



Hochschule der Medien

Bachelorthesis

im Studiengang Audiovisuelle Medien

Traditionelles und KI-generiertes Sounddesign im Vergleich anhand ausgewählter Filmproduktionen

vorgelegt von:
Louis Kuschnir

zur Erlangung des akademischen Grades
Bachelor of Engineering

Ausgabedatum: 17.12.2024

Abgabedatum: 16.03.2025

Erstgutachter: Prof. Michael Felten
Zweitgutachter: Prof. Oliver Curdt

Ehrenwörtliche Erklärung

Hiermit versichere ich, Louis Kuschnir, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Traditionelles und KI-generiertes Sounddesign im Vergleich anhand ausgewählter Filmproduktionen“ selbstständig und ohne fremde Hilfe verfasst habe und keine anderen als die angegebenen Hilfsmittel benutzt wurden.

Alle Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind unter Angabe der Quelle kenntlich gemacht.

Zur sprachlichen und strukturellen Unterstützung, so wie für die Recherche wurde ChatGPT verwendet. Die wissenschaftlichen Inhalte, Analysen und Schlussfolgerungen wurden eigenständig erarbeitet, geprüft und kritisch reflektiert.

Die Arbeit wurde weder veröffentlicht noch in anderer Form als Prüfungsleistung vorgelegt. Ich habe die Bedeutung der ehrenwörtlichen Versicherung sowie die prüfungsrechtlichen Folgen (§ 24 Abs. 2 Bachelor-SPO, § 23 Abs. 2 Master-SPO (Vollzeit)) einer unrichtigen oder unvollständigen Versicherung zur Kenntnis genommen.

Unterschrift 

Kurzfassung

Diese Bachelorarbeit untersucht den Vergleich zwischen traditionellem und KI-gestütztem Sounddesign in der Filmproduktion. Ziel ist es, die Unterschiede in kreativen Prozessen, Qualität, Effizienz und Flexibilität beider Ansätze herauszuarbeiten. Dabei wird analysiert, wie Künstliche Intelligenz die Erstellung von Sounddesigns erleichtert und welche Potenziale sowie Grenzen sie im Vergleich zu traditionellen Methoden bietet.

Der theoretische Teil der Arbeit erläutert die Grundlagen des Sounddesigns und der Künstlichen Intelligenz, während im praktischen Teil Filmbeispiele, wie der Klassiker *Rocky* und ein KI-generierter Boxfilm, gegenübergestellt werden. Aspekte wie Produktionszeit, Klangqualität und die emotionale Wirkung auf das Publikum werden bewertet. Abschließend wird diskutiert, wie eine Kombination aus beiden Ansätzen zukünftige kreative Prozesse in der Filmindustrie bereichern könnte.

Die Arbeit zeigt, dass KI-gestütztes Sounddesign eine vielversprechende Ergänzung darstellt, jedoch die menschliche Kreativität und Expertise nicht vollständig ersetzen kann.

Abstract

This bachelor thesis explores the comparison between traditional and AI-driven sound design in film production. The objective is to analyze the differences in creative processes, quality, efficiency, and flexibility between the two approaches. It examines how artificial intelligence is revolutionizing the creation of sound designs and highlights the potentials and limitations of AI compared to traditional methods.

The theoretical framework provides an overview of sound design and artificial intelligence, while the practical section contrasts examples such as the classic *Rocky* and an AI-generated boxing film. Key aspects like production time, sound quality, and emotional impact on the audience are evaluated. Finally, the discussion addresses how a combination of both approaches could enhance future creative processes in the film industry.

The study concludes that AI-driven sound design offers promising opportunities as a complement to traditional methods but cannot fully replace human creativity and expertise.

Inhaltsverzeichnis

Ehrenwörtliche Erklärung	I
Zusammenfassung	II
Abstract	III
Inhaltsverzeichnis	IV
Abkürzungsverzeichnis	V
Abbildungsverzeichnis	VI
Medienverzeichnis	VII
1. Einleitung.....	1
1.1 Was ist Sounddesign.....	2
1.2 Warum ist Sounddesign ein zentraler Bestandteil von Filmproduktionen?.....	2
1.3 Zielsetzung der Arbeit: Untersuchung und Vergleich von traditionellem und KI-generiertem Sounddesign.....	2
1.4 Forschungsfragen.....	3
1.5 Aufbau der Arbeit.....	3
2. Traditionelles Sounddesign: Wie wurde bisher gearbeitet?.....	4
2.1 Definition und Bedeutung von Sounddesign in der Filmproduktion.....	4
2.2 Arbeitsmethoden im traditionellen Sounddesign.....	5
2.3 Rolle von Foley, Musik und atmosphärischen Klängen.....	6
2.4 Beispiele aus klassischen Filmproduktionen.....	7
2.5 Herausforderungen und Grenzen traditioneller Methoden.....	9
3. Grundlagen der Künstlichen Intelligenz (KI).....	11
3.1 Was ist Künstliche Intelligenz?.....	11
3.2 Technologien wie Machine Learning und Deep Learning im Kontext der Audiogenerierung.....	13
3.3 Anwendungsbereiche von KI in der kreativen Industrie.....	14
4. KI-generiertes Sounddesign: Wie verändert sich der Prozess?.....	17
4.1 Einführung in KI im Sounddesign.....	17
4.2 Technologien und Tools, die KI nutzen.....	17
4.2.1 StabilityAI/Stable Audio Open: Generative Audio-Erstellung.....	17
4.2.2 Meta AudioGen.....	20
4.2.3 Suno und AVIA: KI-basierte Musikkomposition.....	21
4.2.4 ElevenLabs: Sprachsynthese und Soundgeneration.....	23
4.3 Symbolische und nicht-symbolische Audiogenerierung.....	25
4.4 Vorteile und neue Möglichkeiten durch KI.....	26
4.5 Herausforderungen und technische Grenzen von KI im Sounddesign.....	28

5. Unterschiede zwischen traditionellem und KI-gestütztem Sounddesign: Was macht den Unterschied?.....	29
5.1 Vergleich der Herangehensweisen.....	29
5.1.2 Zeitaufwand und Produktionskosten.....	29
5.1.1 Kreative Freiheit und menschlicher Input vs. Automatisierung und Algorithmen.....	30
5.1.2 Zeitaufwand und Produktionskosten.....	30
5.2 Qualitätsvergleich: Wie wirken die Ergebnisse auf das Publikum?.....	30
5.3 Technische Analyse: Klangqualität, Vielfalt und Anpassungsfähigkeit.....	31
6. Filmbeispiele: Wie schneiden Tradition und KI im direkten Vergleich ab?.....	33
6.1 Beschreibung der analysierten Filmszenen.....	33
6.1.1 Klassischer Boxfilm („Rocky“).....	33
6.1.2 KI-generierter Boxfilm (Studioproduktion).....	34
6.2.2 KI-generierte Sounds und ihre technische Umsetzung.....	37
6.3 Ergebnisvergleich.....	39
6.3.1 Effizienz.....	39
6.3.2 Qualität.....	40
6.3.3 Flexibilität.....	41
6.3.4 Anwendungsbereiche.....	42
6.4 Fazit zum Vergleich von traditionellem und KI-gestütztem Sounddesign.....	43
7. Potenziale und Grenzen: Was kann KI im Sounddesign leisten?.....	45
7.1 Chancen für die Filmindustrie.....	45
7.1.1 Effizienzsteigerung und Kostenreduktion.....	45
7.1.2 Neue kreative Möglichkeiten durch KI.....	45
7.2 Grenzen der KI.....	46
7.2.1 Verlust der menschlichen Kreativität?.....	48
7.2.2 Ethische Fragen und Herausforderungen.....	48
8. Fazit und Ausblick: Wohin geht die Reise?.....	50
8.1 Zusammenfassung der zentralen Erkenntnisse.....	50
8.2 Zukunftsaussichten.....	51
8.2.1 Wie wird KI das Sounddesign langfristig beeinflussen?.....	52
8.2.2 Zusammenarbeit von Mensch und Maschine im kreativen Prozess.....	54
8.3 Schlussbetrachtung.....	55
9. Literaturverzeichnis.....	VIII
9.1 Internetquellen.....	IX

Abkürzungsverzeichnis

- AI - Artificial Intelligence
- AR - Augmented Reality
- CNNs - Convolutional Neural Networks
- DAW - Digital Audio Workstation
- DL - Deep Learning
- GANs - Generative Adversarial Networks
- KI - Künstliche Intelligenz
- ML - Machine Learning
- RNNs - Recurrent Neural Networks
- TTS - Text-to-speech
- VR - Virtual Reality

Abbildungsverzeichnis

Abbildung 1: Vereinfachte Darstellung eines neuronalen Netzes (Quelle: Wikipedia, https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz#/media/Datei:Artificial_neural_network.svg) (Zugriff am: 24. Januar 2025)

Abbildung 2: Benutzeroberfläche von Stable Audio Open (Quelle: Screenshot von <https://stableaudio.com/generate>, Zugriff am 22. Januar 2025).

Abbildung 3: Benutzeroberfläche von Stable-Audio-Open-Zero (Quelle: Screenshot von <https://huggingface.co/spaces/artificialguybr/Stable-Audio-Open-Zero>, Zugriff am 22. Januar 2025).

Abbildung 4: Benutzeroberfläche von Suno (Quelle: Screenshot von <https://suno.com/>, Zugriff am 22. Januar 2025)

Abbildung 5: Benutzeroberfläche von AVIA (Quelle: Screenshot von <https://creators.aiva.ai/>, Zugriff am 22. Januar 2025)

Abbildung 6: Benutzeroberfläche von ElevenLabs (Quelle: Screenshot von <https://elevenlabs.io/app/sound-effects/generate>, Zugriff am 22. Januar 2025)

Abbildung 7: Übersicht von AudioLDM design für tex-zu-audio generation (Quelle: Screenshot von <https://audioldm.github.io/>, Zugriff am 02 März 2025)

Abbildung 8: Text-basierte Audio Manipulation (Quelle: Screenshot von <https://audioldm.github.io/>, Zugriff am 02 März 2025)

Abbildung 9: Benutzeroberfläche von Soothe 2 (Quelle: <https://www.gearnews.de/soothe-2-alternativen-vsts-kostenlos-guenstiger-und-besser/>, Zugriff am 02.03.2025)

Abbildung 10: Ergebnisse der Umfrage zu den Grenzen von KI-gestütztem Sounddesign (Quelle: S. Oh, M. Kang, H. Moon, K. Choi, and B. S. Chon, "A Demand-Driven Perspective on Generative Audio AI," *arXiv preprint arXiv:2307.04292*, 2023. [Online]. Verfügbar: <https://arxiv.org/pdf/2307.04292>) (Zugriff am: 22. Jan. 2025)

Medienverzeichnis

Soundbeispiel 1: Rocky Balboa Sound Design (traditionelles Sounddesign)

Soundbeispiel 2: IM Stupro Knockout (KI-Sounddesign)

Soundbeispiel 3: elevenlabs-Punch - KI-generiertes Schlaggeräusch („Punch“)

Soundbeispiel 4: Stable-Audio-Open-Zero Punch - KI-generiertes Schlaggeräusch („Punch“)

Alle Soundbeispiele sind abrufbar unter:

<https://www.dropbox.com/scl/fo/6t4cl8dyk7x4sf8wjhcpp/AOdojjkJzo0HRFQzAXbTjXU?rlkey=rdgw00xohcn6i7hubftkydpwj&st=3ruqfvvj&dl=0>

(Zugriff am: 22. Januar 2025).

1. Einleitung

1.1 Was ist Sounddesign

Sounddesign bezeichnet die kreative und technische Gestaltung von Klängen, die gezielt in audiovisuellen Medien eingesetzt werden, um die narrative und emotionale Wirkung einer Produktion zu unterstützen. Es umfasst eine Vielzahl von Prozessen, darunter die Konzeption, Aufnahme, Bearbeitung und Integration von Geräuschen, Musik und Sprache. Diese Prozesse zielen darauf ab, die visuelle Ebene eines Mediums wie Film, Theater, Videospiele oder Musikproduktionen akustisch zu ergänzen und zu verstärken. Thomas Görne definiert Sounddesign als eine interdisziplinäre Disziplin, die sowohl künstlerische als auch technische Kompetenzen vereint und sich an der Schnittstelle von Akustik, Wahrnehmungspsychologie und Ästhetik bewegt.

Er hebt hervor, dass Sounddesign weit über die bloße Erstellung und Platzierung von Klängen hinausgeht. Es erfordert ein tiefes Verständnis für die physikalischen Eigenschaften von Schall sowie für die Art und Weise, wie Menschen Klänge wahrnehmen und emotional darauf reagieren. Diese Disziplin nutzt bewusst die enge Verbindung zwischen Klang und Emotionen, da Töne und Geräusche tief im menschlichen Unterbewusstsein verankert sind und somit starke emotionale Reaktionen hervorrufen können. Ein einfaches Geräusch, wie das Knarren einer Tür oder das Rascheln von Blättern, kann gezielt eingesetzt werden, um Spannung, Angst oder eine geheimnisvolle Atmosphäre zu erzeugen.

Ein zentraler Aspekt des Sounddesigns ist die bewusste Gestaltung von Klangwelten, die das visuelle Erlebnis des Publikums erweitern. Dies umfasst sowohl die realistische Nachbildung von Geräuschen aus der Alltagswelt als auch die Schaffung völlig neuer, abstrakter Klangwelten, die jenseits der physikalischen Realität liegen können. Görne beschreibt, dass die Wahrnehmungspsychologie hierbei eine Schlüsselrolle spielt, da sie erklärt, wie akustische Signale im Gehirn verarbeitet werden und wie sie bestimmte Assoziationen oder Stimmungen hervorrufen können. So wird Sounddesign zu einem mächtigen Werkzeug, um die Aufmerksamkeit des Publikums zu lenken und die Handlung eines Films oder einer anderen audiovisuellen Produktion gezielt zu strukturieren.

Darüber hinaus wird unterstrichen, dass Sounddesign nicht nur auf der auditiven Ebene eines Werks wirkt, sondern die Gesamtwirkung eines Mediums maßgeblich beeinflusst. Es ist ein unverzichtbarer Bestandteil der audiovisuellen Kommunikation, da es die visuelle Sprache eines Mediums ergänzt und oft unsichtbare oder unausgesprochene Inhalte vermittelt. Durch die Kombination von Klang und Bild wird eine tiefere Immersion geschaffen, die es dem Publikum ermöglicht, intensiver in die erzählte Geschichte einzutauchen. Sounddesign kann die Wahrnehmung und Interpretation von Szenen maßgeblich steuern, indem es emotionale Nuancen betont oder zusätzliche narrative Ebenen eröffnet.

Zudem wird betont, dass Sounddesign nicht nur eine technische Aufgabe ist, sondern auch eine künstlerische Ausdrucksform darstellt. Es fordert von den Gestaltern ein hohes Maß an Kreativität und ein feines Gespür für die Ästhetik von Klängen. Die Fähigkeit, Klänge so zu arrangieren, dass sie sowohl funktional als auch künstlerisch ansprechend sind, macht Sounddesign zu einer anspruchsvollen und vielseitigen Disziplin. Diese Vielschichtigkeit ist es, das Sounddesign zu einem integralen Bestandteil moderner Filmproduktionen und anderer audiovisueller Werke macht ¹.

¹Vgl. T. Görne, Sounddesign: Klang, Wahrnehmung, Emotion, 2017, Verfügbar: <https://books.google.de/books?id=rOUIDwAAQBAJ> (Zugriff am: 22. Jan. 2025)

1.2 Warum ist Sounddesign ein zentraler Bestandteil von Filmproduktionen?

Wie wichtig die akustische Ebene für das Gesamterlebnis eines Films ist, wird oft betont. Der Ton wird dabei als essenzielles Element beschrieben, das den Bildern emotionale Tiefe und Ausdruck verleiht. Ohne das Dröhnen der Schläge und die jubelnde Menge in *Rocky* würde der Zuschauer kaum die gleiche emotionale Intensität erleben.

Filmproduktionen nutzen Sounddesign nicht nur, um Geschichten zu erzählen, sondern sie für das Publikum erlebbar zu machen. Klänge, Musik und Geräusche sind nicht bloß begleitende Elemente, sondern entscheidende Werkzeuge, um Emotionen zu erzeugen, Spannung aufzubauen und die Immersion zu verstärken. Ein gut gestaltetes Sounddesign kann den Unterschied zwischen einer eindrucksvollen Szene und einer, die kaum Eindruck hinterlässt, ausmachen.

Besonders in genrespezifischen Produktionen wie Action-, Horror- oder Sportfilmen spielt Sounddesign eine tragende Rolle. Filme wie *Rocky* zeigen eindrucksvoll, wie durch gezielte akustische Gestaltung eine emotionale Bindung zum Publikum geschaffen werden kann. Die dröhnenden Schläge im Boxring, das Jubeln der Menge und der ikonische Score bilden eine Klangwelt, die den Zuschauer mitten ins Geschehen zieht und die Intensität der Handlung verstärkt.

Die Bedeutung von Sounddesign hat sich seit den Anfängen des Films stetig weiterentwickelt. Während Stummfilme noch vollständig auf Musik angewiesen waren, um Emotionen zu vermitteln, hat die Einführung des Tonfilms eine neue Dimension der Erzählkunst eröffnet. Heutzutage ermöglichen digitale Technologien und innovative Werkzeuge eine bisher nie dagewesene Präzision und Vielfalt in der Gestaltung von Klangwelten.

Im digitalen Zeitalter ist Sounddesign vielseitiger denn je. Neben klassischen Methoden kommen zunehmend neue Technologien zum Einsatz, die die Möglichkeiten der Klanggestaltung erweitern. Die Integration von Künstlicher Intelligenz (KI) bietet eine innovative Perspektive, wie Sounddesign effizienter und kreativer gestaltet werden kann.

1.3 Zielsetzung der Arbeit: Untersuchung und Vergleich von traditionellem und KI-generiertem Sounddesign

Künstliche Intelligenz eröffnet neue Wege, komplexe Klangwelten zu erschaffen, die mit traditionellen Mitteln nur schwer oder gar nicht umsetzbar wären. Durch Algorithmen wie Generative Adversarial Networks (GANs) oder Deep Learning können Klänge generiert werden, die individuell angepasst und in Echtzeit verändert werden können. Diese Effizienz und Flexibilität stellt eine große Chance dar, den kreativen Prozess neu zu definieren. Mit den Fortschritten in der KI-Technologie entstehen neue Werkzeuge und Methoden, die nicht nur die Produktionszeit reduzieren, sondern auch völlig neue kreative Ansätze ermöglichen. Diese Entwicklungen werfen jedoch grundlegende Fragen auf: Inwiefern kann KI den kreativen Prozess des Sounddesigns unterstützen oder sogar ersetzen? Welche qualitativen Unterschiede gibt es zwischen traditionellen und KI-gestützten Ansätzen? Die vorliegende Arbeit zielt darauf ab, diese Fragen zu beantworten, indem sie einen Vergleich zwischen traditionellem Sounddesign und KI-generierten Ansätzen durchführt. Diese Untersuchung bildet die Grundlage, um die Stärken und Schwächen beider Ansätze im kreativen und technischen Kontext zu bewerten und potenzielle Synergien zwischen traditionellem Handwerk und KI-basierten Methoden aufzuzeigen.

1.4 Forschungsfragen

Die zentrale Forschungsfrage dieser Arbeit lautet: Was ist der Unterschied zwischen traditionellem und KI-gestütztem Sounddesign? Dabei wird untersucht, wie sich die kreativen Prozesse der beiden Ansätze unterscheiden, welche qualitativen und zeitlichen Unterschiede bestehen und welche Potenziale sowie Grenzen Künstliche Intelligenz im Vergleich zu traditionellem Sounddesign bietet. Der Fokus liegt darauf, sowohl die technischen als auch die kreativen Aspekte dieser beiden Herangehensweisen zu analysieren und deren Auswirkungen auf die Filmproduktion zu bewerten.

1.5 Aufbau der Arbeit

Die Arbeit beginnt mit einer Untersuchung des traditionellen Sounddesigns, bei der die klassischen Techniken und Herausforderungen sowie Beispiele aus ikonischen Produktionen wie *Rocky* beleuchtet werden. Anschließend wird der Fokus auf KI-gestütztes Sounddesign gelegt, wobei die Prozesse und Methoden analysiert werden, die sich von traditionellen Ansätzen unterscheiden. Der Vergleich von traditionellem und KI-generiertem Sounddesign bildet den Kern der Arbeit. Dabei werden kreative Freiheit, Produktionszeit, Klangqualität und der emotionale Einfluss auf das Publikum betrachtet. Die Ergebnisse des Vergleichs werden abschließend in einer Diskussion zusammengeführt, die die Potenziale und Grenzen beider Ansätze beleuchtet, bevor die Arbeit mit einem Fazit und einem Ausblick auf zukünftige Entwicklungen im Sounddesign endet.

2. Traditionelles Sounddesign: Wie wurde bisher gearbeitet?

2.1 Definition und Bedeutung von Sounddesign in der Filmproduktion

Sounddesign ist ein wesentlicher Bestandteil der Filmproduktion, der dazu dient, visuelle Erzählungen akustisch zu ergänzen und die emotionale Wirkung von Szenen zu verstärken. Es umfasst die kreative und technische Gestaltung von Klängen, Musik und Geräuschen, die zusammen eine immersive Klanglandschaft schaffen. Diese unterstützt die Narration, steigert die Spannung und ermöglicht es dem Publikum, tiefer in die gezeigte Welt einzutauchen.

Konzepte und Funktionen des Sounddesigns

Der Begriff „Sounddesign“ beschreibt die bewusste Komposition und Bearbeitung von Klängen, die das visuelle Erlebnis eines Films auf mehreren Ebenen bereichern. Laut Schätzlein geht es dabei nicht nur um die Nachbildung realer Klänge, sondern auch um die Erzeugung symbolischer oder emotionaler Akzente, die den dramaturgischen Aufbau eines Films unterstützen². In der Praxis umfasst Sounddesign mehrere Ebenen, die jeweils spezifische Funktionen übernehmen.

Dialoge spielen eine zentrale Rolle, da sie die Handlung transportieren und dabei stets klar verständlich sein sollten. Foley-Geräusche, wie Schritte oder das Rascheln von Kleidung, stellen synchrone, alltägliche Klänge dar und verleihen der gezeigten Welt Authentizität. Atmosphärische Klänge schaffen den Hintergrund eines Films und verstärken das Setting, etwa durch das Rauschen eines Waldes oder das Summen eines belebten Marktplatzes. Musik schließlich dient als emotionale Verstärkung und wird gezielt eingesetzt, um spezifische Gefühle wie Spannung, Freude oder Trauer zu erzeugen. Diese Kombination aus realistischen und symbolischen Akzenten machen Sounddesign zu einem integralen Bestandteil der filmischen Narration.

Die Bedeutung von Sounddesign für die Narration

Die Bedeutung des Sounddesigns liegt vor allem in seiner Fähigkeit, Geschichten emotional und räumlich greifbar zu machen. Ein herausragendes Beispiel ist die Arbeit von Walter Murch in *Apocalypse Now* (1979). Durch die kreative Verwendung von Helikoptergeräuschen, gemischt mit dröhnenden Bässen und fließenden Übergängen in die Filmmusik, schuf er eine unverwechselbare akustische Identität, die den Zuschauer in die beklemmende Atmosphäre des Films eintauchen lässt³.

Auch in Filmen wie *Rocky* (1976) wird deutlich, wie essenziell Sounddesign für die emotionale Bindung des Publikums ist. Die dröhnenden Schläge im Boxring, das Keuchen der Kämpfer und die jubelnde Menge erzeugen eine Klanglandschaft, die die physische und emotionale Intensität der Szenen untermalt. Dies zeigt, dass Sounddesign nicht nur eine unterstützende Funktion hat, sondern aktiv an der Erzählung beteiligt ist.

²Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, H. Segeberg und F. Schätzlein, Eds., Marburg: Schüren, 2005, pp. 24–40, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

³Vgl. P. Moormann und F. Hentschel, Filmmusik: Ein alternatives Kompendium, Springer Fachmedien Wiesbaden, 2018, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)

Historische Entwicklung des Sounddesigns

Die Bedeutung von Sounddesign hat sich im Laufe der Filmgeschichte kontinuierlich weiterentwickelt. Während der Ära der Stummfilme war die Musik das einzige akustische Mittel, um Emotionen zu transportieren. Mit der Einführung des Tonfilms in den 1920er-Jahren änderte sich dies grundlegend: Geräusche und Dialoge wurden Teil der filmischen Sprache und eröffneten neue Möglichkeiten der akustischen Gestaltung ⁴.

In den folgenden Jahrzehnten entwickelten sich spezialisierte Techniken wie Foley-Geräusche und Mehrkanalton, die es Sounddesignern ermöglichten, realistischere und immersivere Klangwelten zu schaffen. Die digitale Revolution brachte schließlich Werkzeuge wie DAWs (Digital Audio Workstations) und spezialisierte Software, die den kreativen Prozess automatisierten und erweiterten ⁵.

2.2 Arbeitsmethoden im traditionellen Sounddesign

Die Arbeitsmethoden im traditionellen Sounddesign haben sich im Laufe der Filmgeschichte kontinuierlich weiterentwickelt. Vom Beginn der Tonfilmzeit bis zur Einführung digitaler Technologien basierte Sounddesign auf einem handwerklichen Ansatz, der sowohl kreative als auch technische Fähigkeiten verlangte. Hierbei wurden spezifische Techniken entwickelt, um die Klangwelten von Filmen zu gestalten und das Publikum emotional zu binden.

Frühe Techniken: Beginn des Sounddesigns

Mit der Einführung des Tonfilms in den 1920er-Jahren begann die Integration von Ton in die filmische Narration. Zu Beginn lag der Fokus auf der Synchronisierung von Dialogen und der Einbindung musikalischer Elemente. Die Aufnahmen erfolgten oft live am Set, was technische Herausforderungen wie Umgebungsgeräusche oder begrenzte Aufnahmequalität mit sich brachte. Diese frühen Methoden legten jedoch den Grundstein für das moderne Sounddesign ⁶.

Die Entwicklung der Foley-Techniken

Eine bedeutende Innovation im Sounddesign war die Entwicklung der Foley-Techniken. Jack Foley, der Namensgeber dieser Methode, führte die Praxis ein, Geräusche manuell zu erzeugen, die synchron zu den Bewegungen der Charaktere oder Objekte im Film hinzugefügt wurden. Beispiele hierfür sind Schritte, das Rascheln von Kleidung oder das Klirren von Glas. Diese Techniken waren revolutionär, da sie es ermöglichten, Klänge präzise auf visuelle Ereignisse abzustimmen, wodurch eine realistische akustische Darstellung geschaffen wurde ⁷.

Foley-Geräusche wurden in speziellen Studios aufgenommen, in denen Geräuschkünstler Materialien und Objekte manipulierten, um authentische Klänge zu erzeugen. Ein bemerkenswertes Beispiel für die Anwendung von Foley-Techniken ist der Film Rocky (1976). Die Geräusche der Schläge im Boxring wurden durch das Schlagen auf Leder oder das Werfen von Sandsäcken simuliert, was die Intensität der Kämpfe unterstrich ⁸.

⁴Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, H. Segeberg und F. Schätzlein, Eds., Marburg: Schüren, 2005, pp. 24–40, Verfügbar:

<https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

⁵Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar:

<https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

⁶Vgl. J. Martin und F. Butzmann, „Sound Design,“ in Filmmusik, F. Hentschel und P. Moormann, Eds., Springer Fachmedien Wiesbaden, 2018, pp. 112–123, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)

⁷Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar:

<https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

⁸Vgl. J. Martin und F. Butzmann, „Sound Design,“ in Filmmusik, F. Hentschel und P. Moormann, Eds., Springer Fachmedien Wiesbaden, 2018, pp. 112–123, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)

Musik als emotionaler Verstärker

Neben Geräuschen spielte die Musik eine zentrale Rolle im traditionellen Sounddesign. Sie wurde oft eigens für Filme komponiert und in Zusammenarbeit mit Orchestern aufgenommen. Filmmusik diente nicht nur der Untermalung, sondern war ein essenzielles Werkzeug, um die emotionale Wirkung von Szenen zu verstärken. Ein Beispiel hierfür ist der Film *Psycho* (1960), in dem Bernard Herrmanns schrille Streicherklänge die Spannung der Duschszene verstärkten⁹.

Mehrkanalton und räumliche Klanggestaltung

In den 1970er- und 1980er-Jahren setzte sich die Verwendung von Mehrkanalton durch, der es ermöglichte, Klänge räumlich zu verteilen und somit ein immersives Erlebnis zu schaffen. Filme wie *Star Wars* (1977) nutzten diese Technik, um die Klanglandschaft der Weltraumschlachten realistisch und eindrucksvoll zu gestalten. Dies erforderte die Zusammenarbeit zwischen Sounddesignern und Toningenieuren, die mithilfe von Mehrspuraufnahmen und Mischtechniken einzigartige Klangwelten schufen¹⁰.

Die Ära der analogen Technologien

Bis zur Einführung digitaler Technologien war das Sounddesign stark durch analoge Werkzeuge geprägt. Magnetbänder wurden verwendet, um Klänge aufzunehmen, zu bearbeiten und zu synchronisieren. Diese Technologie war zwar flexibel, aber zeitaufwendig und fehleranfällig, weshalb oft ganze Abschnitte neu aufgenommen werden mussten. Sounddesigner arbeiteten eng mit Regisseuren zusammen, um harmonische Klanglandschaften zu schaffen. Diese Ära prägte nicht nur die Technik, sondern auch die kreative Zusammenarbeit im Sounddesign¹¹.

2.3 Rolle von Foley, Musik und atmosphärischen Klängen

Sounddesign basiert auf mehreren essentiellen Elementen, die die audiovisuelle Erfahrung eines Films prägen. Foley, Musik und atmosphärische Klänge spielen dabei zentrale Rollen, indem sie die Immersion des Publikums verstärken, die Handlung unterstützen und Emotionen gezielt hervorrufen.

Foley: Die Kunst der Geräuscherzeugung

Foley ist eine der wichtigsten Techniken im Sounddesign, um realistische und synchrone Geräusche zu erzeugen, die die visuelle Darstellung ergänzen. Jack Foley, der Begründer dieser Technik, entwickelte Methoden, um Geräusche wie Schritte, das Rascheln von Kleidung oder das Klirren von Glas manuell zu erstellen und mit den Bewegungen der Charaktere abzustimmen. Foley-Künstler arbeiten in speziell ausgestatteten Studios, in denen sie Alltagsgegenstände manipulieren, um die gewünschten Klänge zu erzeugen. Ein Beispiel für den Einsatz von Foley ist der Film *Rocky* (1976), in dem die Geräusche der Boxschläge durch das Schlagen auf Leder oder Sandsäcke erzeugt wurden.

Diese Technik trug dazu bei, die physische Intensität der Kämpfe realistisch darzustellen und das Publikum in die Handlung zu ziehen. Foley bietet eine präzise Möglichkeit, Klänge zu

⁹Vgl. F. Hentschel und P. Moormann, *Filmmusik: Ein alternatives Kompendium*, Springer Fachmedien Wiesbaden, 2018., Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)

¹⁰Vgl. L. Duffner, *Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise*, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

¹¹Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in *Sound. Zur Technologie und Ästhetik des Akustischen in den Medien*, H. Segeberg und F. Schätzlein, Eds., Marburg: Schüren, 2005, pp. 24–40, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

synchronisieren, und verleiht Filmen eine authentische Klangwelt. Diese Methode war besonders in der analogen Ära von unschätzbarem Wert, da digitale Technologien zur Erzeugung von Geräuschen noch nicht verfügbar waren ¹².

Musik: Der emotionale Verstärker

Musik ist ein unverzichtbares Element des Sounddesigns, das die emotionale Tiefe eines Films entscheidend beeinflusst. Sie kann die Stimmung einer Szene verstärken, Charaktere thematisch untermalen und das Publikum gezielt emotional leiten. Laut Schätzlein trägt Musik zur „symbolischen und emotionalen Akzentuierung“ bei und wird oft als narrative Erweiterung verwendet.

Ein Paradebeispiel für den Einsatz von Musik im Film ist die Duschszene aus Psycho (1960). Bernard Herrmanns schrille Streicherklänge intensivieren die Spannung und erzeugen beim Publikum ein Gefühl des Unbehagens, das die visuelle Brutalität verstärkt. Ebenso wurde in Rocky der Score von Bill Conti („Gonna Fly Now“) eingesetzt, um Rockys Entschlossenheit und den emotionalen Höhepunkt seiner Trainingsmontage zu untermalen. Musik wird häufig eigens für Filme komponiert, wobei Regisseure und Komponisten eng zusammenarbeiten, um die dramaturgischen Anforderungen des Films zu erfüllen. Die Einführung von Mehrkanalton in den 1970er-Jahren ermöglichte es zudem, Musik räumlich zu verteilen und die Immersion weiter zu steigern ¹³.

Atmosphärische Klänge: Die akustische Kulisse

Atmosphärische Klänge schaffen den auditiven Hintergrund eines Films und tragen dazu bei, die gezeigte Welt glaubhaft und immersiv zu gestalten. Beispiele sind das Rauschen des Windes, das Summen eines Marktplatzes oder das Plätschern von Wasser. Diese Klänge dienen dazu, den Zuschauer in die Szene zu versetzen und die visuelle Erzählung zu ergänzen. Walter Murch nutzte atmosphärische Klänge meisterhaft in Apocalypse Now (1979). Das dröhnende Summen von Hubschraubern, kombiniert mit subtilen Hintergrundgeräuschen, erzeugte eine beklemmende Atmosphäre, die die psychologische Tiefe des Films verstärkte. Atmosphärische Klänge sind auch in Filmen wie Blade Runner (1982) von zentraler Bedeutung, wo sie die dystopische Stimmung und die futuristische Umgebung unterstreichen ¹⁴.

Die Kombination dieser Elemente im Sounddesign

Die Kraft von Foley, Musik und atmosphärischen Klängen liegt in ihrer Kombination. Diese drei Komponenten arbeiten zusammen, um eine kohärente Klanglandschaft zu schaffen, die die visuelle Erzählung unterstützt und das Publikum emotional involviert. Sounddesigner nutzen diese Elemente, um Spannungsbögen zu erzeugen, Übergänge zu gestalten und den narrativen Fluss eines Films zu fördern. Die analoge Ära erforderte dabei ein hohes Maß an Präzision und Kreativität, da die Synchronisation und Bearbeitung aller Elemente manuell erfolgte ¹⁵.

¹²Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar:

<https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

¹³Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

¹⁴Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

¹⁵Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

2.4 Beispiele aus klassischen Filmproduktionen

Klassische Filmproduktionen demonstrieren eindrucksvoll, wie Sounddesign die emotionale Wirkung von Szenen verstärken und die Narration eines Films unterstützen kann. Besonders hervorzuheben sind Filme wie *Rocky* (1976), *Apocalypse Now* (1979) und *Blade Runner* (1982), die jeweils einzigartige Klangwelten geschaffen haben und als Meilensteine des Sounddesigns gelten.

Rocky (1976): Die Klangwelt des Boxrings

Der Film *Rocky*, inszeniert von John G. Avildsen, ist ein Paradebeispiel dafür, wie Sounddesign die Intensität und Authentizität einer Szene erhöhen kann. Die Boxkämpfe im Film wurden akustisch mit einer Kombination aus Live-Aufnahmen und nachträglich erzeugten Foley-Geräuschen gestaltet. Die Schläge im Ring wurden durch das Schlagen auf Leder oder Sandsäcke simuliert, um ein dumpfes, realistisches Klangbild zu erzeugen. Die Kombination aus realistischen Geräuschen und einem emotional aufgeladenen Score führte dazu, dass Zuschauer die Kämpfe nicht nur sahen, sondern auch akustisch miterlebten. Zusätzlich unterstrich die jubelnde Menge die emotionale Intensität der Szenen. Die musikalische Untermalung, insbesondere Bill Contis „Gonna Fly Now“, ist ein weiteres herausragendes Element. Dieses Lied unterstützt Rockys Trainingsmontagen und vermittelt ein Gefühl von Hoffnung, Kampfgeist und Triumph, das die emotionale Bindung des Publikums an den Protagonisten stärkt ¹⁶.

Apocalypse Now (1979): Psychologische Tiefe durch Klang

Francis Ford Coppolas *Apocalypse Now* ist für seine außergewöhnliche Klanggestaltung bekannt, die maßgeblich von Walter Murch entwickelt wurde. Der Einsatz von Helikoptergeräuschen in Kombination mit düsteren atmosphärischen Klängen erzeugt eine beklemmende Atmosphäre, die die psychologische Dimension des Films verstärkt. Murch kombinierte hierbei natürliche Geräusche mit elektronisch bearbeiteten Klängen, um eine bedrohliche und zugleich surreal wirkende Klanglandschaft zu schaffen ¹⁷.

Blade Runner (1982): Dystopische Klangwelten

Ridley Scotts *Blade Runner* ist ein weiteres Meisterwerk, das zeigt, wie Sounddesign die visuelle Ästhetik eines Films ergänzen kann. Vangelis' futuristische elektronische Musik, kombiniert mit subtilen atmosphärischen Klängen wie Regen, Donner und Straßenlärm, schafft eine düstere, dystopische Welt, die die Handlung des Films perfekt unterstützt. Diese Klangwelt war entscheidend dafür, die visuell dargestellte Zukunft glaubwürdig und immersiv zu machen ¹⁸.

Star Wars (1977): Innovation durch Mehrkanalton

Star Wars (1977) setzte neue Maßstäbe im Sounddesign, insbesondere durch die Nutzung von Mehrkanalton. Ben Burtt, der für die Klanggestaltung verantwortlich war, kreierte einzigartige Soundeffekte wie das charakteristische Summen der Lichtschwerter oder das Brüllen der TIE-Fighter, das durch die Bearbeitung von Tiergeräuschen entstand. Diese

¹⁶Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in *Sound. Zur Technologie und Ästhetik des Akustischen in den Medien*, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

¹⁷Vgl. P. Moormann, „Filmmusik: Theorie und Praxis,“ in *Handbuch der Filmmusik*, Springer Fachmedien Wiesbaden, 2018, pp. 110–130.

¹⁸Vgl. L. Duffner, *Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise*, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

Innovationen haben die immersive Wirkung der Weltraumkämpfe entscheidend geprägt und das Sounddesign als eigenständige Kunstform etabliert ¹⁹.

Bedeutung für die Filmproduktion

Diese klassischen Filmproduktionen zeigen, dass Sounddesign weit mehr ist als eine technische Ergänzung. Es ist ein künstlerisches Mittel, das Geschichten emotional verstärkt, Welten erschafft und das Publikum in die Handlung eintauchen lässt. Die genannten Beispiele verdeutlichen, wie die bewusste Gestaltung von Klängen und Musik die visuelle Ebene eines Films erweitert und bereichert.

2.5 Herausforderungen und Grenzen traditioneller Methoden

Das traditionelle Sounddesign, obwohl handwerklich und künstlerisch anspruchsvoll, war durch verschiedene technische, organisatorische und kreative Herausforderungen geprägt. Diese Grenzen spiegeln die technologische Entwicklung der Filmindustrie wider und beeinflussten, wie Klangwelten in Filmen geschaffen wurden.

Technologische Einschränkungen in der analogen Ära

In der frühen Phase des Sounddesigns war die analoge Technik ein wesentlicher Bestandteil der Produktionsprozesse. Tonaufnahmen wurden auf Magnetbändern gespeichert, die empfindlich gegenüber mechanischen Schäden und Qualitätsverlusten waren. Jede Bearbeitung erforderte präzises Schneiden und Zusammenfügen der Bänder, was zeitaufwendig war und kaum Raum für Fehler ließ ²⁰.

Die Synchronisation von Ton und Bild stellte ebenfalls eine erhebliche Herausforderung dar. Ohne digitale Hilfsmittel mussten Geräusche und Musik manuell auf die Bilder abgestimmt werden. Dies verlangte nicht nur technisches Können, sondern auch eine enorme Geduld seitens der Sounddesigner.

Logistische Herausforderungen bei Foley-Aufnahmen

Foley-Aufnahmen erforderten den Einsatz spezieller Studios, in denen Geräusche synchron zu den Filmaufnahmen erstellt wurden. Das Problem lag in der begrenzten Flexibilität: Geräusche mussten oft mehrfach aufgenommen werden, bis sie exakt mit den gezeigten Bewegungen übereinstimmten. Zudem war die Anzahl der verfügbaren Foley-Künstler begrenzt, was den Produktionsprozess verlangsamen konnte ²¹.

Ein weiteres Hindernis war die Beschränkung auf physische Materialien. Geräusche mussten durch Objekte erzeugt werden, die die gewünschten akustischen Eigenschaften aufwiesen, was nicht immer einfach zu realisieren war. Beispielsweise wurden Boxschläge in Rocky durch Schläge auf Leder simuliert, was zwar authentisch klang, jedoch eine erhebliche Expertise bei der Klangmanipulation erforderte.

¹⁹Vgl. J. Martin und F. Butzmann, „Sound Design,“ in Filmmusik, F. Hentschel und P. Moormann, Eds., Springer Fachmedien Wiesbaden, 2018, pp. 112–123, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)

²⁰Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

²¹Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

Kreative Begrenzungen und Abhängigkeit von Musik

Traditionelles Sounddesign war stark von der Zusammenarbeit mit Komponisten und Musikern abhängig. Während Musik eine zentrale Rolle bei der emotionalen Untermalung spielte, waren Sounddesigner oft darauf angewiesen, die Musik nachträglich an die Handlung anzupassen. Diese Hierarchie konnte die kreative Freiheit der Sounddesigner einschränken ²².

Filme wie Blade Runner zeigen jedoch, wie wichtig die Integration von Musik und Sounddesign für die Gesamtwirkung eines Films ist. Die fehlende Möglichkeit, Klänge flexibel zu modifizieren oder dynamisch an die Handlung anzupassen, stellte jedoch eine Einschränkung dar.

Begrenzte Kapazitäten durch analoge Werkzeuge

Die analogen Werkzeuge der damaligen Zeit boten nur eine begrenzte Anzahl an Spuren für die Aufnahme und Bearbeitung von Klängen. Dies führte dazu, dass Sounddesigner häufig Prioritäten setzen mussten, welche Elemente in den Vordergrund treten sollten.

Mehrspurrekorder wie die in den 1970er-Jahren eingeführten Geräte boten zwar eine Verbesserung, waren jedoch teuer und schwer zugänglich ²³.

Die Bearbeitung von Mehrkanalton für Filme wie Star Wars war ebenfalls aufwendig und erforderte spezialisierte Ingenieure, um die Klänge räumlich zu verteilen und die gewünschten Effekte zu erzielen.

Zeit- und Kostenaufwand

Das traditionelle Sounddesign war durch einen erheblichen Zeit- und Kostenaufwand gekennzeichnet. Jede Szene erforderte sorgfältige Planung, Aufnahme und Bearbeitung, was bei größeren Produktionen Wochen oder Monate in Anspruch nehmen konnte. Die hohe Abhängigkeit von menschlichem Können und die begrenzte Effizienz der Technologien führten dazu, dass Sounddesign oft einen erheblichen Anteil am Produktionsbudget einnahm.

Zusammenfassung

Trotz der genannten Herausforderungen legte das traditionelle Sounddesign den Grundstein für moderne Techniken und bewies, wie wichtig akustische Gestaltung für die Wirkung eines Films ist. Die Grenzen dieser Methoden spiegeln jedoch die Notwendigkeit wider, technologische Fortschritte zu nutzen, um den kreativen Prozess effizienter und flexibler zu gestalten. Dies bildet die Grundlage für den Übergang zu KI-gestützten Ansätzen, die viele dieser Einschränkungen überwinden könnten.

²²Vgl. J. Martin und F. Butzmann, „Sound Design,“ in Filmmusik, F. Hentschel und P. Moormann, Eds., Springer Fachmedien Wiesbaden, 2018, pp. 112–123, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)

²³Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

3. Grundlagen der Künstlichen Intelligenz (KI)

3.1 Was ist Künstliche Intelligenz?

Künstliche Intelligenz (KI) ist ein interdisziplinäres Forschungsfeld innerhalb der Informatik, das sich mit der Entwicklung von Systemen und Algorithmen befasst, die menschenähnliches Verhalten und Intelligenz imitieren können. Sie umfasst eine Vielzahl von Technologien und Methoden, die es Maschinen ermöglichen, Aufgaben wie Lernen, Problemlösen, Sprachverarbeitung, Wahrnehmung und Entscheidungsfindung autonom durchzuführen. Ziel der KI ist es, Maschinen zu schaffen, die menschliches Denken und Handeln nicht nur nachahmen, sondern in bestimmten Bereichen auch übertreffen können.

Unterteilung der Künstlichen Intelligenz

Die KI wird klassischerweise in symbolische und nicht-symbolische Ansätze unterteilt. Symbolische Ansätze basieren auf klar definierten Regeln, Logiken und Wissensrepräsentationen. Sie verwenden Algorithmen, die auf expliziten Anweisungen beruhen, um Probleme zu lösen. Beispiele hierfür sind Expertensysteme und regelbasierte Entscheidungsbäume. Diese Ansätze waren in den frühen Tagen der KI-Forschung dominant und eignen sich besonders für Aufgaben mit festen Regeln und begrenztem Problemraum, wie Schachspielen oder mathematische Berechnungen. Nicht-symbolische Ansätze, wie Machine Learning (ML) und Deep Learning (DL), setzen auf datengetriebene Modelle. Anstatt explizite Regeln zu befolgen, lernen diese Systeme aus großen Datenmengen, um Muster und Zusammenhänge zu erkennen. Sie nutzen neuronale Netzwerke, die von der Funktionsweise des menschlichen Gehirns inspiriert sind, um Informationen zu verarbeiten und Entscheidungen zu treffen. Diese Methoden haben in den letzten Jahren erhebliche Fortschritte gemacht und werden in Bereichen wie Sprachverarbeitung, Bildverarbeitung und autonomem Fahren eingesetzt ²⁴.

Eigenschaften und Vorteile moderner KI-Systeme

Ein wesentlicher Vorteil moderner KI-Systeme liegt in ihrer Fähigkeit, sich an neue Herausforderungen und Umgebungen anzupassen. Anders als klassische algorithmische Systeme, die nur innerhalb vorgegebener Regeln operieren können, sind KI-Systeme in der Lage, eigenständig Wissen zu generieren und komplexe Muster zu identifizieren. Beispielsweise kann ein KI-System für die Bildverarbeitung lernen, verschiedene Objekte wie Autos, Menschen oder Tiere zu erkennen, indem es mit einer großen Menge an markierten Bildern trainiert wird. Es ist dann in der Lage, diese Objekte in neuen, zuvor unbekanntem Bildern korrekt zu identifizieren. Diese Fähigkeit, Wissen aus Erfahrungen zu generieren und auf neue Szenarien anzuwenden, macht KI zu einer Schlüsseltechnologie der digitalen Transformation.

Neuronale Netze

Ein grundlegender Bestandteil moderner KI-Systeme sind neuronale Netze, die von der Funktionsweise des menschlichen Gehirns inspiriert sind. Sie bestehen aus mehreren Schichten: einer Eingabeschicht, einer oder mehreren versteckten Schichten und einer Ausgabeschicht. Jede Schicht enthält künstliche Neuronen, die miteinander durch gewichtete Verbindungen verbunden sind. Diese Gewichte werden während des Trainings angepasst, um Muster und Zusammenhänge in den Daten zu erkennen.

²⁴Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

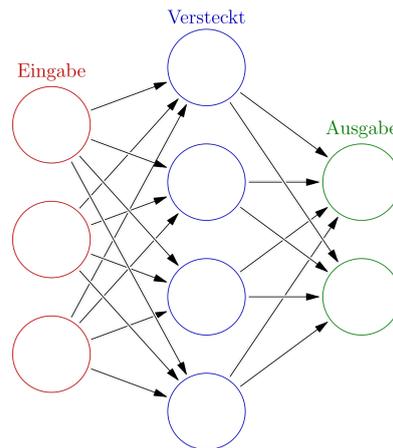


Abbildung 1: vereinfachte Darstellung eines neuronalen Netzes

Das Backpropagation-Verfahren ist ein zentrales Element des Lernprozesses. Es ermöglicht die Optimierung der Gewichte durch die Rückverfolgung von Fehlern, die während der Vorhersagen entstehen. Dieses Verfahren erlaubt es neuronalen Netzen, ihre Genauigkeit iterativ zu verbessern, indem sie aus den Trainingsdaten lernen.

Neuronale Netze sind äußerst vielseitig und werden in einer Vielzahl von Anwendungsbereichen eingesetzt. Convolutional Neural Networks (CNNs) sind auf die Bildverarbeitung spezialisiert und analysieren Merkmale wie Kanten, Texturen oder Formen. Sie finden Anwendung in der Gesichtserkennung, der medizinischen Bildanalyse und in autonomen Fahrzeugen. Recurrent Neural Networks (RNNs) hingegen eignen sich für sequenzielle Daten wie Sprache oder Zeitreihen, da sie Informationen aus früheren Eingaben im Kontext behalten können. Sie werden unter anderem in der Sprachverarbeitung, bei Übersetzungen und in der Musikgenerierung genutzt. Transformermodelle wie GPT oder DALL-E, die ebenfalls auf neuronalen Netzen basieren, haben in der Verarbeitung natürlicher Sprache und der Generierung von Text- und Bildinhalten neue Maßstäbe gesetzt.

Die Fähigkeit neuronaler Netze, komplexe Muster in großen Datenmengen zu erkennen und sich an verschiedene Aufgabenstellungen anzupassen, macht sie zu einem der wichtigsten Treiber moderner KI. Sie ermöglichen Fortschritte in Bereichen wie der Bild- und Sprachverarbeitung, im autonomen Fahren und sogar in der kreativen Audiogestaltung. Durch ihre Vielseitigkeit und Leistungsfähigkeit spielen sie eine Schlüsselrolle in der digitalen Transformation und treiben die Entwicklung neuer Technologien entscheidend voran.

Historische Entwicklung der KI

Die historische Entwicklung der KI begann in den 1950er-Jahren mit der Vorstellung, menschliche Intelligenz durch Maschinen zu simulieren. Die frühen Tage der KI-Forschung waren geprägt von symbolischen Ansätzen. Pioniere wie John McCarthy, Alan Turing und Marvin Minsky legten den Grundstein für die heutige KI. In den 1980er-Jahren führten Fortschritte im Bereich der Expertensysteme zu einer ersten Welle der KI-Euphorie.

Ein entscheidender Wendepunkt war die Einführung datengetriebener Ansätze in den 1990er- und 2000er-Jahren, ermöglicht durch Fortschritte in der Hardware, insbesondere der Rechenleistung und Speicherkapazität. Algorithmen wie neuronale Netzwerke und Methoden wie Backpropagation haben es Maschinen ermöglicht, große Datenmengen schnell und effizient zu

verarbeiten. Dadurch können sie selbst sehr komplexe Probleme analysieren und lösen, die früher nur mit großem Aufwand oder gar nicht bewältigt werden konnten.

Heute hat KI dank Fortschritten in Bereichen wie Deep Learning, Natural Language Processing (NLP) und Computer Vision ein neues Niveau erreicht. Technologien wie Sprachassistenten (z. B. Alexa oder Siri), selbstfahrende Autos und generative Modelle (z. B. GPT und DALL-E) zeigen, wie KI zunehmend in den Alltag integriert wird und einen wesentlichen Einfluss auf Wirtschaft, Wissenschaft und Gesellschaft ausübt.

Herausforderungen und ethische Aspekte

Neben den Vorteilen bringt KI jedoch auch Herausforderungen mit sich. Die Black-Box-Natur vieler nicht-symbolischer Modelle erschwert die Nachvollziehbarkeit ihrer Entscheidungen, was insbesondere in sicherheitskritischen Bereichen wie Medizin oder Justiz problematisch sein kann. Zudem werfen Fragen der Datensicherheit, des Datenschutzes und der ethischen Verantwortung neue Diskussionen auf.

Zusammenfassend ist Künstliche Intelligenz nicht nur ein technologisches Werkzeug, sondern eine transformative Kraft, die viele Bereiche des menschlichen Lebens beeinflusst. Die fortschreitende Entwicklung von KI wird in den kommenden Jahren eine zentrale Rolle in der Gestaltung unserer Zukunft spielen.

3.2 Technologien wie Machine Learning und Deep Learning im Kontext der Audiogenerierung

Machine Learning ist ein Teilbereich der KI, der es Systemen ermöglicht, Muster in Daten zu erkennen und basierend darauf Entscheidungen zu treffen. Deep Learning, ein spezialisierter Ansatz des ML, verwendet neuronale Netzwerke mit mehreren Schichten, um komplexe Datenstrukturen zu analysieren und zu generieren. Diese Technologien haben die Audiogenerierung revolutioniert, insbesondere durch den Einsatz generativer Modelle wie Generative Adversarial Networks (GANs) und rekurrenter neuronaler Netzwerke (RNNs)²⁵. Im Kontext der Audiogenerierung ermöglichen Deep-Learning-Modelle die Erstellung realistischer Klänge und Musikstücke, die sowohl in Filmproduktionen als auch in der Musikindustrie Anwendung finden. Beispielsweise können GANs genutzt werden, um neue Klänge zu generieren, die den Stil bestehender Kompositionen nachahmen, ohne jedoch einfache Kopien zu sein.

Ein weiterer Durchbruch in der Generierung von Audio ist die Verwendung von Transformer-Modellen, wie sie in Anwendungen wie OpenAI Jukebox eingesetzt werden. Diese Modelle kombinieren Text- und Klanginformationen, um Musikstücke zu komponieren, die sowohl harmonisch als auch stilistisch ansprechend sind.

Ein Beispiel für die praktische Anwendung von ML und DL ist die KI-Software AIVA (Artificial Intelligence Virtual Artist), die eigenständig Musikstücke komponieren kann. Diese Technologie wird bereits in der Filmindustrie genutzt, um Produktionsprozesse zu beschleunigen und innovative Klangerlebnisse zu schaffen.

²⁵Vgl. M. Huzaifah und L. Wyse, „Deep Generative Models for Musical Audio Synthesis,“ Preprint for Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, Springer, 2020, Verfügbar: <https://arxiv.org/pdf/2006.06426> (Zugriff am: 22. Jan. 2025)

3.3 Anwendungsbereiche von KI in der kreativen Industrie

Die kreative Industrie hat Künstliche Intelligenz (KI) in den letzten Jahren zunehmend als unverzichtbares Werkzeug integriert, um sowohl neue kreative Möglichkeiten zu erschließen als auch bestehende Prozesse effizienter zu gestalten. KI ist besonders in Bereichen wie Film, Musik, Videospiele und Werbung von Bedeutung, da sie nicht nur Zeit- und Kosteneinsparungen ermöglicht, sondern auch neue Formen der Kreativität eröffnet.

Ein zentraler Vorteil der KI liegt in ihrer Fähigkeit, repetitive und zeitintensive Aufgaben zu automatisieren. Dies umfasst beispielsweise die automatische Bearbeitung von Inhalten, wie das Erstellen von Musikstücken oder die Generierung von Soundscapes, sowie die Unterstützung bei der Datenanalyse, um präzisere kreative Entscheidungen treffen zu können. Auf diese Weise können sich kreative Fachleute stärker auf innovative und künstlerische Aspekte konzentrieren, während KI ihnen einen effizienten Workflow ermöglicht.

Darüber hinaus bieten KI-Technologien die Möglichkeit, bestehende kreative Prozesse durch datengetriebene Ansätze zu verbessern. Beispielsweise können durch Machine-Learning-Modelle Trends analysiert, Zielgruppen besser verstanden und Inhalte entwickelt werden, die genau auf die Bedürfnisse und Präferenzen des Publikums abgestimmt sind. Diese Verbindung von automatisierter Effizienz und kreativer Innovation zeigt, wie KI traditionelle Arbeitsmethoden transformiert und neue Horizonte in der kreativen Industrie eröffnet

²⁶

KI im Sounddesign

Im Bereich des Sounddesigns ermöglicht Künstliche Intelligenz (KI) eine präzisere und schnellere Erstellung von Klangwelten, die zuvor manuell und zeitaufwendig gestaltet werden mussten. KI-Systeme analysieren Geräusche in Echtzeit und passen diese dynamisch an visuelle Szenen an, um eine perfekte Synchronität zu gewährleisten.

Technologien wie generative Modelle und neuronale Netzwerke kommen hierbei zum Einsatz, um individuelle und spezifisch auf Szenen abgestimmte Klangwelten zu erschaffen. Oh et al. (2023) heben hervor, dass generative Audio-KI-Systeme nicht nur repetitive Aufgaben automatisieren, sondern auch kreative Prozesse durch innovative Ansätze bereichern. Ein zentraler Vorteil dieser Technologien liegt in ihrer Fähigkeit, aus textlichen oder visuellen Beschreibungen Klänge zu erzeugen, die sowohl qualitativ hochwertig als auch vielseitig einsetzbar sind. Beispielsweise können generative Modelle wie neuronale Netzwerke genutzt werden, um Soundscapes für dynamische Umgebungen in Echtzeit zu erstellen, wodurch die Produktionszeit signifikant verkürzt wird ²⁷.

Ein weiterer Vorteil ist die Flexibilität generativer Audio-KI. Diese Systeme bieten nicht nur eine hohe Präzision bei der Klanggestaltung, sondern ermöglichen auch die Anpassung an die spezifischen Anforderungen unterschiedlicher Projekte, wie etwa Filme, Spiele oder Werbung. Die Fähigkeit, komplexe akustische Muster und Texturen zu erzeugen, hebt diese Ansätze deutlich von traditionellen Sounddesign-Methoden ab. Diese Kombination aus Effizienz und kreativer Vielseitigkeit zeigt das Potenzial von KI, das Sounddesign auf ein neues Niveau zu heben.

²⁶Vgl. M. Neu, M. Müller, B. Pothen, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst. Wiesbaden, Germany: Springer Fachmedien, 2022, Verfügbar:

<https://link.springer.com/book/10.1007/978-3-658-38891-1> (Zugriff am: 22. Jan. 2025)

²⁷Vgl. S. Oh, M. Kang, H. Moon, K. Choi, und B. S. Chon, „A Demand-Driven Perspective on Generative Audio AI,“ Workshop on Challenges in Deployable Generative AI at the International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA, 2023, Verfügbar:

<https://arxiv.org/pdf/2307.04292>, (Zugriff am: 22. Jan. 2025)

KI in der Musikproduktion

Ein bedeutender Anwendungsbereich der KI ist die Musikproduktion. Systeme wie AIVA (Artificial Intelligence Virtual Artist) erstellen eigenständig Musikstücke, die auf die Präferenzen eines Zielpublikums zugeschnitten sind. Solche Technologien werden häufig in der Film- und Spieleindustrie eingesetzt, um Soundtracks zu komponieren, die sowohl stilistisch passend als auch innovativ sind. Dadurch kann KI kreative Prozesse durch datenbasierte Entscheidungen unterstützen und gleichzeitig neue künstlerische Möglichkeiten eröffnen, indem sie große Datenmengen analysiert und daraus neue Muster ableitet.

Ein bedeutender Anwendungsbereich der Künstlichen Intelligenz (KI) liegt in der Musikproduktion. Moderne KI-Systeme, wie beispielsweise AIVA (Artificial Intelligence Virtual Artist), sind in der Lage, eigenständig Musikstücke zu erstellen, die auf die Präferenzen eines Zielpublikums zugeschnitten sind. Diese Technologien werden zunehmend in der Film- und Spieleindustrie genutzt, um Soundtracks zu komponieren, die sowohl stilistisch passend als auch innovativ sind.

Deep Generative Models haben sich als besonders leistungsfähig für die Erstellung musikalischer Audioinhalte erwiesen. Diese Modelle basieren auf neuronalen Netzwerken, die durch das Training mit großen Datenmengen in der Lage sind, komplexe Muster und Strukturen in der Musik zu lernen. So können sie nicht nur bestehende Stile und Genres reproduzieren, sondern auch völlig neue musikalische Ausdrucksformen schaffen.

Ein wesentlicher Vorteil dieser KI-gestützten Technologien ist ihre Fähigkeit, datenbasierte Entscheidungen zu treffen, um kreative Prozesse zu unterstützen. Durch die Analyse großer Datensätze können solche Systeme musikalische Trends erkennen und innovative Ansätze vorschlagen, die menschliche Kreativität ergänzen. Zudem ermöglichen sie es, in kürzester Zeit hochwertige Kompositionen zu generieren, was den Produktionsprozess erheblich beschleunigt.

Die Arbeit von Huzaifah und Wyse zeigt, dass KI nicht nur als Werkzeug zur Automatisierung betrachtet werden sollte, sondern auch als eine Plattform, die neue künstlerische Möglichkeiten eröffnet. Die Integration von KI in die Musikproduktion erweitert die Grenzen traditioneller Kompositionsmethoden und bietet Musikern und Produzenten neue Wege, um kreative Visionen umzusetzen²⁸.

KI im Film

Die Nutzung von Künstlicher Intelligenz (KI) in der Filmproduktion eröffnet eine Vielzahl von Möglichkeiten, die sowohl kreative Prozesse bereichern als auch bestehende Abläufe optimieren. KI wird in verschiedenen Phasen der Filmproduktion eingesetzt, von der Vorproduktion über die Dreharbeiten bis hin zur Postproduktion.

In der Vorproduktion kann KI bei der Analyse von Drehbüchern und der Planung von Szenen helfen. Algorithmen analysieren frühere Produktionen, um Muster zu identifizieren, die den Erfolg eines Films beeinflussen können. Sie unterstützen Autoren und Regisseure bei der Entwicklung von Handlungssträngen und Charakteren, die besonders ansprechend für das Publikum sind.

Während der Produktion sorgt KI für eine effizientere Umsetzung der Dreharbeiten. Intelligente Kamerasysteme können Bewegungen von Schauspielern automatisch verfolgen

²⁸Vgl. M. Huzaifah und L. Wyse, „Deep Generative Models for Musical Audio Synthesis,“ Preprint for Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, Springer, 2020, Verfügbar: <https://arxiv.org/pdf/2006.06426> (Zugriff am: 22. Jan. 2025)

und Einstellungen optimieren. Zudem können Lichtverhältnisse mithilfe von KI berechnet und in Echtzeit angepasst werden, um eine optimale visuelle Qualität zu gewährleisten. In der Postproduktion spielt KI eine zentrale Rolle bei der Bearbeitung von Bild und Ton. Visuelle Effekte lassen sich mit KI schneller und realistischer umsetzen, was die Erstellung komplexer Szenen deutlich erleichtert. Auch der Schnittprozess wird durch KI beschleunigt, da Szenen automatisch sortiert und synchronisiert werden können. Im Bereich des Sounddesigns ermöglicht KI die präzise Gestaltung individueller Klangwelten, die nahtlos auf die jeweilige Szene abgestimmt sind.

Zusätzlich findet KI Anwendung in der Vermarktung von Filmen. Algorithmen analysieren Zuschauerpräferenzen und generieren gezielte Trailer sowie Werbematerialien, die auf spezifische Zielgruppen zugeschnitten sind. Dies ermöglicht eine personalisierte Ansprache des Publikums und steigert die Reichweite und Attraktivität eines Films.

Durch den Einsatz von KI werden nicht nur Arbeitsabläufe effizienter gestaltet, sondern auch neue kreative Ansätze ermöglicht, die die Qualität und Vielfalt der Filmproduktion auf ein neues Niveau heben. Die Technologie bietet der Filmindustrie somit die Chance, sich weiterzuentwickeln und den wachsenden Ansprüchen des Publikums gerecht zu werden.

KI im kreativen Design

Künstliche Intelligenz (KI) hat die Arbeitsweise im kreativen Design in den letzten Jahren grundlegend verändert und erweitert. KI-gestützte Systeme werden in verschiedenen Bereichen eingesetzt, darunter Grafikdesign, Animation, Produktdesign und visuelle Kommunikation.

Im Grafikdesign können KI-Tools automatisch Layouts, Farbpaletten und Schriftkombinationen erstellen, die ästhetisch ansprechend und auf die Bedürfnisse eines Projekts abgestimmt sind. Dies beschleunigt den Designprozess erheblich und ermöglicht es Designern, sich stärker auf die kreative Konzeption und Feinabstimmung zu konzentrieren. Ein weiterer bedeutender Bereich ist die Erstellung von Animationen. KI-Algorithmen können Bewegungsabläufe analysieren und realistische Animationen generieren, die zuvor manuell erstellt werden mussten. Dies reduziert nicht nur die Produktionszeit, sondern eröffnet auch neue Möglichkeiten für die Erschaffung dynamischer und interaktiver Inhalte.

Im Produktdesign hilft KI bei der Analyse von Markttrends und Kundenpräferenzen, um Designs zu entwickeln, die sowohl funktional als auch ansprechend sind. Mithilfe von KI können Designer neue Konzepte visualisieren und testen, bevor sie in die physische Produktion gehen, was Zeit und Ressourcen spart.

Ein weiterer Vorteil von KI im kreativen Design liegt in der Personalisierung von Inhalten. KI-Modelle können große Mengen an Daten analysieren, um Inhalte zu erstellen, die auf die individuellen Vorlieben und Bedürfnisse von Nutzern zugeschnitten sind. Dies ist besonders im Bereich des Marketings und der Werbung von Bedeutung, wo maßgeschneiderte Designs die Aufmerksamkeit und das Engagement der Zielgruppe erhöhen können.

KI erweitert nicht nur die kreativen Möglichkeiten, sondern steigert auch die Effizienz und Präzision im Designprozess. Sie ermöglicht es, innovative Ideen schneller umzusetzen und hilft, sich wiederholende und zeitintensive Aufgaben zu automatisieren. Dadurch bleibt Designern mehr Raum, sich auf die kreative Vision und die Umsetzung einzigartiger Projekte zu konzentrieren. KI im kreativen Design stellt somit eine wertvolle Ergänzung dar, die sowohl die Qualität der Ergebnisse als auch die Arbeitsweise in der Branche nachhaltig verändert.

4. KI-generiertes Sounddesign: Wie verändert sich der Prozess?

4.1 Einführung in KI im Sounddesign

Künstliche Intelligenz (KI) verändert den Prozess des Sounddesigns grundlegend. Im Gegensatz zu traditionellen Methoden, die auf manuellen Anpassungen und iterativen Prozessen beruhen, bieten KI-gestützte Ansätze eine automatisierte und präzisere Gestaltung von Klanglandschaften. Generative Modelle wie Generative Adversarial Networks (GANs) und Variational Autoencoders (VAEs) ermöglichen nicht nur die Analyse bestehender Klänge, sondern auch die kreative Generierung neuer akustischer Welten. Anwendungen wie AudioLDM zeigen, wie KI Text in Klang umwandeln kann, wodurch völlig neue Werkzeuge für Sounddesigner entstehen ²⁹.

4.2 Technologien und Tools, die KI nutzen

Die Integration von Künstlicher Intelligenz (KI) in das Sounddesign hat zu einer Vielzahl innovativer Technologien und Tools geführt, die es ermöglichen, Klänge effizient und kreativ zu generieren. Diese Werkzeuge nutzen fortschrittliche KI-Modelle wie Generative Adversarial Networks (GANs) und Transformer-basierte Ansätze, um realistische und anpassbare Klanglandschaften zu schaffen. Im Folgenden werden einige der führenden Technologien kurz vorgestellt.

4.2.1 StabilityAI/Stable Audio Open: Generative Audio-Erstellung

Stable Audio Open von StabilityAI ist ein hochentwickeltes Tool, das auf generativen Modellen basiert und speziell für die Erstellung von Audioinhalten aus Texteingaben entwickelt wurde. Dieses innovative System nutzt fortschrittliche Machine-Learning-Technologien, um komplexe und qualitativ hochwertige Klanglandschaften zu generieren, die speziell auf die Anforderungen von Film, Medien und anderen kreativen Projekten zugeschnitten sind ³⁰.

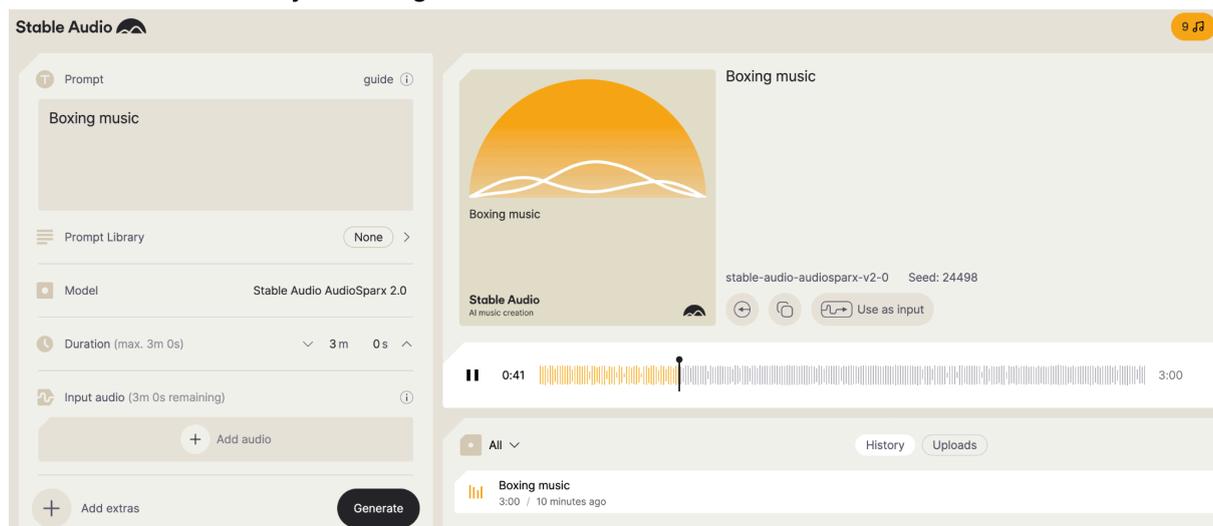


Abbildung 2: Benutzeroberfläche von Stable Audio Open

²⁹Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

³⁰Stable Audio Open by StabilityAI: <https://stability.ai/stable-audio>, <https://huggingface.co/spaces/artificialguybr/Stable-Audio-Open-Zero>

Funktionsweise von Stable Audio Open

Stable Audio Open bietet ein modernes und benutzerfreundliches Interface, das Sounddesignern und Kreativen eine präzise Steuerung über ihre Audioerstellung ermöglicht. Ein zentrales Element des Interfaces ist das Textfeld für Prompts, in dem Nutzer ihre Beschreibungen eingeben können, um Klänge zu erzeugen. Ein Beispiel hierfür wäre „Boxing music“, wodurch das Tool passende musikalische Inhalte generiert, die sich direkt in Projekte integrieren lassen.

Zusätzlich bietet Stable Audio Open die Möglichkeit, ein spezifisches Modell auszuwählen, wie zum Beispiel „Stable Audio AudioSparx v2.0“. Diese Modelle sind auf unterschiedliche Anforderungen optimiert und erlauben es Nutzern, gezielt Klänge für verschiedene Zwecke wie Hintergrundmusik, thematische Klanglandschaften oder spezielle Effekte zu erstellen. Über das Dropdown-Menü für die Dauer (Duration) kann die gewünschte Länge des erzeugten Audiosignals präzise festgelegt werden, mit einer maximalen Begrenzung von 3 Minuten.

Eine weitere innovative Funktion ist die Möglichkeit, Audioinput hochzuladen, der als Grundlage für die Audioerstellung dient. Dies erlaubt es Nutzern, bestehende Audiodateien zu verfeinern oder mit neuen generativen Elementen zu kombinieren. Das Interface bietet außerdem eine „Prompt Library“, die vordefinierte Eingaben enthält, um den kreativen Prozess zu erleichtern.

Dank dieser spezifischen Funktionen ist Stable Audio Open nicht nur äußerst flexibel, sondern auch einfach zu bedienen. Die Plattform ermöglicht es, Klänge effizient und in professioneller Qualität zu erstellen, ohne dass umfangreiche Nachbearbeitungen erforderlich sind. Durch die intuitive Steuerung und die vielseitigen Anpassungsmöglichkeiten können sowohl erfahrene Sounddesigner als auch Einsteiger hochwertige Audioinhalte erstellen, die direkt in Projekten eingesetzt werden können. Diese Kombination aus einfacher Bedienung und leistungsstarken Funktionen macht Stable Audio Open zu einem unverzichtbaren Werkzeug in der modernen kreativen Industrie.

Funktionsweise von Stable Audio Open Zero

Stable Audio Open Zero erweitert die Möglichkeiten der Audio-Generierung durch zusätzliche Parameter, die den kreativen Prozess individualisierbarer und präziser machen. Diese Parameter erlauben Nutzern eine detaillierte Kontrolle über die Generierung der Klänge, sodass spezifische Anforderungen flexibel umgesetzt werden können.

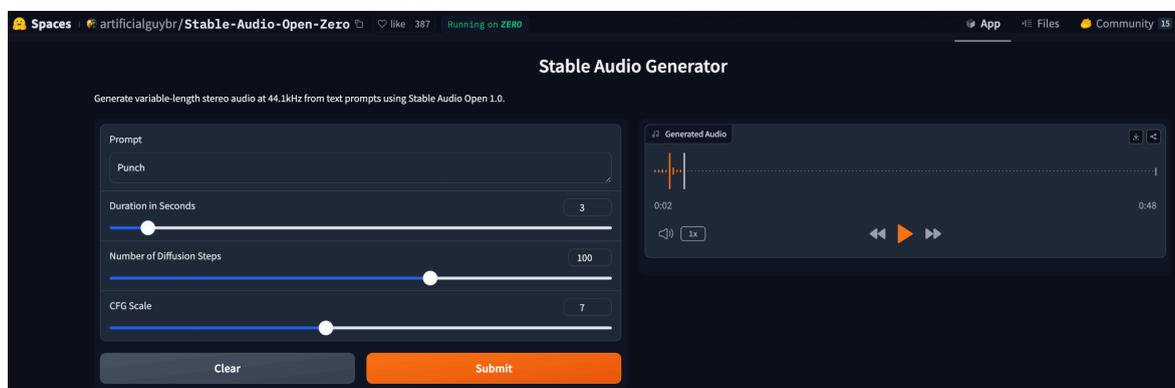


Abbildung 3: Benutzeroberfläche von Stable-Audio-Open-Zero

Stable Audio Open Zero bietet eine Vielzahl von erweiterten Parametern, die Nutzern eine detaillierte Kontrolle über den kreativen Prozess der Klangerzeugung ermöglichen. Der Prompt-Parameter erlaubt es, eine textbasierte Beschreibung einzugeben, die als Grundlage für die Audioerstellung dient. Begriffe wie „Punch“, „Ocean Waves“ oder „Rainstorm“ können eingegeben werden, und die KI interpretiert diese Beschreibungen, um passende Klänge zu generieren.

Mit dem Duration in Seconds-Parameter kann die Länge des erzeugten Audiosignals präzise eingestellt werden. Dies gibt den Nutzern die Möglichkeit, sowohl kurze Soundeffekte wie ein Schlaggeräusch als auch längere Klanglandschaften, beispielsweise atmosphärische Hintergrundgeräusche, zu erstellen.

Ein weiterer wichtiger Parameter ist die Number of Diffusion Steps, die die Anzahl der Diffusionsschritte während des Generierungsprozesses festlegt. Eine höhere Anzahl an Diffusionsschritten sorgt in der Regel für eine klarere und detailliertere Ausgabe, erfordert jedoch mehr Rechenzeit. Dies ermöglicht es den Nutzern, zwischen schnellerer Generierung und höherer Klangqualität abzuwägen.

Die CFG Scale (Classifier-Free Guidance) beeinflusst, wie stark das generierte Audio dem eingegebenen Prompt entspricht. Eine niedrigere CFG-Einstellung führt zu mehr Variationen und Experimentierfreude, während eine höhere Einstellung die Ausgabe präziser auf die Eingabe abstimmt, was besonders bei spezifischen Anforderungen hilfreich ist.

Diese erweiterten Parameter bieten zahlreiche Vorteile. Die Möglichkeit zur Feinsteuerung ermöglicht es Nutzern, Klänge exakt an ihre Anforderungen anzupassen. Die Qualitätsanpassung durch die Kontrolle der Diffusionsschritte gibt ihnen die Flexibilität, den richtigen Kompromiss zwischen Geschwindigkeit und Klangqualität zu finden. Zusätzlich sorgt die Flexibilität der CFG-Skala dafür, dass der kreative Prozess dynamisch bleibt, da experimentelle oder exakte Ergebnisse gezielt gesteuert werden können. Trotz dieser umfangreichen Anpassungsmöglichkeiten bleibt das Interface intuitiv und benutzerfreundlich, sodass auch weniger erfahrene Nutzer in der Lage sind, hochwertige Audioinhalte zu erstellen.

Diese Kombination aus erweiterter Kontrolle und einfacher Bedienung macht Stable Audio Open Zero zu einem mächtigen Werkzeug für professionelle Sounddesigner und Kreative, die maßgeschneiderte Klänge für ihre Projekte benötigen.

Einsatzmöglichkeiten

Stable Audio Open Zero eignet sich ideal für Projekte, bei denen präzise oder individuell gestaltete Klänge erforderlich sind, z. B. in Filmproduktionen, bei der Gestaltung von Soundeffekten oder in der Musikproduktion. Die erweiterten Einstellungen machen es besonders attraktiv für professionelle Sounddesigner, die tiefere Kontrolle über ihre Audioausgaben benötigen.

Herausforderungen und Limitierungen

Trotz der beeindruckenden Möglichkeiten, die Stable Audio Open und Stable Audio Open Zero bieten, gibt es auch einige Herausforderungen und Limitierungen, die Nutzer berücksichtigen müssen. Eine der größten Schwierigkeiten liegt in der Abhängigkeit von der Präzision der Texteingaben. Unklare oder zu allgemeine Prompts führen oft zu ungenauen oder unerwarteten Ergebnissen, die nicht den Anforderungen des Projekts entsprechen. Die

Nutzer müssen also nicht nur kreativ, sondern auch äußerst präzise in der Formulierung ihrer Prompts sein, um brauchbare Klänge zu erhalten.

Darüber hinaus bieten die erweiterten Parameter bei Stable Audio Open Zero zwar eine größere Kontrolle über den Generierungsprozess, setzen jedoch auch ein gutes Verständnis der Funktionsweise voraus. Falsch abgestimmte Einstellungen, wie eine zu hohe Diffusionsstufe oder eine ungeeignete CFG-Skala, können dazu führen, dass die erzeugten Klänge verzerrt, unpassend oder gar verstörend wirken. Nicht selten kommt man zu Ergebnissen, die unerwünschte zusätzliche Sounds enthalten und nicht zur Beschreibung oder zur beabsichtigten Atmosphäre passen. Diese unerwarteten Artefakte erschweren es, konsistente und qualitativ hochwertige Ergebnisse zu erzielen.

Ein weiteres Problem ist die Audioqualität. Obwohl das Tool hochwertige Klänge verspricht, sind diese oft nicht in der Form brauchbar, die für professionelle Produktionen erforderlich ist. Vor allem bei längeren oder komplexeren Klanglandschaften kann es zu hörbaren Verzerrungen, Brüchen oder unnatürlichen Übergängen kommen. Diese erfordern in der Regel zusätzliche Bearbeitungsschritte, was den ursprünglichen Vorteil der Zeitersparnis reduziert.

Zudem ist die Nutzung des Tools von Credits und Systemressourcen abhängig. Besonders bei umfangreichen Projekten kann die begrenzte Verfügbarkeit der Ressourcen dazu führen, dass Nutzer ihre Arbeit unterbrechen oder unvollständig lassen müssen. Diese Einschränkung macht das Tool weniger attraktiv für Projekte, die kontinuierliche und intensive Nutzung erfordern.

Zusammenfassend lässt sich sagen, dass Stable Audio Open und Stable Audio Open Zero zwar ein enormes Potenzial bieten, die Arbeit von Sounddesignern zu revolutionieren, jedoch auch mit Herausforderungen verbunden sind. Nutzer müssen bereit sein, Zeit in die Optimierung ihrer Prompts und Einstellungen zu investieren und sich auf zusätzliche Nachbearbeitungen einzustellen, um das Beste aus dem Tool herauszuholen. Trotz dieser Limitierungen bleibt Stable Audio Open ein wertvolles Werkzeug, das jedoch vor allem in der kreativen und experimentellen Phase seines größten Nutzens entfaltet.

4.2.2 Meta AudioGen

Meta AudioGen ist ein von Meta entwickeltes Tool, das sich so wie StabilityAI auf die Generierung von Audio und Musik aus textbasierten Eingaben spezialisiert hat. Es kombiniert fortschrittliche KI-Algorithmen mit einer benutzerfreundlichen Schnittstelle, die die schnelle Integration von Klängen in narrative und interaktive Medien ermöglicht³¹.

Das Tool richtet sich an eine breite Zielgruppe – von erfahrenen Sounddesignern bis hin zu Einsteigern – und bietet vielseitige Einsatzmöglichkeiten in den Bereichen Film, Videospiele, Werbung und Musikproduktion. Durch die präzise Verarbeitung von Texteingaben können individuelle und maßgeschneiderte Klänge erzeugt werden, die auf die spezifischen Anforderungen eines Projekts abgestimmt sind.

Ein zentraler Vorteil von Meta AudioGen liegt in seiner Effizienz und Qualität: Die erzeugten Audiodateien sind professionell und können direkt in Produktionen verwendet werden, wodurch zeitaufwändige Nachbearbeitungen reduziert werden. Diese Kombination aus leistungsfähigen KI-Technologien und einfacher Bedienbarkeit macht Meta AudioGen zu einem unverzichtbaren Werkzeug für die moderne kreative Industrie.

³¹AudioGen by Meta AI: <https://audiocraft.metademolab.com/audiogen.html>

4.2.3 Suno und AVIA: KI-basierte Musikkomposition

Suno ³² und AVIA ³³ sind Werkzeuge zur KI-basierten Musikkomposition. Sie bieten Funktionen, die es ermöglichen, Musik zu erstellen, die sich stilistisch an verschiedene Genres anpasst.

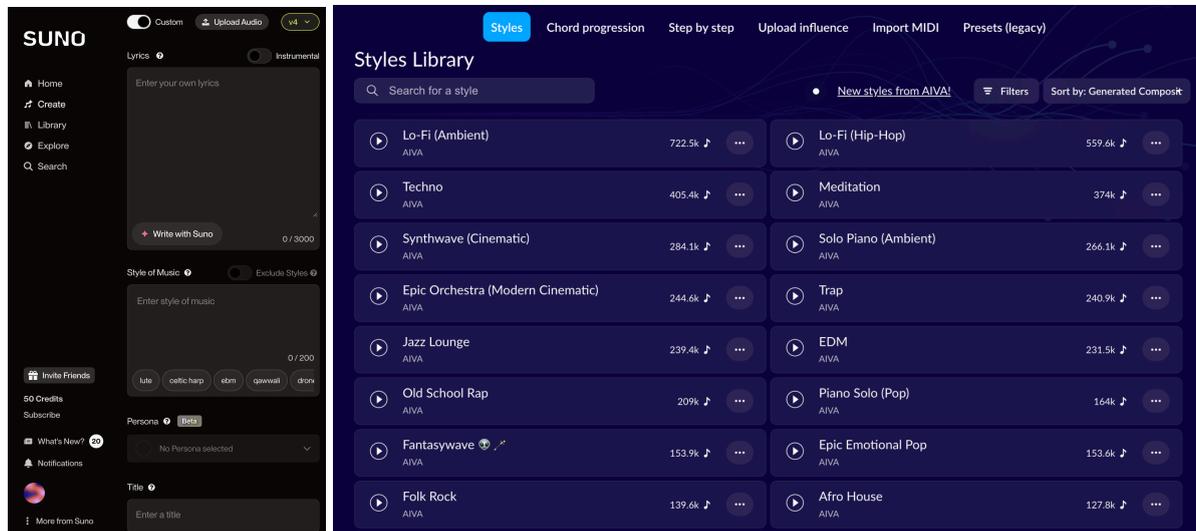


Abbildung 4: Benutzeroberfläche von Suno

Abbildung 5: Benutzeroberfläche von AVIA

Suno ist darauf spezialisiert, hochwertige Hintergrundmusik zu generieren, die sich ideal für Filme, Videospiele, Werbung und andere audiovisuelle Medien eignet. Das System analysiert die vom Nutzer eingegebenen Text-Prompts oder vordefinierten Stimmungen und erstellt darauf basierend Musikstücke, die perfekt auf die gewünschte Atmosphäre abgestimmt sind. Dabei liegt ein besonderer Fokus auf der Fähigkeit, harmonische und unaufdringliche Klanglandschaften zu gestalten, die ein Projekt ergänzen, ohne dabei abzulenken. Diese Eigenschaft macht Suno besonders geeignet für Anwendungen wie ruhige Filmszenen, Präsentationen oder den Gaming-Bereich, wo dynamische Soundscapes benötigt werden.

Ein herausragender Vorteil von Suno ist seine intuitive Benutzeroberfläche, die auch weniger erfahrenen Nutzern den Zugang zu professioneller Musikkomposition ermöglicht. Mit wenigen Eingaben können maßgeschneiderte Musikstücke generiert werden, die direkt in Projekte integriert werden können. Besonders die Möglichkeit, Genres zu wählen oder sogar Audioinputs hochzuladen, erhöht die Flexibilität des Tools. Nutzer können zudem Lyrics eingeben, um komplette Songs zu erstellen, wobei jedoch die generierten Stimmen oft etwas unnatürlich wirken. Instrumentalstücke hingegen überzeugen durch bessere Ergebnisse.

Ein weiterer Pluspunkt von Suno ist die relativ schnelle Generierung der Musik, was den kreativen Prozess erheblich beschleunigt. Allerdings zeigt sich in der Praxis, dass die Klangqualität häufig suboptimal ist und eine Nachbearbeitung erforderlich sein kann, um die Ergebnisse zu optimieren. Dennoch lässt sich mit den generierten Inhalten gut arbeiten, insbesondere bei Projekten, die schnelle und effektive Lösungen erfordern.

Insgesamt ist Suno ein vielseitiges und benutzerfreundliches Werkzeug, das sowohl für Einsteiger als auch für erfahrene Nutzer zahlreiche Möglichkeiten bietet. Trotz kleinerer Schwächen, wie der Qualität von Stimmen oder gelegentlicher Nachbearbeitung der Instrumentals, bleibt es eine wertvolle Ressource für Produzenten und Designer, die auf der Suche nach flexiblen und effizienten Erstellung von Musik sind.

³²Suno AI: <https://suno.com/>

³³AVIA AI Music: <https://www.aiva.ai/>

AVIA (Artificial Intelligence Virtual Artist) ist ein vielseitiges und leistungsstarkes Werkzeug, das speziell für die Erstellung komplexer Musikkompositionen entwickelt wurde. Im Gegensatz zu Suno, das sich auf die Generierung von Hintergrundmusik konzentriert, bietet AVIA eine breite Palette an musikalischen Möglichkeiten, die von orchestralen Werken bis hin zu moderner Pop- und elektronischer Musik reichen. Die Flexibilität und Anpassungsfähigkeit machen AVIA zu einem idealen Werkzeug für Produzenten und Komponisten, die sowohl klassische als auch zeitgenössische Musikstile erkunden möchten. Im Bereich der orchestralen Musik zeigt AVIA seine besondere Stärke. Es ist in der Lage, detaillierte und emotionale Kompositionen zu generieren, die oft in Film- und Videospielproduktionen zum Einsatz kommen. Das System erstellt authentische und dynamische Arrangements mit einer Vielzahl von Instrumenten und bietet Nutzern die Möglichkeit, Parameter wie Tempo, Instrumentierung und Stimmung individuell anzupassen, um die Musik perfekt auf ihr Projekt abzustimmen.

Neben orchestralen Werken bietet AVIA umfangreiche Möglichkeiten für moderne Musikgenres wie Techno, Lofi, Pop oder experimentelle elektronische Musik. Im Gegensatz zu textbasierten Prompts, wie sie bei Suno verwendet werden, ermöglicht AVIA eine flexible Eingabe durch verschiedene Optionen. Nutzer können einen vorgegebenen Stil wählen, beispielsweise Techno oder Lofi, oder mit Akkordprogressionen arbeiten, um harmonische Grundlagen zu schaffen. Darüber hinaus besteht die Möglichkeit, eigene Audioinputs hochzuladen oder MIDI-Dateien einzubinden, die als Basis für die weitere Komposition dienen.

Ein entscheidender Vorteil von AVIA ist die hohe Qualität der generierten Musik. Im Vergleich zu Suno liefert AVIA oft klarere und professionellere Ergebnisse, was es besonders für Projekte geeignet macht, die höchste klangliche Ansprüche stellen. Diese Kombination aus Qualität, Vielseitigkeit und benutzerfreundlicher Oberfläche macht AVIA zu einem unverzichtbaren Werkzeug in der modernen Musikproduktion.

AVIA ermöglicht es Produzenten, effizient und kreativ zu arbeiten, indem es die nahtlose Verbindung von traditioneller und moderner Musik schafft. Ob für Filme, Videospiele, Werbung oder experimentelle Musikprojekte – AVIA bietet eine zukunftsweisende Lösung, die den kreativen Prozess bereichert und neue Horizonte in der Musikkomposition eröffnet.

Beide Tools haben jedoch auch ihre Herausforderungen. Obwohl sie effizient Musik generieren können, fehlt oft die vollständige Kontrolle, wie sie bei einer eigenständigen Komposition gegeben wäre. Die generierten Ergebnisse überzeugen nicht immer zu 100 %, und die Qualität kann variieren, was ihre Integration in professionelle Projekte erschwert. Wie bei allen generativen Modellen sind die Möglichkeiten durch begrenzte Generationen und Prompts eingeschränkt, was den kreativen Spielraum reduzieren kann.

Trotz dieser Einschränkungen bieten Suno und AVIA leistungsstarke Werkzeuge, um schnell und effizient Musikstücke zu erstellen, die sich für eine Vielzahl kreativer Anwendungen eignen. Während Suno durch seine Einfachheit und Geschwindigkeit überzeugt, glänzt AVIA durch seine Vielseitigkeit und die Möglichkeit, anspruchsvolle Kompositionen zu erstellen. Beide Tools bereichern die moderne Musikproduktion und eröffnen neue Horizonte für kreative Prozesse.

4.2.4 ElevenLabs: Sprachsynthese und Soundgeneration

ElevenLabs ist ein führendes Tool im Bereich der Sprachsynthese und Soundgeneration, das Sounddesignern und Produzenten eine Vielzahl von innovativen Funktionen bietet. Es ist besonders darauf spezialisiert, realistische und anpassbare Stimmen zu generieren, die sich in verschiedenen Szenarien einsetzen lassen. Dank seiner fortschrittlichen Technologie, intuitiven Bedienung und hohen Qualität der Sounds hat ElevenLabs eine zentrale Rolle in der Studioproduktion "Knockout" eingenommen und sich als unverzichtbares Werkzeug etabliert ³⁴.

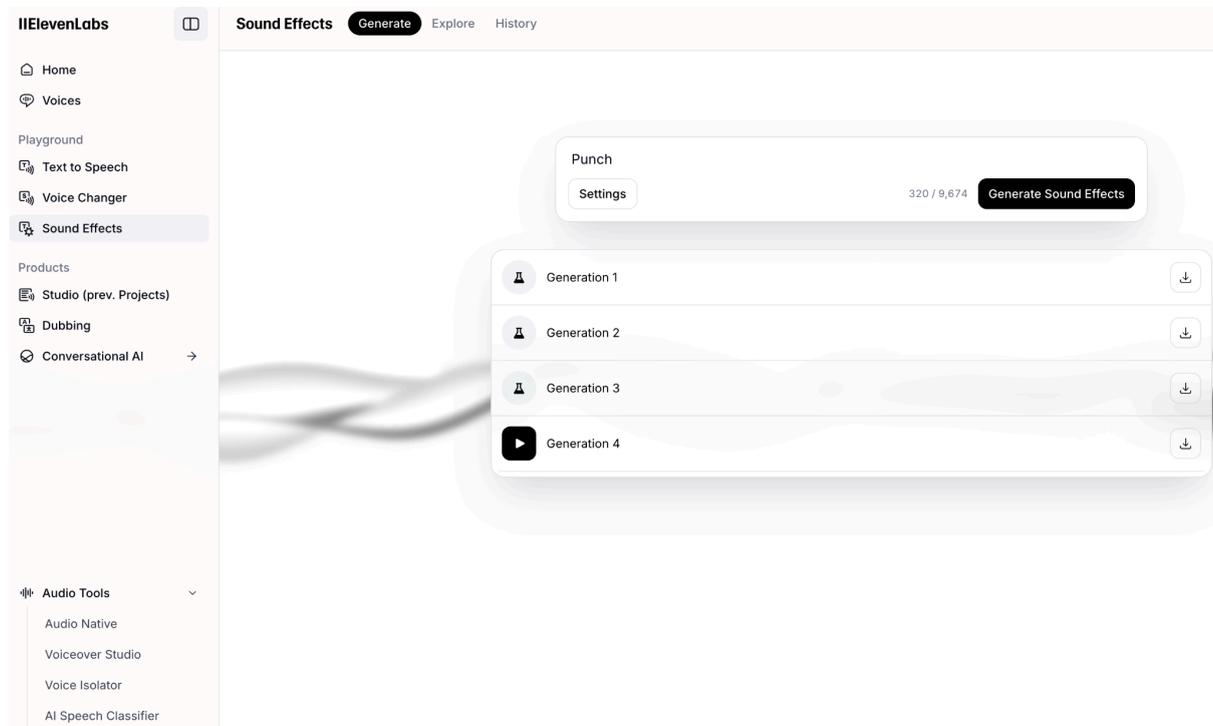


Abbildung 6: Benutzeroberfläche von Elevenlabs

Kernfunktionen von ElevenLabs

ElevenLabs bietet eine Vielzahl innovativer Funktionen, die speziell für die professionelle Sprachsynthese und Soundgeneration entwickelt wurden. Diese Tools ermöglichen es Nutzern, realistische und anpassbare Audioinhalte zu erstellen, die vielseitig in der Medienproduktion einsetzbar sind.

Text-to-Speech (TTS)

Das Text-to-Speech-Tool von ElevenLabs ist eine der zentralen Funktionen und erlaubt die Erstellung realistischer Sprachaufnahmen basierend auf Texteingaben. Nutzer haben die Möglichkeit, aus einer Vielzahl hochwertiger Stimmen in mehreren Sprachen zu wählen, die individuell in Tonalität, Persönlichkeit und Ausdruck angepasst werden können. Diese Funktion ist ideal für Voiceovers, Synchronisationen und die Erstellung von Dialogen in Filmen, Spielen oder anderen Medienprojekten. Besonders hervorzuheben ist die Fähigkeit, Emotionen wie Freude, Trauer oder Wut in den Stimmen zu transportieren, wodurch die Immersion in Projekten erheblich gesteigert wird.

³⁴ElevenLabs – Speech Synthesis: <https://elevenlabs.io/>

Voice Changer

Der Voice Changer ermöglicht es, bestehende Sprachaufnahmen flexibel zu manipulieren. Stimmen können beispielsweise in Geschlecht, Alter oder Tonalität verändert werden, was die Vielseitigkeit der Inhalte erhöht. Dies ist besonders nützlich, um dieselbe Stimme für verschiedene Charaktere zu verwenden oder bestehende Aufnahmen an neue Szenarien anzupassen. Der Voice Changer eröffnet dadurch neue kreative Möglichkeiten und spart gleichzeitig Ressourcen bei der Produktion.

Sound Effects

Mit dem Soundeffekt-Tool lassen sich gezielt Geräusche basierend auf Text-Prompts generieren. Begriffe wie „Punch“ oder „Explosion“ können eingegeben werden, und das System liefert mehrere Variationen des gewünschten Sounds. Nutzer können aus bis zu vier generierten Versionen auswählen, um die beste Option für ihr Projekt zu finden. Diese Funktion ist ideal, um schnell und unkompliziert Soundeffekte zu erstellen, die oft nur minimal nachbearbeitet werden müssen.

Weitere Funktionen

Zusätzlich zu den Kernfunktionen bietet ElevenLabs eine Reihe weiterer Tools, die es zu einer umfassenden Lösung für die Audioproduktion machen. Dazu gehört ein Dubbing-Tool, das speziell für Synchronisationsprojekte entwickelt wurde, sowie ein Voice Isolator, der Stimmen präzise von Hintergrundgeräuschen trennt. Darüber hinaus steht ein Conversational AI-Feature zur Verfügung, das interaktive Anwendungen wie Chatbots oder dynamische Dialogsysteme unterstützt. Diese erweiterten Funktionen machen ElevenLabs zu einer flexiblen und leistungsstarken Plattform für unterschiedlichste Anforderungen in der Audio- und Medienproduktion.

Mit dieser Kombination aus fortschrittlicher Technologie, benutzerfreundlichem Interface und vielseitigen Anpassungsmöglichkeiten stellt ElevenLabs eine leistungsstarke Sammlung an Werkzeugen bereit, welche den kreativen Prozess in der Audioproduktion erheblich erleichtern und erweitern.

4.3 Symbolische und nicht-symbolische Audiogenerierung

Die Audiogenerierung kann grundsätzlich in zwei Ansätze unterteilt werden: symbolische und nicht-symbolische Methoden. Diese Ansätze unterscheiden sich durch die Art der Repräsentation und die Verarbeitung von Audioinformationen.

Symbolische Audiogenerierung

Symbolische Audiogenerierung arbeitet auf der Grundlage abstrakter Repräsentationen wie Noten, MIDI-Daten oder Partituren, die spezifische musikalische Informationen kodieren. Diese Methode abstrahiert von den physikalischen Klangwellen und ermöglicht die Modellierung von Musik auf einer konzeptionellen Ebene. Symbolische Ansätze fokussieren sich auf die Erstellung von Melodien, Harmonien und rhythmischen Mustern. Ein prominentes Beispiel ist PerformanceRNN, ein Modell, das MIDI-Noten generiert und dabei Dynamik und Timing berücksichtigt, um realistische musikalische Darstellungen zu erzeugen ³⁵.

Symbolische Ansätze sind besonders nützlich für die Komposition von Musikstücken, da sie mit musikalischen Konzepten arbeiten, die leicht von Musikern und Produzenten interpretiert und bearbeitet werden können. Sie werden häufig in Musikproduktionssoftware integriert und ermöglichen es Künstlern, neue Kompositionen zu erstellen, die sich an bestehenden Stilen orientieren.

Nicht-symbolische Audiogenerierung

Nicht-symbolische Audiogenerierung arbeitet direkt mit Rohdaten wie Audio-Wellenformen oder spektralen Darstellungen. Diese Methoden nutzen fortschrittliche neuronale Netzwerke, um Klänge realistisch und detailliert zu erzeugen, ohne auf abstrakte Symbole angewiesen zu sein. Modelle wie WaveNet und Jukebox von OpenAI verwenden autoregressive Ansätze, um hochqualitative Audioinhalte zu generieren. Diese Modelle analysieren und reproduzieren feine Nuancen von Klang, einschließlich Timbre und Textur, und sind in der Lage, komplexe Audio-Szenarien zu simulieren ³⁶.

Ein wesentlicher Vorteil der nicht-symbolischen Audiogenerierung liegt in ihrer Fähigkeit, Klänge zu erstellen, die vollständig original und nicht an vordefinierte musikalische Regeln gebunden sind. Dies macht sie ideal für die Produktion von Soundeffekten und die Gestaltung immersiver Klanglandschaften in Film und Videospielen.

Vergleich und Anwendungen

Symbolische und nicht-symbolische Ansätze ergänzen sich in der Praxis oft. Während symbolische Methoden eine höhere Abstraktionsebene für die Erstellung von Musikstrukturen bieten, ermöglichen nicht-symbolische Modelle eine detaillierte klangliche Ausgestaltung. Ein häufiges Anwendungsszenario ist die Kombination beider Ansätze, etwa indem ein symbolisches Modell die Struktur eines Musikstücks generiert und ein nicht-symbolisches Modell die Klänge realistischer gestaltet.

Diese Kombination findet sich in modernen Anwendungen wie generativen Audiotools, die sowohl strukturierte musikalische Elemente als auch natürliche Klangeigenschaften benötigen, um realistische und kreative Audioinhalte zu liefern.

³⁵Vgl. M. Huzaifah und L. Wyse, „Deep Generative Models for Musical Audio Synthesis,“ Preprint for Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, Springer, 2020, Verfügbar: <https://arxiv.org/pdf/2006.06426> (Zugriff am: 22. Jan. 2025)

³⁶Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

4.4 Vorteile und neue Möglichkeiten durch KI

Künstliche Intelligenz (KI) bietet im Sounddesign eine Vielzahl von Vorteilen, die traditionelle Methoden grundlegend verändern. Durch den Einsatz moderner KI-Technologien werden Prozesse nicht nur effizienter gestaltet, sondern auch die kreativen Möglichkeiten erheblich erweitert. Diese Entwicklungen sind nicht nur für die Film- und Musikindustrie relevant, sondern finden auch in der Spieleentwicklung und der Werbung Anwendung. Die Vorteile lassen sich in mehreren Bereichen zusammenfassen.

Effizienzsteigerung durch Automatisierung

Eine der zentralen Stärken von KI im Sounddesign liegt in der Automatisierung repetitiver und zeitaufwendiger Aufgaben. Tätigkeiten wie die Synchronisation von Klängen mit visuellen Szenen oder die Erzeugung von Geräuschen, die zuvor manuell erstellt werden mussten, können durch KI-Modelle erheblich beschleunigt werden. Insbesondere generative Modelle wie AudioLDM zeigen, wie KI-gestützte Ansätze die Erstellung von Klängen in Echtzeit ermöglichen. Dies reduziert die Produktionszeit erheblich und gibt Sounddesignern mehr Raum für kreative Entscheidungen ³⁷.

Darüber hinaus können KI-Systeme bestehende Audioinhalte analysieren und präzise an spezifische Anforderungen anpassen. Dies ist besonders bei groß angelegten Projekten hilfreich, in der zahlreiche Klangkomponenten integriert und synchronisiert werden müssen. KI-Tools übernehmen hierbei die Feinabstimmung, was den menschlichen Aufwand minimiert und gleichzeitig die Konsistenz der Ergebnisse erhöht.

Erweiterung kreativer Möglichkeiten

Ein weiterer zentraler Vorteil von KI im Sounddesign ist die Möglichkeit, neue Klanglandschaften zu erschaffen, die mit traditionellen Methoden schwer oder gar nicht umsetzbar wären. Generative Modelle wie Variational Autoencoders (VAEs) und Generative Adversarial Networks (GANs) ermöglichen die Synthese origineller Klänge. Diese Modelle können nicht nur bestehende Klangmuster analysieren, sondern auch völlig neue Audioelemente generieren, die den kreativen Horizont erweitern.

Die Anwendung von Transformer-Modellen in der Audioproduktion, wie bei AudioLDM, zeigt, wie Textinformationen genutzt werden können, um spezifische Klänge zu erstellen. Diese Technologie bietet eine völlig neue Ebene der kreativen Kontrolle, da Sounddesigner ihre Ideen direkt in Textform ausdrücken können, um entsprechende Klänge zu generieren ³⁸.

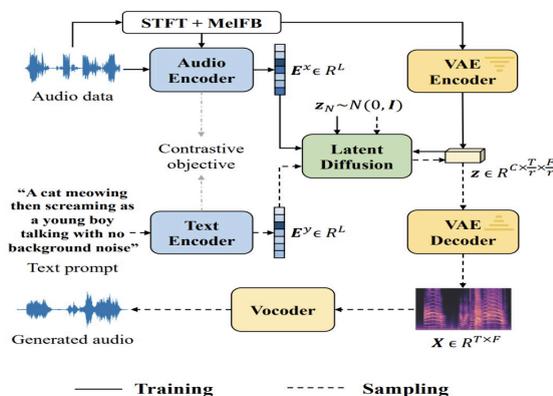


Abbildung 7: Übersicht von AudioLDM design für tex-zu-audio generation

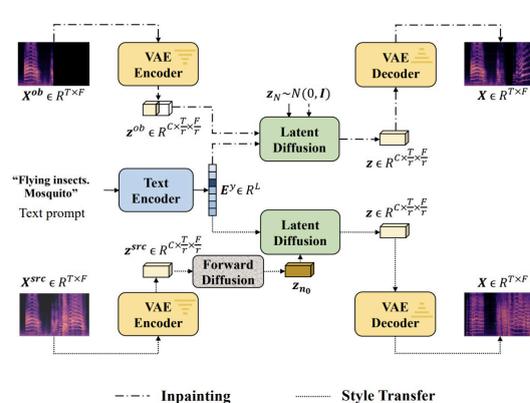


Abbildung 8: Text-basierte Audio Manipulation

³⁷Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

³⁸Vgl. K. Tatar, K. Cotton, und D. Bisig, „Sound Design Strategies for Latent Audio Space Explorations Using Deep Learning Architectures,“ in Proceedings of the Sound and Music Computing Conference (SMC), Stockholm, Sweden, 2023, Verfügbar: <https://arxiv.org/pdf/2305.15571>, (Zugriff am: 22. Jan. 2025)

Verbesserung der Qualität und Präzision

KI-Systeme bieten ein beispielloses Maß an Präzision und Qualität in der Audiogenerierung. Im Vergleich zu traditionellen Methoden, die oft durch manuelle Eingriffe und subjektive Entscheidungen geprägt sind, liefern KI-Modelle konsistente und qualitativ hochwertige Ergebnisse. Dies ist besonders bei der Erstellung von Klanglandschaften für immersive Anwendungen von Vorteil, wo realistische Klänge entscheidend für die Benutzererfahrung sind.

Ein Beispiel ist die Möglichkeit, Klangprofile für verschiedene Szenarien zu optimieren, wie es in der Film- und Spieleindustrie häufig erforderlich ist. KI-gestützte Tools können hierbei spezifische Klangparameter analysieren und anpassen, um ein realistisches oder ansprechendes Klangerlebnis zu schaffen.

Ein Beispiel eines solchen KI-Tools ist Soothe 2, ein intelligenter, dynamischer Equalizer von Oeksound, der mithilfe von künstlicher Intelligenz unerwünschte Resonanzen und harsche Frequenzen automatisch erkennt und reduziert. Während herkömmliche Equalizer manuell auf störende Frequenzbereiche abgestimmt werden müssen, analysiert Soothe 2 das eingehende Audiosignal in Echtzeit und wendet gezielte Anpassungen an, um einen ausgeglichenen und natürlichen Klang zu erzeugen.



Abbildung 9: Benutzeroberfläche von Soothe 2

Soothe 2 nutzt maschinelles Lernen, um störende Frequenzen gezielt zu identifizieren und zu dämpfen, ohne den Klangcharakter zu verfälschen. Der Algorithmus analysiert das Spektrum in Echtzeit und reduziert problematische Resonanzen nur dann, wenn sie tatsächlich auftreten. Dadurch bleibt der natürliche Klang erhalten, während schrille oder harte Frequenzen gezielt abgeschwächt werden.

Besonders nützlich ist Soothe 2 in der Audiotbearbeitung für Vocals, wo es Zischlaute minimiert, sowie für Gitarren und Streicher, indem es unangenehme Härten reduziert. Drums und Percussion profitieren von einer gezielten Kontrolle resonanter Peaks, ohne dass wichtige Transienten verloren gehen. In der Film- und Audioproduktion sorgt es für eine harmonische Gesamtmischung und entfernt störende Klangspitzen.

Insgesamt liegen die Vorteile von KI im Sounddesign in der Automatisierung, der Erweiterung kreativer Möglichkeiten und der Verbesserung der Klangqualität. Trotz dieser Fortschritte bleiben Herausforderungen wie die Abhängigkeit von großen Datensätzen und hohe Rechenanforderungen, die zukünftig weiter optimiert werden müssen.

4.5 Herausforderungen und technische Grenzen von KI im Sounddesign

Die Integration von Künstlicher Intelligenz (KI) in das Sounddesign bringt nicht nur Vorteile, sondern auch eine Vielzahl von Herausforderungen und technischen Grenzen mit sich. Diese betreffen sowohl die Datenqualität und Modellleistung als auch ethische und praktische Aspekte der Implementierung.

Abhängigkeit von Datensätzen

Ein zentrales Problem bei der Nutzung von KI im Sounddesign ist die Abhängigkeit von qualitativ hochwertigen und umfangreichen Datensätzen. Viele Modelle benötigen große Mengen an Trainingsdaten, um realistische und kreative Klänge zu erzeugen. Allerdings ist die Verfügbarkeit solcher Datensätze begrenzt. Besonders im Vergleich zu anderen Bereichen wie der Bildverarbeitung fehlen Audio-Datensätze, die eine ähnliche Größe und Vielfalt aufweisen. Dies schränkt nicht nur die Leistungsfähigkeit der Modelle ein, sondern erhöht auch das Risiko von Verzerrungen und begrenzter Generalisierbarkeit ³⁹.

Begrenzte Kontrolle und Interpretierbarkeit

Ein weiteres Hindernis ist die eingeschränkte Kontrolle über die generierten Klänge. Viele aktuelle KI-Modelle, wie generative Adversarial Networks (GANs), können zwar beeindruckende Ergebnisse erzielen, sind jedoch oft schwer steuerbar. Dies ist besonders problematisch im Sounddesign, wo Präzision und Anpassungsfähigkeit entscheidend sind. Sounddesigner benötigen Tools, die intuitive und präzise Eingriffe in die generierten Klänge ermöglichen. Laut einer Umfrage unter Filmtoneingenieuren ist die mangelnde Kontrolle einer der Hauptkritikpunkte bei der Nutzung von KI im Audibereich ⁴⁰.

Technologische Barrieren

Die technologische Infrastruktur stellt eine weitere Grenze dar. Die meisten KI-Systeme erfordern erhebliche Rechenressourcen, insbesondere wenn sie in Echtzeit arbeiten sollen. Dies kann den Einsatz solcher Technologien in kleineren Produktionen oder bei unabhängigen Sounddesignern erschweren. Zudem sind viele aktuelle Systeme noch nicht in der Lage, hochqualitative Audiodaten in der für professionelle Produktionen erforderlichen Geschwindigkeit und Qualität zu liefern.

Ethische und rechtliche Herausforderungen

Neben technischen Einschränkungen gibt es auch ethische und rechtliche Herausforderungen. Dazu gehören Fragen des Urheberrechts bei der Nutzung von Trainingsdaten und die potenzielle Ersetzung menschlicher Kreativität durch KI-gestützte Prozesse. Der Einsatz von KI wirft zudem Bedenken hinsichtlich der Authentizität auf, da vollständig generierte Klänge möglicherweise als „künstlich“ wahrgenommen werden. Insbesondere der Schutz der Rechte von Künstlern, deren Werke als Trainingsdaten verwendet werden, bleibt ein ungelöstes Problem ⁴¹.

Trotz ihrer beeindruckenden Fortschritte stehen KI-gestützte Technologien im Sounddesign vor erheblichen Herausforderungen. Die Abhängigkeit von hochwertigen Datensätzen, die eingeschränkte Kontrolle, hohe Rechenanforderungen und ethische Fragen sind Barrieren, die überwunden werden müssen, um das volle Potenzial dieser Technologien auszuschöpfen. Dennoch zeigen aktuelle Entwicklungen, dass KI weiterhin eine Schlüsselrolle in der Zukunft des Sounddesigns spielen wird.

³⁹Vgl. M. Neu, M. Müller, B. Pothen, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst. 2022, Verfügbar: <https://link.springer.com/book/10.1007/978-3-658-38891-1>

⁴⁰Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

⁴¹Vgl. K. Tatar, K. Cotton, und D. Bisig, „Sound Design Strategies for Latent Audio Space Explorations Using Deep Learning Architectures,“ 2023, Verfügbar: <https://arxiv.org/pdf/2305.15571>, (Zugriff am: 22. Jan. 2025)

5. Unterschiede zwischen traditionellem und KI-gestütztem Sounddesign: Was macht den Unterschied?

5.1 Vergleich der Herangehensweisen

Die Herangehensweisen im traditionellen und KI-gestützten Sounddesign unterscheiden sich grundlegend. Während traditionelle Methoden stark von menschlichem Input und kreativen Entscheidungen geprägt sind, nutzen KI-gestützte Ansätze Algorithmen, um Prozesse zu automatisieren und neue kreative Möglichkeiten zu eröffnen.

5.1.1 Kreative Freiheit und menschlicher Input vs. Automatisierung und Algorithmen

Traditionelles Sounddesign erfordert ein hohes Maß an kreativem Input und technischer Expertise. Sounddesigner arbeiten eng mit Regisseuren und anderen kreativen Teams zusammen, um Klänge zu erstellen, die die visuelle Erzählung unterstützen. Dieser Ansatz ist durch manuelle Prozesse geprägt, wie die Erstellung von Foley-Geräuschen und die Synchronisation von Klängen mit Szenen. Diese Arbeit verlangt nicht nur handwerkliches Geschick, sondern auch ein tiefes Verständnis für die narrative Funktion von Klang⁴². Im Gegensatz dazu bietet KI-gestütztes Sounddesign die Möglichkeit, automatisierte Prozesse in den kreativen Workflow zu integrieren. Modelle wie Generative Adversarial Networks (GANs) oder Variational Autoencoders (VAEs) analysieren große Mengen an Daten und erzeugen realistische Klänge, die mit minimalem menschlichen Input angepasst werden können. Die Automatisierung erlaubt es Sounddesignern, sich stärker auf die künstlerischen Aspekte zu konzentrieren, während die technischen Aufgaben von Algorithmen übernommen werden. Laut Liu et al. eröffnet diese Automatisierung völlig neue kreative Möglichkeiten, da Klänge in Echtzeit generiert und an spezifische Anforderungen angepasst werden können⁴³.

5.1.2 Zeitaufwand und Produktionskosten

Traditionelle Sounddesign-Methoden sind oft zeitintensiv und ressourcenaufwendig. Die Erstellung von Geräuschen durch Foley-Künstler oder die Aufnahme von Umgebungsgeräuschen erfordert präzise Planung und Durchführung. Dies führt zu hohen Produktionskosten, insbesondere bei aufwendigen Filmproduktionen, bei denen komplexe Klanglandschaften benötigt werden⁴⁴.

KI-gestützte Ansätze bieten hier deutliche Vorteile. Durch den Einsatz von Algorithmen können Soundelemente schneller erstellt und bearbeitet werden. Die Kosten für die Produktion von Klängen sinken, da weniger manuelle Eingriffe erforderlich sind. Neu et al. zeigen, dass KI nicht nur die Produktionszeit erheblich reduziert, sondern auch die Effizienz im gesamten Workflow steigert, was besonders bei großen Projekten von Vorteil ist⁴⁵.

Die Herangehensweisen im traditionellen und KI-gestützten Sounddesign spiegeln die unterschiedlichen Möglichkeiten und Herausforderungen beider Ansätze wider. Während traditionelle Methoden durch die manuelle Arbeit und die kreative Kontrolle des Menschen geprägt sind, bieten KI-gestützte Ansätze eine Automatisierung, die sowohl Effizienz als auch kreative Freiheit ermöglicht. Dennoch bleiben die kreativen Entscheidungen des Menschen auch im KI-gestützten Sounddesign ein unverzichtbarer Bestandteil des Prozesses.

⁴²Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, H. Segeberg und F. Schätzlein, Eds., Marburg: Schüren, 2005, pp. 24–40, Verfügbar:

<https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

⁴³Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

⁴⁴Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

⁴⁵Vgl. M. Neu, M. Müller, B. Pothén, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst. 2022, Verfügbar: <https://link.springer.com/book/10.1007/978-3-658-38891-1> (Zugriff am: 22. Jan. 2025)

5.1.1 Kreative Freiheit und menschlicher Input vs. Automatisierung und Algorithmen

Traditionelles Sounddesign setzt stark auf die kreative Kontrolle und den direkten Input des Sounddesigners. Die Erstellung von Klängen, die visuelle Szenen ergänzen, erfordert ein tiefes Verständnis für narrative Strukturen und akustische Wirkung. Diese Methoden ermöglichen eine detaillierte Anpassung jedes Klangdetails an die spezifischen Bedürfnisse der Produktion, was jedoch zeitintensiv ist und ein hohes Maß an Fachwissen voraussetzt. Im Gegensatz dazu bietet KI-gestütztes Sounddesign eine neue Dimension der Automatisierung. Durch den Einsatz von generativen Modellen wie GANs (Generative Adversarial Networks) oder VAEs (Variational Autoencoders) können große Datenmengen analysiert und in Echtzeit in Klänge umgewandelt werden. Diese Modelle übernehmen technische Aufgaben wie die Anpassung von Klangparametern, sodass sich Sounddesigner auf die kreativen Aspekte konzentrieren können. Obwohl die Automatisierung den kreativen Prozess nicht vollständig ersetzt, ermöglicht sie eine engere Zusammenarbeit zwischen Mensch und Maschine, um innovative Klanglandschaften zu erschaffen ⁴⁶.

5.1.2 Zeitaufwand und Produktionskosten

Traditionelle Sounddesign-Methoden sind oft zeitaufwendig und ressourcenintensiv. Die Aufnahme von Geräuschen durch Foley-Künstler oder das Bearbeiten von Klangmaterial erfordert nicht nur technisches Können, sondern auch eine präzise Planung und Koordination. Bei aufwendigen Produktionen können diese Prozesse Wochen oder Monate in Anspruch nehmen, was zu hohen Produktionskosten führt ⁴⁷.

Im Vergleich dazu reduzieren KI-gestützte Ansätze den Zeitaufwand erheblich, indem sie Prozesse wie die Synchronisation von Geräuschen automatisieren und Klänge aus Textbeschreibungen generieren. Dies senkt nicht nur die Kosten, sondern macht hochwertige Produktionsergebnisse auch für kleinere Studios zugänglich. Trotz der anfänglichen Investitionen in die Implementierung von KI-Technologien bieten diese langfristig signifikante Effizienzgewinne und ermöglichen es Sounddesignern, sich stärker auf kreative Aspekte zu konzentrieren.

5.2 Qualitätsvergleich: Wie wirken die Ergebnisse auf das Publikum?

Die Qualität von Sounddesign, sei es traditionell oder KI-gestützt, spielt eine entscheidende Rolle für die Wirkung eines Films auf das Publikum. Klänge beeinflussen die emotionale Wahrnehmung von Szenen, verstärken die visuelle Darstellung und tragen wesentlich zur Immersion bei. Der Vergleich zwischen traditionellem und KI-generiertem Sounddesign zeigt sowohl Stärken als auch Herausforderungen beider Ansätze.

Emotionale Wirkung und Authentizität

Traditionelles Sounddesign ist bekannt für seine emotionale Tiefe und Authentizität, da die Klänge individuell auf die narrativen Anforderungen eines Films abgestimmt werden. Foley-Geräusche, die von Hand erstellt und präzise synchronisiert werden, erzeugen ein Gefühl von Realismus, das beim Publikum oft unbewusst mitschwingt. Diese manuelle Herangehensweise erlaubt es Sounddesignern, jeden Klangschritt bewusst zu kontrollieren und an die gewünschte emotionale Wirkung anzupassen.

⁴⁶Vgl. F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, H. Segeberg und F. Schätzlein, Eds., Marburg: Schüren, 2005, pp. 24–40, Verfügbar:

<https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)

⁴⁷Vgl. M. Neu, M. Müller, B. Pothen, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst. Wiesbaden, Germany: Springer Fachmedien, 2022, Verfügbar:

<https://link.springer.com/book/10.1007/978-3-658-38891-1> (Zugriff am: 22. Jan. 2025)

KI-generiertes Sounddesign hingegen bietet beeindruckende Flexibilität und Geschwindigkeit bei der Erstellung von Klängen. Generative Modelle wie AudioLDM oder WaveNet können in Echtzeit Klänge erzeugen, die den Stil und die Atmosphäre bestimmter Szenen widerspiegeln. Allerdings gibt es Diskussionen über die emotionale Tiefe dieser Klänge. Einige Studien deuten darauf hin, dass KI-generierte Sounds in ihrer Wirkung auf das Publikum manchmal als weniger authentisch wahrgenommen werden, insbesondere wenn sie nicht ausreichend an den narrativen Kontext angepasst sind ⁴⁸.

Immersion und Klangvielfalt

Ein zentraler Aspekt des Sounddesigns ist die Fähigkeit, das Publikum in die gezeigte Welt eintauchen zu lassen. Traditionelle Ansätze setzen dabei auf akribisch erstellte Klanglandschaften, die sorgfältig geschichtet und abgemischt werden. Diese Detailgenauigkeit trägt dazu bei, dass das Publikum die gezeigte Welt nicht nur visuell, sondern auch akustisch erlebt.

KI-Modelle bieten hier neue Möglichkeiten, insbesondere durch die Erzeugung von Klängen, die mit traditionellen Methoden nur schwer umzusetzen wären. Dank Algorithmen können einzigartige Klanglandschaften geschaffen werden, die die Immersion verstärken. Gleichzeitig ermöglichen KI-Systeme die Anpassung der Klänge an individuelle Vorlieben oder dynamische Szenarien, wie sie in interaktiven Medien oder Spielen häufig vorkommen

⁴⁹.

Herausforderungen im Qualitätsvergleich

Trotz der beeindruckenden Möglichkeiten gibt es auch Herausforderungen bei der Qualität von KI-generiertem Sounddesign. Eine davon ist die fehlende Nuancierung bei komplexen emotionalen Szenen. Während KI-Systeme technisch präzise arbeiten, fehlt ihnen oft das intuitive Verständnis, das erfahrene Sounddesigner bei der Schaffung emotionaler Klanglandschaften einsetzen. Zusätzlich kann die Wahrnehmung von KI-generierten Klängen vom Publikum kritisch betrachtet werden, wenn diese als „synthetisch“ oder „künstlich“ empfunden werden.

Zukünftige Integration in die Praxis

Die Erfahrungen aus der Produktion des KI-gestützten Boxfilms werden zeigen, wie gut KI-generiertes Sounddesign in der Praxis funktioniert und wie es vom Publikum wahrgenommen wird. Diese Erkenntnisse könnten dazu beitragen, die Akzeptanz und die Wahrnehmung von KI im Sounddesign weiter zu verbessern.

5.3 Technische Analyse: Klangqualität, Vielfalt und Anpassungsfähigkeit

Die technische Analyse von Klangqualität, Vielfalt und Anpassungsfähigkeit zeigt deutliche Unterschiede zwischen traditionellem und KI-gestütztem Sounddesign. Diese Aspekte sind entscheidend, um die Leistungsfähigkeit und den kreativen Spielraum der beiden Ansätze zu bewerten.

⁴⁸Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf> (Zugriff am: 22. Jan. 2025)

⁴⁹Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)

Klangqualität

Traditionelles Sounddesign zeichnet sich durch eine hohe Klangqualität aus, da jeder Ton manuell erstellt, aufgenommen und bearbeitet wird. Foley-Künstler und Sounddesigner verwenden präzise Techniken, um Klänge realistisch und immersiv zu gestalten. Diese Herangehensweise erlaubt eine außergewöhnliche Detailgenauigkeit und Authentizität, die besonders bei Filmen mit hohen Produktionsstandards geschätzt wird ⁵⁰.

KI-gestützte Systeme wie AudioLDM und Variational Autoencoders (VAEs) bieten ebenfalls eine beeindruckende Klangqualität. Sie nutzen große Datensätze, um Klänge zu generieren, die natürlichen Klangquellen sehr nahe kommen. Laut Liu et al. sind die generierten Klänge oft kaum von realen Geräuschen zu unterscheiden. Allerdings kann es bei hochkomplexen oder emotional anspruchsvollen Szenen zu Herausforderungen kommen, da KI-Modelle Nuancen oft nicht vollständig erfassen können ⁵¹.

Klangvielfalt

Im traditionellen Sounddesign wird Klangvielfalt durch die manuelle Aufnahme und Bearbeitung von Geräuschen erreicht. Diese Prozesse sind jedoch zeitaufwendig und erfordern erhebliche Ressourcen, um eine breite Palette von Klängen zu erstellen. Die Vielfalt ist zudem durch die Verfügbarkeit physischer Aufnahmematerialien und Umgebungen begrenzt.

KI-gestützte Systeme übertreffen diese Grenzen, indem sie in der Lage sind, unzählige neue Klänge aus bestehenden Daten zu generieren. Generative Modelle wie GANs können neue Klangmuster erstellen, die mit traditionellen Methoden nicht realisierbar wären. Dies eröffnet Sounddesignern neue kreative Möglichkeiten, da Klänge für spezifische Szenarien angepasst oder völlig neue akustische Welten erschaffen werden können ⁵².

Anpassungsfähigkeit

Die Anpassungsfähigkeit traditioneller Methoden ist stark von der Erfahrung und Expertise der Sounddesigner abhängig. Geräusche müssen oft manuell synchronisiert und bearbeitet werden, um sie perfekt an visuelle Szenen anzupassen. Dieser Prozess ist präzise, jedoch zeitintensiv.

Im Vergleich dazu bieten KI-gestützte Ansätze eine dynamische Anpassungsfähigkeit. Systeme wie AudioLDM können Klänge in Echtzeit generieren und an wechselnde Szenarien anpassen. Diese Flexibilität ist besonders in interaktiven Medien wie Videospielen oder Virtual Reality von Vorteil, wo sich akustische Elemente kontinuierlich an die Handlungen der Benutzer anpassen müssen. Die Automatisierung solcher Anpassungen reduziert den Aufwand erheblich und erhöht die Effizienz ⁵³.

Die technische Analyse zeigt, dass KI-gestütztes Sounddesign in vielen Bereichen wie Klangvielfalt und Anpassungsfähigkeit gegenüber traditionellen Methoden überlegen ist. Während die Klangqualität in beiden Ansätzen hoch ist, bieten KI-Modelle eine Effizienz und Flexibilität, die im traditionellen Sounddesign schwer zu erreichen sind. Dennoch bleibt die menschliche Expertise ein unverzichtbarer Bestandteil, um die Ergebnisse zu verfeinern und sicherzustellen, dass sie den narrativen und emotionalen Anforderungen gerecht werden.

⁵⁰Vgl. L. Duffner, Der Foley Artist – Historische Entwicklung und heutige Arbeitsweise, Hochschule der Medien, 2020, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf?utm.com> (Zugriff am: 22. Jan. 2025)

⁵¹Vgl. H. Liu et al., „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ USA, 2023, Verfügbar: <https://arxiv.org/pdf/2301.12503>,

⁵²Vgl. K. Tatar, K. Cotton, und D. Bisig, „Sound Design Strategies for Latent Audio Space Explorations Using Deep Learning Architectures,“ Stockholm, Sweden, 2023, Verfügbar: <https://arxiv.org/pdf/2305.15571>, (Zugriff am: 22. Jan. 2025)

⁵³Vgl. M. Neu, M. Müller, B. Pothen, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst, 2022, Verfügbar: <https://link.springer.com/book/10.1007/978-3-658-38891-1> (Zugriff am: 22. Jan. 2025)

6. Filmbeispiele: Wie schneiden Tradition und KI im direkten Vergleich ab?

6.1 Analyse der Filmszenen

6.1.1 Klassischer Boxfilm („Rocky“)

Der Film *Rocky* (1976), inszeniert von John G. Avildsen, ist ein Meilenstein des traditionellen Sounddesigns und bietet zahlreiche Einblicke in die kreative Gestaltung von Klanglandschaften. Die akustische Untermalung des Films spielt eine zentrale Rolle, um die Intensität und Authentizität der Szenen zu verstärken. Im Fokus steht die Verbindung von Musik, Geräuschen und Raumklang, die das Publikum in die Welt des Protagonisten eintauchen lässt.

Analyse des Sounddesigns

Das Sounddesign in *Rocky* ist geprägt von realistischen und kraftvollen akustischen Elementen, die das Geschehen im Boxring authentisch darstellen. Besonders in der ersten Kampfszene wird deutlich, wie die Geräuschkulisse die physische Intensität der Kämpfe untermalt. Laute, repetitive Schläge dominieren die akustische Gestaltung und vermitteln die Härte und Dynamik des Boxens. Auffällig ist, dass diese Schläge vor allem in den älteren Teilen oft übertrieben und teilweise unrealistisch wirken, was die dramatische Wirkung verstärkt und die Szenen für das Publikum intensiver erscheinen lässt. Gleichzeitig sorgen die Stimmen der Kämpfer und das Durcheinander der Gespräche im Publikum für eine chaotische und lebendige Atmosphäre. Elemente wie die Glocke und schwere Schritte der Kämpfer setzen klare Akzente und unterstützen die Struktur der Szene ⁵⁴.

Begleitet wird das Geschehen durch eine urbane Klanglandschaft, die durch Geräusche wie Züge, Autos und andere Hintergrundgeräusche angereichert wird. Diese akustischen Details unterstreichen die Authentizität der Umgebung und stärken die Immersion. Obwohl die Klangqualität aufgrund des Alters des Films nicht modernen Standards entspricht, trägt sie zur rohen und realistischen Darstellung der Szenen bei ⁵⁵.

Emotionale Wirkung und Narration

Das Sounddesign von *Rocky* ist darauf ausgelegt, Nähe zum Protagonisten zu erzeugen. Die präzise Kombination aus Klängen und Musik verstärkt die emotionale Verbindung des Publikums zu den Charakteren. Insbesondere die Mischung aus lauten, physischen Geräuschen und subtileren Hintergrundgeräuschen lässt die Zuschauer die Strapazen und Emotionen des Protagonisten hautnah miterleben.

Die ikonische musikalische Untermalung von Bill Conti, darunter das berühmte Stück „Gonna Fly Now“, ist ein zentraler Bestandteil der Klanglandschaft des Films. Sie verstärkt die emotionale Tiefe von Schlüsselszenen und verleiht dem Film eine bleibende akustische Identität. Diese musikalischen Elemente, kombiniert mit den präzise gestalteten Geräuschen, machen *Rocky* zu einem herausragenden Beispiel für traditionelles Sounddesign ⁵⁶.

⁵⁴ Eigene Beobachtungen basierend auf der Analyse des Films *Rocky* (1976).

⁵⁵ *Rocky Sounddesign-Analyse in der ersten Kampfszene*, YouTube (zuletzt abgerufen am 16. Januar 2025)

⁵⁶ Dustin Ogburn, „*Rocky* (1976) Sounddesign“, WordPress (zuletzt abgerufen am 16. Januar 2025)

6.1.2 KI-generierter Boxfilm (Studioproduktion)

Das Sounddesign für den KI-Boxfilm "Knockout", entwickelt im Rahmen der Studioproduktion des Wintersemesters 24/25, basiert maßgeblich auf der Nutzung moderner KI-Tools, insbesondere ElevenLabs. Durch den Einsatz dieser Technologie wurde eine innovative Herangehensweise umgesetzt, bei der Klänge automatisch generiert und gezielt an die jeweiligen Szenen angepasst wurden.

Diese Methodik ermöglichte nicht nur eine signifikante Zeitersparnis im Vergleich zu traditionellen Verfahren, sondern eröffnete auch neue kreative Freiräume. Die generierten Klänge reichen von realistischen Umgebungsgeräuschen, Stimmsynthese und Atmosphären bis hin zu akustischen Effekten, die mit klassischen Techniken nur schwer umsetzbar wären. Die Studioproduktion von "Knockout" dient als praxisnahes Beispiel, um die Potenziale und Grenzen von KI-gestütztem Sounddesign zu analysieren.

Analyse des KI-Sounddesigns in Knockout

Die Studioproduktion Knockout stellt ein bemerkenswertes Beispiel für den Einsatz künstlicher Intelligenz im Sounddesign dar. Durch die Nutzung moderner KI-Tools wie ElevenLabs konnten in kurzer Zeit Klänge generiert werden, die gezielt auf die Szenen abgestimmt wurden. Die Analyse des Sounddesigns zeigt sowohl die Stärken als auch die Herausforderungen dieses innovativen Ansatzes auf.

Audioqualität und Klangfülle

Ein zentraler Aspekt des Sounddesigns von Knockout ist die bemerkenswert hohe Audioqualität. Die KI-generierten Sounds weisen eine klare und dynamische Klangfülle auf, die besonders in den atmosphärischen Elementen zur Geltung kommt.

Hintergrundgeräusche, wie das Rauschen in der Umkleidekabine, das Rascheln des Briefes oder die Geräusche der Raben, wirken immersiv und tragen zur dichten Stimmung des Films bei. Die Qualität der generierten Sounds steht traditionellen Aufnahmen kaum nach und lässt sich nahtlos in die Gesamtkomposition des Films integrieren.

Gestaltung der Schlaggeräusche

Eine der auffälligsten Klangkomponenten in Knockout sind die Schlaggeräusche, die durch KI erstellt wurden. Diese weisen eine übertriebene, fast cineastische Qualität auf, die an die Klanggestaltung in klassischen Boxfilmen wie Rocky erinnert. Während in traditionellen Produktionen oft physische Objekte wie Fleischstücke oder Sandkissen verwendet werden, um die Wirkung von Schlägen zu verstärken, erzeugte die KI in Knockout eine eigene Interpretation dieser Klänge. Die Schläge klingen intensiver und kräftiger als in der Realität, wodurch sie eine dramatische und kraftvolle Wirkung entfalten. Dies verstärkt die emotionale Intensität der Kampfszenen, kann aber zugleich als unnatürlich oder übertrieben empfunden werden. Die folgenden Soundbeispiele verdeutlichen die Unterschiede zwischen traditionell erzeugten Schlaggeräuschen in *Rocky* und den KI-generierten Klängen in *Knockout* (Siehe Soundbeispiel 1 & 2 im Medienverzeichnis).

Traditionelle Aufnahmen von Schlaggeräuschen wurden durch physische Objekte erzeugt, wodurch eine natürliche und organische Klangfülle entstand. Im Gegensatz dazu zeigen die KI-generierten Sounds deutliche Unterschiede in ihrer Klangfülle und Authentizität. Während die von Menschen produzierten Sounds eine physische Präsenz besitzen, wirken die

KI-generierten Varianten oft stilisierter und intensiver. Besonders auffällig ist dies bei den mit ElevenLabs und Stable Audio Open generierten Schlaggeräuschen, die zwar eine hohe Lautstärke und Dynamik aufweisen, jedoch nicht immer die gewünschte Natürlichkeit erreichen. Die digitale Erzeugung kann zu übertriebenen oder synthetisch klingenden Ergebnissen führen, die sich vom gewohnten Höreindruck realer Schläge unterscheiden (Siehe Soundbeispiel 3 & 4 im Medienverzeichnis).

Sprachsynthese und Stimmgestaltung

Die durch KI generierten Stimmen in *Knockout* stellen eine der größten Herausforderungen des Sounddesigns dar. Während die Tonqualität hoch ist und sich die Stimmen klar und verständlich in das Klangbild des Films einfügen, bleibt ein merklicher Unterschied zu echten menschlichen Stimmen bestehen. Die synthetischen Stimmen wirken gelegentlich etwas steif und weniger organisch, was besonders in emotional geladenen Szenen auffällt. Während menschliche Sprecher Nuancen wie Betonung, Rhythmus und emotionale Tiefe intuitiv anpassen können, bleibt die KI in diesen Aspekten noch hinter klassischen Synchronisierungen zurück. Dennoch bietet diese Methode den Vorteil einer schnellen und kostengünstigen Umsetzung von Dialogen, die in traditionellen Produktionen aufwendige Studioaufnahmen und Schauspieler erfordern würden.

Atmosphärische Klänge und räumliche Tiefe

Ein zentraler Aspekt des KI-gestützten Sounddesigns in *Knockout* ist die Erzeugung atmosphärischer Klänge und räumlicher Tiefe, die das visuelle Geschehen intensivieren. Obwohl der Film, anders wie bei *Rocky*, ohne Zuschauer auskommt, trägt die gezielt generierte Geräuschkulisse maßgeblich zur Atmosphäre bei. Die KI-gesteuerte Klanggestaltung umfasst eine Vielzahl subtiler, aber wirkungsvoller akustischer Elemente, die die emotionale Wirkung der Szenen verstärken und dem Publikum ein immersives Klangerlebnis bieten. Besonders in spannungsgeladenen Momenten spielen Hintergrundgeräusche eine wesentliche Rolle. So erzeugen tiefe *Riser*-Sounds bei bestimmten Szenen eine wachsende Anspannung, während *Impacts* punktuell eingesetzt werden, um die Intensität bestimmter Bewegungen oder Treffer zu unterstreichen. Diese akustischen Effekte sorgen für eine dynamische Klanglandschaft, die dem Zuschauer das Gefühl gibt, direkt im Geschehen zu sein. Zu Beginn dient das Krächzen der Raben als unterschwelliger akustischer Hinweis auf das bevorstehende Unheil und verstärkt die düstere Atmosphäre der Friedhofsszene. In der Umkleidekabine kommen gezielt dezente, atmosphärische Geräusche zum Einsatz, um die Einsamkeit und innere Anspannung des Protagonisten zu verdeutlichen. Das leise *Rauschen* der Umgebung verstärkt das Gefühl der Isolation und der Trauer die der Protagonist durch den Verlust seines Vaters verspürt. Besonders eindrucksvoll ist der hochfrequente Piepston und das rhythmische Herzschlagen nach dem Knockout-Moment, der nicht nur die Desorientierung des Kämpfers verdeutlicht, sondern auch eine audiovisuelle Überleitung in die animierte Welt schafft. Neben diesen gezielten Klangakzenten spielen auch alltägliche Geräusche eine wichtige Rolle in der klanglichen Erzählung. Die Kombination dieser gezielt eingesetzten Soundelemente zeigt, wie KI-generierte Klänge nicht nur als Ergänzung dienen, sondern aktiv dazu beitragen, die emotionale Wirkung eines Films zu formen. Durch die Möglichkeit, präzise abgestimmte Atmosphären zu erzeugen und einzelne Elemente flexibel zu kombinieren, eröffnet das KI-gestützte Sounddesign neue kreative Möglichkeiten für die audiovisuelle Gestaltung.

Gesamteindruck und Bewertung

Insgesamt bietet das KI-gestützte Sounddesign von Knockout eine eindrucksvolle Demonstration der aktuellen Möglichkeiten und Grenzen von KI in der Audioproduktion. Während die generierten Klänge in vielen Bereichen eine hohe Qualität aufweisen und sich nahtlos in das Filmgeschehen einfügen, gibt es weiterhin Herausforderungen bei der Natürlichkeit von Stimmen und der Authentizität bestimmter Klangeffekte. Die Schlaggeräusche, obwohl kraftvoll und dramatisch, zeigen, dass KI-Modelle teilweise zu stilisierten und überzeichneten Ergebnissen tendieren.

Dennoch beweist Knockout, dass KI-basierte Tools eine wertvolle Ergänzung im modernen Sounddesign darstellen und besonders für Produktionen mit knappen Budgets oder straffen Zeitplänen erhebliche Vorteile bieten. Die Fähigkeit, schnell hochwertige Klänge zu generieren, ermöglicht eine effiziente Arbeitsweise und eröffnet neue kreative Spielräume für zukünftige Filmproduktionen.

Im Rahmen der Studioproduktion des KI-Boxfilms "Knockout" spielte das Tool ElevenLabs eine zentrale Rolle bei der Gestaltung des Sounddesigns. Die vielseitigen Funktionen von ElevenLabs boten die Möglichkeit, verschiedene Elemente des Films akustisch zu untermalen und zu optimieren, was sowohl kreative als auch technische Vorteile mit sich brachte.

Ein zentraler Aspekt war die Generierung von Soundeffekten. Mithilfe präzise formulierter Prompts konnten spezifische Klänge wie Schlaggeräusche, Glockentöne oder andere für Boxkämpfe typische Sounds erzeugt werden. Diese Klänge wurden in hoher Qualität bereitgestellt, was die nahtlose Integration in die Szenen des Films erleichterte. Ergänzend dazu ermöglichte ElevenLabs die Erstellung von Soundscapes, die für die atmosphärische Untermalung eingesetzt wurden. Diese Hintergrundgeräusche trugen dazu bei, das Setting des Films immersiver zu gestalten und das Publikum emotional stärker in die Szenen einzubinden.

Eine weitere herausragende Funktion war die Text-to-Speech-Technologie, die für die Vertonung der animierten Charaktere verwendet wurde. Mit dieser Technologie konnten gesprochene Dialoge in ausgewählter Stimme generiert werden, die nicht nur eine hohe Audioqualität boten, sondern auch in wenigen Sekunden verfügbar waren. Dies erwies sich als besonders hilfreich, um die animierten Figuren glaubwürdig und lebendig darzustellen.

Vorteile der Nutzung von ElevenLabs

Die Nutzung von ElevenLabs brachte zahlreiche Vorteile mit sich. Ein entscheidender Faktor war die hohe Effizienz: Klänge und Texte konnten innerhalb von etwa zehn Sekunden generiert werden, was den Produktionsprozess erheblich beschleunigte. Außerdem bot das Tool die Möglichkeit, vier Variationen eines Sounds zu erstellen, was die kreative Auswahl und Anpassung erleichterte. Die Ausgabedateien in WAV-Format überzeugten durch ihre hohe Audioqualität und waren sofort einsatzbereit. Auch die geringen Kosten für die Nutzung des Tools machten es zu einer attraktiven Option für die Produktion.

Herausforderungen und Einschränkungen

Trotz der vielen Vorteile gab es auch einige Herausforderungen bei der Nutzung von ElevenLabs. Ein entscheidender Faktor war die Qualität der Prompts, die maßgeblich das Ergebnis beeinflusste. Ungenau formulierte Eingaben führten oft zu unpassenden oder qualitativ minderwertigen Ergebnissen, was einen erhöhten Aufwand für die Optimierung

erforderte. Zudem war eine umfangreiche Auslese der generierten Sounds nötig, um die besten Ergebnisse auszuwählen. Eine weitere Einschränkung stellten die begrenzten Credits dar, die eine sparsame und gezielte Nutzung des Tools erforderlich machten. Die Studioproduktion von "Knockout" verdeutlicht die vielseitigen Einsatzmöglichkeiten von ElevenLabs im Sounddesign. Die Effizienz und Qualität des Tools ermöglichen es, den Produktionsprozess zu optimieren und gleichzeitig neue kreative Wege zu gehen. Trotz kleinerer Einschränkungen bietet ElevenLabs eine zukunftsweisende Lösung, die die Arbeit von Sounddesignern erheblich erleichtern und bereichern kann. Konkrete Beispiele für die Generierung von Klängen und Stimmen werden im weiteren Verlauf der Arbeit detailliert dargestellt, um die Praxisanwendung des Tools zu illustrieren.

6.2.2 KI-generierte Sounds und ihre technische Umsetzung

Die Erstellung von KI-generierten Sounds basiert auf fortschrittlichen Algorithmen, die speziell dafür entwickelt wurden, Audioinhalte effizient und präzise zu erzeugen. Dabei kommen maschinelle Lernmodelle wie neuronale Netzwerke zum Einsatz, die große Mengen an Audio- und Textdaten analysieren und daraus Klangmuster ableiten. Diese Prozesse ermöglichen die Generierung realistischer und innovativer Sounds, die sowohl kreativ als auch funktional in der Audioproduktion genutzt werden können.

Latent Diffusion Modeling

Latent Diffusion Modeling (LDM) ist eine innovative Methode, die den Generierungsprozess in einen latenten Raum verlagert, um die Effizienz und Qualität zu maximieren. Dabei wird der ursprüngliche hochdimensionale Datenraum, wie etwa Pixel in Bildern oder Audiodaten, durch einen komprimierten latenten Raum ersetzt, der mithilfe von Autoencodern erstellt wird. Dieser Ansatz reduziert die Rechenanforderungen erheblich, da die Modellierung in einem niedrigdimensionalen Raum erfolgt, ohne wichtige semantische Informationen zu verlieren.

LDMs arbeiten, indem sie verrauschte Daten schrittweise verfeinern. Der Prozess beginnt mit einem stark verrauschten Signal und endet mit einem klaren und detaillierten Ergebnis. Durch diese iterative Verfeinerung können KI-Systeme hochpräzise und realistische Klänge oder Bilder erzeugen. Ein besonderer Vorteil von LDMs ist ihre Fähigkeit, sich auf semantisch relevante Informationen zu konzentrieren, während irrelevante Details herausgefiltert werden.

Diese Methode ist besonders effektiv in der Audiogenerierung, da sie ermöglicht, komplexe Klanglandschaften effizient zu erstellen. Der wissenschaftliche Artikel von Rombach et al. zeigt, wie LDMs mithilfe von Autoencodern eine Balance zwischen Komplexitätsreduktion und Detailgenauigkeit erreichen. Die Autoren heben hervor, dass die Verwendung latenter Räume die Rechenkosten erheblich senkt, ohne Kompromisse bei der Qualität einzugehen

⁵⁷.

Technologische Herausforderungen

Obwohl die Nutzung von KI-basierten Technologien wie Latent Diffusion Modeling bemerkenswerte Ergebnisse liefert, gibt es auch spezifische Herausforderungen, die berücksichtigt werden müssen. Eine der größten Hürden liegt im hohen Rechenaufwand, der für die Verarbeitung großer Datenmengen und die Durchführung mehrerer Iterationen erforderlich ist. Dies kann die Anwendung solcher Modelle auf umfangreiche Projekte oder weniger leistungsfähige Hardware einschränken.

⁵⁷Vgl. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, „High-Resolution Image Synthesis with Latent Diffusion Models,“ arXiv, 2022, Verfügbar: <https://arxiv.org/abs/2112.10752> (Zugriff am: 22. Jan. 2025)

Ein weiteres Problem stellt die Schwierigkeit dar, emotionale Nuancen in Klängen authentisch darzustellen. Während KI-Modelle bereits in der Lage sind, realistische und präzise Klänge zu erzeugen, bleibt die genaue Wiedergabe subtiler emotionaler Details, wie variierende Intensitäten oder dynamische Übergänge, eine technische Herausforderung. Darüber hinaus hängt die Qualität der generierten Ergebnisse stark von der Präzision der Eingaben ab. Unklare oder zu allgemeine Beschreibungen können zu unbrauchbaren oder unerwarteten Ergebnissen führen, was zusätzliche Nachbearbeitungen oder erneute Generierungen erforderlich macht. Diese Abhängigkeit von präzisen Eingaben macht es erforderlich, dass Nutzer ein gutes Verständnis für die Funktionsweise der Modelle haben, um optimale Ergebnisse zu erzielen.

Zusätzlich können Einschränkungen in der Anpassungsfähigkeit der Modelle auftreten. Obwohl generative KI-Systeme beeindruckende Flexibilität zeigen, stoßen sie an Grenzen, wenn extrem spezifische oder komplexe Anforderungen gestellt werden. In solchen Fällen ist oft manuelle Nachbearbeitung oder die Kombination mit traditionellen Methoden notwendig.

Trotz dieser Herausforderungen bleibt die Entwicklung von KI-basierten Modellen ein entscheidender Schritt, um innovative Ansätze in der Audioproduktion zu fördern und neue kreative Möglichkeiten zu eröffnen. Die kontinuierliche Verbesserung der Modelle wird dazu beitragen, bestehende Hürden zu überwinden und ihre Anwendung noch vielseitiger und zugänglicher zu machen.

Beispiele für erfolgreiche Anwendungen

KI-basierte Tools wie ElevenLabs und Suno haben in der Praxis ihre Effizienz und Vielseitigkeit unter Beweis gestellt. In einem Kurzfilmprojekt wurde beispielsweise die gesamte Hintergrundmusik mit Suno erstellt. Dieser Ansatz ermöglichte es dem Produktionsteam, Musikstücke in kurzer Zeit zu generieren, die perfekt auf die Atmosphäre des Films abgestimmt waren. Durch den Einsatz von KI konnte das Projekt ohne aufwendige Studioaufnahmen oder externe Komponisten realisiert werden, was sowohl Zeit als auch Kosten sparte.

Ein weiteres Beispiel ist die Nutzung von ElevenLabs zur schnellen Erstellung realistischer Soundeffekte. In einer Szene, die einen Boxkampf darstellt, wurde mit einfachen Text-Prompts wie „Punch“ eine Vielzahl an Schlaggeräuschen generiert. Die Möglichkeit, aus mehreren Varianten zu wählen, erleichterte die Auswahl der passendsten Effekte für die gewünschte Wirkung. Dadurch konnte der kreative Prozess erheblich beschleunigt werden, ohne Kompromisse bei der Qualität einzugehen.

Darüber hinaus erwies sich ElevenLabs als wertvoll für die Vertonung von Charakteren. Mit dem Text-to-Speech-Tool konnten realistische Dialoge in verschiedenen Tonalitäten und Sprachen erzeugt werden. Dies war besonders nützlich, um schnell und kosteneffizient Voiceovers für die Charaktere zu erstellen, ohne auf professionelle Sprecher angewiesen zu sein.

Diese Beispiele zeigen, wie KI-gestützte Tools nicht nur den kreativen Prozess erleichtern, sondern auch neue Möglichkeiten in der Audioproduktion eröffnen. Sie ermöglichen es, hochwertige Inhalte effizient zu erstellen und dabei die Flexibilität zu bewahren, die in der modernen Medienproduktion unerlässlich ist.

6.3 Ergebnisvergleich

Der Vergleich zwischen traditionellen und KI-gestützten Methoden im Sounddesign zeigt sowohl Stärken als auch Schwächen beider Ansätze. Die Analyse konzentriert sich auf zentrale Kriterien wie Effizienz, Qualität, Flexibilität und Anwendungsbereiche, um die jeweiligen Ergebnisse zu bewerten. Beobachtungen aus der Produktion des Films *Rocky* und des KI-Films *Knockout* werden herangezogen, um die Vor- und Nachteile beider Methoden im praktischen Kontext zu beleuchten.

6.3.1 Effizienz

Die Effizienz im Sounddesign beschreibt, wie schnell und ressourcenschonend Klänge erstellt, bearbeitet und in Projekte integriert werden können. Der Vergleich zwischen traditionellen Methoden, wie sie im Film *Rocky* angewandt wurden, und KI-gestützten Ansätzen, wie sie im Film *Knockout* zum Einsatz kamen, zeigt deutliche Unterschiede in den Arbeitsprozessen und dem Zeitaufwand.

Traditionelles Sounddesign

Traditionelle Methoden sind bekannt für ihre Präzision und die individuelle Gestaltung von Klängen, jedoch ist dies mit einem erheblichen Zeitaufwand verbunden. Im Fall von *Rocky* mussten Foley-Künstler beispielsweise Geräusche gezielt aufnehmen, um die physische Härte des Boxsports realistisch darzustellen. Ein ikonisches Beispiel sind die Schlaggeräusche, die durch das Schlagen auf Fleischstücke erzeugt wurden. Dieser Prozess erforderte eine sorgfältige Planung, spezielle Materialien und wiederholte Aufnahmen, um die bestmögliche Klangqualität zu gewährleisten.

Die Nachbearbeitung dieser Klänge war ebenfalls äußerst zeitintensiv. Jeder Sound musste präzise auf die jeweilige Szene abgestimmt werden, um die gewünschte emotionale Wirkung zu erzielen. Insbesondere bei groß angelegten Produktionen oder Szenen mit komplexen Klanglandschaften, wie in Trainingsmontagen oder Kämpfen, konnte dieser Prozess mehrere Tage oder Wochen in Anspruch nehmen.

Zusätzlich erfordert traditionelles Sounddesign eine enge Zusammenarbeit zwischen verschiedenen Abteilungen, etwa Regisseuren, Foley-Künstlern und Sounddesignern. Diese kooperativen Prozesse, während sie die kreative Qualität steigern, verlängern gleichzeitig die Produktionszeit.

KI-gestütztes Sounddesign

Im Gegensatz dazu bietet KI-gestütztes Sounddesign eine erhebliche Zeitersparnis, da viele der manuellen Schritte automatisiert werden. Im Film *Knockout* wurde dies deutlich: Sounds wie Schlaggeräusche und atmosphärische Klanglandschaften konnten innerhalb weniger Sekunden mit Tools wie ElevenLabs und Stable Audio Open generiert werden. Die Möglichkeit, durch präzise Text-Prompts schnell passende Klänge zu erzeugen, eliminierte viele der zeitaufwendigen manuellen Arbeiten.

Ein weiterer Vorteil des KI-gestützten Ansatzes liegt in der Geschwindigkeit, mit der verschiedene Varianten eines Sounds getestet werden können. Während im traditionellen Sounddesign jede Aufnahme einzeln erstellt und bearbeitet werden muss, können KI-Tools in kurzer Zeit mehrere Versionen eines Geräusches generieren. Dies ermöglichte es dem Produktionsteam von *Knockout*, schnell und effizient die besten Optionen auszuwählen und anzupassen.

Trotz dieser Effizienz gibt es auch Einschränkungen. Die Qualität der Ergebnisse hängt stark von der Präzision der Eingaben und den Parametern ab. Unklare Prompts oder ungenaue Spezifikationen können zu unbrauchbaren Klängen führen, die zusätzliche Iterationen erfordern. In solchen Fällen kann die Zeitersparnis durch den Mehraufwand bei der Nachbearbeitung oder der erneuten Generierung von Sounds relativiert werden.

Während traditionelle Methoden mehrere Tage oder Wochen in Anspruch nehmen können, um eine einzelne Szene mit authentischen Klängen zu versehen, reduzieren KI-Tools diese Zeit auf wenige Minuten bis Stunden. Besonders bei Projekten mit engen Zeitplänen oder begrenztem Budget erweisen sich KI-basierte Ansätze als unschlagbar. Dennoch bleibt die Abhängigkeit von präzisen Eingaben und der begrenzte Spielraum bei der kreativen Kontrolle eine Herausforderung.

Traditionelles Sounddesign bietet durch seinen sorgfältigen Ansatz eine unübertroffene kreative Kontrolle und Detailtiefe, erfordert jedoch erheblichen zeitlichen und personellen Aufwand. KI-gestütztes Sounddesign hingegen punktet durch Schnelligkeit und Flexibilität, was es besonders für zeitkritische oder Projekte mit beschränktem Budget attraktiv macht. Beide Ansätze haben ihre spezifischen Vorzüge und können je nach Projektanforderung effektiv eingesetzt werden. In der Praxis zeigt sich, dass eine Kombination beider Methoden – die Effizienz der KI gepaart mit der Präzision traditioneller Nachbearbeitung – das Potenzial hat, optimale Ergebnisse zu liefern.

6.3.2 Qualität

Die Qualität eines Sounds ist ein zentraler Faktor für seine Wirkung auf das Publikum und seine Integration in die visuelle Ebene. Ein hochwertiges Sounddesign trägt maßgeblich dazu bei, die Atmosphäre und emotionale Wirkung eines Films oder Projekts zu verstärken. Traditionelle und KI-gestützte Methoden weisen dabei unterschiedliche Stärken und Schwächen auf.

Traditionelles Sounddesign

Die manuelle Erstellung von Sounds bietet eine außergewöhnliche Kontrolle über die Details und Nuancen. In *Rocky* wurde diese Qualität durch individuell gestaltete Klänge erreicht, die perfekt auf die Szenen abgestimmt waren. Ein Beispiel dafür sind die realistischen Schlaggeräusche, die durch das Schlagen auf Fleischstücke erzeugt wurden, um die physische Intensität der Boxkämpfe glaubwürdig darzustellen. Diese präzise Anpassung erlaubte es, gezielt emotionale Reaktionen beim Publikum hervorzurufen und die Wirkung der Szenen zu intensivieren.

Allerdings hängt die Qualität traditioneller Methoden stark von der Erfahrung und dem Können des Sounddesigners sowie von der verwendeten Technik ab. In älteren Produktionen zeigt sich häufig, dass bestimmte Klänge aufgrund technischer Begrenzungen verzerrt oder verrauscht wirken. Dies betrifft vor allem Filme, in denen die Aufnahmequalität nicht mit den heutigen Standards vergleichbar war. Trotz dieser Herausforderungen bieten traditionelle Ansätze eine unvergleichliche Tiefe und Authentizität, die in vielen Fällen schwer durch automatisierte Prozesse zu erreichen ist.

KI-gestütztes Sounddesign

KI-Systeme wie ElevenLabs und Stable Audio Open können ebenfalls beeindruckende Qualität liefern, insbesondere bei der Erstellung von generischen Sounds oder atmosphärischen Hintergrundgeräuschen. Im Film *Knockout* wurden Schlaggeräusche und Soundscapes durch KI-Modelle generiert, die eine starke emotionale Wirkung entfalteten und effektiv in die Szenen integriert werden konnten. Diese automatisierten Ansätze bieten den Vorteil, dass sie innerhalb kurzer Zeit hochwertige Ergebnisse erzielen können. Allerdings gibt es auch klare Einschränkungen. Die Qualität der KI-generierten Klänge hängt stark von der Präzision der Eingaben und den Einstellungen der Parameter ab. Unklare oder zu allgemeine Prompts können zu unerwarteten oder unbrauchbaren Ergebnissen führen. Zudem ist es für KI-Modelle schwierig, spezifische emotionale Nuancen oder subtile Dynamiken zu erzeugen, die für besonders eindringliche Szenen erforderlich sind. Dies kann die Tiefe und Authentizität der Klänge beeinträchtigen und zusätzliche Nachbearbeitung erforderlich machen, um die gewünschten Ergebnisse zu erzielen.

Vergleich der Qualität

Während traditionelles Sounddesign durch seine Detailgenauigkeit und emotionale Tiefe überzeugt, bietet KI-gestütztes Sounddesign eine effiziente Alternative, die in vielen Fällen zufriedenstellende Ergebnisse liefert. Allerdings erreicht die KI bei komplexen oder emotional anspruchsvollen Anforderungen oft nicht die gleiche Qualität wie ein erfahrener Sounddesigner.

Die Wahl zwischen traditionellem und KI-gestütztem Sounddesign hängt maßgeblich von den Qualitätsanforderungen des Projekts ab. Für Szenen, die eine hohe Präzision und emotionale Tiefe erfordern, bleibt das traditionelle Sounddesign die bessere Wahl. KI-Tools hingegen eignen sich ideal für schnelle und kosteneffiziente Lösungen, insbesondere bei generischen oder atmosphärischen Klängen. Beide Methoden haben ihre Berechtigung und ergänzen sich in hybriden Ansätzen, die die Stärken beider Welten kombinieren.

6.3.3 Flexibilität

Die Flexibilität im Sounddesign bezieht sich auf die Fähigkeit, Klänge an unterschiedliche Anforderungen und kreative Visionen anzupassen. Sowohl traditionelle als auch KI-gestützte Methoden bieten Ansätze zur individuellen Gestaltung, unterscheiden sich jedoch in ihrem Vorgehen und den Möglichkeiten der Anpassung.

Traditionelles Sounddesign

Traditionelle Methoden zeichnen sich durch eine außergewöhnliche Flexibilität aus, da jeder Sound individuell und von Grund auf gestaltet werden kann. In Produktionen wie *Rocky* ermöglichten manuelle Techniken eine präzise Anpassung der Klänge an die Szenen. Beispielsweise konnten spezifische Schlaggeräusche so gestaltet werden, dass sie die physische Intensität der Boxkämpfe realistisch und emotional verstärkten. Diese Herangehensweise erlaubt es, kreative Visionen umzusetzen und auch komplexe oder einzigartige Szenarien mit individuell gestalteten Klängen zu unterlegen.

Allerdings hat diese Flexibilität ihren Preis: Sie erfordert nicht nur eine enge Zusammenarbeit zwischen Regisseuren, Foley-Künstlern und Sounddesignern, sondern auch einen erheblichen Zeit- und Ressourcenaufwand. Besonders bei Projekten mit

komplexen Klanglandschaften oder einer großen Anzahl an individuellen Sounds können die manuelle Erstellung und Anpassung die Produktionszeit und -kosten stark erhöhen

KI-gestütztes Sounddesign

KI-Systeme bieten ebenfalls eine hohe Flexibilität, vor allem durch ihre Geschwindigkeit und die Möglichkeit, mehrere Variationen eines Sounds in kurzer Zeit zu erstellen. Im Film *Knockout* konnten durch präzise Text-Prompts spezifische Klänge für verschiedene Szenen generiert werden, was den kreativen Prozess erheblich beschleunigte. Tools wie ElevenLabs und Stable Audio Open erlaubten es, Klänge dynamisch zu erzeugen und bei Bedarf direkt anzupassen.

Dennoch stößt die Flexibilität von KI-Systemen an Grenzen, insbesondere bei extrem spezifischen oder komplexen Anforderungen. Manche Klänge sind nur schwer oder gar nicht synthetisch generierbar, da die Modelle auf vorab trainierten Datensätzen basieren und daher nur eine begrenzte Bandbreite an Klängen erzeugen können. Zudem können die generierten Klänge generisch wirken und erfordern häufig eine Nachbearbeitung, um die gewünschte Wirkung zu erzielen.

Vergleich der Flexibilität

Traditionelle Methoden bieten eine unvergleichliche Anpassungsfähigkeit, da sie nicht durch algorithmische oder datenbasierte Einschränkungen begrenzt sind. Sie erfordern jedoch mehr Zeit und Ressourcen, um diese Flexibilität voll auszuschöpfen. KI-gestützte Ansätze hingegen punkten durch ihre schnelle und kosteneffiziente Anpassung an wechselnde Anforderungen, sind jedoch weniger geeignet für Projekte mit extrem spezifischen Klanganforderungen oder einer hohen Komplexität.

Während das traditionelle Sounddesign eine tiefgreifende und kreative Gestaltung ermöglicht, bietet das KI-gestützte Sounddesign eine schnelle und effiziente Anpassung an vielfältige Anforderungen. Beide Ansätze haben ihre Vorzüge und ergänzen sich hervorragend, insbesondere in hybriden Produktionsansätzen, die die Geschwindigkeit und Flexibilität der KI mit der Detailtiefe und Präzision traditioneller Methoden kombinieren.

6.3.4 Anwendungsbereiche

Der Einsatzbereich einer Methode hängt stark von den Anforderungen des Projekts, den verfügbaren Ressourcen und dem gewünschten Ergebnis ab. Sowohl traditionelle als auch KI-gestützte Ansätze bieten spezifische Vorteile, die je nach Projekt unterschiedlich relevant sind.

Traditionelles Sounddesign

Traditionelle Methoden sind besonders für Projekte mit hohen kreativen und künstlerischen Ansprüchen geeignet. Filme wie *Rocky* demonstrieren, wie detailreich gestaltete Klänge die narrative Tiefe eines Films verstärken können. Schlaggeräusche, die durch Foley-Techniken erzeugt wurden, oder die dynamischen Klanglandschaften in Trainingsmontagen sind Beispiele dafür, wie präzises Sounddesign emotionale Reaktionen des Publikums gezielt beeinflussen kann.

Auch in Videospiele mit immersiven Klangwelten oder in Blockbuster-Produktionen, die eine außergewöhnliche Qualität und realistische Klanglandschaften erfordern, ist das

traditionelle Sounddesign unverzichtbar. Die Fähigkeit, Klänge individuell auf komplexe visuelle Szenen abzustimmen, macht diesen Ansatz besonders für große, aufwendig produzierte Projekte ideal, bei denen kreative Kontrolle und hohe Qualität oberste Priorität haben.

KI-gestütztes Sounddesign

KI-Systeme wie ElevenLabs und Stable Audio Open sind ideal für zeitkritische und budgetorientierte Projekte. In Produktionen wie dem Film *Knockout* konnte die Effizienz der KI genutzt werden, um Soundeffekte und Dialoge innerhalb kürzester Zeit zu generieren. Dies erleichterte die Produktion erheblich, da aufwendige Studioaufnahmen und Foley-Arbeiten vermieden werden konnten.

Besonders in Werbespots oder Kurzfilmen, bei denen enge Zeitpläne und begrenzte Ressourcen häufig eine Rolle spielen, bietet KI-gestütztes Sounddesign eine schnelle und kosteneffiziente Lösung. Auch in Videospielen mit dynamischen Klanglandschaften kann KI effektiv eingesetzt werden, um Soundscapes und Effekte zu generieren, die flexibel auf die Spielumgebung reagieren.

Vergleich der Anwendungsbereiche

Das traditionelle Sounddesign ist optimal für Projekte, die hohe Qualität und kreative Kontrolle erfordern, wie Blockbuster-Filme oder große Videospieldproduktionen. Es eignet sich besonders, wenn spezifische Klanglandschaften oder emotionale Nuancen erzielt werden sollen, die eine manuelle Gestaltung erfordern.

KI-gestütztes Sounddesign hingegen ist die bevorzugte Wahl für Projekte, die schnelle und effiziente Ergebnisse benötigen, wie etwa in der Werbung, kleineren Videospielen oder Kurzfilmen. Während die generierten Klänge in vielen Fällen überzeugend wirken, kann es bei komplexen oder emotional anspruchsvollen Szenarien zu Einschränkungen kommen.

Der Einsatzbereich von traditionellem und KI-gestütztem Sounddesign hängt stark von den spezifischen Anforderungen eines Projekts ab. Während traditionelle Methoden durch Qualität und emotionale Tiefe überzeugen, bieten KI-Tools eine unschlagbare Effizienz und Flexibilität. Beide Ansätze ergänzen sich hervorragend, insbesondere in hybriden Ansätzen, die die Stärken beider Welten nutzen, um qualitativ hochwertige und gleichzeitig effiziente Ergebnisse zu erzielen.

6.4 Fazit zum Vergleich von traditionellem und KI-gestütztem Sounddesign

Der Vergleich zwischen traditionellem und KI-gestütztem Sounddesign verdeutlicht, dass beide Ansätze spezifische Stärken und Schwächen aufweisen, die sie in unterschiedlichen Kontexten wertvoll machen. Während das traditionelle Sounddesign durch seine hohe Qualität, kreative Kontrolle und emotionale Tiefe besticht, bietet KI-gestütztes Sounddesign vor allem Effizienz und Flexibilität, die in zeitkritischen oder budgetorientierten Projekten entscheidend sein können.

Traditionelle Methoden erlauben eine präzise Gestaltung von Klängen, die individuell an die Szenen und die narrative Struktur eines Projekts angepasst werden können. Dies ermöglicht eine einzigartige Klanglandschaft, die tief in die visuelle und emotionale Ebene eines Films oder Spiels eingebettet ist. Produktionen wie *Rocky* zeigen, wie detailreiche und realistische Klänge, die durch manuelle Techniken wie Foley erzeugt wurden, das Publikum nachhaltig beeindruckt und in die Handlung eintauchen lassen können.

Ein weiterer Vorteil liegt in der kreativen Kontrolle, die durch die manuelle Gestaltung und die enge Zusammenarbeit zwischen Sounddesignern und Regisseuren gegeben ist. Dadurch können spezifische Visionen umgesetzt und auch komplexe Szenarien mit großer Präzision untermalt werden. Dieser Ansatz eignet sich besonders für Blockbuster-Filme, aufwendige Videospielproduktionen oder künstlerisch anspruchsvolle Projekte, bei denen Qualität und emotionale Tiefe im Vordergrund stehen.

Im Gegensatz dazu zeichnet sich das KI-gestützte Sounddesign durch Geschwindigkeit und Effizienz aus. Tools wie ElevenLabs und Stable Audio Open ermöglichen es, Sounds innerhalb von Sekunden zu generieren, ohne aufwendige Aufnahmeprozesse oder teures Equipment. Diese Effizienz ist besonders in Projekten wie Knockout sichtbar, wo KI verwendet wurde, um Soundeffekte und Dialoge schnell und kosteneffizient zu erstellen. Darüber hinaus bieten KI-Systeme eine bemerkenswerte Flexibilität, da Nutzer in kurzer Zeit mehrere Varianten eines Sounds testen und anpassen können. Dies eröffnet kreative Möglichkeiten, insbesondere in der Vorproduktion, wo schnelle Iterationen und Anpassungen erforderlich sind. Für kleinere Projekte wie Werbespots, Kurzfilme oder einfache Spiele ist KI-gestütztes Sounddesign eine ideale Lösung, da es eine hohe Qualität bei gleichzeitig geringerem Zeit- und Ressourcenaufwand bietet.

Schwächen beider Ansätze

Trotz ihrer Stärken sind beide Methoden nicht frei von Schwächen. Das traditionelle Sounddesign ist oft zeit- und kostenintensiv, da es auf manuelle Prozesse und eine enge Zusammenarbeit zwischen mehreren Abteilungen angewiesen ist. Besonders bei groß angelegten Projekten kann dies die Produktionszeit erheblich verlängern.

KI-gestütztes Sounddesign hingegen hat seine Grenzen in der kreativen Kontrolle und der Tiefe der Klänge. Die Qualität der Ergebnisse hängt stark von der Präzision der Eingaben und den zugrunde liegenden Algorithmen ab. Komplexe oder spezifische emotionale Nuancen sind oft schwer zu erzeugen, und die Ergebnisse können gelegentlich generisch wirken.

Die Zukunft des Sounddesigns könnte in hybriden Ansätzen liegen, die die Stärken beider Welten kombinieren. KI-Tools könnten beispielsweise zur schnellen Generierung von Grundsounds oder atmosphärischen Klanglandschaften genutzt werden, während traditionelle Methoden die finale Feinarbeit übernehmen. Dadurch könnten Effizienz und Qualität gleichermaßen gesteigert werden.

Die Wahl zwischen traditionellem und KI-gestütztem Sounddesign sollte immer auf die spezifischen Anforderungen und Ziele eines Projekts abgestimmt sein. Während traditionelle Methoden für Projekte mit hohen kreativen Ansprüchen unverzichtbar bleiben, bieten KI-Tools eine wertvolle Alternative für zeitkritische oder budgetorientierte Produktionen. Letztlich ergänzen sich beide Ansätze und bieten gemeinsam die Möglichkeit, innovative und hochwertige Klangwelten zu schaffen, die die Audioproduktion auf ein neues Niveau heben können.

7. Potenziale und Grenzen: Was kann KI im Sounddesign leisten?

7.1 Chancen für die Filmindustrie

Künstliche Intelligenz (KI) bietet der Filmindustrie neue Möglichkeiten, die sowohl bestehende Produktionsprozesse optimieren als auch kreative Freiheiten erweitern. Besonders im Sounddesign zeigt sich das Potenzial dieser Technologie, indem sie Zeit und Kosten spart und gleichzeitig innovative Klanglandschaften ermöglicht.

Während traditionelle Methoden oft ressourcenintensiv sind, eröffnen KI-gestützte Tools schnelle und flexible Wege, um Klänge zu generieren und dynamisch an Szenen anzupassen. Diese Entwicklungen machen KI zu einem wichtigen Bestandteil moderner Filmproduktionen und schaffen eine Grundlage für Effizienzsteigerungen sowie kreative Ansätze, die in den folgenden Abschnitten detaillierter betrachtet werden.

7.1.1 Effizienzsteigerung und Kostenreduktion

Eine der größten Stärken von Künstlicher Intelligenz (KI) im Sounddesign liegt in der deutlichen Effizienzsteigerung und Kostensenkung. Durch den Einsatz KI-gestützter Tools können viele zeitaufwendige Prozesse wie Foley-Arbeiten, Studioaufnahmen oder aufwendige Nachbearbeitungen automatisiert werden. Klänge lassen sich mit Hilfe präziser Text-Prompts oder generativer Modelle in kürzester Zeit erzeugen, ohne dass teure Produktionsmittel oder umfangreiches Personal erforderlich sind.

Laut Neu et al. kann KI nicht nur repetitive Aufgaben übernehmen, sondern auch den kreativen Prozess unterstützen, indem sie schnell Variationen eines Sounds erstellt, die sofort einsatzbereit sind. Dies spart nicht nur Zeit, sondern reduziert auch den Bedarf an Ressourcen, was besonders für kleinere Filmproduktionen oder Projekte mit begrenztem Budget von Vorteil ist ⁵⁸.

Die Automatisierung durch KI trägt außerdem dazu bei, die Nachbearbeitungszeit erheblich zu verkürzen. So können Sounddesigner in kurzer Zeit mehrere Versionen eines Sounds testen und direkt an die Anforderungen einer Szene anpassen. Diese Effizienz macht KI zu einem unverzichtbaren Werkzeug für moderne Filmproduktionen, die schnelle Ergebnisse und Kosteneffizienz erfordern.

7.1.2 Neue kreative Möglichkeiten durch KI

Künstliche Intelligenz (KI) eröffnet der Filmindustrie neue kreative Horizonte, insbesondere im Bereich des Sounddesigns und der Sprachsynthese. Fortschritte bei Technologien wie Text-to-Speech (TTS) ermöglichen es, realistische Stimmen in mehreren Sprachen und Tonalitäten zu generieren. Laut einem Whitepaper der Hochschule der Medien Stuttgart hat sich die Qualität von TTS-Tools so weit verbessert, dass diese inzwischen nicht nur für die Vertonung von Animationsfilmen und Videospiele, sondern auch für die Nachvertonung von Live-Action-Filmen genutzt werden können ⁵⁹. Diese Technologie ermöglicht es, Dialoge schnell und flexibel zu erstellen, ohne dass auf teure Studioaufnahmen oder Sprecher zurückgegriffen werden muss.

⁵⁸Vgl. M. Neu, M. Müller, B. Pothen, and M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst. Wiesbaden, Germany: Springer Fachmedien, 2022, Verfügbar:

<https://link.springer.com/book/10.1007/978-3-658-38891-1> (Zugriff am: 22. Jan. 2025)

⁵⁹Vgl. M. Schill, M. Hepp, und B. Böröcz, „Künstliche Intelligenz in der Audiowelt,“ Whitepaper, Hochschule der Medien, Wintersemester 2018/19, Verfügbar: https://ai.hdm-stuttgart.de/downloads/student-white-paper/Winter-1819/KI_in_der_Audiowelt.pdf, (Zugriff am: 22. Jan. 2025)

Darüber hinaus erlaubt die KI-gestützte Generierung von Soundeffekten die Erschaffung von Klängen, die in der realen Welt schwer oder gar nicht aufzunehmen wären. Futuristische oder surreale Szenarien, wie etwa Geräusche eines außerirdischen Raumschiffs oder einer postapokalyptischen Stadt, können durch KI-Modelle problemlos realisiert werden. Dies erweitert die kreativen Möglichkeiten und ermöglicht es Sounddesignern, akustische Welten zu erschaffen, die bisher unvorstellbar waren.

Eine weitere Anwendungsmöglichkeit ist die dynamische Anpassung von Klängen an visuelle Szenen. Mithilfe von Text-Prompts oder visuellen Analysen kann die KI spezifische Sounds erzeugen, die exakt auf die Stimmung oder Atmosphäre einer Szene abgestimmt sind. Dies erleichtert es, in kürzerer Zeit mehrere Varianten zu testen und die besten Ergebnisse auszuwählen.

KI bietet auch neue Ansätze für die Personalisierung von Klängen. Designer können Parameter wie emotionale Intensität, Tempo oder Klangfarbe individuell einstellen, um den Sound perfekt an Charaktere oder Situationen anzupassen. Besonders in Videospielen wird diese Funktion zunehmend genutzt, um interaktive und dynamische Klanglandschaften zu erzeugen, die sich in Echtzeit an die Aktionen des Spielers anpassen. Zusätzlich macht die KI-gestützte Generierung experimenteller Klänge neue künstlerische Ansätze möglich. Modelle können Klänge erschaffen, die auf keinen realen Vorlagen basieren, und damit abstrakte oder avantgardistische Szenarien unterstützen. Dies eröffnet insbesondere im Bereich der experimentellen Kunst oder Indie-Filmproduktionen spannende neue Möglichkeiten.

Insgesamt trägt KI dazu bei, die kreativen Grenzen im Sounddesign zu erweitern. Sie ermöglicht es, schnell und flexibel auf die Anforderungen eines Projekts zu reagieren, und bietet gleichzeitig die Möglichkeit, innovative und einzigartige Klangerlebnisse zu schaffen, die die narrative Tiefe und emotionale Wirkung moderner Filme und Spiele verstärken.

7.2 Grenzen der KI

Künstliche Intelligenz bietet im Bereich des Sounddesigns beeindruckende Möglichkeiten, doch ihre Nutzung ist mit signifikanten Einschränkungen verbunden. Eine detaillierte Analyse zeigt, dass diese Grenzen vor allem in den Bereichen Qualität, Kreativität, Editierbarkeit und Textsteuerung liegen, wie aus einer Umfrage hervorgeht.

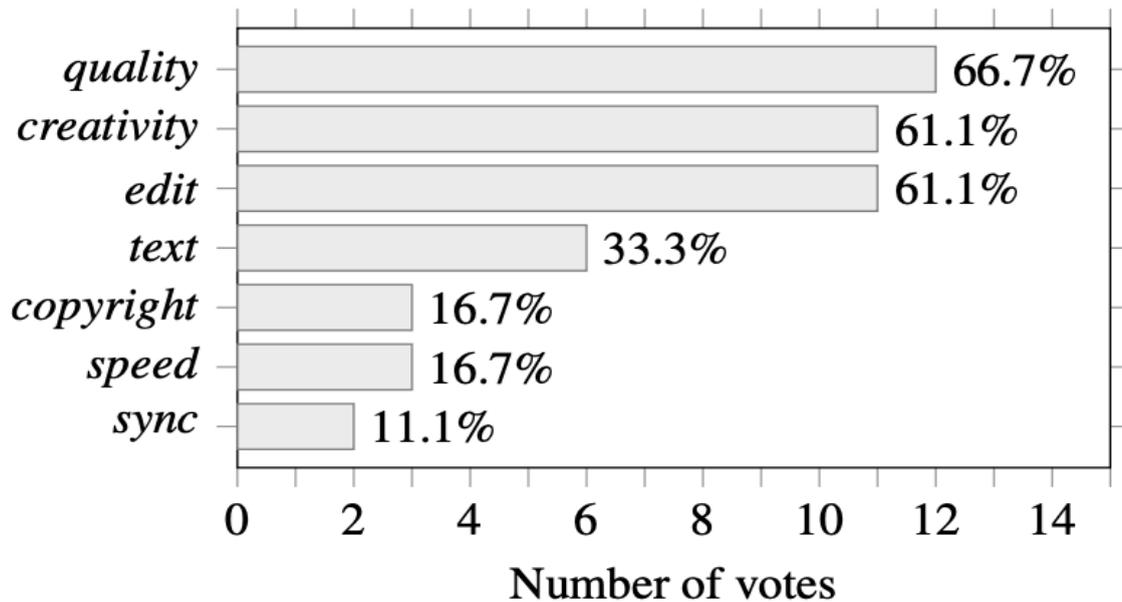


Abbildung 10: Ergebnisse der Umfrage zu den Grenzen von KI-gestütztem Sounddesign

Ein zentrales Problem stellt die Audioqualität dar. Trotz technologischer Fortschritte arbeiten viele Systeme weiterhin mit begrenzten Samplingraten von 16kHz bis 24kHz, während die Industrie Full-Band-Signale von 48kHz oder höher benötigt. Diese Begrenzung führt zu einer geringeren Klangtreue, wahrnehmbarer Rauheit und insgesamt niedrigerer Audioqualität, was die praktische Nutzung einschränkt. Besonders in professionellen Produktionen, wo höchste Ansprüche an die Klangqualität gestellt werden, stoßen KI-Modelle hier an ihre Grenzen. Laut einer Umfrage zu textbasierten Audiogenerierungsmodellen sehen Nutzer die niedrige Samplingrate und die damit verbundene geringe Klangqualität als die größte Herausforderung für eine praktische Nutzung ⁶⁰.

Ein weiteres Hindernis ist die Kreativität der KI. Obwohl KI-Modelle in der Lage sind, neue Klänge zu generieren, fehlt es ihnen oft an der Fähigkeit, spezifische künstlerische Visionen vollständig zu erfüllen. Zum Beispiel wird die Erzeugung eines ikonischen Sounds, wie dem Lichtschwert in Star Wars, häufig als Herausforderung beschrieben. Die künstlerische Intuition und Erfahrung eines menschlichen Sounddesigners kann durch KI nur schwer ersetzt werden, was in komplexen und einzigartigen Projekten deutlich wird.

Ein zusätzlicher Aspekt ist die Editierbarkeit der generierten Sounds. Nutzer berichten häufig von Schwierigkeiten, die Klänge nachträglich zu bearbeiten oder präzise an spezifische Anforderungen anzupassen. KI-Systeme bieten zwar eine Vielzahl von generierten Varianten, jedoch fehlt es oft an Kontrollmöglichkeiten, um die Ergebnisse nachträglich fein abzustimmen.

Auch die Textsteuerung, die über Text-Prompts erfolgt, ist ein kritischer Punkt. Die Ergebnisse sind stark von der Eingabegenauigkeit und der Formulierung des Prompts abhängig. Unklare oder zu allgemeine Beschreibungen führen häufig zu ungenauen oder unbrauchbaren Klängen. Dies begrenzt die Flexibilität und erfordert eine zusätzliche Lernkurve für Nutzer, die diese Systeme einsetzen möchten.

Zusammenfassend zeigt sich, dass KI trotz ihrer beeindruckenden Fortschritte im Sounddesign weiterhin signifikante Einschränkungen aufweist. Die Qualität der Ergebnisse, die begrenzte kreative Freiheit, die eingeschränkte Editierbarkeit und die Abhängigkeit von

⁶⁰Vgl. S. Oh, M. Kang, H. Moon, K. Choi, and B. S. Chon, "A Demand-Driven Perspective on Generative Audio AI," arXiv preprint arXiv:2307.04292, 2023, Verfügbar: <https://arxiv.org/pdf/2307.04292> (Zugriff am: 22. Jan. 2025)

präzisen Eingaben stellen Herausforderungen dar, die gelöst werden müssen, um das volle Potenzial der KI auszuschöpfen. Diese Grenzen verdeutlichen die Notwendigkeit einer Weiterentwicklung der Systeme, insbesondere für professionelle und künstlerisch anspruchsvolle Anwendungen.

7.2.1 Verlust der menschlichen Kreativität?

Eine zentrale Herausforderung beim Einsatz von Künstlicher Intelligenz (KI) im Sounddesign ist die potenzielle Gefährdung der menschlichen Kreativität. Während KI-Systeme beeindruckende Ergebnisse erzielen können, stellt sich die Frage, ob und in welchem Maße sie den kreativen Schaffensprozess ersetzen oder sogar einschränken könnten.

KI-Systeme sind darauf ausgelegt, durch die Analyse großer Datenmengen und die Nutzung generativer Modelle Sounds zu erstellen, die oft qualitativ hochwertig und effizient sind.

Doch die Fähigkeit, neue, originelle und emotional tiefgehende Klänge zu erschaffen, basiert häufig auf menschlicher Intuition, Erfahrung und künstlerischem Ausdruck – Aspekte, die KI nur simulieren, aber nicht eigenständig entwickeln kann. Dies wird besonders in Szenarien deutlich, in denen ein Sound eine spezifische emotionale Botschaft oder künstlerische Vision verkörpern soll.

Wie der Deutsche Ethikrat betont, darf der Einsatz von KI den Menschen nicht ersetzen und sollte vielmehr dazu dienen, die menschliche Entfaltung zu erweitern. Laut Alena Buyx, Vorsitzende des Deutschen Ethikrates, sollte der Einsatz von KI die menschliche Entfaltung erweitern und nicht vermindern ⁶¹.

Diese ethische Perspektive verdeutlicht, dass KI ein Werkzeug bleiben sollte, das die Kreativität unterstützt, anstatt sie zu verdrängen.

Ein weiterer Aspekt ist die Gefahr der Standardisierung. KI-Modelle basieren auf Trainingsdaten und neigen dazu, Ergebnisse zu erzeugen, die auf bestehenden Mustern und Trends beruhen. Dies kann dazu führen, dass Innovation und Originalität in der kreativen Arbeit leiden, da die erzeugten Klänge möglicherweise generisch oder vorhersehbar wirken.

Besonders in der Film- und Musikindustrie, in der einzigartige und individuelle Klangerlebnisse gefragt sind, kann dies eine erhebliche Einschränkung darstellen.

KI eine wertvolle Unterstützung im kreativen Prozess, birgt jedoch das Risiko, die menschliche Kreativität zu verdrängen, wenn sie unreflektiert eingesetzt wird. Der Schlüssel liegt darin, KI als Ergänzung zu verstehen, die menschliche Fähigkeiten erweitert und Raum für innovative, von Menschen geschaffene Kunst lässt.

dass die Integration von KI in künstlerische Prozesse das Risiko birgt, menschliche Kreativität zu standardisieren und einzigartige kulturelle Ausdrucksformen zu untergraben. Der Einsatz von KI im Sounddesign bedeutet nicht nur technische Innovation, sondern auch eine sorgfältige Auseinandersetzung mit ethischen und rechtlichen Fragen. Klare Regelungen, Transparenz und ein bewusster Umgang mit der Technologie sind notwendig, um die Potenziale von KI verantwortungsvoll zu nutzen und ihre Risiken zu minimieren.

7.2.2 Ethische Fragen und Herausforderungen

Der Einsatz von Künstlicher Intelligenz (KI) im Sounddesign birgt nicht nur technologische und kreative Potenziale, sondern auch eine Vielzahl ethischer Herausforderungen. Diese betreffen insbesondere die Aspekte Verantwortung, Transparenz und gesellschaftliche Auswirkungen. Um die Technologie verantwortungsvoll und nachhaltig einzusetzen, bedarf es klarer Prinzipien und Leitlinien.

⁶¹Vgl. Deutscher Ethikrat, „Künstliche Intelligenz darf menschliche Entfaltung nicht vermindern,“ 2021, Verfügbar: <https://www.ethikrat.org/presse/mitteilungen/ethikrat-kuenstliche-intelligenz-darf-menschliche-entfaltung-nicht-vermindern/> (Zugriff am: 22. Jan. 2025)

Verantwortung und Transparenz

Die Gestaltung und der Einsatz von KI-Systemen erfordern ein hohes Maß an Verantwortung seitens der Entwickler und Anwender. KI kann täuschend echte Klänge und Stimmen erzeugen, die potenziell missbraucht werden könnten, beispielsweise für die Verbreitung von manipulierten Medien oder Täuschungsversuchen. Wie Vatter und Butscher betonen, sollte die menschenzentrierte Nutzung von KI stets im Vordergrund stehen. Sie empfehlen, ethische Leitlinien zu entwickeln, die sicherstellen, dass die Technologie die menschliche Kreativität ergänzt und nicht ersetzt⁶².

Ein weiterer kritischer Punkt ist die Transparenz. Nutzer und Konsumenten müssen in der Lage sein, KI-generierte Inhalte klar von menschlich erstellten zu unterscheiden. Dies erfordert transparente Prozesse und klare Kennzeichnungen von KI-Inhalten. Ohne diese Maßnahmen könnten Vertrauen und Glaubwürdigkeit von Medieninhalten langfristig geschädigt werden.

Rechtliche und soziale Aspekte

Alpers et al. heben hervor, dass die rechtlichen und sozialen Aspekte des KI-Einsatzes oft hinter den technologischen Entwicklungen zurückbleiben. Ein zentrales Problem ist das geistige Eigentum an KI-generierten Inhalten. Es ist unklar, wem die Rechte an einem durch KI erstellten Klang oder Musikstück zustehen – dem Entwickler des Modells, dem Anwender oder den Urhebern der Trainingsdaten. Um Konflikte zu vermeiden, ist eine rechtliche Klärung erforderlich, die sowohl den Schutz des geistigen Eigentums als auch die Förderung von Innovationen gewährleistet⁶³.

Darüber hinaus könnte die Automatisierung durch KI zu einer Entwertung kreativer Berufe führen. Berufe im Bereich des Sounddesigns, die stark auf menschliche Kreativität angewiesen sind, könnten durch KI-Systeme unter Druck geraten. Dies hat nicht nur wirtschaftliche Konsequenzen, sondern verändert auch das kulturelle Verständnis von Kunst und Kreativität.

Prinzipien für den ethischen Einsatz von KI

Barton und Pöppelbuß betonen die Wichtigkeit von Prinzipien für die ethische Nutzung von KI, um gesellschaftliche und individuelle Werte zu schützen. Sie schlagen vor, dass KI-Systeme so gestaltet werden, dass sie menschliche Bedürfnisse und Werte respektieren und fördern. Dazu gehört auch, dass KI-Tools so entwickelt werden, dass sie nicht nur funktional, sondern auch sozial verträglich sind. Sie unterstreichen, dass ethische Prinzipien nicht als statisch betrachtet werden dürfen, sondern kontinuierlich an neue technologische Entwicklungen angepasst werden müssen⁶⁴.

Ethische Fragen und Herausforderungen sind ein integraler Bestandteil der Diskussion um den Einsatz von KI im Sounddesign. Klare Prinzipien, rechtliche Rahmenbedingungen und transparente Prozesse sind entscheidend, um die Technologie verantwortungsvoll zu nutzen. KI kann kreative Prozesse bereichern und effizienter machen, doch ihr Einsatz muss stets im Einklang mit ethischen und gesellschaftlichen Werten stehen, um langfristig akzeptiert und erfolgreich zu sein.

⁶²Vgl. P. Vatter und R. Butscher, „Ethische Aspekte der künstlichen Intelligenz: Implikationen und Empfehlungen für eine menschenzentrierte Nutzung,“ in Faktor Mensch, K. Weber und S. Reinheimer, Eds., Wiesbaden, Germany: Springer Fachmedien, 2022, pp. 115–130, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-37524-9_10 (Zugriff am: 22. Jan. 2025)

⁶³Vgl. S. Alpers, B.-J. Krings, C. Becker, M. Rill, und M. Weinreuter, „Ethische, rechtliche und soziale Aspekte (ELSA) der Gestaltung von KI-Systemen: Systematisierung der Betrachtung durch Vorgehensmodelle und Leitfäden,“ Zeitschrift für Arbeitswissenschaft, Bd. 78, Nr. 4, S. 160–169, Mai 2024, Verfügbar: <https://doi.org/10.1007/s41449-024-00422-9>

⁶⁴Vgl. M.-C. Barton und J. Pöppelbuß, „Prinzipien für die ethische Nutzung künstlicher Intelligenz,“ HMD Praxis der Wirtschaftsinformatik, Bd. 59, Nr. 4, S. 468–481, 2022, Verfügbar: <https://doi.org/10.1365/s40702-022-00850-3>

8. Fazit und Ausblick: Wohin geht die Reise?

Die rasante Entwicklung von Künstlicher Intelligenz (KI) hat das Sounddesign grundlegend verändert und neue Möglichkeiten für die kreative und technische Gestaltung von Klängen geschaffen. Gleichzeitig wirft die Technologie zahlreiche Fragen auf, die sowohl die Qualität und Steuerbarkeit als auch die Rolle des Menschen im kreativen Prozess betreffen. In diesem abschließenden Kapitel werden die zentralen Erkenntnisse der Arbeit zusammengefasst und ein Ausblick auf die zukünftigen Entwicklungen und Herausforderungen gegeben, die den Einsatz von KI im Sounddesign prägen könnten. Ziel ist es, die Potenziale und Grenzen der Technologie in einen breiteren Kontext zu stellen und eine Perspektive für den weiteren Weg aufzuzeigen.

8.1 Zusammenfassung der zentralen Erkenntnisse

Die vorliegende Arbeit hat sich umfassend mit den Potenzialen und Grenzen von Künstlicher Intelligenz (KI) im Sounddesign auseinandergesetzt und dabei wesentliche Erkenntnisse herausgearbeitet. Die Analyse hat gezeigt, dass der Einsatz von KI nicht nur bestehende Prozesse revolutionieren, sondern auch völlig neue Möglichkeiten für die Film- und Musikproduktion eröffnen kann. Gleichzeitig wurden aber auch klare Herausforderungen und Grenzen identifiziert, die eine bewusste und verantwortungsvolle Nutzung der Technologie erfordern.

Potenziale von KI im Sounddesign

Ein zentrales Potenzial von KI liegt in der Effizienzsteigerung und Kostensenkung. Durch den Einsatz KI-gestützter Tools können viele zeitaufwendige und manuelle Prozesse automatisiert werden. Klänge, die traditionell durch aufwendige Foley-Arbeiten oder Studioaufnahmen erstellt wurden, lassen sich nun innerhalb von Sekunden generieren. Dies spart nicht nur Zeit, sondern macht professionelles Sounddesign auch für kleinere Produktionen mit begrenztem Budget zugänglich. Besonders in Projekten mit engen Zeitrahmen erweist sich KI als wertvolles Werkzeug.

Darüber hinaus eröffnet KI neue kreative Möglichkeiten, die mit traditionellen Methoden nur schwer umsetzbar wären. Generative Modelle ermöglichen es, einzigartige und experimentelle Klänge zu erzeugen, die nicht auf realen Vorlagen basieren. Dies ist besonders in futuristischen oder surrealen Szenarien von Vorteil. Fortschritte in der Sprachsynthese (Text-to-Speech) haben zudem die Nachvertonung von Filmen, Animationsprojekten und Videospiele revolutioniert. Die Fähigkeit, Stimmen in verschiedenen Tonalitäten, Emotionen und Sprachen zu generieren, bietet Film- und Spielproduzenten eine beispiellose Flexibilität.

Grenzen und Herausforderungen

Trotz dieser Fortschritte gibt es signifikante Einschränkungen, die den Einsatz von KI im Sounddesign begrenzen. Die Qualität der generierten Klänge bleibt in vielen Fällen hinter den Anforderungen professioneller Produktionen zurück. Insbesondere die niedrigen Samplingraten und die begrenzte Klangtreue der aktuellen Modelle schränken ihre praktische Anwendbarkeit ein.

Ein weiteres zentrales Problem ist die begrenzte kreative Freiheit der KI. Während sie in der Lage ist, neue Klänge zu generieren, basiert diese Kreativität auf den vorhandenen Trainingsdaten. Die Erzeugung wirklich origineller, ikonischer Klänge – wie dem Lichtschwert

aus Star Wars – bleibt eine Herausforderung, die menschliches Know-how und künstlerische Intuition erfordert.

Die Editierbarkeit und Steuerbarkeit von KI-generierten Klängen stellt eine weitere Hürde dar. Viele Modelle bieten nur begrenzte Möglichkeiten, die Ergebnisse nachträglich zu bearbeiten oder präzise auf spezifische Anforderungen abzustimmen. Dies kann den kreativen Prozess einschränken und zusätzliche Iterationen erforderlich machen.

Ethische und rechtliche Fragen stellen ebenfalls bedeutende Herausforderungen dar. Die Verantwortung für die generierten Inhalte, die Wahrung des geistigen Eigentums und die potenziellen Auswirkungen auf kreative Berufe sind zentrale Themen, die durch klare Rahmenbedingungen und Leitlinien adressiert werden müssen.

Rolle der menschlichen Kreativität

Ein zentraler Punkt der Analyse war die Frage, wie KI die menschliche Kreativität beeinflusst. Die Ergebnisse zeigen, dass KI ein wertvolles Werkzeug sein kann, um kreative Prozesse zu unterstützen und zu erweitern. Sie darf jedoch nicht als Ersatz für die menschliche Intuition und künstlerische Vision verstanden werden. Die Synergie zwischen Mensch und Maschine bietet das größte Potenzial, innovative und hochwertige Ergebnisse zu erzielen.

Fazit der Erkenntnisse

Zusammenfassend lässt sich festhalten, dass KI im Sounddesign sowohl immense Chancen als auch klare Grenzen mit sich bringt. Ihre Effizienz und Flexibilität machen sie zu einem unverzichtbaren Werkzeug für moderne Produktionen, während die Herausforderungen in den Bereichen Qualität, Steuerbarkeit und Ethik eine bewusste Nutzung erfordern. Die Zukunft des Sounddesigns liegt in einer ausgewogenen Kombination aus KI-gestützter Technologie und menschlicher Kreativität, die zusammen eine neue Ära der audiovisuellen Produktion einläuten können.

8.2 Zukunftsaussichten

Die Zukunft des Sounddesigns wird maßgeblich durch die fortschreitende Entwicklung von Künstlicher Intelligenz (KI) geprägt. Dabei eröffnen sich neue Möglichkeiten in der kreativen Gestaltung und der Einbindung von KI in innovative Technologien. Gleichzeitig bleiben Herausforderungen bestehen, die eine nachhaltige Weiterentwicklung der Technologie erforderlich machen.

Integration in kreative Prozesse

Ein zentraler Fokus der Weiterentwicklung liegt in der Optimierung der Zusammenarbeit zwischen Mensch und Maschine. Laut Neu et al. wird KI künftig stärker als „intelligenter Assistent“ fungieren, der in der Lage ist, iterative Eingaben zu verarbeiten und diese direkt in kreative Ergebnisse umzusetzen. Diese Form der Echtzeit-Interaktion könnte den kreativen Prozess nicht nur beschleunigen, sondern auch erweitern, indem KI neue Perspektiven und Inspirationen einbringt ⁶⁵.

Zusätzlich wird die Demokratisierung des Sounddesigns durch KI-gestützte Tools eine größere Vielfalt an kreativen Stimmen ermöglichen. Durch den niedrigeren Zugang zu professionellen Technologien werden insbesondere unabhängige Kreative und kleinere Produktionsstudios von dieser Entwicklung profitieren. Dies wird die kreative Landschaft weiter diversifizieren und Innovationen fördern.

⁶⁵Vgl. M. Neu, M. Müller, B. Pothen, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst. Wiesbaden, Germany: Springer Fachmedien, 2022, Verfügbar: <https://link.springer.com/book/10.1007/978-3-658-38891-1> (Zugriff am: 22. Jan. 2025)

Neue Anwendungsfelder und immersive Technologien

Neben der traditionellen Film- und Musikproduktion wird KI zunehmend in immersiven Technologien wie Virtual Reality (VR) und Augmented Reality (AR) eingesetzt. Alpers et al. betonen, dass KI dazu beitragen kann, dynamische und interaktive Klanglandschaften zu erzeugen, die sich in Echtzeit an die Bewegungen und Aktionen der Nutzer anpassen. Solche Entwicklungen könnten das Potenzial von VR- und AR-Anwendungen erheblich erweitern und vollständig neue audiovisuellen Erlebnisse ermöglichen⁶⁶.

Ein weiteres vielversprechendes Einsatzgebiet ist die personalisierte Audioproduktion. KI-Modelle könnten genutzt werden, um individualisierte Klangerlebnisse zu schaffen, die sich an die Vorlieben und das Verhalten einzelner Nutzer anpassen. Ob maßgeschneiderte Soundtracks oder interaktive Effekte – diese Entwicklungen haben das Potenzial, die Art und Weise, wie wir Audio wahrnehmen und nutzen, grundlegend zu verändern.

Ethische und soziale Herausforderungen

Die Weiterentwicklung der Technologie bringt jedoch auch ethische und soziale Herausforderungen mit sich. Barton und Pöppelbuß weisen darauf hin, dass die Automatisierung kreativer Prozesse das Risiko birgt, die Authentizität von künstlerischen Werken zu gefährden. Sie argumentieren, dass klare Prinzipien erforderlich sind, um sicherzustellen, dass KI als unterstützendes Werkzeug eingesetzt wird, das menschliche Kreativität ergänzt, statt sie zu ersetzen. Transparenz und eine verantwortungsvolle Gestaltung der Systeme sind dabei unerlässlich⁶⁷.

Darüber hinaus bleibt die gesellschaftliche Akzeptanz der Technologie ein entscheidender Faktor. Die Befürchtung, dass KI kreative Berufe verdrängen oder standardisierte Inhalte fördern könnte, erfordert eine klare Kommunikation über die Vorteile und Grenzen der Technologie. Gleichzeitig müssen rechtliche Rahmenbedingungen geschaffen werden, die sowohl den Schutz geistigen Eigentums als auch die Verantwortung für generierte Inhalte gewährleisten.

Fazit zu den Zukunftsaussichten

Die Zukunft von KI im Sounddesign ist von spannenden Möglichkeiten geprägt, aber auch von Herausforderungen, die es zu bewältigen gilt. Fortschritte in der Integration von KI in immersive Technologien und personalisierte Audioproduktion versprechen neue kreative Horizonte. Gleichzeitig wird es darauf ankommen, ethische Prinzipien und rechtliche Rahmenbedingungen zu schaffen, die den verantwortungsvollen Einsatz der Technologie sicherstellen. Die Synergie zwischen menschlicher Kreativität und technologischer Innovation wird der Schlüssel sein, um die Möglichkeiten von KI im Sounddesign voll auszuschöpfen.

8.2.1 Wie wird KI das Sounddesign langfristig beeinflussen?

Die langfristigen Auswirkungen von Künstlicher Intelligenz (KI) auf das Sounddesign sind weitreichend und betreffen sowohl die technischen Möglichkeiten als auch die kreative Gestaltung von Klängen. Laut einer Analyse von Pietschke über die Relevanz und Anwendbarkeit aktueller KI-Technologien in der Audiosamplenerzeugung wird deutlich, dass

⁶⁶Vgl. S. Alpers, B.-J. Krings, C. Becker, M. Rill, und M. Weinreuter, „Ethische, rechtliche und soziale Aspekte (ELSA) der Gestaltung von KI-Systemen: Systematisierung der Betrachtung durch Vorgehensmodelle und Leitfäden,“ Zeitschrift für Arbeitswissenschaft, Bd. 78, Nr. 4, S. 160–169, Mai 2024, Verfügbar: <https://doi.org/10.1007/s41449-024-00422-9>

⁶⁷Vgl. M.-C. Barton und J. Pöppelbuß, „Prinzipien für die ethische Nutzung künstlicher Intelligenz,“ HMD Praxis der Wirtschaftsinformatik, Bd. 59, Nr. 4, S. 468–481, 2022 Verfügbar: <https://doi.org/10.1365/s40702-022-00850-3>

⁶⁸Vgl. B. Pietschke, Eine Analyse der Relevanz und Anwendbarkeit aktueller Künstlicher Intelligenz in der Generierung von Audiosamples für Dubstep Produzenten, Hochschule der Medien, 2024, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Pietschke.pdf> (Zugriff am: 22. Jan. 2025)

KI nicht nur Arbeitsprozesse optimiert, sondern auch die Art und Weise verändert, wie Klänge produziert und wahrgenommen werden ⁶⁸.

Effizienz und Automatisierung

KI wird langfristig eine zentrale Rolle bei der Automatisierung und Beschleunigung von Sounddesign-Prozessen spielen. Durch den Einsatz generativer Modelle können komplexe Klänge, die früher stundenlange manuelle Bearbeitung erforderten, in Sekunden erstellt werden. Dies wird besonders in Genres wie Dubstep, wo eine Vielzahl von hochspezifischen und oft experimentellen Klängen benötigt wird, zu einer Standardpraxis. Pietschke argumentiert, dass diese Effizienzsteigerung die Produktionszeit erheblich reduziert und gleichzeitig die Flexibilität erhöht, da Sounddesigner in kürzester Zeit zahlreiche Varianten eines Sounds testen können.

Neue kreative Möglichkeiten

Ein weiterer langfristiger Einfluss von KI liegt in der Erweiterung der kreativen Möglichkeiten. KI-Modelle sind in der Lage, Klänge zu generieren, die über das hinausgehen, was mit traditionellen Methoden machbar ist. Beispielsweise können Sounds erzeugt werden, die völlig abstrakt sind oder sich nicht an physikalischen Gesetzen orientieren. Pietschke hebt hervor, dass diese Fähigkeit insbesondere in elektronischen Musikgenres wie Dubstep, aber auch in Filmproduktionen mit futuristischen oder surrealen Szenarien, bahnbrechend sein kann. Diese Entwicklung wird es Sounddesignern ermöglichen, künstlerische Visionen mit einer Präzision und Vielfalt umzusetzen, die bislang undenkbar war.

Personalisierung und Interaktivität

KI wird auch die Interaktivität und Personalisierung von Klangerlebnissen fördern. In der Analyse wird betont, dass KI-gestützte Modelle in der Lage sind, Sounds dynamisch an die Präferenzen oder Handlungen eines Nutzers anzupassen. Diese Entwicklung könnte besonders in Videospielen und immersiven Technologien wie Virtual Reality (VR) und Augmented Reality (AR) entscheidend sein. Personalisierte Klanglandschaften, die sich in Echtzeit an die Aktionen eines Spielers oder an die Umgebung in einer VR-Simulation anpassen, könnten zu einer neuen Form der audiovisuellen Immersion führen.

Herausforderungen und Verantwortung

Trotz der zahlreichen Vorteile gibt es auch Herausforderungen, die berücksichtigt werden müssen. Laut Pietschke birgt die zunehmende Automatisierung durch KI die Gefahr, dass kreative Prozesse standardisiert werden und die Authentizität von Klängen verloren geht. Es wird entscheidend sein, den menschlichen Einfluss im kreativen Prozess zu bewahren und KI als unterstützendes Werkzeug zu nutzen, das menschliche Intuition und Kreativität ergänzt.

Langfristig wird KI das Sounddesign in vielerlei Hinsicht bereichern und transformieren. Sie wird Prozesse effizienter gestalten, neue kreative Möglichkeiten eröffnen und die Personalisierung von Klangerlebnissen fördern. Gleichzeitig bleibt es essenziell, die Technologie verantwortungsvoll zu nutzen und ihre Entwicklung so zu lenken, dass sie menschliche Kreativität unterstützt, anstatt sie zu verdrängen. Die Zukunft des Sounddesigns liegt in einer Synergie aus KI und menschlichen Einfallsreichtum, die zusammen eine neue Ära der Audioproduktion einläuten können.

8.2.2 Zusammenarbeit von Mensch und Maschine im kreativen Prozess

Die Zusammenarbeit von Mensch und Maschine im kreativen Prozess birgt enormes Potenzial, insbesondere im Bereich der Musik- und Soundproduktion. KI-gestützte Tools haben sich nicht nur als effiziente Werkzeuge etabliert, sondern erweitern auch die kreativen Möglichkeiten, indem sie Aufgaben übernehmen, die bisher zeitintensiv oder technisch herausfordernd waren. Die Interaktion zwischen Mensch und Maschine eröffnet eine neuartige Synergie, bei der die Stärken beider Seiten optimal genutzt werden können.

Rollenverteilung im kreativen Prozess

Ein zentraler Aspekt der Zusammenarbeit von Mensch und Maschine liegt in der klaren Rollenverteilung. Während KI-Modelle Aufgaben wie die Generierung von Klängen, die Verarbeitung großer Datenmengen oder die Automatisierung repetitiver Prozesse übernehmen, bleibt der Mensch für die kreative Leitung, die Definition von Zielen und die ästhetische Bewertung verantwortlich. KI dient hier als „kreativer Assistent“, der Vorschläge liefert und Inspirationen bietet, während der Mensch entscheidet, welche Elemente in das endgültige Werk integriert werden.

Frieler et al. betonen, dass die Rolle des Menschen in diesem Prozess nicht marginalisiert wird, sondern vielmehr eine erweiterte kreative Freiheit erhält. Durch die Unterstützung von KI können sich Kreative stärker auf konzeptionelle und künstlerische Aspekte konzentrieren, da die technische Umsetzung von der Maschine übernommen wird. Diese Arbeitsteilung eröffnet neue Möglichkeiten, besonders in Bereichen wie der Filmmusik, wo komplexe Klanglandschaften oft unter Zeitdruck entstehen müssen ⁶⁹.

Kollaboration und Inspiration

Ein wichtiger Vorteil dieser Zusammenarbeit liegt in der Fähigkeit der KI, als Inspirationsquelle zu dienen. Sie generiert oft unkonventionelle Klänge und kombiniert musikalische Elemente auf innovative Weise, was Ideen hervorbringt, die ohne maschinelle Unterstützung möglicherweise nicht entstanden wären. Diese unvorhersehbaren Ergebnisse bereichern den kreativen Prozess und eröffnen neue künstlerische Horizonte. Darüber hinaus ermöglichen datenbasierte Analysen der KI eine neue Perspektive auf bestehendes Material, was die Transformation und Wiederverwendung von Ideen erleichtert.

Interaktivität und dynamische Zusammenarbeit

Ein entscheidender Fortschritt moderner KI-Tools ist ihre Fähigkeit, in Echtzeit auf Feedback zu reagieren. Diese Interaktivität ermöglicht Sounddesignern, ihre Arbeit iterativ zu gestalten, mit der KI zu experimentieren und Ergebnisse kontinuierlich zu verfeinern. Solche dynamischen Arbeitsprozesse fördern nicht nur die Effizienz, sondern führen auch zu Ergebnissen, die sowohl technisch präzise als auch künstlerisch innovativ sind. Die Zusammenarbeit wird dadurch zu einem flexiblen und kreativen Dialog zwischen Mensch und Maschine, der die Grenzen traditioneller Arbeitsweisen erweitert.

⁶⁹Vgl. K. Frieler, W.-G. Zaddach, and S. Meyer, „Künstliche Intelligenz in der Musikproduktion,“ in Musik und Internet, N. Ruth and P. Moormann, Eds., Wiesbaden, Germany: Springer Fachmedien Wiesbaden, 2023, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-39145-4_1 (Zugriff am: 22. Jan. 2025)

Herausforderungen der Zusammenarbeit

Trotz der zahlreichen Vorteile der Mensch-Maschine-Interaktion gibt es auch Herausforderungen, die berücksichtigt werden müssen. Der Einsatz von KI erfordert ein gewisses Maß an technischer Expertise, die nicht jedem Kreativen sofort zugänglich ist. Diese Hürde kann dazu führen, dass das volle Potenzial der Technologie nicht ausgeschöpft wird. Darüber hinaus besteht das Risiko, dass die Ergebnisse von KI-Tools generisch oder austauschbar wirken, insbesondere wenn der kreative Einfluss des Menschen zu gering bleibt. Um dies zu vermeiden, müssen KI-Modelle so entwickelt werden, dass sie flexibel und anpassbar sind, damit sie den individuellen Stil und die Vision des Nutzers optimal unterstützen.

Die Zusammenarbeit von Mensch und Maschine im kreativen Prozess bietet dennoch eine vielversprechende Perspektive für die Zukunft des Sounddesigns. Während die KI technische Aufgaben automatisiert und innovative Ideen liefert, bleibt der Mensch die entscheidende treibende Kraft hinter der künstlerischen Vision. Diese Synergie ermöglicht effizientere Arbeitsprozesse und gleichzeitig Ergebnisse, die den kreativen Horizont erweitern. Damit diese Zusammenarbeit jedoch nachhaltig erfolgreich ist, ist es entscheidend, dass die Technologie benutzerfreundlich bleibt und den kreativen Freiraum des Menschen respektiert. Nur so kann das volle Potenzial dieser Partnerschaft ausgeschöpft werden.

8.3 Schlussbetrachtung

Die Arbeit hat den Vergleich zwischen traditionellem und KI-gestütztem Sounddesign in der Filmproduktion in den Mittelpunkt gestellt und dabei zentrale Unterschiede, Stärken und Schwächen beider Ansätze aufgezeigt. Traditionelles Sounddesign beeindruckt durch seine künstlerische Präzision, die Fähigkeit, emotionale Nuancen gezielt zu gestalten, und die enge Verbindung zwischen Kreativen und Werk. Es erfordert jedoch oft einen hohen Zeit- und Ressourcenaufwand, was seine Flexibilität und Effizienz begrenzt. Im Gegensatz dazu bietet KI-gestütztes Sounddesign eine moderne Alternative, die sich durch Geschwindigkeit, Anpassungsfähigkeit und die Möglichkeit, komplexe Klangwelten automatisch zu generieren, auszeichnet. Diese Technologien ermöglichen eine deutliche Effizienzsteigerung und eröffnen neue kreative Möglichkeiten, insbesondere in zeitkritischen Produktionsumgebungen. Dennoch zeigt die Arbeit, dass KI-Ansätze häufig von der kreativen Leitung durch den Menschen abhängen, um ihre Ergebnisse individuell und einzigartig zu gestalten.

Der Vergleich verdeutlicht, dass weder traditionelles noch KI-gestütztes Sounddesign allein die optimale Lösung darstellen. Stattdessen liegt die Zukunft in der Kombination beider Ansätze, bei der die künstlerische Expertise des Menschen durch die technologische Unterstützung der KI ergänzt wird. Diese Synergie ermöglicht es, sowohl die emotionale Tiefe traditioneller Methoden als auch die Effizienz und Flexibilität moderner Technologien zu nutzen.

Abschließend lässt sich sagen, dass die Zusammenarbeit von Mensch und Maschine das Potenzial hat, das Sounddesign grundlegend zu verändern. Die in dieser Arbeit gewonnenen Erkenntnisse liefern eine Grundlage für zukünftige Entwicklungen und betonen die Bedeutung eines bewussten Einsatzes von KI, um den kreativen Prozess zu bereichern und gleichzeitig die künstlerische Authentizität zu bewahren.

9. Literaturverzeichnis

- A. Hoffmann, „Prinzipien für die ethische Nutzung künstlicher Intelligenz,“ Ethik in der Technik, vol. 14, no. 2, pp. 101–115, 2022, Verfügbar: <https://link.springer.com/article/10.1365/s40702-022-00850-3> (Zugriff am: 22. Jan. 2025)
- B. Pietschke, Eine Analyse der Relevanz und Anwendbarkeit aktueller Künstlicher Intelligenz in der Generierung von Audiosamples für Dubstep Produzenten, Hochschule der Medien, 2024, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Pietschke.pdf>, (Zugriff am: 22. Jan. 2025)
- E. Guizzo und A. Novello, „A Neural Network Based Framework for Archetypical Sound Synthesis,“ 6. März 2020, Verfügbar: <https://arxiv.org/pdf/2003.03160>, (Zugriff am: 22. Jan. 2025)
- F. Schätzlein, „Sound und Sounddesign in Medien und Forschung,“ in Sound. Zur Technologie und Ästhetik des Akustischen in den Medien, H. Segeberg und F. Schätzlein, Eds., Marburg: Schüren, 2005, Verfügbar: <https://mediarep.org/server/api/core/bitstreams/ff7aef42-6cb7-46be-a90c-79d1d6eb604f/content>, (Zugriff am: 22. Jan. 2025)
- H. Liu, Z. Chen, Y. Yuan, X. Mei, X. Liu, D. Mandic, W. Wang, und M. D. Plumbley, „AudioLDM: Text-to-Audio Generation with Latent Diffusion Models,“ in Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023. Mit Soundbeispielen: <https://audioldm.github.io/> <https://arxiv.org/pdf/2301.12503>, (Zugriff am: 22. Jan. 2025)
- J. Martin und F. Butzmann, „Sound Design,“ in F. Hentschel und P. Moormann (Hrsg.), Filmmusik, Springer Fachmedien Wiesbaden GmbH, 2018, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11, (Zugriff am: 22. Jan. 2025)
- J. Müller und S. Braun, „Ethische, rechtliche und soziale Aspekte (ELSA) der Gestaltung von KI-Systemen,“ KI & Gesellschaft, vol. 12, no. 3, pp. 45–58, 2024, Verfügbar: <https://link.springer.com/article/10.1007/s41449-024-00422-9> (Zugriff am: 22. Jan. 2025)
- J.-P. Briot, G. Hadjeres, und F.-D. Pachet, Deep Learning Techniques for Music Generation – A Survey, Computational Synthesis and Creative Systems, Springer, 2019, Verfügbar: <https://arxiv.org/pdf/1709.01620>, (Zugriff am: 22. Jan. 2025)
- K. Frieler, W.-G. Zaddach, und S. Meyer, „Künstliche Intelligenz in der Musikproduktion,“ in Musik und Internet, N. Ruth und P. Moormann, Eds., Springer Fachmedien Wiesbaden, 2023, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-39145-4_1, (Zugriff am: 22. Jan. 2025)

- K. Tatar, K. Cotton, und D. Bisig, „Sound Design Strategies for Latent Audio Space Explorations Using Deep Learning Architectures,“ in Proceedings of the Sound and Music Computing Conference (SMC), Stockholm, Sweden, 2023, Verfügbar: <https://arxiv.org/pdf/2305.15571>, (Zugriff am: 22. Jan. 2025)
- L. Duffner, „Der Foley Artist – historische Entwicklung und heutige Arbeitsweise,“ Fachartikel, 2020. Zugriff am 5. Januar 2025, Verfügbar: <https://curdt.home.hdm-stuttgart.de/PDF/Foley%20Artist.pdf>, (Zugriff am: 22. Jan. 2025)
- L. Münter, Künstliche Intelligenz im (räumlichen) Sounddesign, Hochschule der Medien, 2022, Verfügbar: https://ai.hdm-stuttgart.de/downloads/student-white-paper/Winter-2122/KI_im_Sounddesign.pdf, (Zugriff am: 22. Jan. 2025)
- L. Prechelt, Künstliche Intelligenz (KI), 2023, Verfügbar: https://www.inf.fu-berlin.de/inst/ag-se/teaching/V-AdI-2022/10_KI.pdf, (Zugriff am: 22. Jan. 2025)
- M. Božić und M. Horvat, „A survey of deep learning audio generation methods,“ University of Zagreb, Faculty of Electrical Engineering and Computing, 2023, Verfügbar: <https://arxiv.org/pdf/2406.00146>, (Zugriff am: 22. Jan. 2025)
- M. Huzafah und L. Wyse, „Deep Generative Models for Musical Audio Synthesis,“ Preprint for Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, Springer, 2020, Verfügbar: <https://arxiv.org/pdf/2006.06426>, (Zugriff am: 22. Jan. 2025)
- M. Neu, M. Müller, B. Pothen, und M. Zingel, Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz: Wie KI die Gesellschaft und Marketing beeinflusst, Wiesbaden, Germany: Springer Fachmedien, 2022, Verfügbar: <https://link.springer.com/book/10.1007/978-3-658-38891-1>, (Zugriff am: 22. Jan. 2025)
- M. Schill, M. Hepp, und B. Böröcz, „Künstliche Intelligenz in der Audiowelt,“ Whitepaper, Hochschule der Medien, Wintersemester 2018/19, Verfügbar: https://ai.hdm-stuttgart.de/downloads/student-white-paper/Winter-1819/KI_in_der_Audiowelt.pdf, (Zugriff am: 22. Jan. 2025)
- M. Schmidt, „Ethische Aspekte der künstlichen Intelligenz: Implikationen und Herausforderungen,“ in Künstliche Intelligenz und Gesellschaft, Springer, Wiesbaden, Germany, pp. 95–108, 2022, Verfügbar: https://link.springer.com/chapter/10.1007/978-3-658-34524-2_7 (Zugriff am: 22. Jan. 2025)
- N. Anantrasirichai und D. Bull, „Artificial intelligence in the creative industries: a review,“ Artificial Intelligence Review, vol. 55, pp. 589–656, Jul. 2021, Verfügbar: <https://link.springer.com/article/10.1007/s10462-021-10039-7>, (Zugriff am: 22. Jan. 2025)

P. Aichroth und J. Liebetrau, „KI-basiertes akustisches Monitoring: Herausforderungen und Lösungsansätze für datengetriebene Innovationen auf Basis audiovisueller Analyse,“ in Entrepreneurship der Zukunft, L. Heim und S. Gerth, Eds., Springer, 2023. Verfügbar: [\[https://link.springer.com/chapter/10.1007/978-3-658-42060-4_4\]](https://link.springer.com/chapter/10.1007/978-3-658-42060-4_4), (Zugriff am: 22. Jan. 2025)

P. Barth and A. Ulges, Künstliche Intelligenz, Hochschule RheinMain, 2023
<https://pdfs.semanticscholar.org/567f/fa0e1421e1937b9488b5475f159c8761b5cb.pdf?utm.com>, (Zugriff am: 22. Jan. 2025)

P. Moormann und F. Hentschel, Filmmusik – Ein alternatives Kompendium, Springer Fachmedien Wiesbaden GmbH, 2018.
https://link.springer.com/chapter/10.1007/978-3-658-11237-0_11?utm.com, (Zugriff am: 22. Jan. 2025)

Q. Shang, Künstliche Intelligenz und Musikproduktion, KITopen Library, 2023
<https://publikationen.bibliothek.kit.edu/1000165222?utm.com>, (Zugriff am: 22. Jan. 2025)

R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, „High-Resolution Image Synthesis with Latent Diffusion Models,“ arXiv, 2022. [Online]. Verfügbar: <https://arxiv.org/abs/2112.10752>, (Zugriff am: 22. Jan. 2025)

S. Oh, M. Kang, H. Moon, K. Choi, and B. S. Chon, “A Demand-Driven Perspective on Generative Audio AI, 2023
<https://arxiv.org/pdf/2307.04292>, (Zugriff am: 22. Jan. 2025)

T. Görne, “Sounddesign, Klang, Wahrnehmung, Emotion” 2017
https://books.google.de/books?id=rOUIDwAAQBAJ&printsec=frontcover&hl=de&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false, (Zugriff am: 22. Jan. 2025)

T. Wagner, “Sound-Design für Filmemacher,“ filmpraxis.de, 2023
<https://filmpraxis.de/wp-content/uploads/LK-Sound-Design-Lektionen.pdf?utm.com>, (Zugriff am: 22. Jan. 2025)

Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,“ Nature, vol. 521, pp. 436–444, May 2015, Verfügbar: https://www.researchgate.net/publication/277411157_Deep_Learning, (Zugriff am: 22. Jan. 2025)

Z. Zhao and L. Zhang, “Design of artificial intelligence cultural creative industry based on machine learning,“ Soft Computing, vol. 27, no. 13, pp. 8693–8703, 2023, Verfügbar: <https://link.springer.com/article/10.1007/s00500-023-08693-w?utm.com>, (Zugriff am: 22. Jan. 2025)

9.1 Internetquellen

AVIA, Verfügbar: <https://www.aiva.ai/> (Zugriff am: 22. Jan. 2025)

Deutscher Ethikrat, „Künstliche Intelligenz darf menschliche Entfaltung nicht vermindern,“ 2020, Verfügbar: <https://www.ethikrat.org/presse/mitteilungen/ethikrat-kuenstliche-intelligenz-darf-menschliche-entfaltung-nicht-vermindern/> (Zugriff am: 22. Jan. 2025)

Elevenlabs, Verfügbar: <https://elevenlabs.io/> (Zugriff am: 22. Jan. 2025)

Meta AudioGen, Verfügbar: <https://audiocraft.metademolab.com/audiogen.html> (Zugriff am: 22. Jan. 2025)

ROCKY (1976) Sound, Verfügbar: <https://dustinogburn.wordpress.com/2014/07/13/rocky-1976-sound/> (Zugriff am: 22. Jan. 2025)

Rocky Balboa - sound design, Verfügbar: https://www.youtube.com/watch?v=U_CTB_I8zEE (Zugriff am: 22. Jan. 2025)

Stable-Audio-open, Verfügbar: <https://stable-audio-open.com/?utm.com> (Zugriff am: 22. Jan. 2025)

Stable-Audio-open-zero, Verfügbar: <https://huggingface.co/spaces/artificialguybr/Stable-Audio-Open-Zero> (Zugriff am: 22. Jan. 2025)

Suno, Verfügbar: <https://suno.com/> (Zugriff am: 22. Jan. 2025)