf die Wahrgenommene Qualität gewinnen. Es könnte sich für weitere Forschung auf diesem Gebiet als eminent erweisen, exakt zu klären, welche Tools als Qualitätssteigernd wahrgenommen werden.

Auch eine stärkere Orientierung an realen Produktionskontexten wäre sehr interessant. Das ließe sich z.B durch den Vergleich kommerziell veröffentlichter Tracks mit stark variierendem Bearbeitungsaufwand realisieren. Dazu müssten Projektdaten von professionell produzierten Songs vorliegen, um diese miteinander vergleichen zu können.

Eine solche Analyse verschiedener Songs könnte helfen, die Wichtigkeit des Einsatzes bestimmter Tools aus Sicht der Praxis noch genauer zu beurteilen. Ergänzend dazu wäre auch ein Langzeitvergleich interessant, bei dem untersucht wird, wie sich Bearbeitungsintensitäten auf Wiedererkennungswert, emotionale Bindung oder Streaming-Performance eines Songs auswirken.

Auch aktuelle Entwicklungen im Bereich künstlicher Intelligenz (KI) eröffnen neue Möglichkeiten und Herausforderungen: Tools wie KI-gestützte Vocal-Chains, Mixing-Systeme oder AI-basierte Stimmrekonstruktionen und Stimmerzeuger ermöglichen eine immer schnellere, präzisere und teils stilisierte Bearbeitung – mit potenziell geringerem Zeitaufwand. Zukünftige Studien könnten untersuchen, ob diese Verfahren das Effizienz-Wirkungs-Verhältnis weiter verschieben und ob die Resultate mit traditionellen Bearbeitungsschritten vergleichbar oder vielleicht überlegen sind. Auch das könnte im Hinblick auf subjektive Wahrnehmung durch verschiedene Zielgruppen interessant sein.

Auch aus pädagogischer Perspektive ergeben sich relevante Anschlussfragen. Für angehende Produzierende, Studierende und Lehrende im Bereich Musikproduktion stellt sich zunehmend die Herausforderung, sich bei einer Vielzahl an Produktionsmöglichkeiten auf das Wesentliche zu fokussieren. Ein besseres Verständnis dafür, welche Bearbeitungsschritte in der Praxis wirklich entscheidend sind, könnte dazu beitragen, Ausbildungsprozesse zielgerichteter zu gestalten und produktivere Workflows zu vermitteln. Die Ergebnisse dieser Arbeit legen nahe, dass eine fundierte Reduktion auf wesentliche Tools und Entscheidungen nicht nur ausreicht, sondern mitunter effizienter zum Ziel führen kann als maximaler Bearbeitungsaufwand.

Nicht zuletzt könnte auch der Diskurs um Klangästhetik, Wahrnehmung und Produktionsethik neu befeuert werden: In welchem Maße soll Klang „optimiert“ werden, und wo beginnt kreative oder sogar künstliche Überformung?

Insgesamt zeigt sich, dass die Frage nach der „richtigen“ oder „ausreichenden“ Vocal-Bearbeitung nicht nur ästhetisch, sondern auch ökonomisch und rezeptionspsychologisch hoch relevant ist. Die vorliegenden Ergebnisse stellen einen explorativen Einstieg in ein bislang kaum systematisch untersuchtes Feld dar. Sie liefern damit eine Grundlage und zahlreiche Anhaltspunkte für weiterführende Forschung im Bereich von Musikproduktion, Wahrnehmungsästhetik sowie Konsum und Nutzung von Audiovisuellen Medien.

7 Quellenverzeichnis

- [1] G. Gensch, E. M. Stöckler, und P. Tschmuck, Hrsg., *Musikrezeption, Musikdistribution und Musikproduktion: Der Wandel des Wertschöpfungsnetzwerks in der Musikwirtschaft*. in SpringerLink Bücher. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH, Wiesbaden, 2009. doi: 10.1007/978-3-8349-8045-8.
- [2] A. Endreß und H. Wandjo, Hrsg., *Musikwirtschaft im Zeitalter der Digitalisierung: Handbuch für Wissenschaft und Praxis*. Nomos Verlagsgesellschaft mbH & Co. KG, 2021. doi: 10.5771/9783845276939.
- [3] R. Seifert, *Popmusik in Zeiten der Digitalisierung: Veränderte Aneignung - veränderte Wertigkeit*, 1st ed. in Studien zur Populärmusik. Bielefeld: Transcript Verlag, 2018.
- [4] M. Pfeiderer, „Vocal pop pleasures. Theoretical, analytical and empirical approaches to voice and singing in popular music“, *IASPM@Journal*, Bd. 1, Nr. 1, S. 1–16, Apr. 2011, doi: 10.5429/2079-3871(2010)v1i1.7en.
- [5] A. Daniel und F. Hillebrandt, Hrsg., *Die Praxis der Popmusik: Soziologische Perspektiven*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019. doi: 10.1007/978-3-658-22714-2.
- [6] K. Gerdes und K. Siedenburg, „Lead-vocal level in recordings of popular music 1946–2020“, *JASA Express Lett.*, Bd. 3, Nr. 4, S. 043201, Apr. 2023, doi: 10.1121/10.0017773.
- [7] Equipment-Geschichte, „6 Echo Chambers That Shaped the Sound of Pop Music“, reverb.com. Zugegriffen: 17. Juni 2025. [Online]. Verfügbar unter: <https://reverb.com/news/6-echo-chambers-that-shaped-the-sound-of-popular-music>
- [8] deutschlandfunkkultur.de, „Streaming-Ökonomie 2020 - Im Rausch der Aufmerksamkeit“, Deutschlandfunk Kultur. Zugegriffen: 4. Juni 2025. [Online]. Verfügbar unter: <https://www.deutschlandfunkkultur.de/streaming-oekonomie-2020-im-rausch-der-aufmerksamkeit-100.html>
- [9] K. Pisanski, J. Raine, und D. Reby, „Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries“, *R. Soc. Open Sci.*, Bd. 7, Nr. 2, S. 191642, Feb. 2020, doi: 10.1098/rsos.191642.
- [10] S. R, A. D. S, S. B. M, S. M. P, N. Sharma, und P. Singh, „Enhancing Singing Performances: Novel Method for Automatic Vocal Pitch Correction“, in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India: IEEE, Sep. 2023, S. 1095–1102. doi: 10.1109/ICOSEC58147.2023.10276251.
- [11] L. Bohn, *Moderne Musikproduktion in der Praxis: die besten Tipps und Tricks für Produzenten von Pop, Dance und Hip Hop*, 3. Auflage. Bergkirchen: PPVMEDIEN, 2020.
- [12] *Handbuch der Tonstudioteknik*, 8., Überarb. und erw. Aufl. in Handbuch der Tonstudioteknik. Berlin: De Gruyter Saur, 2014.

- [13] S. Weinzierl, Hrsg., *Handbuch der Audiotechnik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-34301-1.
- [14] M. Schwarzer, „Popmusik: Warum Songs immer kürzer werden – und was Tiktok damit zu tun hat“. Zugegriffen: 4. Juni 2025. [Online]. Verfügbar unter: <https://www.rnd.de/kultur/popmusik-warum-songs-immer-kuerzer-werden-und-was-tiktok-damit-zu-tun-hat-K6WA2YZDEBFLN6SNUD5IZJ2GU.html>
- [15] „Wie Streaming den Sound der Popmusik beeinflusst“. Zugegriffen: 4. Juni 2025. [Online]. Verfügbar unter: <https://www.fr.de/wissen/streaming-sound-popmusik-beeinflusst-10938174.html>
- [16] M. Anastasiadis, *Social-Pop-Media: Zur Mediatisierung von Popmusik und digitaler Musikpraxis in sozialen Medien*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019. doi: 10.1007/978-3-658-27882-3.
- [17] Produzent 6, „Experteninterview Musikproduzenten mit Produzent 6“, 1. April 2025.
- [18] Produzent 2, „Experteninterview Musikproduzenten mit Produzent 2“, 25. Februar 2025.
- [19] Produzent 5, „Experteninterview Musikproduzenten mit Produzent 5“, 2. März 2025.
- [20] Produzent 4, „Experteninterview Musikproduzenten mit Produzent 4“, 4. März 2025.
- [21] Produzent 3, „Experteninterview Musikproduzenten mit Produzent 3“, 17. März 2025.
- [22] Produzent 1, „Experteninterview Musikproduzenten mit Produzent 1“, 25. Februar 2025.
- [23] A. Körner und F. Strack, „Articulation posture influences pitch during singing imagery“, *Psychon. Bull. Rev.*, Bd. 30, Nr. 6, S. 2187–2195, Dez. 2023, doi: 10.3758/s13423-023-02306-1.
- [24] J. Gilder, „The Psychology Of Studio Recording“, ProSoundWeb. Zugegriffen: 17. Juni 2025. [Online]. Verfügbar unter: <https://www.prosoundweb.com/the-psychology-of-recording/>
- [25] M. Senior, *Mixing secrets for the small studio*. Burlington, MA: Focal Press, 2011.
- [26] S. Fortune, „What Is Direct Monitoring Audio Interface“, AudioLover. Zugegriffen: 17. Juni 2025. [Online]. Verfügbar unter: <https://audiolover.com/production-technology/audio-interface/what-is-direct-monitoring-audio-interface/>
- [27] „Direct monitoring with effects“, Antelope Audio. Zugegriffen: 17. Juni 2025. [Online]. Verfügbar unter: <https://en.antelopeaudio.com/2022/04/direct-monitoring-with-effects-and-the-synergy-core-advantage/>
- [28] „Gain Staging In Your DAW Software“. Zugegriffen: 17. Juni 2025. [Online]. Verfügbar unter: https://www.soundonsound.com/techniques/gain-staging-your-daw-software?utm_source=chatgpt.com

- [29] *Recording Engineer's Handbook*, 2nd ed. Boston: Course Technology / Cengage Learning, 2014.
- [30] M. Neal, „The Little-Known Recording Trick That Makes Singers Sound Perfect“, VICE. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://www.vice.com/en/article/the-little-known-recording-trick-that-makes-singers-sound-perfect/>
- [31] M. Monster, „Vocal Comping Guide 2025 | Compiling Vocals Like A Pro“. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: <https://mixingmonster.com/vocal-comping/>
- [32] „A Brief History of Vocal Doubling“. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://www.izotope.com/en/learn/a-brief-history-of-vocal-doubling>
- [33] A. Kody, „Mastering the art of vocal layering: a step-by-step guide“, Native Instruments Blog. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://blog.native-instruments.com/vocal-layering/>
- [34] A. Putri Puspitasari und Kasiyan, „The Ethical Dilemma of Pitch Correction and Its Influence on Singer Quality in the Popular Music Industry“, *Int. J. Multicult. Multireligious Underst.*, Feb. 2025.
- [35] M. Skidmore, „Pitch-Correcting Vocals With Melodyne“. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: https://www.soundonsound.com/techniques/pitch-correcting-vocals-melodyne?utm_source=chatgpt.com
- [36] J. Kedves, „20 Jahre Auto-Tune: Das Öl, die Cher und der Heartbeat“, Musikexpress. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: <https://www.musikexpress.de/20-jahre-auto-tune-das-oel-die-cher-und-der-heartbeat-1147739/>
- [37] S. Team, „Common Pitch Correction Tools for Home Producers“, Flypaper. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://flypaper.soundfly.com/produce/common-pitch-correction-tools-for-home-producers/>
- [38] D. Miraglia, „Vocal EQ Chart: The Ultimate Vocal EQ Cheat Sheet (2024)“, Unison. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: <https://unison.audio/eq-chart/>
- [39] V. Smyrek, *Tontechnik für Veranstaltungstechniker in Ausbildung und Praxis: mit 703 Abbildungen, 74 Tabellen und 152 Übungsaufgaben*, 3., Korrigierte Auflage. Stuttgart: S. Hirzel Verlag, 2016.
- [40] U. Zölzer, *Digitale Audiosignalverarbeitung*, 4., erw. Auflage 2022. Wiesbaden: Springer Fachmedien Wiesbaden GmbH Springer Vieweg, 2022.
- [41] O. A. Somefun, K. F. Akingbade, und F. M. Dahunsi, „Uniformly-Damped Binomial Filters: Five-percent Maximum Overshoot Optimal Response Design“, 2020, doi: 10.48550/ARXIV.2007.00890.
- [42] „Was ist Zischlaute in Audio?“ Zugegriffen: 8. Juli 2025. [Online]. Verfügbar unter: <https://www.antarestech.com/community/what-is-sibilance-in-audio>

- [43] M. Schweitzer, „Gesang abmischen: 5 elementare Frequenzen für gute Vocals“, delamar.de. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://www.delamar.de/mixing/gesang-abmischen-die-fuenf-elementaren-frequenzen-fuer-gute-vocals-4330/>
- [44] „Dynamische Equalizer“, Musikhaus Thomann. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: https://www.thomann.de/at/onlineexpert_page_mastering_multiband_equalizer.html
- [45] „Diese vier Kompressor-Arten musst Du kennen“, Bonedo. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://www.bonedo.de/artikel/diese-vier-kompressor-arten-musst-du-kennen/>
- [46] „How to NAIL Vocal Compression: One Compressor or Two? - Waves Audio“, waves.com. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://www.waves.com/how-to-nail-vocal-compression>
- [47] R. Mayzes, „Vocal Compression: Learn How to Mix Like the Pros, 06.06.2016“. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: https://mastering.com/vocal-compression-how-to-compress-vocals/?utm_source=chatgpt.com
- [48] E. Keeley, „What is a Limiter in Audio? When & How to use in Mastering, 19.07.2018“. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: <https://www.masteringbox.com/de/learn/audio-limiter>
- [49] J. Sturgis, „Limiting Versus Compressing Vocals, 15.05.2019“, ProSoundWeb. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: <https://www.prosoundweb.com/limiting-versus-compressing-vocals/>
- [50] D. Miraglia, „De-essing 101: Get Pristine Audio Quality Every Single Time“, Unison. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: <https://unison.audio/de-essing/>
- [51] N. Rogers, „Saturation Strategies, 05.2021“. Zugegriffen: 5. Mai 2025. [Online]. Verfügbar unter: https://www.soundonsound.com/techniques/saturation-strategies?utm_source=chatgpt.com
- [52] R. Koch, *The 80/20 Principle, Third Edition: The Secret to Achieving More with Less*. Westminster: Crown/Archetype, 2011.
- [53] J. M. Juran, Hrsg., *Juran's quality control handbook*, 4. ed., Rev. New York: McGraw-Hill, 1988.
- [54] M. Newman, „Power laws, Pareto distributions and Zipf's law“, *Contemp. Phys.*, Bd. 46, Nr. 5, S. 323–351, Sep. 2005, doi: 10.1080/00107510500052444.
- [55] A. Przyborski und M. Wohlrab-Sahr, „Forschungsdesigns für die qualitative Sozialforschung“, in *Handbuch Methoden der empirischen Sozialforschung*, N. Baur und J. Blasius, Hrsg., Wiesbaden: Springer Fachmedien Wiesbaden, 2019, S. 105–123. doi: 10.1007/978-3-658-21308-4_7.

- [56] M. Rewak, „How Loud Should You Master Your Music? (2024 Guide)“, Blog | Splice. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://splice.com/blog/how-loud-master-music/>
- [57] „Mastering trends in 2024: what you should be doing“. Zugegriffen: 19. Juni 2025. [Online]. Verfügbar unter: <https://www.izotope.com/en/learn/mastering-trends>
- [58] P. Mayring, *Qualitative Inhaltsanalyse: Grundlagen und Techniken*, 13., Überarbeitete Auflage. Weinheim Basel: Beltz, 2022.
- [59] *Si Seulement Tessa B Deconstructing a Mix #57 - Videos - Mix With The Masters*, (17. Juli 2025). Zugegriffen: 29. Juli 2025. [Online Video]. Verfügbar unter: <https://mixwiththemasters.com/de/videos/jaycen-joshua-tessa-b-si-seulement>
- [60] *BIRDS OF A FEATHER Billie Eilish Inside the Track #123 - Videos - Mix With The Masters*, (18. März 2025). Zugegriffen: 29. Juli 2025. [Online Video]. Verfügbar unter: <https://mixwiththemasters.com/de/videos/jon-castelli-aron-forbes-billie-eilish-birds-of-a-feather>
- [61] A. Barna und C. McLaughlin, „Vocal Production, Mimesis, and Social Media in Bedroom Pop“, *Music Theory Online*, Bd. 30, Nr. 4, Dez. 2024, doi: 10.30535/mt0.30.4.1.
- [62] Profi 1, „Transkript: Interview Profi 1 (Gruppe: Professionals)“, 22. April 2025.
- [63] Profi 4, „Transkript: Interview Profi 4 (Gruppe: Professionals)“, 19. April 2025.
- [64] Laie 5, „Transkript: Interview Laie 5 (Gruppe: Laien)“, 19. April 2025.
- [65] Laie 2, „Transkript: Interview Laie 2 (Gruppe: Laien)“, 14. April 2025.
- [66] Audiophile Person 2, „Transkript: Interview Audiophile Person 2 (Gruppe: Audiophile)“, 15. April 2025.
- [67] Profi 5, „Transkript: Interview Profi 5 (Gruppe: Professionals)“, 15. April 2025.
- [68] Profi 2, „Transkript: Interview Profi 2 (Gruppe: Professionals)“, 19. April 2025.
- [69] Profi 3, „Transkript: Interview Profi 3 (Gruppe: Professionals)“, 18. April 2025.
- [70] Audiophile Person 1, „Transkript: Interview Audiophile Person 1 (Gruppe: Audiophile)“, 7. Mai 2025.
- [71] Laie 1, „Transkript: Interview Laie 1 (Gruppe: Laien)“, 19. April 2025.
- [72] Laie 3, „Transkript: Interview Laie 3 (Gruppe: Laien)“, 8. Mai 2025.
- [73] Laie 4, „Transkript: Interview Laie 4 (Gruppe: Laien)“, 14. April 2025.
- [74] Audiophile Person 5, „Transkript: Interview Audiophile Person 5 (Gruppe: Audiophile)“, 10. April 2025.
- [75] Audiophile Person 3, „Transkript: Interview Audiophile Person 3 (Gruppe: Audiophile)“, 7. Mai 2025.

Digitaler Anhang

Aufgrund des Umfangs der erhobenen Materialien, insbesondere der vollständigen Interviewtranskripte, Audio-Demos und Dokumente zum Hörversuch - wurden diese als digitale Anlagen auf einem separaten Datenträger beigefügt. Diese Vorgehensweise ermöglicht eine strukturierte, vollständige und zugleich ressourcenschonende Abgabe der Arbeit, ohne unnötigen Papierverbrauch zu verursachen.

Sämtliche digitalen Anlagen wurden in der angegebenen Reihenfolge dem beiliegenden Speichermedium hinzugefügt.

Digitaler Anhang 1: Experteninterviews zur Ermittlung des Mindestmaßes

Digitaler Anhang 2: Audioaufzeichnungen der Interviews zum Hörversuch

Digitaler Anhang 3: Transkripte der Interviews des Hörversuchs

Digitaler Anhang 4: Demomaterial des Hörversuchs

Digitaler Anhang 5: Aussagen der Testpersonen des Hörversuchs Stichpunktartig dargestellt. (Dient als Übersicht. Statistisch nicht repräsentativ).