

BACHELORARBEIT

KI-gestützte Kunst- und Musiktherapie in VR

vorgelegt am 28. März 2025
Anh Sang Tran

Erstprüferin: Prof. Dr. Sabine Schumann
Zweitprüfer: Prof. Dr. Eike Langbehn

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**

Department Medientechnik
Finkenau 35
20081 Hamburg

Zusammenfassung

Virtual Reality (VR) und Künstliche Intelligenz (KI) bieten neue Möglichkeiten zur Unterstützung von kreativen und entspannungsfördernden Anwendungen. In dieser Arbeit wurde ein VR-Prototyp entwickelt, der zentrale Aspekte von kunst- und musiktherapeutischen Ansätzen miteinander kombiniert. Dieser erlaubt es Nutzern in einer dreidimensionalen VR-Umgebung zu zeichnen und dabei Gesangsstimmen erklingen zu lassen, die zuvor durch KI generiert wurden. Ziel war es, eine benutzerfreundliche Anwendung zu gestalten, die zur Entspannung beiträgt. Die Wirksamkeit wurde in einer Nutzerstudie untersucht, bei der unter anderem das subjektive Entspannungsgefühl vor und nach der Nutzung erfasst wurde. 80% der Teilnehmer berichteten von einer Steigerung ihrer Entspannung. Zudem wurde die intuitive Steuerung sowie das Zusammenspiel von künstlerischem Ausdruck und klanglicher Rückmeldung des Prototyps positiv bewertet. Die Ergebnisse zeigen das Potenzial von VR, kunst- und musiktherapeutische Elemente in einem interaktiven, immersiven Umfeld zu vereinen. Während kleine Optimierungspotenziale, insbesondere in der KI-gestützten Audiogenerierung identifiziert wurden, bestätigen die positiven Rückmeldungen die Wirksamkeit des Konzepts und seine Eignung als unterstützende Methode zur Förderung von Entspannung und kreativem Ausdruck.

Abstract

Virtual Reality (VR) and Artificial Intelligence (AI) offer new opportunities to support creative and relaxation-promoting applications. In this work, a VR prototype was developed that combines central aspects of art and music therapy approaches. It allows users to draw in a three-dimensional VR environment while having singing voices, previously generated by AI, play in the background. The goal was to create a user-friendly application that contributes to relaxation. The effectiveness was examined in a user study, which included measuring the subjective feeling of relaxation before and after use. 80% of the participants reported an increase in their relaxation. Additionally, the intuitive control as well as the interplay of artistic expression and auditory feedback of the prototype were positively evaluated. The results demonstrate the potential of VR to combine art and music therapy elements in an interactive, immersive environment. While small optimization potentials, particularly in AI-supported audio generation, were identified, the positive feedback confirms the effectiveness of the concept and its suitability as a supportive method for promoting relaxation and creative expression.

Danksagung

Ich möchte an dieser Stelle meine Dankbarkeit gegenüber allen ausdrücken, die mich während der Arbeit an meiner Bachelorarbeit unterstützt und begleitet haben.

Ich danke meiner Erstprüferin Prof. Dr. Sabine Schumann für die Hilfe bei der Themenfindung und für die Betreuung und Unterstützung, während des gesamten Prozesses. Ich bedanke mich ebenfalls bei Prof. Dr. Eike Langbehn, der ebenfalls bei der Wahl des Forschungsthemas unterstützt hat und sich bereit erklärt hat, die Rolle des Zweitprüfers zu übernehmen.

Ich möchte mich beim Stadtorchester Winsen Luhe bedanken, vor allem bei allen Mitgliedern, die sich trotz stressiger Tagesprobe bereit erklärt haben bei meiner Nutzerstudie teilzunehmen.

Ich möchte mich ebenfalls bei allen Mitarbeitenden des GKV-IT und PM-Teams von Check24 bedanken, die ihren stressigen Arbeitstag unterbrochen haben, um bei meiner Nutzerstudie teilzunehmen.

Ich danke Vivien Laasch für die wundervollen Modelle des Pinsels und der Palette - meine eigenen waren nicht halb so schön.

Ich möchte zuletzt meinen Eltern danken für ihr Verständnis und ihre tolle Unterstützung, die ich während dieser intensiven Zeit erhalten habe.

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	IV
Abbildungsverzeichnis.....	V
Tabellenverzeichnis.....	VII
1 Einleitung	1
1.1 Vorwort.....	1
1.2 Motivation.....	1
1.3 Zielsetzung.....	2
1.4 Aufbau	3
2 Grundlagen	4
2.1 Therapie	4
2.1.1 Kunsttherapie	4
2.1.2 Musiktherapie.....	6
2.2 Virtual Reality	8
2.2.1 Therapie in VR.....	9
2.3 Vergleichbare Arbeiten.....	10
2.3.1 Blob Opera	10
2.3.2 Tilt Brush.....	11
2.3.3 Virtual Harmony	12
2.4 Künstliche Intelligenz.....	13
2.4.1 Was ist Künstliche Intelligenz?.....	13
2.4.2 Maschinelles Lernen.....	13
2.4.3 Tiefes Lernen.....	15
2.4.4 Neuronale Netzwerke	15
2.4.5 Singing Voice Synthesis.....	18
2.4.6 DNN-basierte Gesangssynthese	20
2.5 Weitere Grundlagen.....	23
2.5.1 Grundlagen der Farblehre.....	23
2.5.2 Grundlagen der Musiktheorie.....	24

3	Konzept	26
3.1	Kernkonzept.....	26
3.2	Anforderungen.....	27
3.3	Design- und Bedienkonzept.....	27
3.3.1	Auswahl der Hardware.....	27
3.3.2	Nutzungserfahrung und Design.....	28
3.3.3	Auswahl der Töne	31
3.3.4	Auswahl der Farben.....	33
3.4	Unity als VR-Entwicklungsumgebung.....	34
3.4.1	OpenXR.....	34
3.4.2	XR Interaction Toolkit	35
3.4.3	Line Renderer	35
3.4.4	Audiokomponenten	36
3.4.5	Physikkomponenten	36
3.5	Erstellung der Töne.....	37
3.5.1	MuseScore Studio	37
3.5.2	Sinsy	37
3.5.3	Audacity	38
4	Umsetzung.....	39
4.1	Vorbereitungen	39
4.1.1	Erstellung der 3D-Modelle.....	39
4.1.2	Erstellung der Musikpartituren.....	40
4.1.3	Generierung der Gesangsstimmen.....	42
4.1.4	Schneiden der Audiodateien.....	43
4.2	Entwicklung in Unity.....	44
4.2.1	VR-Interaktion	44
4.2.2	Zeichenfunktion.....	45
4.2.3	Sounds	46
4.2.4	Mischmodus	47
5	Evaluation.....	50

5.1	Fragebogen	50
5.2	Durchführung.....	51
5.3	Ergebnisse bezüglich des Entspannungsgefühls.....	53
5.3.1	Auswertung der Ergebnisse zur Entspannung.....	55
5.4	Ergebnisse bezüglich des Prototyps.....	57
5.4.1	Auswertung der Ergebnisse zum Prototyp	59
5.5	Auswertung der Kommentare und Beobachtungen.....	63
5.6	Limitationen.....	65
6	Fazit	66
6.1	Ausblick.....	67
	Literaturverzeichnis.....	68
	Anhang	76
	Eigenständigkeitserklärung	84

Abkürzungsverzeichnis

BiLSTM	Bidirektionale Long Short-Term Memory
CNN	Convolutional Neural Network
DL	Deep Learning (Tiefes Lernen)
DNN	Deep Neural Network (Tiefes neuronales Netzwerk)
FEE	Fragebogen zur Erfassung von Entspannung
FFNN	Feedforward Neural Network
HMD	Head-Mounted Display
HMM	Hidden Markov Model
KI	Künstliche Intelligenz
LSTM	Long Short-Term Memory
M	Mittelwert
MDN	Mixture Density Network
ML	Maschinelles Lernen
RNN	Recurrent Neural Network
RSQ	Relaxation State Questionnaire
SD	Standardabweichung
SVS	Singing Voice Synthesis
VR	Virtual Reality

Abbildungsverzeichnis

Abbildung 1: Kunsttherapeutische Triade	5
Abbildung 2: Komplettes VR-System nach LaVelle	8
Abbildung 3: Das von Sutherland entwickelte HMD	9
Abbildung 4: Blob Opera	10
Abbildung 5: Tilt Brush in aktiver Nutzung, aufgenommen in mehreren Winkeln	11
Abbildung 6: Xylophon, Klavier und Schlagzeug in Virtual Harmony.....	12
Abbildung 7: Methoden der KI.....	13
Abbildung 8: Unterschied von Maschinellern Lernen zu Traditionellen Programmen	14
Abbildung 9: Vereinfachte Darstellung eines KNNs mit Rückkoppelung.....	17
Abbildung 10: Vereinfachte Darstellung eines SVS mit Waveform-Synthesizer	18
Abbildung 11: Vereinfachte Darstellung eines SVS mit konkatentivem Verfahren.....	19
Abbildung 12: Aufbau des akustischen Modells von Sinsy.....	21
Abbildung 13: Komplettes Sinsy-System	22
Abbildung 14: Beispielhafter Farbkreis mit Primärfarben (Rot, Blau, Gelb) und den Zwischenstufen	23
Abbildung 15: C-Dur Tonleiter.....	24
Abbildung 16: Tonika-Dreiklang, Subdominanten-Dreiklang, Dominanten-Dreiklang und Parallele in C-Dur.....	24
Abbildung 17: Der vierstimmige Satz am Beispiel eines C-Dur Akkords.	25
Abbildung 18: Stimmreichweiten der vier Stimmlagen Sopran, Alt, Tenor und Bass	25
Abbildung 19: Startansicht: Das Erste, was Nutzer sehen, nachdem das Programm gestartet ist.....	28
Abbildung 20: Pinsel und Palette von oben (Ansicht im Editor).....	29
Abbildung 21: Gezeichnetes Wort OPER mit Benutzung aller vier Farben der kleinen Farbkleckse ..	29
Abbildung 22: Gezeichnetes Wort Hamburg mit allen sieben Farben Quelle: Eigene Darstellung.....	30
Abbildung 23: Tonumfang der vier Stimmlagen mit Startton in der Mitte	31
Abbildung 24: Einige mögliche Akkord-Kombinationen, wenn alle vier Stimmen zu hören sind	32
Abbildung 25: Audiokomponenten: Ein Audio Listener und zwei Audio Sources	36
Abbildung 26: Ursprüngliche Skizze (links) und die fertigen Modelle (rechts).....	39
Abbildung 27: Ausschnitt aus der Japanisch-Phonem-Tabelle des Sinsy Reference Guides.....	40
Abbildung 28: Fertige Partitur am Beispiel Bass und der Silbe „Wah“	41
Abbildung 29: Fertige Partitur am Beispiel Sopran und der Silbe „Do“	41
Abbildung 30: Kompletter Audio-Clip nach Trennung	43
Abbildung 31: „Bereinigter“ Audio-Clip, der nur aus den benötigten Tönen besteht.....	43
Abbildung 32: Subjektives Entspannungsgefühl vor und nach der Nutzung pro Teilnehmer (Gruppe 1)	53

Abbildung 33: Subjektives Entspannungsgefühl vor und nach der Nutzung pro Teilnehmer (Gruppe 2)	53
Abbildung 34: Mittelwerte (beider Gruppen) und Standardabweichungen der Ergebnisse vor der Nutzung	54
Abbildung 35: Mittelwerte (beider Gruppen) und Standardabweichungen der Ergebnisse nach der Nutzung	55
Abbildung 36: Vergleich der Mittelwerte der Diagramme Abb. 34 und 35 (ohne Standardabweichung)	56
Abbildung 37: Verteilung der Antworten zur Wahrnehmung der KI-generierten Stimmen	57
Abbildung 38: Verteilung der Antworten zur Zeichen- und Musizierfunktion	57
Abbildung 39: Verteilung der Antworten zum Zusammenspiel von Kunst und Musik	58
Abbildung 40: Verteilung der Antworten bezüglich des therapeutischen Nutzens des Prototyps	58
Abbildung 41: Mittelwerte und Standardabweichung bei Antworten zur Benutzerfreundlichkeit und Interaktivität	59
Abbildung 42: Bearbeitung der Aufgabe Sonne und Blume zeichnen mit ergänzendem Fluss und Wiese von Teilnehmer 5 der Gruppe 2 (Screenshot innerhalb des HMDs)	64
Abbildung 43: Musikalische Experimente von Teilnehmer 3 der Gruppe 1 (Screenshot vom Livestream auf Smartphone)	64

Tabellenverzeichnis

Tabelle 1: Parameter der einzelnen Stimmen.....	42
Tabelle 2: Demographische und kontextuelle Daten der Gruppe 1 (Orchester).....	51
Tabelle 3: Demographische und kontextuelle Daten der Gruppe 1 (Büro)	51
Tabelle 4: Mittelwerte und Standardabweichung der Fragen bezüglich der KI-generierten Stimmen..	60
Tabelle 5: Mittelwerte und Standardabweichung der Fragen bezüglich der erzeugten Akkorde	60
Tabelle 6: Mittelwerte und Standardabweichungen zum Zeichnen und Malen und der Kombination aus beiden	61
Tabelle 7: Mittelwerte und Standardabweichungen zu Aussagen zum Prototyp als Alternative	62
Tabelle 8: Mittelwerte und Standardabweichung aller Einzelaussagen zur Kategorie Steuerung.....	63

1 Einleitung

1.1 Vorwort

Zur besseren Lesbarkeit dieser Bachelorarbeit wird bei Personenbezeichnungen das generische Maskulinum verwendet. Selbstverständlich gelten sämtliche Personenbezeichnungen auf alle Geschlechter.

1.2 Motivation

Wo Worte fehlen, spricht die Musik. Dieses Zitat wird häufig Hans Christian Andersen zugesprochen und es ist selbst in der heutigen Zeit noch sehr passend. Musik ist für viele Menschen nicht mehr vom Alltag wegzudenken und kann dank neusten technologischen Fortschritten beinahe zu jederzeit und überall gehört oder praktiziert werden (Bullerjahn, 2004). Doch Musik kann weitaus mehr als nur unterhalten. Es wurde nachgewiesen, dass Musik einen äußerst positiven Effekt auf die Psyche und den Körper hat, weshalb sie häufig als Medizin in Form von Musiktherapie eingesetzt wird (Bernatzky & Kreutz, 2015; de Witte et al., 2022). Aber nicht nur Musik kann positive Emotionen und Gefühle auslösen, sondern auch Farben und Kunst im Allgemeinen. Die Kunsttherapie ist sich dessen bewusst und wurde in den letzten Jahren immer häufiger eingesetzt (Obermeier, 2019).

Gleichzeitig ist jedoch die Nachfrage nach psychotherapeutischer Behandlung in den letzten Jahren insbesondere wegen stressbedingter Beschwerden deutlich gestiegen (American Psychological Association, 2021). Stress kann schwerwiegende gesundheitliche Folgen haben. Stresspsychologen Rasoul Tanghatar (2012) zufolge gilt er unter anderem als der Auslöser von Schlafstörungen, Burnout, Hörsturz, Bluthochdruck und Hauterkrankungen bis hin zur Schwächung des Autoimmunsystems, Depressionen und Demenz (Tanghatar, 2012).

64 Prozent der Menschen in Deutschland gaben in einer Stressstudie der Techniker Krankenkasse 2021 an, dass sie gelegentlich Stress empfinden. Dabei gab ein gutes Viertel (26 Prozent) sogar an, dass sie häufiger gestresst sind. Das ist ein Anstieg von sechs Prozent im Vergleich zu der Stressstudie aus dem Jahr 2013 (Techniker Krankenkasse, 2021). Dieser Anstieg hat zur Folge, dass das Thema Stress nun mehr von der Gesellschaft anerkannt wird. Das führt wiederum dazu, dass sich vor allem Therapeuten vermehrt mit den Auswirkungen von Stress und ihren Behandlungsmöglichkeiten befassen müssen (M. Heinrichs et al., 2015).

Neben den bereits etablierten Methoden der Kunst- und der Musiktherapie rückt zunehmend auch der Einsatz neuer Technologien in den Fokus der Forschung. Insbesondere Virtual Reality (VR) hat bereits in der Behandlung von Angststörungen und Stress Erfolge zeigen können (Maples-Keller et al., 2017; Syed Abdullah et al., 2021). VR in Kombination mit Musiktherapie hat bereits einige Studien mit moderaten bis guten Ergebnissen (Alexanian, 2021; Brungardt et al., 2021). Laut den Ergebnissen der

Studie von Alexanian (2021) hat Musiktherapie in Kombination mit VR zwar positive Effekte, ähnliche Ergebnisse erzielt Musiktherapie ohne VR seinen Ergebnissen zufolge auch (Alexanian, 2021). Dabei standen meist die Effekte des Musikhörens im Vordergrund, nicht jedoch das eigene Musizieren. Musiktherapeutische Ansätze, die musikalische Improvisation beinhalten, spielen aber auch eine zentrale Rolle in der Musiktherapie (Gold et al., 2009). Eine Studie, die sowohl das Hören als auch das eigene Musizieren in VR untersuchte, zeigte insbesondere bei der Reduzierung von Stress positive Ergebnisse (Lin et al., 2020). Diese Arbeit wird in Kapitel 2.3.3 näher erläutert.

Kunsttherapie in VR hingegen hat bereits eine etablierte Software, welche bereits in diversen Studien getestet wurde (Shamri Zeevi, 2021; Ying-Chun & Chwen-Liang, 2018). Auf diese wird in Kapitel 2.3.2. näher eingegangen.

Eine weitere Technologie, mit der Kunst- und Musiktherapie bereichert werden kann, ist die künstliche Intelligenz (KI). KI-Technologien haben sich schon lange in der Medizin etabliert und eröffnen neue Möglichkeiten für innovative Therapieansätze (Wittpahl, 2019).

Diese Arbeit beschäftigt sich nun damit, all die Vorteile der Technologien und Therapien zu vereinen: Eine KI-gestützte Kunst- und Musiktherapie in VR. Durch die Kombination von VR, KI, Kunsttherapie und Musiktherapie soll ein innovatives Konzept entwickelt werden, das neue Möglichkeiten zur Stressbewältigung und Entspannung schafft.

1.3 Zielsetzung

Ziel dieser Arbeit ist es, ein innovatives Konzept für eine KI-gestützte Kunst- und Musiktherapie in VR zu erarbeiten sowie die Entwicklung eines Prototyps und einer ersten Evaluation. Dabei sollen die positiven Effekte des musikalischen Improvisierens der Musiktherapie mit den gestalterischen Ausdrucksmöglichkeiten der Kunsttherapie in VR verbunden werden. KI wird für die Generierung von Stimmen eingesetzt, um das Erlebnis zu erweitern.

Der VR-Prototyp soll dem Nutzer die Möglichkeit bieten in einem freien virtuellen Raum mit einem Pinsel zu zeichnen und dabei verschiedene Töne von KI-generierten Stimmen erklingen zu lassen. Dabei soll die Stimmlage abhängig von der ausgewählten Farbe und die Tonhöhe abhängig von der Bewegung des Pinsels sein. Durch die Verbindung von KI, Kunsttherapie, Musiktherapie und VR soll ein neuartiges therapeutisches Erlebnis geschaffen werden. Der Fokus der Evaluation des Prototyps soll dabei auf das Erhöhen des subjektiven Entspannungsgefühls der Nutzer und in seiner Benutzerfreundlichkeit und Interaktivität liegen. Die Evaluation des Prototyps basiert auf folgenden Hypothesen:

- H1: Der VR-Prototyp hat einen positiven Einfluss auf das subjektive Entspannungsgefühl unmittelbar nach der Nutzung.

- H2: Der VR-Prototyp erfüllt die im Konzept definierten Anforderungen (Kapitel 3.3) an Benutzerfreundlichkeit und Interaktivität.

Diese Hypothesen werden in einer Nutzerstudie, in der Nutzer den Prototyp ausprobieren, überprüft.

1.4 Aufbau

Das folgende Kapitel 2 behandelt die für diese Arbeit relevanten Grundlagen zu Kunst- und Musiktherapie, Virtual Reality (VR) und Künstliche Intelligenz (KI). Zudem werden grundlegende Konzepte der Farb- und Musiktheorie erläutert, die für die Entwicklung des Prototyps relevant sind. Dort folgt auch eine Übersicht der vergleichbaren Arbeiten, die zum Teil das Konzept inspirierten. Kapitel 3 beschreibt das Konzept, auf dem der Prototyp basiert. Es werden auch die Technologien, die zur Umsetzung des Konzepts verwendet werden, vorgestellt. In Kapitel 4 wird die schrittweise Umsetzung des Konzeptes erklärt. In Kapitel 5 geht es um die Evaluation des Prototyps. Abschließend wird in Kapitel 6 ein Fazit gezogen und ein Ausblick gegeben.

2 Grundlagen

Dieses Kapitel vermittelt die theoretischen Grundlagen der Arbeit. Zu Beginn werden die Therapieformen Kunst- und Musiktherapie vorgestellt. Danach folgt eine Einführung in Virtual Reality (VR) und ihrer Nutzung in psychotherapeutischen Anwendungen. Darauf folgt eine Analyse von vergleichbaren Arbeiten die Kunst- oder Musiktherapie mit VR kombinieren. Anschließend werden Grundlagen zu Künstlichen Intelligenz (KI) erläutert. Dabei liegt der Fokus auf der KI-gestützten Stimmengenerierung (Singing Voice Synthesis). Zum Schluss werden relevante Grundlagen der Farb- und Musiktheorie kurz behandelt.

2.1 Therapie

In der Therapie werden verschiedene Formen der Therapie angewendet. Für diese Arbeit sind jedoch nur die Kunsttherapie und die Musiktherapie von Bedeutung, weshalb diese in den nächsten Unterkapiteln vorgestellt werden.

2.1.1 Kunsttherapie

Die Kunsttherapie ist eine relativ neue Form der Psychotherapie. Sie ist aber aufgrund ihrer positiven Wirkungen eine besonders häufig genutzte Therapiemethode (Obermeier, 2019). Aber wie genau lässt sich Kunsttherapie definieren? Laut Schuster (2014) handelt es sich bei der Kunsttherapie um „eine Psychotherapie, in der Werke aus dem Bereich der bildenden Künste entstehen (also z.B. Werke der Malerei, Bildhauerei oder Fotografie).“ Dabei unterscheidet er die Werke, die in der Kunsttherapie entstehen von Kunstwerken und nennt diese „Therapiekunstwerke“. Laut ihm ist diese Unterscheidung wichtig, da Werke aus der Kunsttherapie nicht die gleichen kulturellen und historischen Anforderungen wie traditionelle Kunstwerke haben (Schuster, 2014).

Die Kunsttherapie hat ihren Ursprung in der Beschäftigungstherapie, die bereits im 19. Jahrhundert in psychiatrischen Anstalten angeboten wurde (Menzen, 2023; Schuster, 2014). Während in der Beschäftigungstherapie jedoch nur künstlerische Tätigkeiten zur reinen Beschäftigung und ohne therapeutische Bearbeitung durchgeführt wurden, beruht die Kunsttherapie auf wissenschaftlich fundierten Methoden, bei der die entstandenen Werke gezielt in den therapeutischen Prozess mit einbezogen werden (Schuster, 2014).

Im Gegensatz zu vielen anderen Therapieformen ist die Kunsttherapie eine nonverbale Therapieform. Unbewusste innere Prozesse sollen durch die bildnerische Gestaltung und visuelle Kommunikation sichtbar gemacht werden (Obermeier, 2019). Ein zentrales Merkmal der Kunsttherapie ist die kunsttherapeutische Triade. Sie setzt sich aus der Beziehung zwischen Klient, Therapeut und künstlerischem Medium zusammen (Obermeier, 2019; Schuster, 2014). Das entstandene Werk dient dabei nicht nur zur äußeren Darstellung innerer Prozesse, sondern auch als Ausgangspunkt für die

Selbstreflexion. Durch die Betrachtung des eigenen Werkes kann der Klient emotionale Zusammenhänge erkennen, die ihm vorher nicht bewusst waren. In der Kunsttherapie geht es daher nicht um eine Interpretation des Bildes durch den Therapeuten, sondern um eine gemeinsame, offene Erkundung. Durch die Verbindung von Kreativität, Emotionen und der anschließenden verbalen Auseinandersetzung soll die Kunsttherapie innere Prozesse anregen und eine positive Wirkung auf den Klienten erzielen (Obermeier, 2019). Eine der positiven Auswirkungen ist die entspannungsfördernde Wirkung der Kunsttherapie. So konnte eine Senkung des Stresshormons Cortisol nachgewiesen werden (Kaimal et al., 2016).



Abbildung 1: Kunsttherapeutische Triade
Quelle: Obermeier, 2019, S. 5

Die bewusste Wahl der Materialien ist in der Kunsttherapie entscheidend. Das Material beeinflusst das künstlerische Arbeiten und sollte daher gezielt genutzt werden und von Klient zu Klient unterschiedlich sein. Ölmalkreiden ermöglichen vielseitige Techniken, erfordern aber einen kraftvollen und entschlossenen Einsatz, während Acryl- und Gouachefarben durch ihre Leuchtkraft und Flüssigkeit emotionale Prozesse verstärken können (Obermeier, 2019). Ton ist ein besonderes Material, das durch seine plastische Formbarkeit nicht nur kreative Freiheit bietet, sondern auch ein intensives haptisches Erlebnis ermöglicht (Schuster, 2014). Bei Materialien, die einen Papieruntergrund benötigen, spielt die Wahl des richtigen Papiers eine entscheidende Rolle. Eine zu hohe Blattgröße kann Klienten überfordern und Ölmalkreiden und Acryl- und Gouachefarben benötigen andere Stärken. Auch die Farbe des Papiers ist wichtig: Gedämpfte Weißfarben oder braunes Papier sollten überlegt werden, da reines weißes Papier verunsichern kann (Obermeier, 2019).

Neben analogen Materialien werden auch digitale Materialien wie Zeichen-Apps auf Laptops, Smartphones und Tablets erfolgreich in der Kunsttherapie eingesetzt (Darewych et al., 2015; Garner, 2016).

Wie in der Einleitung bereits kurz erwähnt, hat auch die Kunsttherapie in Virtual Reality schon gute Erfolge erzielen können (Shamri Zeevi, 2021; Ying-Chun & Chwen-Liang, 2018). Eine detailliertere Betrachtung eines konkreten Beispiels erfolgt in Kapitel 2.3.2.

2.1.2 Musiktherapie

Ähnlich wie die Kunsttherapie gehört auch die Musiktherapie den künstlerischen Therapieformen an (Schuster, 2014). Sie gilt als einer der ältesten Therapieformen und wurde bereits im 15. Jahrhundert angewendet (Kowal-Summek, 2016). Ihre jetzige Form etablierte sich jedoch erst nach dem Zweiten Weltkrieg, als Musik gezielt zur psychotherapeutischen Behandlung genutzt wurde (Bernatzky & Kreutz, 2015). Laut Kowal-Summek (2016) ist es aber trotz der langen Geschichte der Musiktherapie zumindest im deutschsprachigen Raum nicht einfach eine allgemeingültige Definition zu finden. Erst im Jahr 1994 haben sich mehrere musiktherapeutische Verbände zusammengeschlossen und in der Kasseler Konferenz musiktherapeutischer Vereinigungen in Deutschland versucht eine gemeinsame Beschreibung der Musiktherapie zu finden (Kowal-Summek, 2016). Die Deutsche Musiktherapeutische Gesellschaft (DMtG) verwendet auf ihrer Webseite eine Definition, die auf den Kasseler Thesen basiert:

- „Musiktherapie ist der gezielte Einsatz von Musik im Rahmen der therapeutischen Beziehung zur Wiederherstellung, Erhaltung und Förderung seelischer, körperlicher und geistiger Gesundheit.
- Musiktherapie ist eine praxisorientierte Wissenschaftsdisziplin, die in enger Wechselwirkung zu verschiedenen Wissenschaftsbereichen steht, insbesondere der Medizin, den Gesellschaftswissenschaften, der Psychologie, der Musikwissenschaft und der Pädagogik.
- Der Begriff „Musiktherapie“ ist eine summarische Bezeichnung für unterschiedliche musiktherapeutische Konzeptionen, die ihrem Wesen nach als psychotherapeutische zu charakterisieren sind, in Abgrenzung zu pharmakologischer und physikalischer Therapie.
- Musiktherapeutische Methoden folgen gleichberechtigt tiefenpsychologischen, verhaltenstherapeutisch-lerntheoretischen, systemischen, anthroposophischen und ganzheitlich-humanistischen Ansätzen“ (DMtG, o. J.; *Kasseler Thesen zur Musiktherapie*, o. J.).

Ein wesentliches Merkmal der Musiktherapie ist ihr nonverbaler Charakter. Klienten können durch Musik emotionale Inhalte ausdrücken, die verbal schwer zugänglich sind (Bruscia, 1998). Man unterscheidet dabei zwischen rezeptiver Musiktherapie und aktiver Musiktherapie.

Rezeptive Musiktherapie fokussiert sich auf das bewusste Hören von Musik, um Emotionen zu regulieren und eine positive Wirkung auf das psychische Wohlbefinden zu erzielen. Dazu sucht der Musiktherapeut gezielt Musikstücke aus, die er gemeinsam mit dem Klienten hört und anschließend bespricht (Bernatzky & Kreutz, 2015; Kowal-Summek, 2016).

Bei der aktiven Musiktherapie musiziert der Klient selbst und benutzt Töne, Rhythmen und Klänge, um sich auszudrücken. Dies kann durch Singen, Instrumentalspiel oder rhythmischem Klopfen am Körper geschehen (Bernatzky & Kreutz, 2015; Bruscia, 1998; Kowal-Summek, 2016)

Ein zentrales Element der aktiven Musiktherapie ist die Improvisation. Dabei entstehen Klänge, Rhythmen und Melodien ganz spontan – entweder allein oder mit dem Musiktherapeuten. Laut Bruscia

(1998) verfolgt die Improvisation keinen festen Plan, sondern entwickelt sich intuitiv. Der Klient nutzt verschiedene musikalische Mittel, um Klangstrukturen zu formen. Das erlaubt ihm, innere Zustände auszudrücken und Emotionen erlebbar zu machen. Musikalisches Vorwissen ist nicht vonnöten. Stattdessen wird die natürliche Fähigkeit des Menschen mit Klängen zu spielen genutzt. Der Therapeut gibt lediglich eine kleine Einführung und begleitet den Klienten durch den musikalischen Schaffungsprozess. Die entstehende Musik reflektiert oft unbewusste emotionale Prozesse und kann dem Klienten helfen, diese wahrzunehmen und zu verarbeiten (Bruscia, 1998).

Die Wirksamkeit der Musiktherapie wurde in zahlreichen Studien untersucht. Sie wird erfolgreich in der Behandlung von Angststörungen und Depressionen genutzt und trägt zudem bei der Reduktion von Stress und Erhöhung von Entspannung bei (Bradt et al., 2013; De Witte et al., 2020; de Witte et al., 2022; Gold et al., 2009).

Auch in der Musiktherapie wird VR zunehmend erforscht. Bisherige Studien konzentrierten sich dabei hauptsächlich auf die rezeptive Musiktherapie, bei der der Klient Musik in einer virtuellen Umgebung hört. Erste Untersuchungen zeigen, wie in der Einleitung erwähnt, positive Effekte auf die Psyche und auf die Entspannung (Alexanian, 2021; Brungardt et al., 2021). Weniger erforscht ist jedoch die Verbindung von aktiver Musiktherapie in VR. Ein konkretes Beispiel für eine Anwendung wird in Kapitel 2.3.3. näher vorgestellt.

2.2 Virtual Reality

Es wurde oft von Virtual Reality (VR) gesprochen, aber was genau wird unter VR verstanden?

Virtual Reality bezeichnet eine computergenerierte Umgebung, die die physikalische Realität durch künstliche Sinnesreize ersetzt. Der aus dem französischen stammende Begriff *virtuell* bedeutet dabei *scheinbar*. VR vermittelt also eine Realität, die nicht physisch existiert, aber als echt wahrgenommen wird (Brill, 2023). Im Gegensatz zu Medien wie Büchern, Filmen oder Spielen ermöglicht VR nicht nur das Vorstellen einer anderen Welt, sondern das aktive Erleben und Eintauchen in diese Welt (Jones et al., 2022). Wenn der Nutzer die VR als real akzeptiert, also die künstlichen Reize überzeugend genug sind, wird dies als *Immersion* bezeichnet. Dem Nutzer wird das Gefühl gegeben, als sei er präsent in der virtuellen Umgebung (Brill, 2023).

VR wird bereits in vielen Bereichen eingesetzt, die über Unterhaltung und Gaming hinausgehen. VR findet unter anderem auch Anwendung in Architektur und Medizin. Architekten können geplante Gebäude oder Produktionsprozesse bereits in einer virtuellen Umgebung visualisieren, bevor sie in die Realität umgesetzt werden. In der Medizin ermöglicht VR Chirurgen, Eingriffe virtuell zu simulieren und zu üben, bevor sie an echten Patienten durchgeführt werden (Isdale, 1993).

Um VR nutzen zu können und vollständig in die virtuelle Umgebung einzutauchen, wird in der Regel ein Head-Mounted Display (HMD) verwendet. Moderne HMDs nutzen die stereoskopische Darstellung. Dabei werden zwei leicht versetzte Bilder für das linke und rechte Auge berechnet. Das Gehirn setzt diese Bilder dann zu einer dreidimensionalen Szene zusammen und erzeugt den Eindruck von Tiefe (Isdale, 1993; Theis et al., 2016). Zusätzlich wird sogenanntes *Tracking* verwendet. Sensoren und Controller helfen dem HMD dabei Handbewegungen, Fortbewegung in der physikalischen Welt und die Blickrichtung zu verfolgen. Dieses Tracking muss in Echtzeit geschehen, da sonst das Gefühl der Immersion darunter leidet (Brill, 2023). Aus diesen Gründen besteht ein komplettes VR-System nicht nur aus der Hardware und der Software, sondern auch aus dem Nutzer und der physischen Umgebung, in der er sich befindet. Während das HMD künstliche Sinnesreize erzeugt, bleiben reale Einflüsse wie Bewegungsspielraum oder äußere Störfaktoren weiterhin bestehen (LaVelle, 2023).

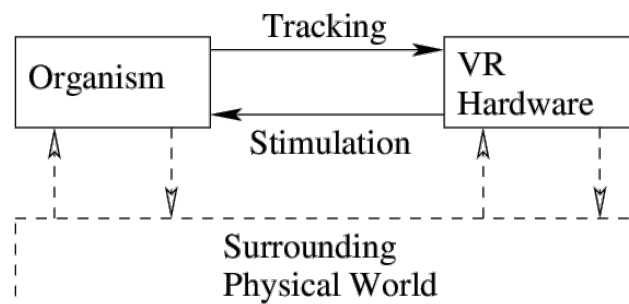


Abbildung 2: Komplettes VR-System nach LaVelle
Quelle: LaVelle, 2023, S. 38

Moderne Systeme berücksichtigen diese Aspekte zunehmend, indem sie Raumgrenzen erfassen oder den Wechsel zwischen virtueller und physischer Realität erleichtern (Wu et al., 2023).

Die Entwicklung von VR-Systemen ist keine technische Innovation der letzten paar Jahre, obwohl dies oft suggeriert wird (Brill, 2023). Bereits im Jahre 1965 beschreibt Ivan E. Sutherland ein Display, das er „Ultimate Display“ nennt. Dieses soll eine computergenerierte Umgebung zeigen, in der Nutzer vollständig in eine virtuelle Welt eintauchen können, in der Objekte real erscheinen (Sutherland, 1965). Diese Ideen legten den Grundstein für die spätere Entwicklung von VR-Technologien, einschließlich des ersten HMDs, welches von Sutherland einige Jahre später selbst entwickelt wurde (Sutherland, 1968). Der Begriff Virtual Reality fiel dabei jedoch nie. Erst fast 20 Jahre später verwendete Jerome Lanier den Begriff auf einer Podiumsdiskussion (Brill, 2023; Conn et al., 1989).

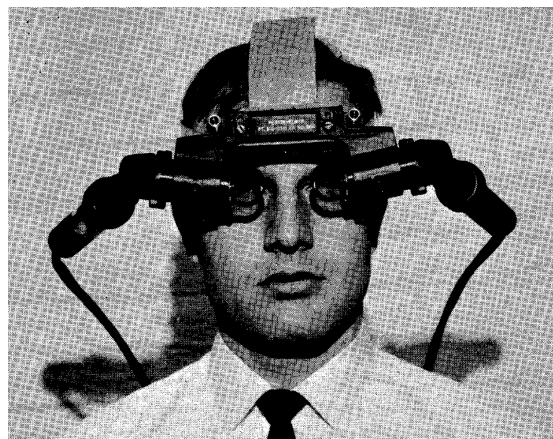


Abbildung 3: Das von Sutherland entwickelte HMD
Quelle: Sutherland, 1968, S. 759

2.2.1 Therapie in VR

Ein weiterer Anwendungsbereich für VR ist neuerdings auch die Psychotherapie. Durch immersive virtuelle Umgebungen können Patienten gezielt mit Angstsituationen konfrontiert werden, ohne realen Risiken ausgesetzt zu sein. Diese Form der Expositionstherapie wird insbesondere bei Angststörungen, Phobien und posttraumatischen Belastungsstörungen (PTBS) eingesetzt (Maples-Keller et al., 2017; Reger et al., 2011).

Ein Beispiel hierfür ist die von Reger et al. (2011) durchgeführte Studie zur Behandlung von PTBS bei aktiven Soldaten, die nach Einsätzen in Irak und Afghanistan an Folgen von Traumata litten. In einer psychiatrischen Klinik wurde Virtual Reality Exposure Therapy (VRE) angewandt, um die Klienten mit ihren traumatischen Erlebnissen zu konfrontieren. Es wurden individualisierte virtuelle Umgebungen verwendet, die das erlebte Kriegsszenario möglichst realitätsnah nachbilden. Die Studie zeigte positive Ergebnisse. 62% der Klienten berichteten eine zuverlässige Besserung ihrer Symptome (Reger et al., 2011). Neben PTBS wird VRE auch erfolgreich bei anderen psychischen Störungen wie sozialen Ängsten oder spezifischen Phobien eingesetzt. Maples-Keller et al. (2017) zeigen auf, dass VR-basierte Therapien praktikabel in der Umsetzung sind, da im Gegensatz zu traditionellen Expositionstherapien

reale Konfrontationen nur schwer oder gar nicht umsetzbar sind. Beispiele dafür sind Flug- oder Höhenangst (Maples-Keller et al., 2017).

Neben der VRE findet VR auch Anwendung in der Stressbewältigung. Virtuelle Umgebungen, insbesondere Natur- und Waldszenarien können positive Effekte auf das Entspannungsgefühl haben (Syed Abdullah et al., 2021).

2.3 Vergleichbare Arbeiten

Im Folgenden werden vergleichbare Arbeiten vorgestellt, die entweder als Inspiration für das Konzept des VR-Prototypen dienen oder kunst- oder musiktherapeutische Ansätze in VR nutzen.

2.3.1 Blob Opera

Blob Opera ist eine der Hauptinspirationen für den entstehenden Prototypen. Es ist ein interaktives Machine-Learning-Experiment von David Li, welches er in Zusammenarbeit mit Google Arts and Culture entwickelt hat (*Blob Opera*, o. J.). In *Blob Opera* können Nutzer ohne musikalisches Vorwissen vier virtuelle Opernstimmen in Echtzeit steuern, indem sie animierte Blobs hoch- und runterziehen. Dabei können auch mehrere Stimmen ausgewählt werden, die dann gemeinsam harmonisieren. Das System basiert auf KI-Model, das mit Stimmen von professionellen Opernsängern trainiert wurde. Die aufgenommenen Stimmdateien werden dabei nicht direkt wiedergegeben, sondern haben dem Modell geholfen, eine eigene Interpretation von Operngesang zu entwickeln (Li, 2020).

Dieses Konzept zeigt das Potenzial von KI-gestützten, interaktiven Klangumgebungen, die nicht nur unterhaltsam sein können, sondern auch therapeutisch genutzt werden können. Im Bereich der aktiven Musiktherapie könnte ein ähnliches Konzept wie *Blob Opera* genutzt werden, um Klienten eine intuitive und spielerische Möglichkeit geben, mit Musik zu improvisieren.



Abbildung 4: Blob Opera
Quelle: (*Blob Opera*, o. J.)

2.3.2 Tilt Brush

Eine weitere große Inspiration für den Prototypen ist *Tilt Brush*. Die ebenfalls von Google entwickelte VR-Anwendung ermöglicht es Nutzern, dreidimensionale Zeichnungen in einer virtuellen Umgebung zu erstellen („Tilt Brush“, o. J.). Im Gegensatz zu *BlobOpera* wird der Einsatz von *Tilt Brush* im kunsttherapeutischen Umfeld getestet. Studien zeigen, dass VR-Kunsttherapie mit *Tilt Brush* jungen Erwachsenen helfen kann, Stress zu bewältigen und Emotionen auszudrücken. Die Immersion in der kreativen Umgebung eröffnet neue Möglichkeiten der emotionalen Verarbeitung (Ying-Chun & Chwen-Liang, 2018). Auch Shamri Zeevi (2021) zeigt auf, dass VR-Kunsttherapie speziell bei Jugendlichen mit Angststörungen oder sozialen Schwierigkeiten wirksam sein kann (Shamri Zeevi, 2021).

Das Konzept von *Tilt Brush* ist also nicht nur als kreatives Medium nutzbar, sondern hat auch wertvolle therapeutische Anwendungsmöglichkeiten.



Abbildung 5: Tilt Brush in aktiver Nutzung, aufgenommen in mehreren Winkeln
Quelle: Shamri Zeevi, 2021, S. 4

2.3.3 Virtual Harmony

Die meisten wissenschaftlichen Arbeiten, die sich mit Musiktherapie in VR auseinandersetzen, behandeln die rezeptive Musiktherapie. Zahlreiche Studien zeigen, dass das Hören von Musik in einer virtuellen Umgebung das Entspannungsgefühl steigert und helfen kann Stress zu reduzieren (Alexanian, 2021; Brungardt et al., 2021; Cabezas et al., 2022; Zheng et al., 2024).

Virtual Harmony hat neben einem rezeptiven Modus auch einen aktiven Modus integriert. Nutzer können in einer immersiven virtuellen Umgebung auf virtuellen Instrumenten wie Klavier, Xylophon und Schlagzeug spielen. Die Studie zeigt, dass Nutzer sich nach der Nutzung deutlich entspannter fühlen als vor der Nutzung (Lin et al., 2020).

Dieses Konzept zeigt das Potenzial von VR in der aktiven Musiktherapie, indem nicht nur eine passive Soundkulisse, sondern aktive und kreative Mitgestaltung geboten wird.

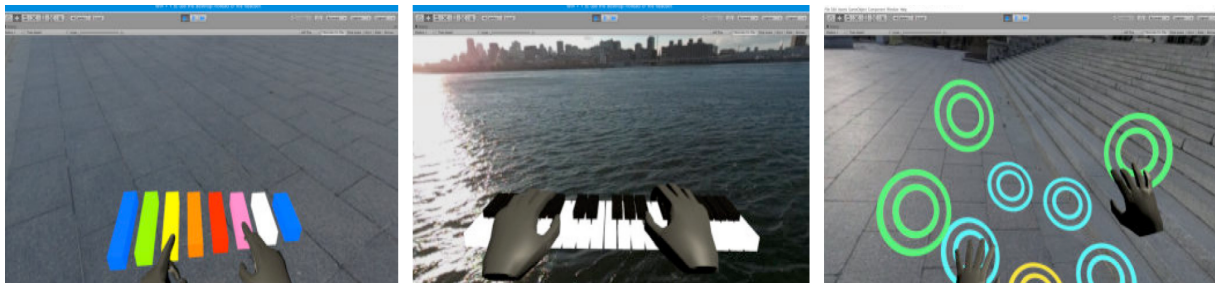


Abbildung 6: Xylophon, Klavier und Schlagzeug in Virtual Harmony
Quelle: Lin et al., 2020, S. 0950

2.4 Künstliche Intelligenz

Im folgenden Kapitel werden Grundlagen der Künstlichen Intelligenz vorgestellt und darauf eingegangen, wie es durch den Einsatz von KI möglich geworden ist, synthetische Stimmen mit hoher Natürlichkeit und Ausdrucksstärke zu erzeugen.

2.4.1 Was ist Künstliche Intelligenz?

Künstliche Intelligenz ist ein Fachbereich der Informatik, der sich mit der Entwicklung intelligenter Systeme befasst. Ziel ist es, Maschinen oder Programme zu entwickeln, die sowohl das Verhalten als auch die Intelligenz des Menschen nachahmen sollen (Strümke & Dalen, 2024).

KI-Systeme sollen also eigenständig in der Lage sein, komplexe Probleme zu bearbeiten (Wittpahl, 2019). Um dies zu erreichen, kommen unter anderem maschinelles Lernen (ML) und tiefe neuronale Netzwerke (DNN) zum Einsatz. Diese Technologien werden bereits bei Sprach- und Bilderkennung, Text- und Bildgenerierung, autonomen Fahren und auch bei der Sprachsynthese verwendet (Hono, Hashimoto, et al., 2021; Strümke & Dalen, 2024).

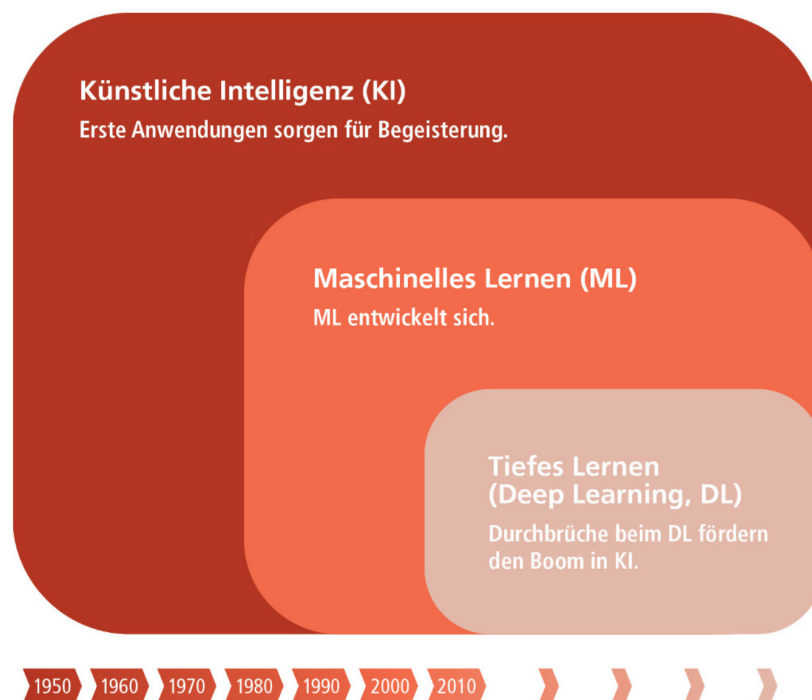


Abbildung 7: Methoden der KI
Quelle: Wittpahl, 2019, S. 22

2.4.2 Maschinelles Lernen

ML ist ein zentraler Teilbereich der KI. Der Name deutet bereits darauf hin, dass es sich um lernende Systeme handelt. Im Gegensatz zu herkömmlichen Programmen, die feste Regeln befolgen, entwickeln

sich ML-Modelle durch Interaktion mit ihrer Umgebung weiter und verbessern kontinuierlich ihre Fähigkeiten (B. Heinrichs et al., 2022). ML-Modelle befolgen ihre Entscheidungsregeln nicht starr, sondern passen sie durch kontinuierliche Rückkoppelung an (Wittpahl, 2019).

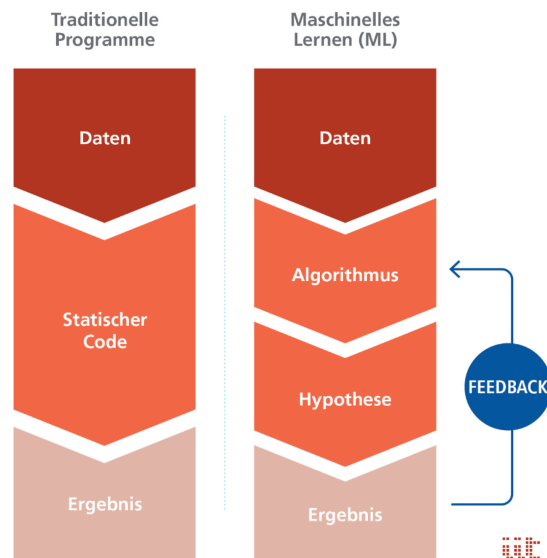


Abbildung 8: Unterschied von Maschinellern Lernen zu traditionellen Programmen
Quelle: Wittpahl, 2019, S. 25

ML lässt sich in drei Grundtypen unterteilen: überwachtes Lernen (Supervised Learning), unüberwachtes Lernen (Unsupervised Learning) und verstärktes Lernen (Reinforcement Learning).

Beim überwachten Lernen erfolgt das Trainieren des Systems mit einer großen Menge klassifizierter Daten. Dabei sind die richtigen Antworten bereits bekannt, sodass das Modell lernen kann, neue Daten richtig einzuordnen. Ein Beispiel ist die Bilderkennung: Das KI-System erhält Bilder, die als Katzen klassifiziert sind. Anschließend kann ein zuverlässig trainiertes System eigenständig neue Bilder als Katzenbilder oder keine Katzenbilder identifizieren (B. Heinrichs et al., 2022; Wittpahl, 2019).

Das unüberwachte Lernen verzichtet auf klassifizierte Daten. Stattdessen werden große Datenmengen analysiert und nach Mustern in Gruppen eingeteilt. Beispielsweise könnte eine KI verschiedene Tierbilder in Gruppen einteilen, ohne vorher zu wissen, welches davon Katzen oder Hunde sind (B. Heinrichs et al., 2022; Wittpahl, 2019).

Beim verstärkten Lernen verbessert sich das System durch Belohnungen. Es probiert in einer bestimmten Umgebung verschiedene Aktionen aus und erhält für erfolgreiche Entscheidungen eine positive Rückmeldung. Durch wiederholtes Testen lernt es, die besten Strategien zu entwickeln. Typische Anwendungsfälle sind Spiele, in denen KI durch wiederholte Versuche die optimale Spielstrategien erlernt (B. Heinrichs et al., 2022; Wittpahl, 2019).

2.4.3 Tiefes Lernen

Während einfache Machine-Learning-Algorithmen viele Probleme lösen können, haben sie Probleme mit Daten, die eine hohe Dimensionalität besitzen. Die erhöhte Anzahl an Merkmalen lässt die Menge der Konfigurationsmöglichkeiten exponentiell steigen. Das macht das Erlernen der Muster erheblich schwerer und den Berechnungsaufwand deutlich höher. Dieses als Fluch der Dimensionalität bezeichnete Phänomen ist einer der Gründe für die Entwicklung des Tiefen Lernens bzw. Deep Learning (DL) (Goodfellow et al., 2018).

In DL-Ansätzen werden neuronale Netzwerke eingesetzt, um komplexe Muster und Strukturen in den Daten erkennen zu können. Was neuronale Netzwerke sind, wird im nächsten Abschnitt erläutert. Dabei bezeichnet das Wort *tief* im Tiefen Lernen die Anzahl der Schichten im neuronalen Netzwerk. Bei einem Netzwerk mit mehr als drei Schichten, spricht man von DL oder einem Deep Neural Network (DNN) (IBMNN, o. J.).

2.4.4 Neuronale Netzwerke

Spricht man in der KI von neuronalen Netzwerken sind damit Künstliche Neuronale Netzwerke (KNN) gemeint. Diese sind mathematische Modelle, die von biologischen Nervensystemen inspiriert sind, aber nicht dieselbe Funktionsweise besitzen. Stattdessen setzen sie Konzepte wie Gewichtung und Schwellenwerte um, um Eingaben in relevante Ausgaben zu transformieren (Wittpahl, 2019).

Ein neuronales Netzwerk besteht aus einer Eingabeschicht, einer oder mehreren versteckten Schichten (Hidden Layers) und einer Ausgabeschicht. Informationen fließen dabei von Schicht zu Schicht, wobei jedes Neuron die Signale der vorherigen Schicht gewichtet und zur nächsten Schicht weiterleitet. Die einfachste Struktur stellt das Feedforward-Netz (FFNN) dar. Bei diesem fließen Informationen ausschließlich in eine Richtung, ohne Rückkopplungen oder Rekursionen. Die Eingabeschicht nimmt die Daten an, die bearbeitet werden sollen. Eingaben werden mit einem Gewicht multipliziert, welches während des Trainings nach und nach angepasst wird. Je höher das Gewicht, desto höher der Einfluss auf das Ergebnis. In der verborgenen Schicht werden die Daten verarbeitet. Die ausgeführten Berechnungen sind nicht für den Nutzer einsehbar, weshalb diese Schicht so genannt wird. Schließlich gibt die Ausgabeschicht das finale Ergebnis aus, wie eine Klassifikation oder eine numerische Vorhersage (Aggarwal, 2018; Goodfellow et al., 2018; Wittpahl, 2019).

Wird nicht nur eine Klassifikation oder eine numerische Vorhersage benötigt, sondern mehrere Vorhersagen oder eine Wahrscheinlichkeitsverteilung möglicher Ergebnisse, können sogenannte Mixture Density Networks (MDNs) verwendet werden. Diese kombinieren ein FFNN mit einem Mixture Density Model, um komplexe mehrdeutige Beziehungen zwischen Eingabe- und Ausgabewerten abzubilden (Bishop, 1994).

Fließen Informationen nicht nur in eine Richtung, sondern gleichzeitig nach vorne und teils nach hinten, spricht man von Recurrent Neural Networks (RNN). RNNs sind aufgebaut wie FFNNs, nutzen jedoch eine Rückkoppelung, um Informationen aus vorherigen Schritten zu speichern und in zukünftige Berechnungen einzubeziehen (Wittpahl, 2019).

Eine spezielle Form von RNNs sind Long Short-Term-Memory (LSTM)-Netzwerke. Anders als RNNs, die oft Schwierigkeiten haben, über längere Sequenzen hinweg relevante Informationen zu behalten, verwenden LSTMs spezielle Speicherzellen, um Informationen über längere Zeit zu speichern. Diese Speicherzellen bestehen aus drei Komponenten:

- Ein Eingabe-Gate, das bestimmt, welche neuen Informationen in die Speicherzelle aufgenommen werden
- Ein Forget-Gate, das bestimmt, welche Informationen beibehalten oder vergessen werden sollen
- Ein Ausgabe-Gate, das bestimmt, welche Informationen weitergegeben werden sollen (*MathworksLSTM*, o. J.)

Dadurch sind LSTMs in der Lage, sowohl kurzfristige als auch langfristige Abhängigkeiten in Daten zu erkennen, was sie beispielsweise besonders effektiv in der Verarbeitung von Sprache macht (*MathworksLSTM*, o. J.).

Zusätzlich gibt es noch Bidirektionale LSTMs (BiLSTMs), die Informationen aus der gesamten Zeitreihe beziehen. BiLSTMs bestehen aus einem Vorwärts-LSTM, welches Informationen vom Anfang der Sequenz bis zum Ende verarbeitet und einem Rückwärts-LSTM, welches die Informationen vom Ende bis zum Anfang der Sequenz verarbeitet (*MathworksLSTM*, o. J.).

Convolutional Neural Networks (CNNs) sind neuronale Netzwerke, die speziell für die Verarbeitung von Daten mit einer rasterartigen Struktur, wie z.B. Bildern oder Zeitreihen, entwickelt wurden. Im Convolution Layer nutzen CNNs Faltung (Convolution), um Merkmale in den Eingabedaten zu erkennen und zu filtern. Eine in vielen CNNs verwendete Pooling-Schicht verdichtet diese Daten dann und reduziert irrelevante Informationen. Die Ausgabeschicht verarbeitet dann zum Schluss die Ergebnisse der Schichten (Goodfellow et al., 2018).

Es gibt noch weitere Formen von KNNs. Diese sind jedoch nicht für diese Arbeit relevant und werden daher nicht weiter erläutert.

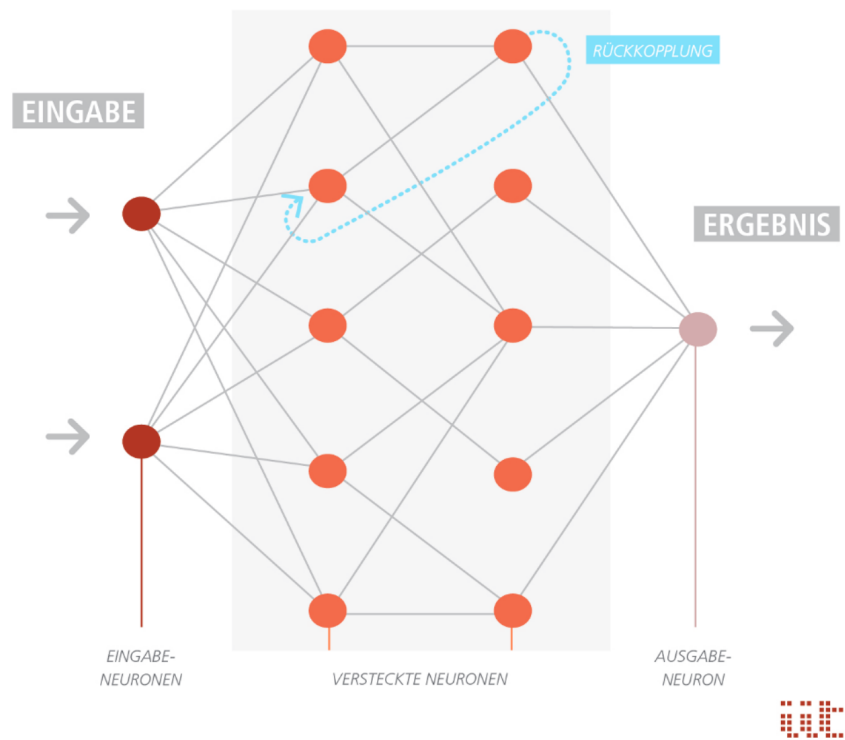


Abbildung 9: Vereinfachte Darstellung eines KNNs mit Rückkoppelung
 Quelle: Wittpahl, 2019, S. 30

2.4.5 Singing Voice Synthesis

Eine mögliche Anwendung von Deep-Learning-Methoden im Bereich der Sprachverarbeitung ist die Singing Voice Synthesis (SVS). SVS bezeichnet die computergestützte Erzeugung gesungener Stimmen auf Basis von musikalischen Partituren. Diese beinhalten Informationen wie Liedtext, Tempo und Tonhöhe (Lu et al., 2020). Damit eine synthetische Gesangsstimme natürlich, ausdrucksvoll und rhythmisch klingt, werden häufig eine Kombination aus diesen drei Modellen verwendet:

- Spektrummodell: Dieses Modell sagt spektrale Merkmale der Stimme vorher und ist entscheidend für klare Aussprache und hohe natürliche Klangqualität
- F0-Modell: Dieses bestimmt die komplexen Muster in der Grundfrequenz (F0) einer Gesangsstimme und ist entscheidend für Intonation und Ausdruck
- Dauer-Modell: Das Dauer-Modell erfasst rhythmische Faktoren beim Singen. Es steuert wie lang jede Silbe gesungen wird (Lu et al., 2020).

Frühe Methoden der SVS begannen mit der Replikation von Wellenformen (Waveform-Synthesizer). Gesang wurde dabei durch die mathematische Modellierung von Formanten – den resonanten Frequenzbereichen der Stimme - erzeugt. Dieser Ansatz erlaubte eine präzise Kontrolle über die Klangfarbe der Stimme (Benett & Rodet, 1989). Diese Methode der signalbasierten Modelle der Wellenform-Synthese modelliert das Sprachsignal durch Parameter wie Formantfrequenzen und spektralen Eigenschaften (Rodet, 2002).

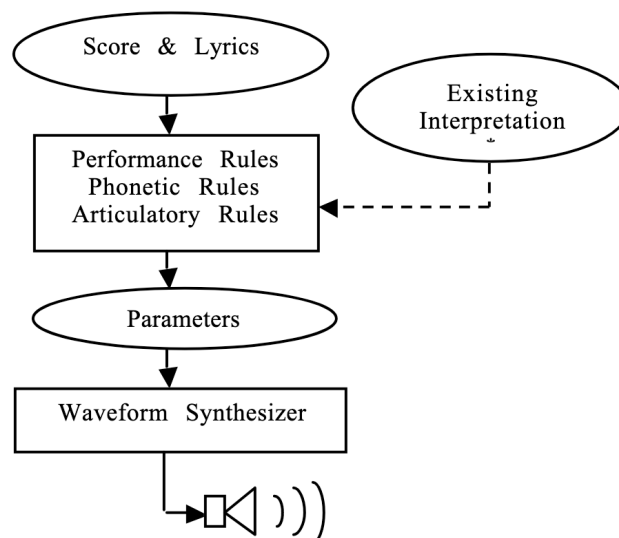


Abbildung 10: Vereinfachte Darstellung eines SVS mit Waveform-Synthesizer
Quelle: Rodet, 2002, S. 101

Mit der Zeit wurden konkatentative Verfahren eingeführt, die eine Datenbank aus vorab aufgenommen Sprach- oder Gesangselementen verwenden, um diese zu modifizieren und kombinieren. Passende

Einheiten (Units) werden basierend auf den zu singenden Phonemen ausgewählt. Anschließend werden diese in Eigenschaften wie Vibrato, Tonhöhe, Lautstärke und Klangfarbe angepasst (Rodet, 2002)

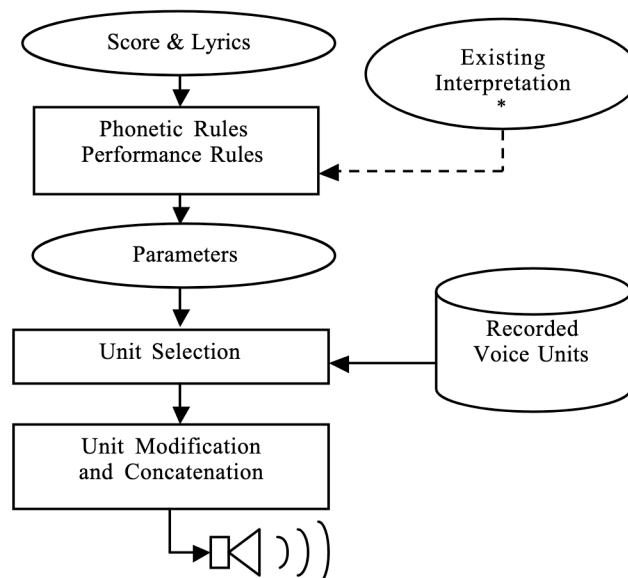


Abbildung 11: Vereinfachte Darstellung eines SVS mit konkatentivem Verfahren
Quelle: Rodet, 2002, S. 101

Ein System, das dieses Verfahren verwendet, ist das SVS *Lyricos* aus dem Jahre 1997. Es setzt dabei erstmalig nach Aussagen des Autors eine datengetriebene Methode zur Modellierung der phonetischen Merkmale der Stimme ein. Dadurch kann es die stimmlichen Charakteristiken eines echten Sängers nachbilden, um natürliche und musikalische Gesangsstimmen zu erzeugen (Macon, 1997).

Eines der ersten SVS-Systeme, das kommerziellen Erfolg feiern konnte, ist *VOCALOID*. Es erlaubt Nutzern, einfaches Generieren von Gesangsstimmen für die Nutzung in der Musikproduktion. *VOCALOID* nutzt dabei ebenfalls ein konkatentatives Verfahren zur Gesangssynthese, bietet Nutzern aber noch Werkzeuge, um die erzeugten Stimmen zu modifizieren (Kenmochi & Ohshita, 2007).

Konkatentative Ansätze haben jedoch den Nachteil, dass die Übergänge zwischen den einzelnen Einheiten nicht immer geschmeidig sind. Zudem sind diese Ansätze wenig flexibel und benötigen große Datenbanken zur Synthese von Gesangsstimmen. Um diese Nachteile zu verhindern, wurden statistisch-parametrische SVS-Ansätze wie Hidden-Markov-Model (HMM)-basierte Systeme entwickelt (Hono, Hashimoto, et al., 2021).

HMMs sind statistische Modelle, die in der Spracherkennung des Öfteren Anwendung finden. Sie eignen sich besonders gut, um Sequenzen zu analysieren und vorherzusagen. Sie bestehen aus einer Anzahl versteckter Zustände, die nicht direkt beobachtbar sind und beobachtbaren Ausgaben, die von diesen Zuständen abhängen. Die Übergänge zwischen den Zuständen erfolgen nach einer bestimmten Wahrscheinlichkeit und jede Sequenz von Zuständen erzeugt eine Abfolge von beobachtbaren

Ausgaben. Mithilfe von Algorithmen kann das wahrscheinlichste Zustandsmuster für eine gegebene Sequenz vorausgesagt werden (Rabiner, 1989).

Ein Beispiel für ein HMM-basiertes SVS-System ist *Sinsy*. *Sinsy* benutzt HMMs, um spektrale Merkmale, Grundfrequenz (F0) und Dauer von Gesangsstimmen zu modellieren. Dabei werden kontextabhängige HMMs aus einer Gesangsdatenbank trainiert. Um Gesangsstimmen zu generieren, werden diese Modelle mit einer musikalischen Partitur kombiniert. HMM-basierte Ansätze brauchen daher weniger Daten als konkatenative Methoden, da die Gesangswellenform direkt aus den HMMs generiert wird (Oura et al., 2010).

HMM-basierte Methoden können zwar die ungeschmeidigen Übergänge der konkatenativen Ansätze verhindern, neigen jedoch zum *Over-Smoothing*. Dadurch verliert die generierte Stimmung an Variationen, die typisch für einen menschlichen Gesang sind (Hono, Hashimoto, et al., 2021).

Daher forschte das Team von *Sinsy* an einem DNN-basierten Ansatz der SVS, welcher im nächsten Abschnitt näher vorgestellt wird.

2.4.6 DNN-basierte Gesangssynthese

Sinsy bietet in seiner neusten Version auch ein DNN-basiertes SVS-System an, welches Gesangsstimmen auf Basis von Musikpartituren generieren kann (Hono, Hashimoto, et al., 2021).

Es besteht im Wesentlichen aus vier Modulen:

- Time-Lag-Modell: Sagt vorher, wann die Note beginnt
- Dauer-Modell: Bestimmt die Länge einzelner Phoneme in einer Note
- Akustisches Modell: Generiert akustische Merkmale auf Basis des vorhergesagten Timings
- Neuraler Vocoder (PeriodNet): Wandelt die akustischen Merkmale in eine Wellenform um

Diese Modelle werden aus einer Datenbank von Gesangsdaten trainiert (Hono, Hashimoto, et al., 2021).

Zudem wird ein *Score Analyzer* benutzt, der eingegebene Musikpartituren in numerische Werte umwandelt. Diese können von neuronalen Netzwerken leichter verarbeitet werden. Die Musikpartituren müssen dabei im MusicXML-Format sein. (Hono, Hashimoto, et al., 2021).

Das Time-Lag-Modell lernt die Abweichung zwischen den idealen Notenstartzeiten (aus der Partitur) und den tatsächlich gesungen Startzeiten (aus Trainingsdaten). Dazu verwendet es ein mehrschichtiges FFNN. Dieser Time-Lag ist notwendig, da im natürlichen Gesang oft etwas vor oder nach der exakten Angabe begonnen wird. Leichte Verzögerungen oder Vorverlagerungen gehören zu den individuellen Gesangsstilen. Das Modell nimmt als Eingabe die Musikpartitur und gibt einen Zeitversatz aus (positiv = Verzögerung/Lag, negativ = früherer Einsatz) (Hono, Hashimoto, et al., 2021).

Das Dauer-Modell ist notwendig, da im Gesang einzelne Phoneme unterschiedlich lang gedehnt werden (z.B. lange Vokale bei bestimmten Noten). Mithilfe eines MDNs nutzt es die Informationen aus der

Musikpartitur in Kombination mit dem Time-Lag, um die Dauer jedes Phonems zu bestimmen (Hono, Hashimoto, et al., 2021).

Das Herzstück von Sinsy ist das Akustik-Modell. Es übersetzt die Timing-Informationen des Dauer- und des Time-Lag-Modells und Noteninformationen aus den Musikpartituren in akustische Parameter. Es besteht aus drei vollständig verbundenen FFNN-Schichten, die die Daten vorverarbeiten, drei CNN-Schichten, die Merkmale extrahieren und zwei BiLSTM-Schichten, um kontinuierliche Parameter zu erzeugen, die fließende Übergänge in Tonhöhe und Klangfarbe ermöglichen. Zuletzt gibt es eine Ausgabe-Schicht, die die vorhergesagten akustischen Merkmale in eine für den neuronalen Vocoder geeignete Form bringt (Hono, Hashimoto, et al., 2021).

Zusätzlich werden automatische Tonhöhen-Korrekturtechniken verwendet, falls die Tonhöhe von der musikalischen Partitur abweicht (Hono, Hashimoto, et al., 2021).

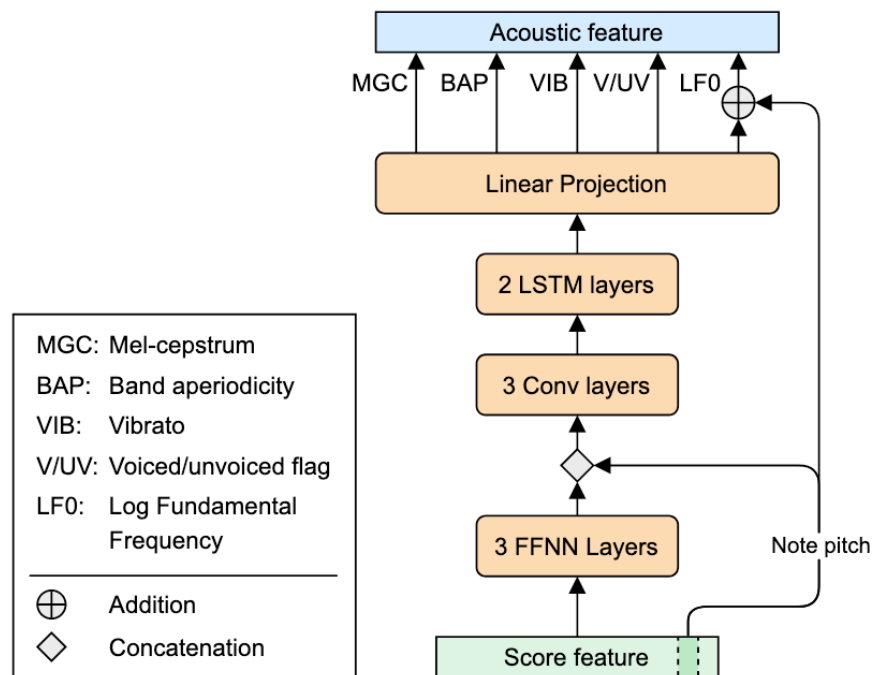


Abbildung 12: Aufbau des akustischen Modells von Sinsy
Quelle: Hono, Hashimoto, et al., 2021, S. 6

Das vierte Modul, der neurale Vocoder, wandelt die akustischen Merkmale in eine hörbare Gesangsstimme. *Sinsy* nutzt dafür den Vocoder PeriodNet (Hono, Takaki, et al., 2021). Dieser zeichnet sich aus durch eine präzise Reproduktion von Tonhöhe und Atemgeräuschen aus (Hono, Hashimoto, et al., 2021).

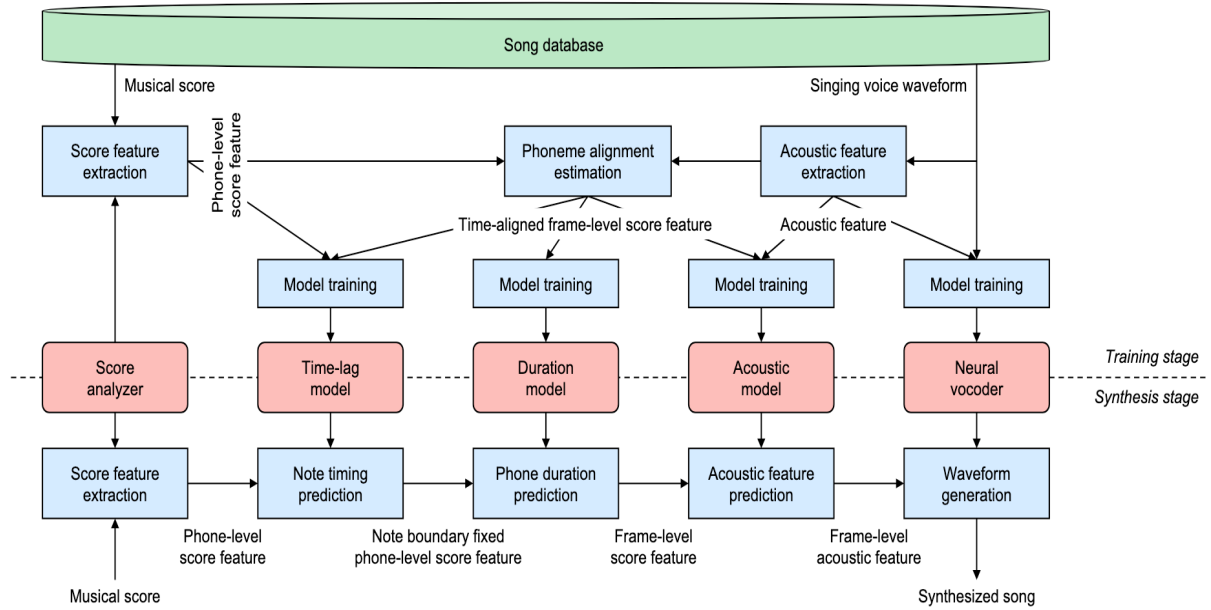


Abbildung 13: Komplettes Sinsy-System
 Quelle: Hono, Hashimoto, et al., 2021, S.3

2.5 Weitere Grundlagen

Es gibt noch weitere Grundlagen, die für das Verständnis relevant werden. Diese werden in den folgenden Unterkapiteln kurz behandelt.

2.5.1 Grundlagen der Farblehre

Farben sind ein Bestandteil unserer visuellen Wahrnehmung und bestimmen unsere Stimmung und unser Verhalten (Wright, 2018).

Man kann Farben anhand von drei Merkmalen beschreiben:

- **Farbton (Hue):** Dies bezeichnet die spezifische Farbe, wie Rot, Blau oder Gelb. Diese Primärfarben dienen als Grundlage, aus denen durch Mischen die Sekundärfarben Grün, Orange und Violett entstehen. Die weitere Zwischenstufe, die Tertiärfarben ergeben sich durch Mischen der Primär- und Sekundärfarben.
- **Helligkeit (Value):** Dies beschreibt, wie hell oder dunkel eine Farbe ist. Durch Hinzufügen von Weiß oder Schwarz wird eine Farbe heller oder dunkler.
- **Sättigung (Intensity):** Dies gibt an, wie lebendig oder gedämpft eine Farbe erscheint. Eine hohe Sättigung bedeutet eine kräftige, leuchtende Farbe und eine niedrige Sättigung eine gedämpfte, verwaschene (Wright, 2018)

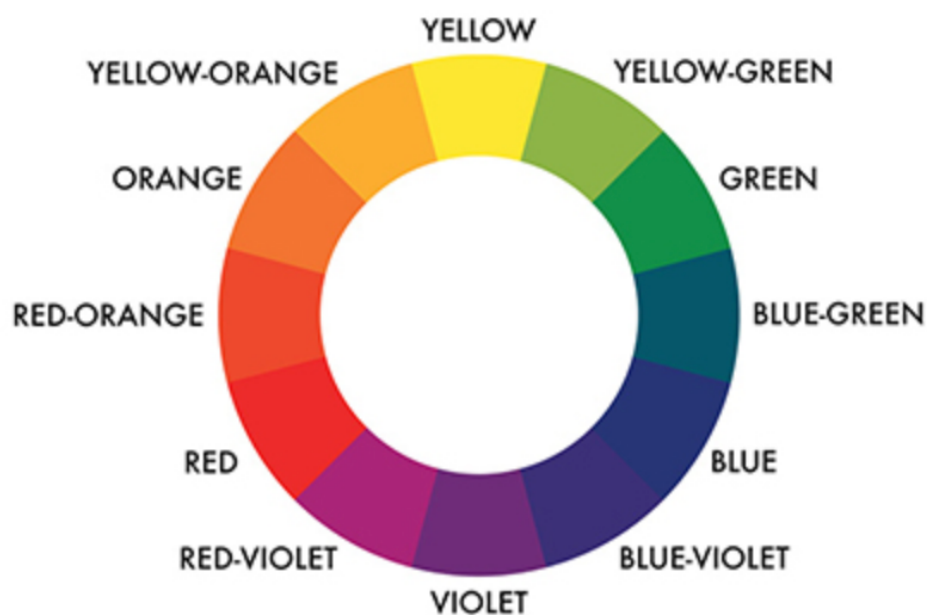


Abbildung 14: Beispielhafter Farbkreis mit Primärfarben (Rot, Blau, Gelb) und den Zwischenstufen
Quelle: Wright, 2018

Besonders die Helligkeit einer Farbe und die Zwischenstufen im Farbkreis werden später im Konzept des Prototyps (Kapitel 3.3.4) eine Rolle spielen.

2.5.2 Grundlagen der Musiktheorie

Es gibt einige Grundlagen der Musiktheorie, die für das Konzept des Prototyps wichtig sind, diese werden im Folgenden erläutert. Eine detaillierte Erklärung dieser Grundlagen ist für das Verständnis jedoch nicht notwendig, daher werden einzelne Themen nur kurz und knapp vorgestellt.

Tonleitern: Werden Töne geordnet hintereinander gereiht, dann spricht man von einer Tonleiter. Töne aus einer jeweiligen Tonleiter klingen immer harmonisch miteinander. Die Abstände zwischen einzelnen Tönen werden Intervalle genannt (Moser, 1968).



Abbildung 15: C-Dur Tonleiter
Quelle: Eigene Darstellung

Unterschiedliche Intervallabfolgen innerhalb einer Tonleiter geben das Tongeschlecht an (Moser, 1968).

In der Musik unterscheidet man zwischen zwei Tongeschlechtern: Dur und Moll. Während Dur eher als fröhlich und optimistisch gilt, hat Moll eher traurigere und pessimistische Charakteristika (Moser, 1968).

Spielt man mindestens drei Töne gleichzeitig, spricht man von Akkorden. Wie genau Akkorde aufgebaut werden, wird in dieser Arbeit nicht behandelt. Die wichtigsten Akkorde in der Musiktheorie sind:

- Tonika oder Grundakkord (1. Stufe der Tonleiter)
- Subdominante (4. Stufe der Tonleiter)
- Dominante (5. Stufe der Tonleiter)
- Tonika-Parallele (6. Stufe der Tonleiter) (Moser, 1968)



Abbildung 16: Tonika-Dreiklang, Subdominanten-Dreiklang, Dominanten-Dreiklang und Parallele in C-Dur
Quelle: Eigene Darstellung

Diese vier Akkorde bilden das harmonische Fundament vieler Musikstücke. Die meisten Musikstücke der Neuzeit basieren fast ausschließlich aus diesen Akkorden (Shenoy & Wang, 2005).

Abschließend muss noch der vierstimmige Satz erläutert werden. Chöre teilen Musiker in ihre Stimmlagen: Sopran, Alt, Tenor und Bass auf. Dadurch ist es Komponisten möglich eine hohe Reichweite an Tönen abzudecken. Während Sopran-Stimmen die höchsten Töne übernehmen, singen Bass-Stimmen die tiefsten. Alt und Tenor decken die Mitte ab. So kann jeder beliebige der oben beschriebenen Dreiklangs-Akkorde gesungen werden, wobei der Grundton dabei gedoppelt wird (*Der vierstimmige Satz*, o. J.).

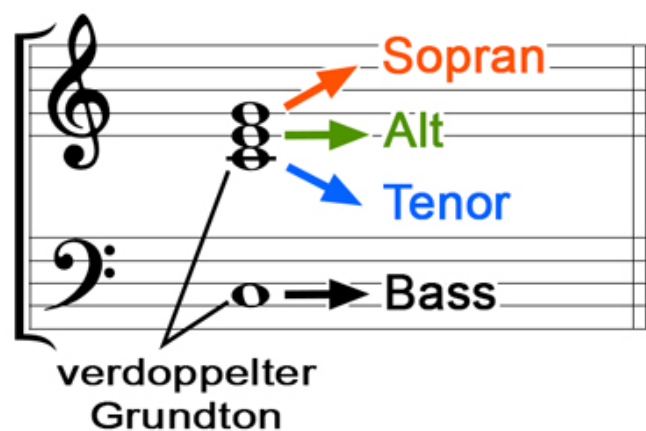


Abbildung 17: Der vierstimmige Satz am Beispiel eines C-Dur Akkords.
Quelle: (*Der vierstimmige Satz*, o.J.)

Die vier Stimmlagen haben meist definierte und abgegrenzte Tonbereiche (Scirea & Brown, 2015). Die folgende Abbildung zeigt die typischen Tonumfänge der jeweiligen Stimmlagen. Der tiefere Ton markiert die untere, der höhere Ton die obere Grenze.

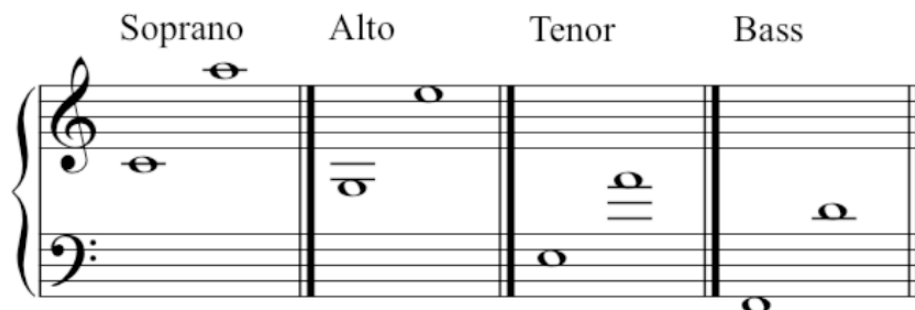


Abbildung 18: Stimmreichweiten der vier Stimmlagen Sopran, Alt, Tenor und Bass
Quelle: Scirea & Brown, 2015, S.2

3 Konzept

Dieses Kapitel stellt das grundlegende Konzept des entwickelten Prototyps vor. Im ersten Unterkapitel 3.1 wird das Kernkonzept dargestellt und in 3.2 werden Anforderungen erarbeitet, die der Prototyp erfüllen soll. Anschließend werden die notwendigen Entwicklungsumgebungen, um in VR zu programmieren, vorgestellt. Abschließend werden noch Anwendungen und Werkzeuge beschrieben, die für die Erstellung der Töne notwendig sind.

3.1 Kernkonzept

Das Kernkonzept des Prototyps basiert auf der Nutzung eines VR-Headsets, um in einer virtuellen Umgebung mithilfe der mitgelieferten Controller des HMDs zu zeichnen und gleichzeitig zu musizieren. Das Zeichenkonzept orientiert sich dabei an dem in Kapitel 2.3.2 beschriebenen Programm *Tilt Brush* („Tilt Brush“, o. J.). Nutzer nutzen den Controller wie einen Pinsel und können dadurch frei in einem dreidimensionalen Raum vor sich, hinter sich und über sich zeichnen. Dabei können sie sich frei im Raum bewegen.

Außerdem können Nutzer Farben über eine Palette auswählen. Während sie in der rechten Hand (bei rechtshändigen Nutzern) den Pinsel haben, haben sie in der linken Hand eine Farbpalette. Diese hat vier kleine Farbkleckse, die durch Berührung des Pinsels mit dem Farbkleck ausgewählt werden können. Die Farben können auch gemischt werden, indem der große fünfte Farbkleck ausgewählt wird. Dadurch gelangen Nutzer in den Mischmodus, in welchem sie die Farben nacheinander kombinieren können. Insgesamt sind dadurch sieben verschiedene Farben möglich.

Während die Nutzer malen, entstehen auch gesungene Töne. Diese sind genau wie im Konzept von *Blob Opera* KI-generiert und sind in vier Stimmlagen: Sopran, Alt, Tenor und Bass eingeteilt (*Blob Opera*, o. J.). Nutzer können durch die Auswahl der Farben gleichzeitig eine Stimmlage wählen, die gesungen wird, wobei jeder Farbe eine Stimmlage zugeordnet ist. Nutzer können ähnlich wie in *Blob Opera* durch Auf- und Abwärtsbewegungen die Tonhöhe beeinflussen. Das Malen nach oben oder unten erhöht oder verringert die Tonhöhe und das Malen in der horizontalen verändert die gesungene Silbe.

Sobald Nutzer in den Mischmodus gehen, werden nicht nur die einzelnen Farben gemischt, sondern auch die Stimmen. Je mehr Farben in den Mix dazukommen, desto mehr Stimmen erklingen auch gleichzeitig bis am Ende alle vier Stimmen gleichzeitig zu hören sind. Diese Stimmen sollen jedoch nicht alle den gleichen Ton singen, sondern singen Akkorde wie bei einem vierstimmigen Satz.

Der Prototyp ermöglicht es Nutzern, sowohl ihre künstlerische als auch ihre musikalische Ausdrucksfähigkeit zu entfalten. Da er das gleichzeitige Malen und Musizieren vereint, bietet das Konzept Potenzial für den Einsatz in der Kunst- und Musiktherapie. Inspiriert von den Anwendungen *Tilt Brush* und *Blob Opera* soll der entwickelte Prototyp den Namen *Brush Opera* erhalten.

3.2 Anforderungen

In diesem Kapitel werden die Anforderungen an den Prototyp, auf denen Hypothese 2 aus Kapitel Zielsetzung 1.3 basiert, aufgestellt. Diese Anforderungen haben das Ziel, einen benutzerfreundlichen Prototyp zu gestalten, der eine harmonische Verbindung zwischen Audio und Klang schafft.

- A1: Die KI-generierten Stimmen sollen die Nutzererfahrung nicht beeinträchtigen
- A2: Die gesungenen Akkorde im Misch-Modus sollen harmonisch klingen und passen
- A3: Es soll genug Möglichkeiten und Motivation geben, im dreidimensionalen Raum nach Belieben zu malen
- A4: Es soll motivieren, frei zu musizieren
- A5: Die Kombination aus Zeichnen und Musizieren soll eine stimmige Atmosphäre schaffen
- A6: Nutzer sollen sich die Anwendung als Alternative zur traditionellen Kunst- und Musiktherapie oder als Freizeitbeschäftigung vorstellen können
- A7: Die Steuerung soll intuitiv und nicht verwirrend sein
- A8: Die Interaktion soll ohne wahrnehmbare technische Verzögerungen oder Störungen erfolgen

In den folgenden Kapiteln wird unter anderem erläutert, wie versucht wird, diese Anforderungen zu erfüllen.

3.3 Design- und Bedienkonzept

Im Folgenden wird erläutert, welche Hardware für den Prototypen genutzt wird, wie die Nutzungserfahrung sein wird und auf das Design eingegangen.

3.3.1 Auswahl der Hardware

Als HMD der Wahl wird für den Prototypen eine Quest 3 von Meta verwendet. Laut Herstellerangaben ist es zum Zeitpunkt der Erstellung dieser Arbeit das leistungsstärkste HMD von Meta, sodass hardwarebedingte technische Einschränkungen unwahrscheinlich sind (*Meta Quest 3*, o. J.).

Zudem erhält man beim Erwerb einer Quest 3 auch zwei Controller, die jeweils vier Knöpfe und einen Joystick haben, wodurch es mehr als genug Möglichkeiten für Benutzereingaben gibt (*Meta Quest 3*, o. J.).

Die Quest 3 besitzt auch ein leicht verstellbares Band, wodurch es schnell und einfach ist, das HMD an verschiedene Kopfformen anzupassen (*Meta Quest 3*, o. J.).

Obwohl der Prototyp auf dieser Hardware entwickelt wurde, funktioniert er auch auf allen anderen gängige HMDs (s. Kapitel 3.4.1).

3.3.2 Nutzungserfahrung und Design

Nachdem die Software des Prototyps gestartet ist, sehen Nutzer auf dem Boden eine riesengroße graue Fläche und vor ihnen einen grauen Würfel. Über ihnen ist ein leerer Himmel und sie können einen Horizont erkennen.

Das minimalistisch gehaltene Design ermöglicht und ermutigt Nutzer, diese leere Welt mit Farbe zu füllen. Wenn ein Nutzer eine Wiese mit Blumen haben will, kann er diese zeichnen und wenn er am Himmel eine Sonne vermisst, kann er diese ebenfalls malerisch hinzufügen.

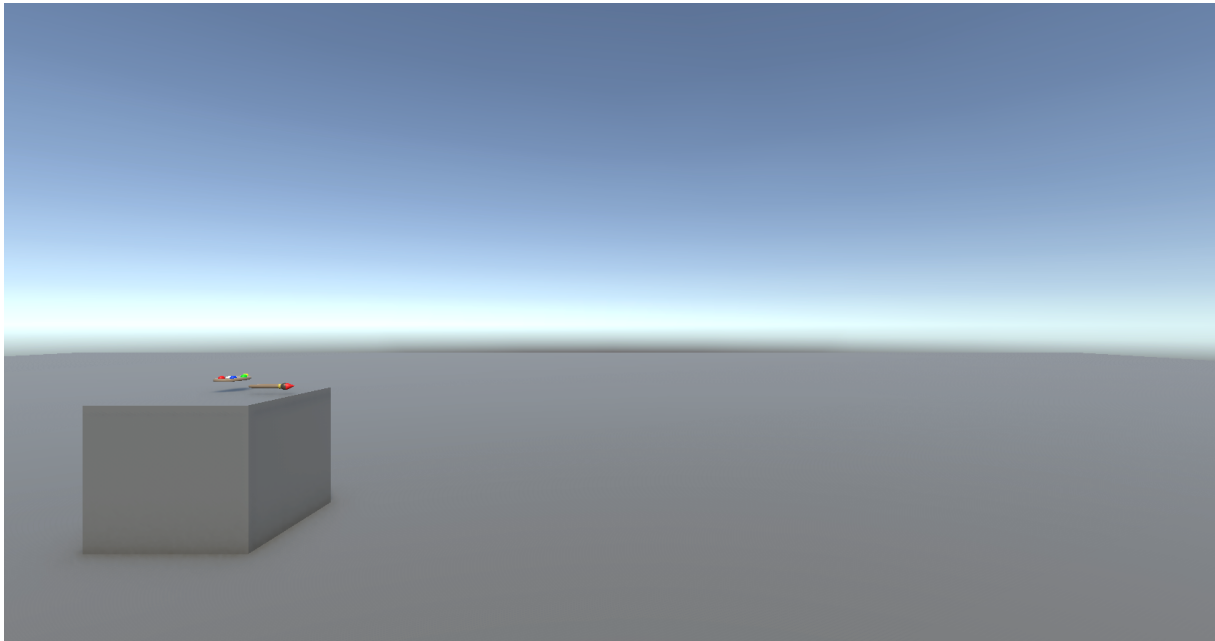


Abbildung 19: Startansicht: Das Erste, was Nutzer sehen, nachdem das Programm gestartet ist
Quelle: Eigene Darstellung

Es wird für den Prototypen und in dieser Studie auf erklärende Benutzeroberfläche verzichtet. Alle Informationen über die Verwendung des Prototyps erhalten Nutzer über eine kleine Einführung vor der Nutzung und kleinen Anmerkungen zum Start der Nutzung vom Autor dieser Arbeit, der auch gleichzeitig der Durchführende der Studie ist. Dabei soll die Steuerung intuitiv genug sein, dass sich jeder danach mit allen Steuerungselementen auskennt.

Um sich an die Bewegung im virtuellen Raum zu gewöhnen, können Nutzer zum grauen Würfel gehen. Dabei können sie sich entweder durch Bewegung im realen physikalischen Raum fortbewegen oder mit dem Joystick am linken Controller.

Am Kasten angekommen sehen sie vor sich zwei Objekte: Ein Pinsel und eine Palette. Beide Objekte sind ebenfalls sehr minimalistisch im Design, entsprechen aber der allgemeinen Vorstellung eines Pinsels und einer Farbpalette. Vor allem der Pinsel ist etwas größer, da die Leinwand (der endloserscheinende dreidimensionale Raum) auch viel größer ist als herkömmliche Leinwände.

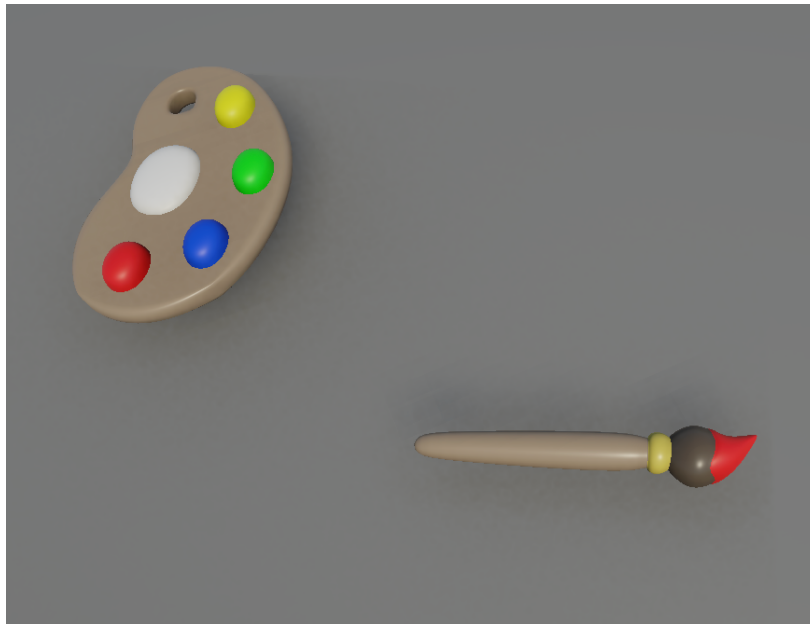


Abbildung 20: Pinsel und Palette von oben (Ansicht im Editor)

Quelle: Eigene Darstellung

Anschließend können Nutzer die Objekte mit der Greif-Taste aufnehmen. Der Pinsel beginnt bereits mit einer vorausgewählten Farbe, weshalb Nutzer direkt mit dem Malen beginnen können.

Um zu malen wird der Pinsel bzw. der rechten Controller in der Luft bewegt, wobei der A-Button gedrückt werden muss. Solange der Button gedrückt gehalten wird, wird gezeichnet, wodurch es ermöglicht wird, Linien und Formen im dreidimensionalen Raum zu zeichnen.

Durch Berühren einer der vier kleinen Farbkleckse auf der Palette, können Nutzer die Farbe des Pinsels und damit die Farbe des gemalten Striches ändern. Die Auswahl der Farben wird in Kapitel 3.3.4 näher erläutert.



Abbildung 21: Gezeichnetes Wort OPER mit Benutzung aller vier Farben der kleinen Farbkleckse

Quelle: Eigene Darstellung

Sobald gemalt wird, erklingen auch gesungene Stimmen, die mithilfe von SVS erstellt wurden (Näheres dazu in Kapitel 3.5.2). Die Höhe des gesungenen Tons kann beeinflusst werden, indem höher oder tiefer gemalt wird. Es erklingt immer derselbe Grundton, unabhängig davon, wo angefangen wird zu malen. Von diesem aus werden dann höhere Töne gespielt, wenn in der vertikalen Achse nach oben gezeichnet wird und tiefere Töne, wenn nach unten gezeichnet wird. Auf die Auswahl der Töne wird in Kapitel 3.3.3 näher eingegangen.

Während die vertikale Bewegung die Tonhöhe bestimmt, bestimmt die Zeichenbewegung in der Horizontalen die gesungene Silbe. Genau wie bei der Tonhöhe, wird zu Beginn immer die Silbe „Wah“ gesungen. Durch Bewegung nach Links oder Rechts beim Zeichnen, ändert sich die gesungene Silbe zu „Do“.

Durch Wechseln der Farbe kann zudem die Stimmlage verändert werden. Jeder Farbe ist eine Stimmlage zugeordnet. Die erste Farbe wählt dabei den Bass aus, die zweite Tenor, die dritte Alt und die vierte Sopran. Bei Bass und Tenor erklingen Männerstimmen und bei Alt und Sopran Frauenstimmen.

Durch Berühren des fünften großen weißen Farbkleckses gelangen Nutzer in den sogenannten Misch-Modus. Der Misch-Modus lässt Nutzer nach Betätigen des B-Buttons am rechten pinselführenden Controller Farben mischen und Stimmen dazuschalten. Der Misch-Modus beginnt mit der ersten Farbe und es erklingt nur der Bass. Nach erstmaligem Betätigen des B-Buttons mischen sich die ersten beiden Farben und es erklingen Bass und Tenor. Ein erneutes Drücken schaltet die Alt-Stimme dazu und die Farbe ändert sich erneut. Ein drittes Drücken auf den Misch-Button und es erklingen beim Zeichnen alle vier Stimmlagen und man zeichnet mit der letzten siebten Farbe. Die Mischung kann zurückgesetzt werden, indem - nach Erreichen der siebten Farbe – der Misch-Button gedrückt wird. Dann erklingt wieder nur der Bass und man zeichnet mit der ersten Farbe. Die Farbe des großen Farbkleckses passt sich dabei ebenfalls an.

Sobald mehrere Stimmen zu hören sind, singen diese nicht dieselben Töne, sondern harmonisieren miteinander und singen mehrstimmige Akkorde. Diese werden im nächsten Kapitel vorgestellt.



Abbildung 22: Gezeichnetes Wort Hamburg mit allen sieben Farben
Quelle: Eigene Darstellung

3.3.3 Auswahl der Töne

Damit eine generell fröhliche und harmonische Klangatmosphäre geschaffen wird, werden ausschließlich Töne einer Dur-Tonleiter gespielt. Die Wahl der zugrundeliegenden Tonart in *Brush Opera* fiel auf A-Dur, da es die Lieblings-Tonart des Autors dieser Arbeit ist. Grundsätzlich hätte es aber auch jede andere Dur-Tonart sein können, da diese nach Moser einen optimistischen Charakter haben (Moser, 1968).

Der Tonumfang, der durch Zeichnen spielbaren Töne, orientiert sich am klassischen Tonumfang der vier Stimm lagen und ist wie folgt (Scirea & Brown, 2015):

- Bass: Von E2 bis A3
- Tenor: Von A2 bis E4
- Alt: Von A3 bis C#5
- Sopran: Von E4 bis A5

Unabhängig von der Zeichenposition beginnt der erste gespielte Ton immer mit einem harmonisch passenden Grundton:

- Bass und Sopran starten auf A, dem Grundton der Tonart
- Tenor und Alt beginnen auf E, der 5. Stufe in A-Dur

Dies stellt sicher, dass zu Beginn immer ein Ton aus dem A-Dur-Dreiklang erklingt und der gesamte verfügbare Tonumfang effizient genutzt wird.

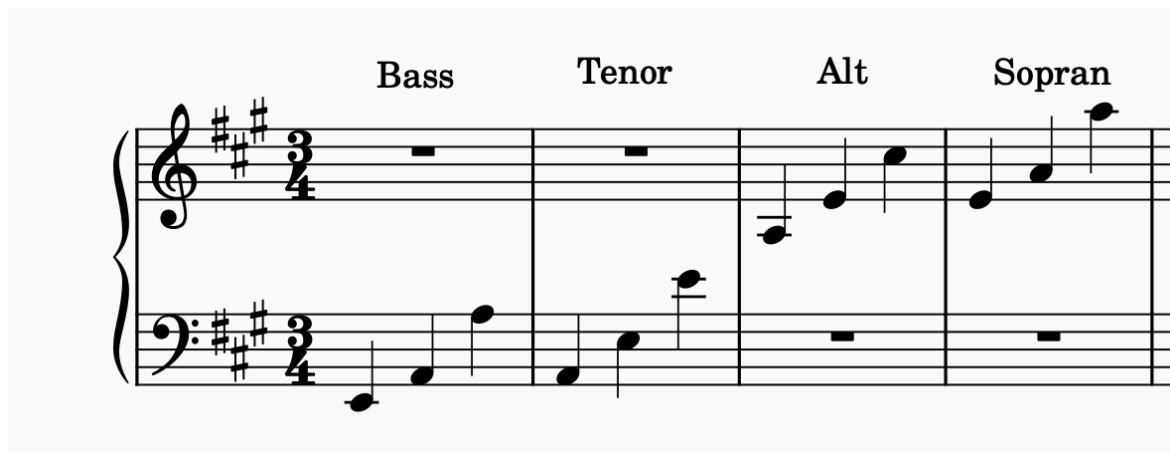


Abbildung 23: Tonumfang der vier Stimm lagen mit Startton in der Mitte
Quelle: Eigene Darstellung

Um im Misch-Modus harmonische Akkorde zu spielen, erklingen folgende Akkorde:

- Tonika: A-Dur (A, C#, E)
- Subdominante: D-Dur (D, F#, A)
- Dominante: E-Dur (E, G#, H)
- Tonika-Parallele F-Moll (F#, A, C#)

Wie bereits erwähnt, sind dies die Akkorde, die in den häufigsten Musikstücken vorkommen (Shenoy & Wang, 2005).

Die oberste Stimme gibt dabei immer die Melodie an und ist in der Tonhöhe beeinflussbar durch das Zeichnen der Nutzer. Die unteren Stimmen spielen dann abhängig von dem Melodieton der obersten Stimme einen Akkord, der den Melodieton verwendet. Dabei ist die Stimmenaufteilung ähnlich wie beim vierstimmigen Satz. Dadurch, dass nur Töne der A-Dur-Tonleiter verwendet werden, kann aus jedem Ton mindestens einer der vier oben genannten Akkorde gebildet werden.

Beispiel: Wenn im Misch-Modus alle vier Stimmen zu hören sind, ist die Tonhöhe des Soprans durch die Bewegungen beeinflussbar. Die unteren Stimmen harmonisieren mit dem Sopran. Spielt der Sopran beispielsweise ein A, dann erklingt A-Dur. Der Bass singt ein tiefes A, der Tenor ein E und der Alt ein C#.



Abbildung 24: Einige mögliche Akkord-Kombinationen, wenn alle vier Stimmen zu hören sind
Quelle: Eigene Darstellung

Manche Töne wie beispielsweise das C# sind in mehreren Akkorden (A-Dur und Fis-Moll) enthalten. In diesem Fall wird durch Zufall entschieden, welcher der Akkorde verwendet wird.

Ein Sonderfall tritt auf, wenn nur zwei Stimmen (Bass und Tenor) gespielt werden. Zwei Stimmen reichen nicht, um einen Akkord zu bilden, daher erklingt dann eine sogenannte „leere Quinte“. Ein Zweiklang, dem der mittlere Ton fehlt, um einen Akkord als Dur oder Moll zu klassifizieren (Burianek, 2021). In Abbildung 24 sind die drei möglichen leeren Quinten im Bass-Schlüssel zu sehen.

3.3.4 Auswahl der Farben

Um eine möglichst große Farbvielfalt zu ermöglichen, wurden als erste drei Farben die in den Grundlagen nach Wright (2018) beschriebenen Grundfarben Rot, Blau und Gelb gewählt (Wright, 2018). Da es vier Stimmlagen gibt, wird die Farbpalette um Grün ergänzt - eine Grundfarbe des RGB-Systems (*HTML RGB and RGBA Colors*, o. J.).

Diese vier Farben werden den Stimmlagen zugeordnet, wobei die Farbzugeordnungen von Personen mit Synästhesie, insbesondere Klang-Farb-Synästhesie, als Orientierung dienten. Studien zeigen, dass sowohl Menschen mit als auch ohne Klang-Farb-Synästhesie helle Farben mit hohen Tonlagen und dunklere Farben mit tiefen Tonlagen assoziieren (Riggs & Karwoski, 1934; Ward et al., 2006).

Basierend auf diesen Erkenntnissen ergibt sich folgende Farbzugeordnung für die Stimmlagen:

- Bass: ein dunkleres Blau
- Tenor: etwas dunkleres Rot
- Alt: Grün
- Sopran: ein helleres Gelb

Für die Farbmischungen wurden die Zwischenstufen zweier Farben aus dem Farbkreis nach Wright (2018) verwendet (Wright, 2018). Daraus ergeben sich diese Mischfarben:

- Bass und Tenor: Rot und Blau ergeben Violett
- Bass, Tenor und Alt: Grün und Blau ergeben Türkis
- Bass, Tenor und Alt: Gelb und Rot ergeben Orange

Es wurde dabei bewusst darauf verzichtet, exakte Farbmischungen zu nutzen. Stattdessen wurden nur grob die Zwischenstufen im Farbkreis gewählt, um eine größere Farbvielfalt zu gewährleisten und dennoch den Eindruck einer Mischung zu bewahren.

3.4 Unity als VR-Entwicklungsumgebung

Unreal Engine und Unity sind zwei der gängigsten Entwicklungsumgebungen, wenn es um die Entwicklung von sowohl Spielen als auch Anwendungen in VR geht (Brill, 2023; *Unity*, o. J.; *Unreal Engine*, o. J.).

Ohne die Vor- und Nachteile beider Umgebungen abzuwägen, wurde sich für Unity entschieden, da der Autor dieser Arbeit bereits Vorerfahrungen mit dieser Plattform hat. Es wird die zum Zeitpunkt der Erstellung des Prototyps aktuellste Version 6000.0.26f1 verwendet.

Unity hat sich in der Entwicklung von 3D-Anwendung bewährt und nutzt dafür ein linkshändiges Weltkoordinatensystem. Wie üblich in der Computergrafik zeigt die Y-Achse nach oben und die X-Achse (in der Vorderansicht) nach links. Die Z-Achse repräsentiert die Tiefe. Die Unity-API hat zudem eine Klasse `Vector3`, die die Einheitsvektoren für die Koordinatenachsen beinhaltet (Brill, 2023). Diese werden in der Umsetzung benötigt, um die Zeichenpositionen zu verfolgen.

Sehr hilfreich für die Entwicklung von VR-Anwendungen ist auch, dass Unity mit ISO-Einheiten arbeitet. Wenn ein Objekt im Editor eine Skalierung von 1 x 1 x 1 hat, dann entspricht dies den Maßen 1 Meter x 1 Meter x 1 Meter (Brill, 2023).

Beispielsweise hat der Würfel, der in Kapitel 3.3.2 beschrieben und abgebildet ist, genau diese Maße und hilft unter anderem bei der Einschätzung der Größenordnungen in der Szene.

In Unity gibt es zudem zahlreiche Plugins und Komponenten, die für die Umsetzung erforderlich sind:

- OpenXR für die Unterstützung verschiedener VR-Geräte
- XR Interaction Toolkit für Interaktionen in VR
- LineRenderer für das Zeichnen der Linien
- Audiokomponenten für das Abspielen der Töne
- Physikkomponenten für Pinsel und Palette

Diese werden in den folgenden Unterkapiteln näher vorgestellt.

3.4.1 OpenXR

Unity unterstützt diverse XR/VR-Plugins darunter OpenXR (*OpenXR*, o. J.). OpenXR ist ein kosten- und lizenzfreies Plugin, das die plattformübergreifende Entwicklung von VR- und AR-Anwendung in Unity erleichtert. VR-Anwendungen, die mithilfe dieses Plugins entwickelt werden, werden von diversen gängigen VR-Geräten wie HoloLens, Oculus und Meta unterstützt (*OpenXR*, o. J.).

Aufgrund dieser breiten Unterstützung und der Unabhängigkeit von spezifischer Hardware wurde OpenXR als Grundlage für die Entwicklung des Prototyps gewählt.

3.4.2 XR Interaction Toolkit

Das XR Interaction Toolkit (*UnityXRIT*, o. J.) stellt eine Sammlung an Komponenten bereit, die für ein Interaktionssystem in VR verwendet werden können. Es bietet dabei ein Basisset aus Interactors (Komponenten, die mit anderen Objekten interagieren, z.B. Controller) und Interactables (Komponenten, die auf Interaktionen reagieren, z.B. aufhebbare Objekte). Diese werden mithilfe eines Interaction Managers verwaltet und verbunden (*UnityXRIT*, o. J.).

In jedem Frame prüfen aktive Interactor, ob sich ein Objekt, mit dem sie interagieren können, in ihrer Reichweite befindet. Wird eines entdeckt, wechselt das Interactable in den sogenannten *Hover-State*. Es können dann Events ausgelöst werden, die Nutzern, die Möglichkeit der Interaktion signalisieren. Wenn ein solches Interactable z.B. durch den Druck eines Buttons ausgewählt wird, wechselt es in den *Selected-State* und weitere Ereignisse können ausgelöst werden. Ein *Grab Interactable* kann beispielsweise aufgenommen werden (*UnityXRIT*, o. J.).

Eine weitere Komponente ist XR Origin. Dieses verwaltet die Kameraposition für stationäre und raumbasierte VR-Erlebnisse. Sie beinhaltet die Main Camera-Komponente, die für das Rendern der VR-Ansicht zuständig ist und sich dynamisch an verschiedene Tracking-Methoden anpassen kann (*UnityXRIT*, o. J.).

Das Toolkit ermöglicht zudem verschiedene Bewegungsmöglichkeiten im virtuellen Raum. Neben der normalen Bewegung im Raum durch physisches Bewegen oder Controller-Eingaben, werden auch Teleportation und Klettern unterstützt (*UnityXRIT*, o. J.).

Unity liefert mit dem XR Interaction Toolkit auch eine schon fertige Demo-Szene, in der alle Interaktionsmöglichkeiten des Toolkits präsentiert werden und ausprobiert werden können.

Die Komponenten des XR Interaction Toolkits sind das Kernstück des Prototyps und sind für alle Interaktionen mit der VR-Umgebung zuständig.

3.4.3 Line Renderer

Unity erlaubt es Linien im dreidimensionalen Raum zu zeichnen. Dazu wird die Line Renderer-Komponente verwendet. Diese malt zwischen zwei oder mehr Punkten eine gerade Linie. Line Renderer unterstützt dabei verschiedene Materialien, Shader und Texturen (Unity Technologies, o. J.-a).

Dadurch kann Line Renderer verwendet werden, um verschiedenfarbige Linien zeichnen zu lassen.

3.4.4 Audiokomponenten

Um die Stimmen abzuspielen und hörbar zu machen, werden Unitys Audio Listener und Audio Source-Komponenten verwendet (Unity Technologies, o. J.-d).

Audio Sources sind an Objekte in der VR-Szene gekoppelt und emittieren von dort aus die Geräusche. Diese Geräusche werden dann vom Audio Listener aufgenommen und der Nutzer kann sie hören. Dabei ist es sogar möglich zu hören, von wo die Geräusche kommen. Der Audio Listener ist standardmäßig mit der Main Camera des XR Origins gekoppelt (Unity Technologies, o. J.-d). In *Brush Opera* können Audio Sources beispielsweise mit dem Pinsel gekoppelt werden.



Abbildung 25: Audiokomponenten: Ein Audio Listener und zwei Audio Sources
Quelle: Unity Technologies, o.J.-d

Unity ist in der Lage Audiodateien in Formaten wie AIFF, WAV und MP3 zu importieren. Dadurch können im Voraus aufgenommene Töne als abspielbare Audio Clips verwendet werden (Unity Technologies, o. J.-d). Mit welchen Technologien die Töne erstellt werden, wird im Kapitel 3.5 erläutert.

3.4.5 Physikkomponenten

In Unity können Objekte mit verschiedenen Komponenten versehen werden, um physikalisches Verhalten zu simulieren. Die Rigidbody-Komponente erlaubt es, einem Objekt eine Masse zu geben, sodass es durch die Physik-Engine beeinflusst wird (Unity Technologies, o. J.-b). Collider-Komponenten definieren die physikalische Begrenzung eines Objektes und bestimmen, ob und wie es mit anderen Objekten kollidiert (Unity Technologies, o. J.-c).

3.5 Erstellung der Töne

Um im fertigen Prototyp die Töne zu hören, müssen diese erst erstellt werden. Die dafür notwendigen Technologien und Anwendungen zur Erstellung der Musikpartitur, Umwandeln der Partitur in Stimmen und Schneiden der Audios werden im Folgenden vorgestellt.

3.5.1 MuseScore Studio

MuseScore Studio ist ein kostenloses, Open-Source-Programm zur Erstellung von Musikpartituren. Es ermöglicht das Komponieren, Bearbeiten und Abspielen von Noten für verschiedene Instrumente und eignet sich sowohl für einfache Melodien als auch für komplexere Arrangements (*MuseScore*, o. J.). In dieser Arbeit wird die aktuellste Version 4.5 verwendet.

Noten können intuitiv durch Mausklick, per Tastatur oder MIDI-Keyboard eingegeben werden. Dabei werden alle gängigen musikalischen Symbole, Dynamiken und Liedtexte unterstützt. Eingegebene Noten können anschließend mit verschiedenen Instrumentensounds abgespielt werden (*MuseScore*, o. J.).

Als Open-Source-Software bietet MuseScore Studio die Möglichkeit, fehlende Features durch eigene Plugins zu ergänzen oder anzupassen.

MuseScore wurde zur Musikpartitur-Erstellung in dieser Arbeit gewählt, da die einfache Bedienung und Flexibilität es möglich machte, viele verschiedene Musikpartituren zu erstellen und auszuprobieren. Zudem unterstützt die Software den Export der Partitur im MusicXML-Format, das für die im nächsten Abschnitt vorgestellte Technologie notwendig ist.

3.5.2 Sinsy

Die im Prototyp gehörten Stimmen werden mithilfe von Sinsy generiert. Es wandelt die zuvor in MuseScore Studio erstellten Musikpartituren in Gesangsstimmen im WAV-Format um. Dabei werden sowohl Tonhöhen als auch Liedtexte erfasst und synthetisch wiedergegeben. Da die genaue Funktionsweise der KI-Modelle im Grundlagen-Kapitel bereits erläutert wurden, folgt hier eine Beschreibung der Benutzeroberfläche.

Auf der Webseite von Sinsy können Nutzer eine MusicXML-Datei hochladen. Nach der Einstellung einiger Parameter wie die Auswahl der Stimme (auf der das Training der Modelle basiert) und dem *Gender Param*, welcher entscheidet, ob sich die Stimme maskuliner oder femininer anhört. Nachdem Nutzer ihre Einstellungen und Parameter ausgewählt haben, kann durch Knopfdruck die Synthese gestartet werden. Nach einer kurzen Zeit wird eine WAV-Datei erzeugt, die direkt auf der Webseite angehört und heruntergeladen werden kann (*Sinsy*, o. J.).

Sinsy wurde aufgrund dieser intuitiven Benutzeroberfläche und der hohen Flexibilität für die Generierung der Gesangsstimmen ausgewählt. Zudem klangen die generierten Stimmen nach subjektiver Einschätzung des Autors dieser Arbeit natürlich, harmonisch und ähnelten den Opernstimmen von *Blob Opera*. Ein weiterer Vorteil ist auch die große Auswahl an Stimmen beider Geschlechter, wodurch die verschiedenen Stimmlagen besser abgebildet werden können.

Die so erstellten WAV-Dateien könnten schon in Unity genutzt werden, lassen sich bei Bedarf aber noch bearbeiten. Für das Zuschneiden und Anpassen der Audiodateien wird die im nächsten Abschnitt vorgestellte Anwendung genutzt.

3.5.3 Audacity

Audacity ist ein kostenloses Open-Source-Programm zur Aufnahme und Bearbeiten von Audio (*Audacity*, o. J.).

Während die Aufnahme-Funktion von Audacity keine Verwendung in dieser Arbeit findet, hilft Audacity dabei, die von Sinsy generierten WAV-Dateien effizient zuzuschneiden.

Die Wahl fiel auf Audacity aufgrund der intuitiven Bedienung und der Möglichkeit, mehrere Dateien schnell und gleichzeitig zu exportieren. Verwendet wurde die Version 3.7.1

4 Umsetzung

Im folgenden Kapitel wird die Umsetzung des Prototyps auf Basis des vorgestellten Konzeptes und mithilfe der vorgestellten Technologien beschrieben. Dabei werden im ersten Schritt alle notwendigen Vorbereitungen wie die Erstellung der Gesangstöne beschrieben und darauffolgend die schrittweise Umsetzung in Unity.

4.1 Vorbereitungen

Dieses Kapitel beschreibt alle Vorbereitungen, die erledigt wurden, bevor die eigentliche Programmierung des Prototyps in Unity losging. Dazu gehören das Erstellen der 3D-Modelle des Pinsels und der Palette sowie die Generierung und Bearbeitung der Töne.

4.1.1 Erstellung der 3D-Modelle

Die Modelle für den Pinsel und die Palette wurden von Vivien Laasch, einer Bekannten des Autors dieser Arbeit, mit der Software Blender erstellt. Blender ist ein kostenloses Open-Source Programm unter anderem zur Erstellung von 3D-Modellen (Blender Foundation, o. J.). Grundlage für die Modellierung war eine handgezeichnete Skizze und eine kurze Beschreibung der Anforderungen.

Beide Objekte bestehen aus mehreren separat bearbeitbaren Materialien, was das spätere Verändern der Farben in Unity vereinfacht.

Da die Erstellung der Modelle nicht im Fokus dieser Arbeit steht, wird auf den detaillierten Modellierungsprozess nicht weiter eingegangen. Mithilfe von Blender konnten beide Objekte im Unity-kompatiblen FBX-Format exportiert werden.

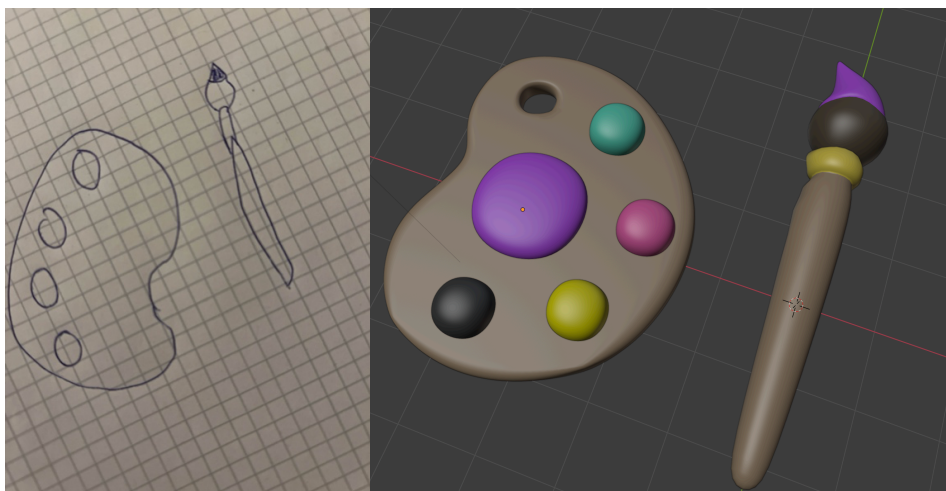


Abbildung 26: Ursprüngliche Skizze (links) und die fertigen Modelle (rechts)

Quelle: Eigene Darstellung

4.1.2 Erstellung der Musikpartituren

Bevor die Gesangsstimmen generiert werden können, müssen erst die Musikpartituren erstellt werden.

Dazu werden in MuseScore Studio für jede der beiden gesungenen Silben („Wah“ und „Do“) Partituren für die vier Stimm lagen erstellt – insgesamt entstehen so acht Partituren.

Im Folgenden wird das Vorgehen für die Silbe „Wah“ und der Bass-Stimm lage beschrieben. Das Vorgehen ist jedoch für alle Stimm lagen und Silben identisch und unterscheidet sich nur in den verwendeten Noten, Voreinstellungen und Liedtexten.

Zu Beginn wurde eine neue Partitur in MuseScore Studio mit den folgenden Voreinstellungen erstellt:

- Instrument: Gesang -> Bass
- Tonart: A-Dur
- Taktart: 4/4
- Tempo: 100

Anschließend werden die leeren Notenlinien mit den im Konzept (s. Kapitel 3.3.3) beschriebenen Tönen der Reihe nach gefüllt. Jeder Ton erhält einen eigenen Takt, um eine ausreichende Länge für die spätere Verarbeitung zu gewährleisten. Die endgültige Anpassung der Tondauer erfolgt später in Audacity.

Da Sinsy mit japanischen Stimmen arbeitet, müssen die Liedtexte in einem für das System verständlichen Format eingegeben werden. Hierzu wurde die von Sinsy bereitgestellte Japanisch-Phonem-Tabelle genutzt (*SinsyReference*, 2021).

フ r,a	リ r,i
ワ w,a	キ i
ガ	ギ

Abbildung 27: Ausschnitt aus der Japanisch-Phonem-Tabelle des Sinsy Reference Guides
Quelle: SinsyReference, 2021

Damit die Silbe „Wah“ korrekt erkannt wird, muss sie als [w,a] eingegeben werden. Die eckigen Klammern sind eine Vorgabe von Sinsy, um das Phonem korrekt zu erkennen (*SinsyReference*, 2021).

Abschließend wird die fertige Partitur im MusicXML-Format exportiert und gespeichert. Dieses Prozedere wird dann anschließend für die anderen drei Stimm lagen Tenor, Alt und Sopran wiederholt. Danach wird der gesamte Prozess für die Silbe „Do“ ([d,o]) erneut für alle vier Stimm lagen durchgeführt, bis alle acht Partituren im MusicXML-Format bereitstehen.

Bass

w,a

Komponist / Arrangeur

♩ = 100

The musical score for the Bass part is written in bass clef with a key signature of two sharps (F# and C#) and a 4/4 time signature. It consists of ten measures, each containing a half note. The notes are: F#2, F#2, F#2, F#2, F#2, F#2, F#2, F#2, F#2, and F#2. Below each note is the phonetic label [w,a]. A yellow highlight is placed under the first [w,a] label. The score ends with a double bar line.

11

The final measure of the Bass part is shown, featuring a half note F#2 with the phonetic label [w,a] below it. A yellow highlight is placed under the [w,a] label. The measure ends with a double bar line.

Abbildung 28: Fertige Partitur am Beispiel Bass und der Silbe „Wah“
 Quelle: Eigene Darstellung

Sopran

d,o

Komponist / Arrangeur

♩ = 100

The musical score for the Soprano part is written in treble clef with a key signature of two sharps (F# and C#) and a 4/4 time signature. It consists of ten measures, each containing a half note. The notes are: F#4, F#4, F#4, F#4, F#4, F#4, F#4, F#4, F#4, and F#4. Below each note is the phonetic label [d,o]. A yellow highlight is placed under the last [d,o] label. The score ends with a double bar line.

11

The final measure of the Soprano part is shown, featuring a half note F#4 with the phonetic label [d,o] below it. A yellow highlight is placed under the [d,o] label. The measure ends with a double bar line.

Abbildung 29: Fertige Partitur am Beispiel Sopran und der Silbe „Do“
 Quelle: Eigene Darstellung

Diese fertigen Partituren werden im nächsten Schritt auf der Webseite von Sinsy hochgeladen (Sinsy, o. J.).

4.1.3 Generierung der Gesangsstimmen

Die zuvor erstellte MusicXML-Datei kann über die Benutzeroberfläche der Sinsy-Webseite ausgewählt und hochgeladen werden. Anschließend können die Parameter eingestellt werden. Um die bestmöglichen Einstellungen zu finden, wurden mehrere Kombinationen ausprobiert. Dafür mussten alle gewünschten Parameter eingegeben und die Synthese per Knopfdruck gestartet werden. Über den integrierten Player auf der Webseite konnten die Ergebnisse direkt angehört und verglichen werden. Da in den zuvor erstellten Musikpartituren bereits alle relevanten Töne enthalten waren, konnten sie direkt in einem Durchgang abgespielt und angehört werden. Dadurch war es nicht notwendig, jeden einzelnen Ton separat zu synthetisieren, was den Prozess erheblich beschleunigte.

Nach mehreren Tests erwiesen sich die in der folgenden Tabelle dargestellten Einstellungen als besonders geeignet:

Tabelle 1: Parameter der einzelnen Stimmen
Quelle: Eigene Darstellung

Stimmlage	Silbe	Vocal	Gender param.	Vibrato intensity	Pitch Shift
Bass	Wah	m01083j_dnn_beta5: Tatsumi	0.55	0	0
Tenor	Wah	f00001j_dnn_beta5: Yoko	0.8	0	0
Alt	Wah	f00001j_dnn_beta5: Yoko	0.55	0	0
Sopran	Wah	f00001j_dnn_beta5: Yoko	0.42	0.5	0
Bass	Do	m01083j_dnn_beta5: Tatsumi	0.55	0	0
Tenor	Do	m01083j_dnn_beta5: Tatsumi	0.55	0	0
Alt	Do	f00001j_dnn_beta5: Yoko	0.55	0	0
Sopran	Do	f00001j_dnn_beta5: Yoko	0.42	0.5	0

Die Sprache *Language* war bei allen Stimmlagen und Silben *Japanese*.

Nachdem die Stimmen synthetisiert und heruntergeladen wurden, müssen die acht langen WAV-Dateien, die alle Töne enthalten, im nächsten Schritt geschnitten werden.

4.1.4 Schneiden der Audiodateien

Im letzten Schritt der Vorbereitung werden in Audacity die langen Gesangs-Dateien in einzelne Töne aufgeteilt.

Dazu werden die einzelnen WAV-Dateien in Audacity geöffnet und mithilfe des Trenn-Werkzeugs in separate Audioclips geschnitten. Dabei werden auch stille Passagen sowie Abschnitte mit Störgeräuschen oder Glitches isoliert. Sobald der lange Audioclip in mehrere separate Clips zerlegt wurde, können alle nicht benötigten Clips entfernt werden, sodass nur die gewünschten Töne übrig bleiben.



Abbildung 30: Kompletter Audio-Clip nach Trennung
Quelle: Eigene Darstellung

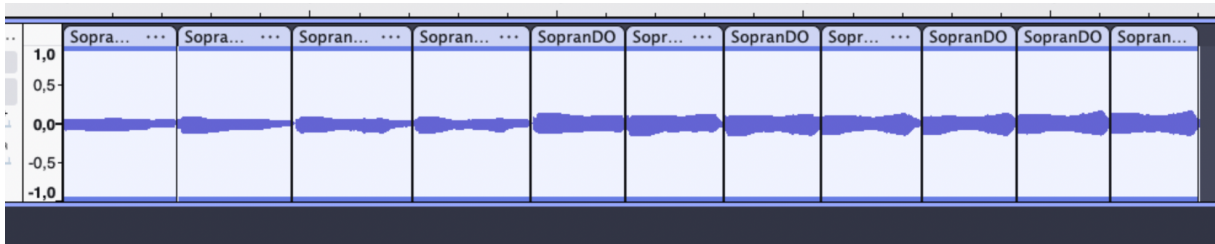


Abbildung 31: „Bereinigter“ Audio-Clip, der nur aus den benötigten Tönen besteht
Quelle: Eigene Darstellung

Dieser Prozess wird für jede der acht generierten Audiodateien von Sinsy durchgeführt, wodurch eine Sammlung einzelner Clips entsteht.

Anschließend werden die benötigten Audioclips einzeln im WAV-Format exportiert, sodass sie für eine weitere Nutzung in Unity bereit sind.

4.2 Entwicklung in Unity

In diesem Kapitel wird die Umsetzung des Prototyps in Unity erläutert. Dabei wird der Prototyp Schritt für Schritt um eine weitere Funktion erweitert. Zunächst wird die Interaktion mit den Objekten in VR implementiert. Anschließend folgt die Zeichenfunktion und dann wird das Abspielen der einzelnen Töne integriert. Am Ende wird der Mischmodus samt Akkord-Funktion hinzugefügt. Da immer ein Schritt nach dem anderen umgesetzt wurde, ist der Code alles andere als optimal. Es wurde nur darauf geachtet, dass alle im Konzept erwähnten Funktionen implementiert werden, ohne, dass es zu technischen Fehlern kommt.

4.2.1 VR-Interaktion

Zu Beginn wird ein neues 3D-Projekt in Unity angelegt. Da es noch keine Voreinstellungen gibt, mit denen VR-Entwicklung unterstützt wird, muss das XR-Plug-in Management installiert werden. Anschließend kann darüber das OpenXR-Plugin installiert werden.

Das OpenXR-Plugin verlangt die Auswahl von Interaction Profiles, diese Profile definieren, welche Eingabegeräte verwendet werden. Für diesen Prototyp wurden vorerst nur die Profile Oculus Touch Controller Profile und Meta Quest Touch Pro Controller Profile, um die Quest 3 zu unterstützen.

In Unitys Package Manager muss nun das XR Interaction Toolkit installiert werden. Dieses Toolkit bietet eine Demoszene mit Starter Assets und Komponenten, die für die Entwicklung genutzt werden.

Nach diesen Vorbereitungen wird eine neue Unity-Szene erstellt. Diese beinhaltet standardmäßig eine Main Camera, Directional Light und eine Standard-Skybox. Da die Standardkamera nicht benötigt wird, wird sie entfernt und stattdessen eine XR Origin (XR Rig)-Komponente aus der Demoszene hinzugefügt. Diese beinhaltet die neue Main Camera, alle notwendigen Komponenten zur Bewegung und definiert die beiden Controller als Interactor. Zudem erstellt Unity automatisch einen XR Interaction Manager, der für die Steuerung der Interaktionen zuständig ist.

Nun wird der virtuelle Boden erstellt, indem ein Plane-Objekt zur Szene hinzugefügt wird. Dieses erhält eine Skalierung von 5x5x5 und ist damit 50m x 50m groß (standardmäßig ist ein Plane 10x10). Dann wird darauf ein Cube-Objekt mit der Skalierung 1x1x1 (entspricht 1m x 1m x 1m) platziert, auf dem dann die beiden Objekte Pinsel und Palette platziert werden.

Nun können sich Nutzer bereits frei in VR bewegen und sich umschauen. Um mit dem Pinsel und der Palette zu interagieren, müssen diese noch spezifische Komponenten erhalten.

Durch das Hinzufügen der XR Grab Interactable-Komponente können Nutzer die Objekte greifen und mit ihnen interagieren. Der Movement Type in der Komponente wird auf Instantaneous gestellt und der Far Attach Mode auf Near. Zudem erhalten beide eine leere GameObject-Komponente, die an den Positionen platziert wird, wo die Objekte später gehalten werden. Dieses wird jeweils *AttachTransform (Brush)* und *AttachTransform (Palette)* genannt. Diese Einstellungen sorgen dafür,

dass sich Pinsel und Palette direkt und synchron mit den Bewegungen des Controllers mitbewegen. Beim Aufnehmen positionieren sie sich exakt an den zuvor definierten Attach Transforms.

Das Hinzufügen der XR Grab Interactable-Komponente fügt den Objekten auch automatisch einen Rigidbody hinzu. Damit Pinsel und Palette sich physikalisch korrekt verhalten, wird ihnen schließlich jeweils ein passender Collider zugewiesen (kapselförmiger Collider für den Pinsel und ein flacher Quader für die Palette).

Nachdem nun die Interaktionsmechanismen erfolgreich implementiert wurden, folgt im nächsten Schritt die Umsetzung der Zeichenfunktion.

4.2.2 Zeichenfunktion

Die LineRenderer-Komponente ist hauptsächlich für das Zeichnen der Linien verantwortlich. Zuerst muss jedoch festgestellt werden, ob der Pinsel gehalten wird und der Zeichen-Button (A) gedrückt wird.

Unity prüft in jedem Frame, ob diese beiden Bedingungen erfüllt sind und führt, falls ja, die Funktion `Paint` aus. Diese Funktion liest die momentane Position der Pinselspitze aus und erstellt einen Punkt. Wird der Pinsel nun bewegt, wird erst geprüft, ob er sich genug bewegt hat (in diesem Fall 0.01 Einheiten) und dann ein neuer Punkt erstellt. Diese Punkte werden in einer Liste gespeichert und dem LineRenderer übergeben, welcher dann eine Linie entlang der Punkte mit einem festgelegten Material erstellt.

Um das Wechseln des Materials durch Eintauchen der Pinselspitze in die Kleckse der Palette zu erlauben, werden den einzelnen Farbklecksen Tags und Collider gegeben. Mithilfe der in Unity-Funktion `OnTriggerEnter` kann der Pinsel bei Berühren des Kleckses die Tags vergleichen und herausfinden, welcher Farbe er zugewiesen ist und sein Material dem Material des Kleckses angleichen.

Codeblock 1: Paint-Funktion (vereinfachte Darstellung)

Quelle: Eigene Darstellung

```
1 private List<Vector3> points = new List<Vector3>();
2 Vector3 brushPosition = brushTip.position;
3 if (points.Count == 0 || Vector3.Distance(points[points.Count - 1],
4 brushPosition) > 0.01f) {
5     points.Add(brushPosition);
6     LineRenderer.material = currentActiveColor;
7     LineRenderer.positionCount = points.Count;
8     LineRenderer.SetPositions(points.ToArray());
```

Codeblock 2: Funktion zum Farbe wechseln (vereinfachte Darstellung)

Quelle: Eigene Darstellung

```
1 void OnTriggerEnter(Collider other) {
2     if (other.CompareTag("Color1") || other.CompareTag("Color2"));
3     int touchedColor = other.tag switch { "Color1" => 0, "Color2" => 1 }
4     currentActiveColor = touchedColor
5     brushTipRenderer.materials = currentActiveColor // Färbt auch die
6     Pinselspitze }
```

4.2.3 Sounds

Als Nächstes folgt die Integration der Audiodateien. Dafür werden acht AudioClip-Arrays erstellt, die mit den einzelnen Tönen von Audacity gefüllt werden. Außerdem wird dem Pinsel eine Audio Source-Komponente hinzugefügt, von der aus diese Töne emittiert werden.

Der Paint-Funktion wird nun eine weitere MapToneToPosition-Funktion hinzugefügt und die Vector3-Variable startPosition. MapToneToPosition berechnet die Position relativ vom Startpunkt für die Achsen X und Y. Bei der Bewegung in der X-Achse wird noch eine Bool-Variable UseDo eingeführt, die entscheidet welche der beiden Silben ausgewählt wird.

Da unabhängig von der Startposition beim Malen immer derselbe Ton anfangs gespielt werden soll, wird eine Variable startingNote als Ausgangsnote eingeführt. Wenn von dort aus eine bestimmte Schwelle (0.15 Unity-Einheiten) vertikal überschritten wird, oder das 1,5-fache der Schwelle in der horizontalen überschritten wird, soll ein neuer Ton oder eine neue Silbe gespielt werden. Die Überprüfung der Schwellenüberschreitung sorgt dafür, dass nicht bei jeder kleinsten Bewegung ein neuer Ton gespielt wird. Sie wird, wie im folgenden Codeblock dargestellt, berechnet:

Codeblock 3: Berechnung von switchTone
Quelle: Eigene Darstellung

```
1 bool switchTone = Mathf.Abs(currentPosition.y - lastPlayedPosition.y) >
toneChangeThreshold || Mathf.Abs(currentPosition.x - lastPlayedPosition.x) >
toneChangeThreshold * 1.5;
```

Um einen Ton zu spielen, wird die Funktion PlayTone verwendet. Diese wählt abhängig davon, ob UseDo wahr oder falsch ist, einen AudioClip mit der entsprechenden Silbe und der ausgewählten Stimmlage aus. Welcher genaue Ton im AudioClip gespielt wird, bestimmt die Variable note. Diese berechnet, wie viele Tonschritte von der Startnote aus gegangen sind, also wie oft die Schwelle überschritten wurde. Als Beispiel: Die Startnote 3 und eine Bewegung von +0,3 Einheiten (die Schwelle ist 0.15) entspricht einer Bewegung von 2 Tonschritten, also wird Ton 3+2=5 gespielt.

Der Ton wird mithilfe von Unitys AudioSource.play()-Funktion abgespielt.

Codeblock 4: MapToneToPosition-Funktion (vereinfachte Darstellung)
Quelle: Eigene Darstellung

```
1 float relativeY = currentPosition.y - startPosition.y
2 float relativeX = currentPosition.x - startPosition.x
3 bool useDO = Mathf.Abs(relativeX) > toneChangeThreshold * 2
4 int startingNote = 3 // im Original ein switch je nach aktiver Farbe
5 int note = startingNote + Mathf.RoundToInt((relativeY /
toneChangeThreshold))
6 if (switchTone){
7     PlayTone(note, useDO);
8     lastPlayedposition = currentPosition; }
```

Codeblock 5: PlayTone-Funktion mit Auswahl des richtigen AudioClips
Quelle: Eigene Darstellung

```
<!--weitere Funktionalitäten-->
1 AudioClip[] activeAudioClip = WaOrDo(useDO)
2 AudioSource.clip = activeAudioClip[note]
3 AudioSource.play
<!--weitere Funktionalitäten-->
4 private AudioClip[] WaOrDo(bool useDO) => currentActiveColor switch {
5 0 => useDO ? bassAudioDO : bassAudioWA,
6 1 => useDO ? tenorAudioDO : tenorAudioWA,
7 2 => useDO ? altoAudioDO : altoAudioWA,
8 3 => useDO ? sopranoAudioDO : sopranoAudioWA }
```

Nun ist bereits die Kernfunktion des Prototyps implementiert. Nutzer können sich in der VR-Umgebung umsehen, bewegen, zeichnen und, abhängig von der gezeichneten Position, musizieren.

4.2.4 Mischmodus

In diesem Schritt wird der Mischmodus implementiert und erläutert, wie die Akkorde zusammengestellt werden.

Dafür wird dem Pinsel und der Palette eine neue Funktion `AddMoreVoices` hinzugefügt. Diese wird immer beim Drücken des Misch-Buttons aufgerufen und erlaubt es, die Variabel `activeVoices` zu erhöhen und die Farbe der Pinselspitze bzw. des großen den Farbklecks der Palette anzupassen.

Codeblock 6: AddMoreVoices-Funktion (vereinfachte Darstellung)
Quelle: Eigene Darstellung

```
1 private void AddMoreVoices(InputAction.CallbackContext context) {
2 if (currentActiveColor == 4) { //Aktiver Mischmodus
3 activeVoices = (activeVoices % 4) + 1;
4 brushTipRenderer.materials = currentActiveColor; }
// Analog dazu existiert ein Skript in der Palette, das die Farbe des Klecks
ebenfalls bei Knopfdruck anpasst
```

Damit die Stimmen gemeinsam Akkorde singen, wird die `PlayTone`-Funktion erweitert. Alle Töne, die gesungen werden können, werden in einem Array aus Strings gespeichert und ein Dictionary `chordMap` wird erstellt, in dem die Akkorde (A-Dur, E-Dur, D-Dur und Fis-Moll) definiert werden. Eine Methode `GetHarmonizingChord` weist jedem Ton einen Akkord (oder entscheidet durch Zufall bei zwei möglichen Akkorden) zu. Der Ton, der den Akkord definiert, wird mit der Variable `melodyNote` übergeben. Schließlich sucht die For-Schleife der Methode `GetNoteFromArray` die Note aus dem Array und gibt diese an die `AudioSource`. Die Funktion `PlayTone` spielt dann abhängig von der Anzahl aktiven Stimmen (`activeVoices`) einen oder mehrere Töne ab.

Während bei den einzelnen Tönen `AudioSource.Play` verwendet wurde, ist für das gleichzeitige Abspielen mehrerer Töne `Unitys AudioSource.PlayOneShot` erforderlich.

Codeblock 7: Array an singbaren Tönen am Beispiel Bass
Quelle: Eigene Darstellung

```
string[] bassNotes = new string[] { "E", "Fis", "Gis", "A", "H", "Cis", "D",  
"E", "Fis", "Gis", "A" };
```

Codeblock 8: chordMap-Dictionary, welches die Akkorde definiert
Quelle: Eigene Darstellung

```
1 readonly Dictionary<string, (string bass, string tenor, string alto)>  
chordMap = new()  
2 {  
3     ["A-Dur"]      = ("A",    "E",    "Cis"),  
4     ["E-Dur"]      = ("E",    "Gis",  "H"),  
5     ["D-Dur"]      = ("D",    "Fis",  "A"),  
6     ["Fis-Moll"]   = ("Fis",  "A",    "Cis")  
7 };
```

Codeblock 9: Methode, um der Melodie-Note einem oder mehreren Akkorden zuzuweisen
Quelle: Eigene Darstellung

```
1 string GetHarmonizingChord(string melodyNote) => melodyNote switch  
2 {  
3     "A"    => "A-Dur",  
4     "H"    => "E-Dur",  
5     "Cis"  => Random.value < 0.5f ? "A-Dur" : "Fis-Moll",  
6     "D"    => "D-Dur",  
7     "E"    => Random.value < 0.5f ? "A-Dur" : "E-Dur",  
8     "Fis"  => Random.value < 0.5f ? "D-Dur" : "Fis-Moll",  
9     "Gis"  => "E-Dur",  
10    _      => "A-Dur"  
11 };
```

Codeblock 10: Methode, um den Index aus dem Array an Notennamen-Strings zu extrahieren
Quelle: Eigene Darstellung

```
1 int GetNoteFromArray(string note, string[] notes) {  
2     for (int i = 0; i < notes.Length; i++) {  
3         if (notes[i] == note) return i;  
4     }  
5     return 0; }
```

Codeblock 11: Erweiterte PlayTone-Funktion. Verkürzte Darstellung mit drei Stimmen
Quelle: Eigene Darstellung

```
<!--Einzelton-Funktionalitäten-->  
1 if (currentActiveColor == 4) { // Aktiver Mischmodus (5. Farbe in Palette)  
<!--Einzelton-Funktionalitäten für nur eine und zwei Farben-->  
2 else if (activeVoices == 3) {  
3     string melodyNote = altoNotes[note]; //Melodie im Alt durch Bewegung  
beeinflusst (MapToneToPosition-Funktion)  
4     string harmonizingChord = GetHarmonizingChord(melodyNote);  
5     var chordTones = chordMap[harmonizingChord];  
6     int bassNote = GetNoteFromArray(chordTones.bass, bassNotes);  
7     int tenorNote = GetNoteFromArray(chordTones.tenor, tenorNotes);  
8     AudioSource.PlayOneShot((useDO ? bassAudioDO : bassAudioWA)[bassNote]);  
9     AudioSource.PlayOneShot((useDO ? tenorAudioDO : tenorAudioWA)[tenorNote]);  
10    AudioSource.PlayOneShot((useDO ? altoAudioDO : altoAudioWA)[note]); }  
11 else if (activeVoices == 4)  
<!--Funktionalitäten für vier Stimmen analog zu drei, aber mit Sopran als  
Melody Tone-->
```

Als Beispiel:

Der Sopran spielt den Ton A. `GetHarmonizingChord(A)` gibt „A-Dur“ zurück. Dieser Akkord ist in der `chordMap` definiert als „A-Dur“ = („A“, „E“, „Cis“) für Bass, Tenor und Alt. Dementsprechend ergibt sich der folgende Akkord:

- Sopran: A
- Alt: Cis
- Tenor: E
- Bass: A

`GetNoteFromArray` sucht dann für den Bass beispielsweise die Note „A“, welche dem Index 3 (s. Codeblock 7) entspricht. `AudioSource.PlayOneShot` spielt dann das vierte Element (Index 3) aus dem `AudioClip`-Array mit den erzeugten Bassnoten, welcher der Note A entspricht.

Durch diese Implementationen hat der Prototyp nun alle im Konzept beschriebenen Funktionen und funktioniert ohne technische Einschränkungen.

Um den Prototyp nun Probanden zu zeigen und zu evaluieren, muss er noch exportiert werden und auf das HMD geladen werden.

5 Evaluation

Es wird im Folgenden der entwickelte Prototyp evaluiert. Ziel ist es, zu untersuchen, inwiefern die Nutzung des Prototyps das subjektive Entspannungsgefühl der Nutzer verbessert. Verwendet wird hierfür ein Fragebogen, der quantitative und qualitative Elemente enthält.

Dieses Kapitel erläutert im ersten Schritt den verwendeten Fragebogen, beschreibt dann die Durchführung und präsentiert zuletzt die Ergebnisse.

5.1 Fragebogen

Der Fragebogen, der in dieser Studie verwendet wird, wurde mithilfe von Google Forms erstellt und besteht aus mehreren zusammengesetzten Fragebögen und Elementen, die im Folgenden erläutert werden (*Google Formulare*, o. J.).

Um die Effekte und Wirksamkeit von Entspannungsaufgaben kurzfristig nach der Ausführung zu bewerten, haben die Forscher Sarah Steghaus und Christian H. Poth den Relaxation State Questionnaire (RSQ) entwickelt, der aus 10 Aussagen besteht. Diese Aussagen sollen auf einer fünfstufigen Skala von 0 (trifft überhaupt nicht zu) bis 4 (trifft voll und ganz zu) bewertet werden und beziehen sich auf vier Hauptfaktoren: Muskelentspannung, Schläfrigkeit, kardiovaskuläre Aktivitäten und allgemeine Entspannung (Steghaus & Poth, 2022). Der RSQ wird ergänzt mit einer angepassten Form des Fragebogens zur Erfassung von Entspannung (FEE), welcher auf der Webseite von Prof. Dr. med. Thomas Loew heruntergeladen werden kann. Dieser Fragebogen lässt Klienten ihr subjektives Entspannungsempfinden auf einer Skala von 0 bis 10 vor und nach einer Entspannungsübung bewerten (*FEE*, o. J.). Die Skala wurde auf 1 bis 10 angepasst, sodass dieselbe Skala wie in der Evaluation von der vergleichbaren Arbeit *Virtual Harmony* verwendet wird (Lin et al., 2020).

Zusätzlich zu den beiden obengenannten Fragebögen enthält der Fragebogen dieser Studie Fragen zu demographischen Informationen, kontextbezogene Fragen (Vorerfahrungen mit VR und Kunst- oder Musiktherapie) und Aussagen zur Bewertung des Prototyps. Die Aussagen zum Prototyp sind inhaltlich in Kategorien gegliedert, die sich an den zuvor definierten Anforderungen orientieren.

Um die Probanden nicht zu sehr zu verwirren und eine einheitliche Bewertungsskala zu schaffen, sollen die Aussagen zur Benutzerfreundlichkeit und Interaktivität des Prototyps auf derselben fünfstufigen Skala, wie die des von Steghaus & Poth (2022) entwickelten Fragebogens, bewertet werden.

Auf dem Fragebogen werden die Probanden zunächst begrüßt und der Ablauf der Studie wird erklärt. Begonnen wird mit einem Prä-Test-Teil, der demographische Fragen hat und den subjektiven Entspannungsgrad mithilfe des RSQs und des FEEs feststellt. Nach der Nutzung des Prototyps kommt der Post-Test-Teil, auf dem dieselben Fragen zum Entspannungsgefühl und die Aussagen zum Prototyp sind. Am Ende ist noch ein Text-Feld, in das die Teilnehmer Kommentare und Verbesserungsvorschläge reinschreiben können.

5.2 Durchführung

Die Nutzerstudie wurde an zwei Tagen an verschiedenen Orten mit unterschiedlichen Gruppen durchgeführt. Beide Gruppen bestanden jeweils aus fünf Personen, die sich freiwillig erklärt haben, bei der Studie zu unterstützen. Die erste Gruppe bestand aus fünf Frauen im Alter zwischen 46 und 64, die an diesem Tag an einer verlängerten Orchester-Tagesprobe teilnahmen. Die zweite Gruppe bestand aus einer Frau und vier Männern im Alter zwischen 19 und 28 (eine Person gab ihr Alter nicht an). Die Personen aus der zweiten Gruppe sind Kollegen des Autors dieser Arbeit und waren am Tag der Studie im Büro. Die beiden Gruppen wurden gewählt, da sich die Mitglieder beider Gruppen in einem unterschiedlichen potenziell stressigen Umfeld befinden (Arbeitsalltag oder intensive Musikprobe).

Tabelle 2: Demographische und kontextuelle Daten der Gruppe 1 (Orchester)

Quelle: Eigene Darstellung

Teilnehmer	Geschlecht	Alter	Erfahrung mit VR	Nutzt Musik zur Entspannung	Nutzt Kunst zur Entspannung	Erfahrung mit Kunst- oder Musiktherapie
1	weiblich	52	Nein	Ja, aktiv	Nein	Nein
2	weiblich	46	Nein	Nein	Ja	Nein
3	weiblich	57	Nein	Nein	Nein	Nein
4	weiblich	57	Nein	Ja, rezeptiv	Ja	Nein
5	weiblich	64	Nein	Ja, aktiv	Nein	Nein

Tabelle 3: Demographische und kontextuelle Daten der Gruppe 1 (Büro)

Quelle: Eigene Darstellung

Teilnehmer	Geschlecht	Alter	Erfahrung mit VR	Nutzt Musik zur Entspannung	Nutzt Kunst zur Entspannung	Erfahrung mit Kunst- oder Musiktherapie
1	männlich	26	Nein	Nein	Nein	Nein
2	männlich	19	Ja	Ja, aktiv und rezeptiv	Nein	Nein
3	männlich	-	Nein	Ja, rezeptiv	Nein	Nein
4	männlich	24	Nein	Ja, rezeptiv	Nein	Nein
5	weiblich	28	Nein	Ja, rezeptiv	Ja	Nein

Der Ablauf für beide Gruppen war derselbe: Am Vortag wurde eine Chat-Nachricht in die jeweiligen Kanäle verfasst, die nach Freiwilligen für die Studie fragt. Da Gruppe 1 ausschließlich aus Frauen bestand, wurde in der Nachricht für Gruppe 2 geschrieben, dass Männer bevorzugt gesucht werden. Am Tag der Studie wurde ein Laptop mit dem Fragebogen und dem HMD mit der fertigen Software des Prototyps an den jeweiligen Ort mitgebracht. Die Probanden von Gruppe 1 wurden in den Pausen und die Probanden von Gruppe 2 zu abgemachten Zeiten der Reihe nach in einen separaten Raum begleitet. In diesem Raum haben die Probanden den Laptop mit dem Fragebogen vor sich, welcher so positioniert ist, dass der Durchführende der Studie keine Sicht darauf hat. Nachdem der Ablauf erklärt und durchgelesen wurde, wurden die Prä-Test-Fragen beantwortet.

Da bis auf Teilnehmer 2 der Gruppe 2 keiner der Teilnehmer angegeben hat, Vorerfahrungen mit VR zu haben, wurde eine kleine, allgemeine Einführung gegeben, bevor der eigentliche Test erklärt wurde.

Zu Beginn jedes Tests wurden die Steuerungsmöglichkeiten erläutert und die Umgebung beschrieben, die die Teilnehmer gleich betreten werden. Anschließend wird das HMD für den Probanden angepasst und aufgesetzt. In der VR-Umgebung angekommen, wird ihm erklärt, dass er sich zu dem Kasten bewegen soll und den Pinsel sowie die Palette aufnehmen soll. Dabei sieht der Durchführende der Studie alles, was die Teilnehmer sehen, auch live auf dem Smartphone. Dann soll erst mit der Farbe für Bass, dann Tenor, dann Alt und zuletzt Sopran etwas experimentiert werden, um mit dem Malen und dem Musizieren vertraut zu werden. Falls die Probanden nach Nachfrage der Auffassung waren, dass diese Interaktionsmöglichkeit beherrscht wurde, wurde der Mischmodus erklärt. Dabei sollen die Probanden alle verschiedenen Möglichkeiten der Zusammensetzung der Stimmen ausprobieren. Wenn auch diese Funktion verstanden wurde, wurden den Probanden kleine Aufgaben gestellt.

Zuerst sollen sie mit einer beliebigen Farbe, die einen einzelnen Ton abspielt, eine vorgegebene Melodie spielen. Die Anweisung dafür lautete sinngemäß: „Spiele die ersten 5 Töne einer Tonleiter“. Das Prinzip einer Tonleiter musste hierbei nicht erklärt werden. Anschließend wurde gefordert: „Spiele mit mehreren Stimmen – egal, wie viele – vier Töne rauf und 5 Töne runter“.

Die letzte Aufgabe bestand darin, eine Blume und eine Sonne zu malen, und zwar mit den Farben, die die Probanden auch normalerweise nutzen würden, um diese Objekte zu malen.

Falls die Probanden nach dem Abschluss der Aufgaben noch weiter experimentieren wollten, wurde ihnen dafür Zeit gegeben, falls nicht, wurden sie gebeten den Post-Test-Fragebogen auszufüllen.

Dafür wurde ihnen das HMD abgesetzt und sie sollten die Fragen, genau wie im Prä-Test, ohne, dass eine andere Person die Antworten sehen konnte, beantworten. Die Teilnehmer hatten zu jederzeit die Möglichkeit, Fragen zum Fragebogen zu stellen und mit dem Durchführenden der Studie zu reden.

Zum Schluss wurde sich bei den Probanden bedankt und der Versuchsaufbau zurückgesetzt.

Die Dauer der Erhebung der Daten durchschnittlich 30 Minuten pro Person.

5.3 Ergebnisse bezüglich des Entspannungsgefühls

Die folgenden Diagramme (Abbildung 32 und 33) zeigen das subjektive Entspannungsgefühl der Probanden. Die Diagramme fassen die Daten des angepassten FEE von Prof. Thomas Loew zusammen. Sie sind in die Gruppen aufgeteilt und vergleichen das angegebene Entspannungsgefühl pro Teilnehmer vor und nach der Nutzung des Prototyps.

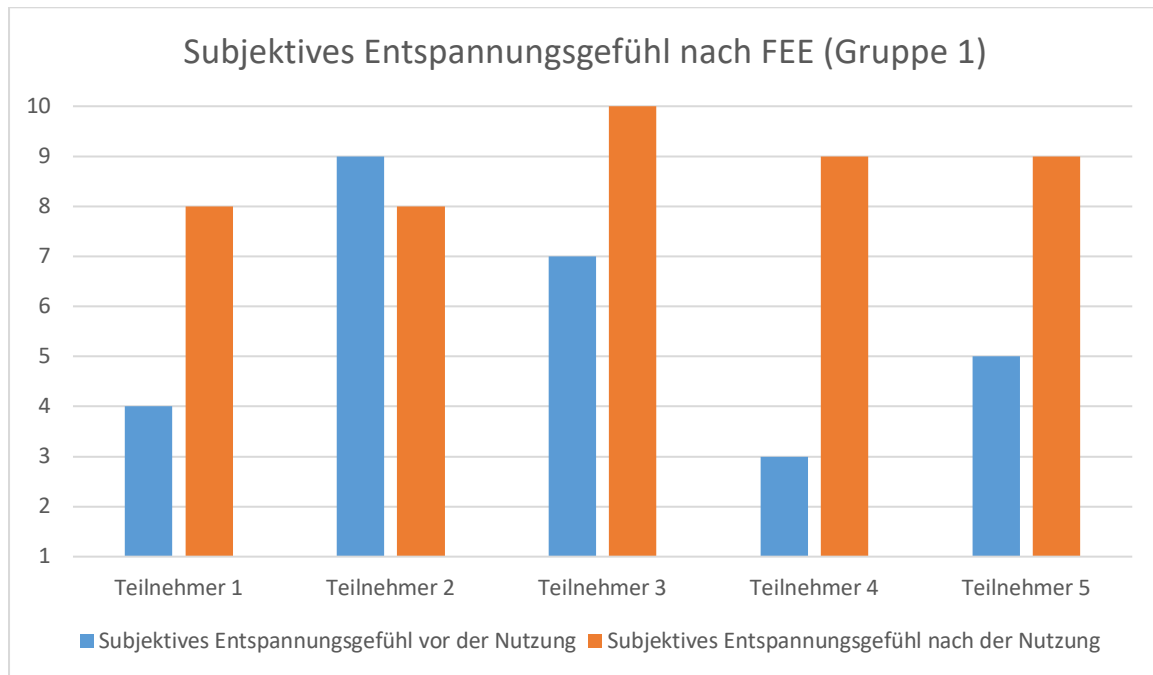


Abbildung 32: Subjektives Entspannungsgefühl vor und nach der Nutzung pro Teilnehmer (Gruppe 1)
Quelle: Eigene Darstellung

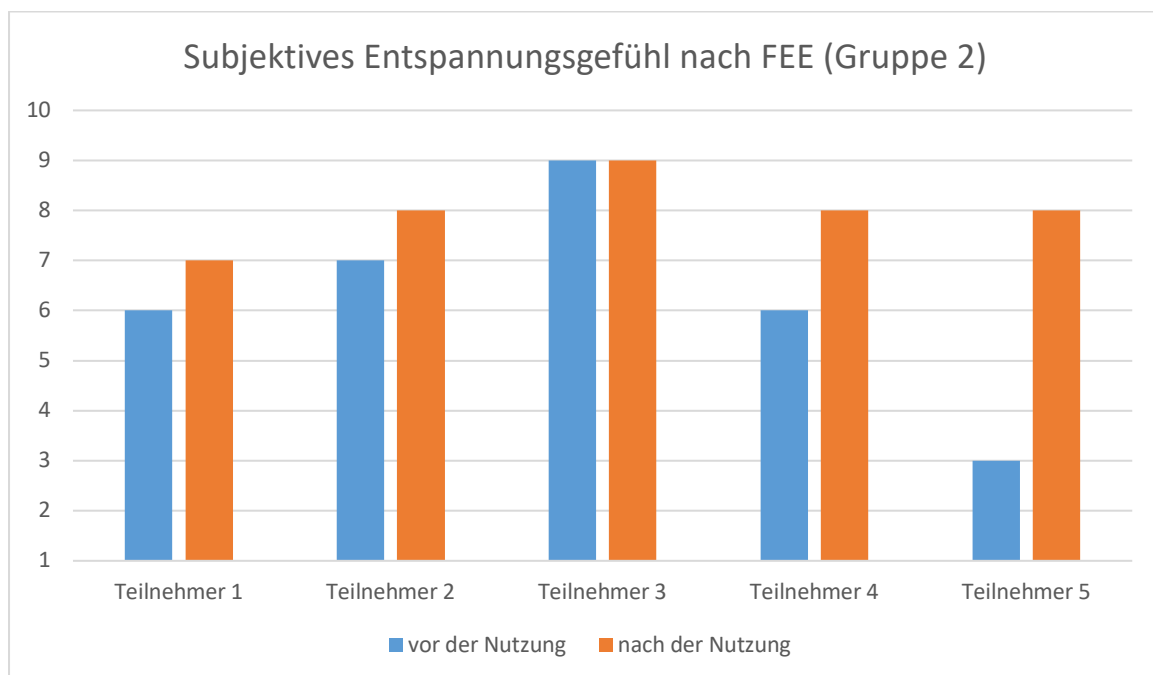


Abbildung 33: Subjektives Entspannungsgefühl vor und nach der Nutzung pro Teilnehmer (Gruppe 2)
Quelle: Eigene Darstellung

Die folgenden Diagramme (Abbildung 34 und 35) zeigen die Mittelwerte (M) und Standardabweichungen (SD) der Antworten aus dem Fragebogenteil, der auf dem RSQ basiert. Das erste Diagramm zeigt die Ergebnisse aller 10 Probanden vor und das zweite Diagramm nach der Nutzung des Prototyps.

Um eine bessere Lesbarkeit der Diagramme zu gewährleisten, wurden die einzelnen Items des Fragebogens abgekürzt. Die ungekürzten Aussagen sind die exakt gleichen Aussagen wie im RSQ und lauten wie folgt:

- Mein Atem ist schneller als gewöhnlich
- Mein Herz schlägt schneller als sonst
- Meine Muskeln fühlen sich angespannt an (Hand zur Faust geballt; Kiefer angespannt, gerunzelte Stirn)
- Meine Muskeln fühlen sich entspannt an
- Meine Muskeln fühlen sich locker an
- Ich fühle mich sehr entspannt
- Ich bin gerade vollkommen ruhig
- Ich bin schläfrig und müde
- Ich bin kurz davor einzunicken
- Ich fühle mich erfrischt und wach (Steghaus & Poth, 2022)

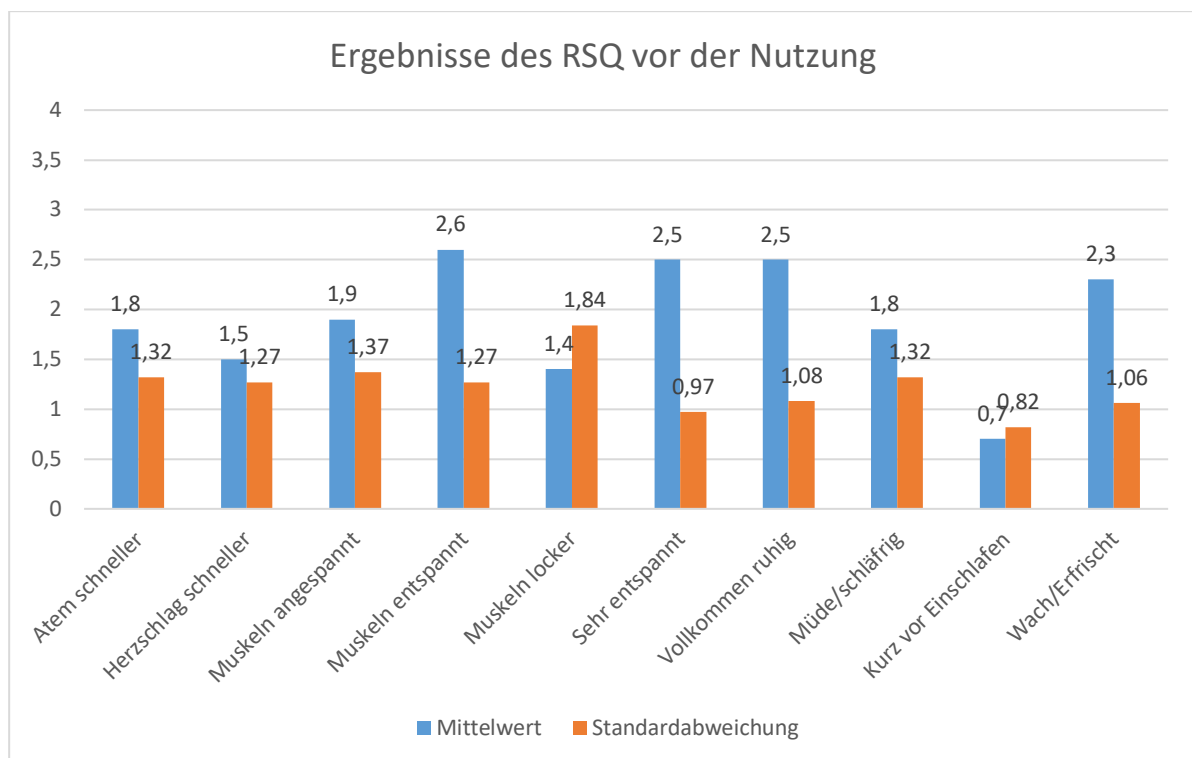


Abbildung 34: Mittelwerte (beider Gruppen) und Standardabweichungen der Ergebnisse vor der Nutzung
Quelle: Eigene Darstellung

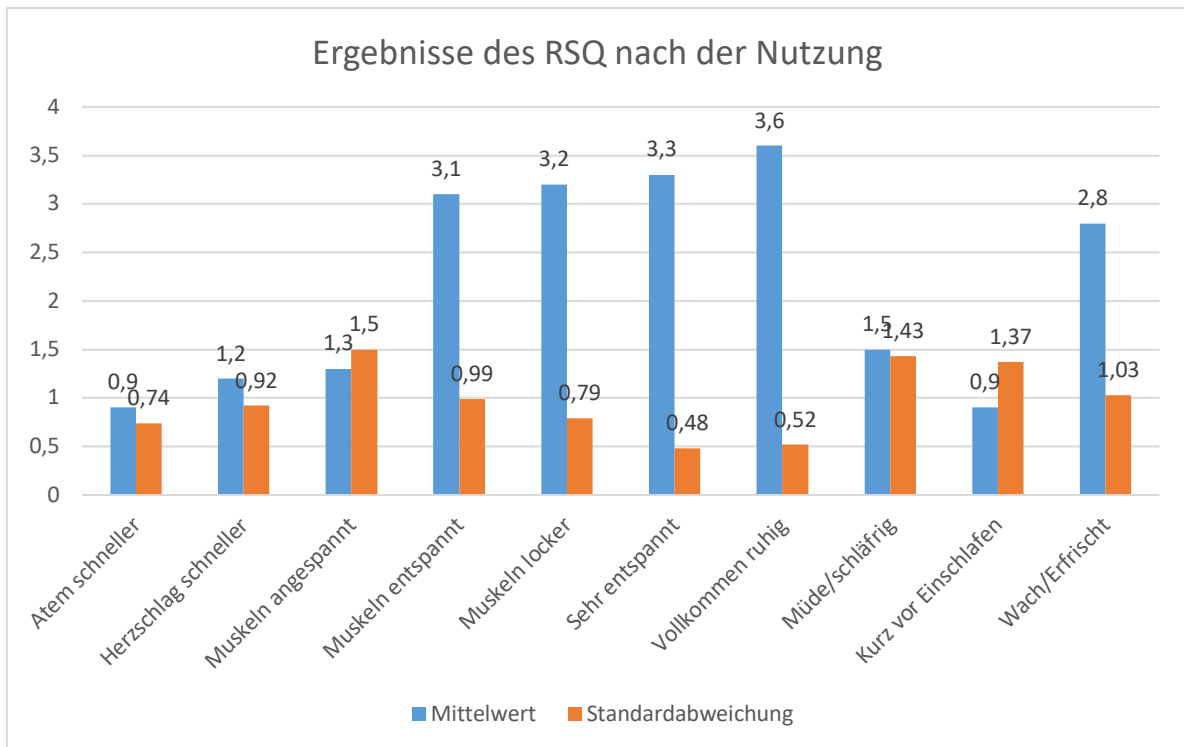


Abbildung 35: Mittelwerte (beider Gruppen) und Standardabweichungen der Ergebnisse nach der Nutzung
Quelle: Eigene Darstellung

5.3.1 Auswertung der Ergebnisse zur Entspannung

Die Diagramme in den Abbildungen 32 und 33 zeigen, dass bei 80% der Teilnehmer das subjektive Entspannungsempfinden nach der Nutzung des Prototyps leicht bis signifikant stieg. Lediglich bei einem Teilnehmer nahm das Entspannungsempfinden ab und bei einem weiteren blieb es auf dem gleichen hohen Niveau.

Hierbei ist anzumerken, dass der Teilnehmer mit der einen Abstieg des Entspannungsgefühls angab, ebenfalls angab, dass eine allgemeine Unruhe und viele Störfaktoren das Empfinden beeinflusst haben. Dies lag daran, dass dieser Teilnehmer (Teilnehmer 2, Gruppe 1) in der großen Mittagspause des Orchesters die Studie durchgeführt hat, wodurch es zu sehr vielen Störungen durch Außenstehende kam.

Auf der Webseite von Prof. Dr. Thomas Loew wird bezüglich des FEEs geschrieben, dass sich Probanden nach einer Entspannungsmaßnahme in der Regel um 2 Punkte entspannter fühlen (LoewDownloads, o. J.). Loew bezieht sich bei diesen Aussagen auf eine Skala von 0 bis 12, während in dieser Studie eine Skala von 1 bis 10 verwendet wurde, wodurch ein Anstieg von 2 Punkten auf seiner Skala einer etwas geringeren Veränderung auf der verwendeten Skala entspricht.

Die Teilnehmer dieser Studie haben im Durchschnitt eine Verbesserung von 2,5 Punkten, wobei der geringste Anstieg bei -1 Punkten (Teilnehmer 2, Gruppe 1) und der höchste Anstieg bei 6 Punkten (Teilnehmer 4, Gruppe 4) lag. Dadurch übersteigt die beobachtete Verbesserung trotz eines erkläraren negativen Ergebnisses leicht den von Prof Loews beschriebenen Durchschnittswert.

Die Verbesserung des subjektiven Entspannungsgefühls bestätigen auch die obigen Diagramme 34 und 35. Bei einem Vergleich der Mittelwerte von vor der Nutzung mit denen von nach der Nutzung kann festgestellt werden, dass sich die subjektiv wahrgenommenen kardiovaskulären Aktivitäten der Nutzer im Schnitt beruhigt haben. Die Mittelwerte der bereits gering bewerteten Aussagen *schnellerer Atem* und *schnellerer Herzschlag* wurden nach der Nutzung noch geringer, wobei bei der ersteren der Wert um 50% (Rückgang von 1,8 auf 0,9) und der zweite um 20% (Rückgang von 1,5 auf 1,2) gesunken ist.

Im Bereich Muskelspannung ist ebenfalls ein positives Ergebnis zu erkennen. Die Probanden fühlten weniger Anspannung in ihren Muskeln (Rückgang von 1,9 auf 1,3) und empfanden ihre Muskeln stattdessen entspannter (Anstieg von 2,6 auf 3,1) und deutlich lockerer (Anstieg von 1,4 auf 3,2).

Ihre allgemeine Entspannung bewerteten Probanden nach der Nutzung ebenfalls positiver. Die Aussage *Ich fühle mich sehr entspannt* zeigte einen Anstieg von 2,5 auf 3,3 (+32%) und die Aussage *Ich bin vollkommen ruhig* einen Anstieg von 2,5 auf 3,6 (+44%), was beides ebenfalls darauf hinweist, dass die Nutzung des Prototyps zu einer spürbaren Verbesserung des Entspannungsgefühls beitrug. Gleichzeitig halbierte sich die Standardabweichung beider Werte, was auf eine einheitlichere Wahrnehmung der Entspannung hindeutet.

Probanden fühlten sich zudem weniger schläfrig, aber dafür erwachter und frischer.

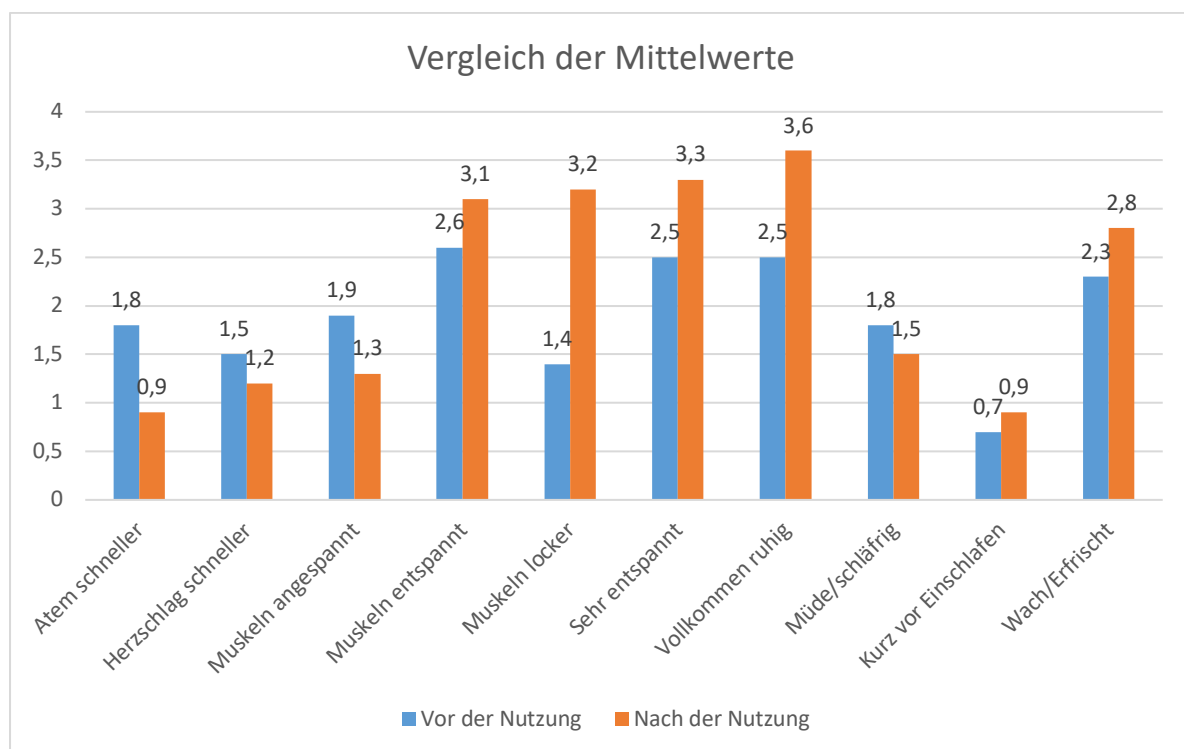


Abbildung 36: Vergleich der Mittelwerte der Diagramme Abb. 34 und 35 (ohne Standardabweichung)
Quelle: Eigene Darstellung

Die Auswertung dieser Ergebnisse zeigt, dass der VR-Prototyp einen positiven Einfluss auf das subjektive Entspannungsgefühl der meisten Nutzer nach der Nutzung hat.

5.4 Ergebnisse bezüglich des Prototyps

Die folgenden Diagramme zeigen die Bewertungen der Aussagen bezüglich der Interaktivität und Benutzerfreundlichkeit des Prototyps. Aufgeteilt sind diese in Kategorien, die auf den Anforderungen an den Prototyp (Kapitel 3.2) basieren. Auch bei diesen Diagrammen wurde die Beschriftung leicht angepasst, um eine bessere Lesbarkeit zu gewähren.

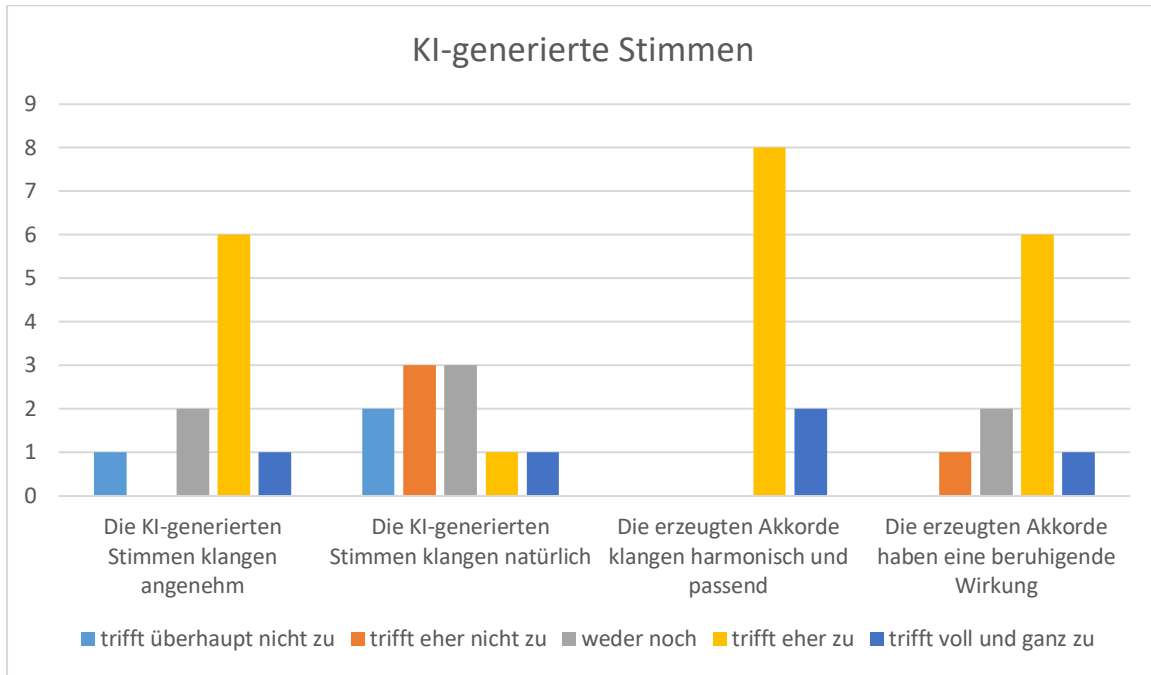


Abbildung 37: Verteilung der Antworten zur Wahrnehmung der KI-generierten Stimmen
Quelle: Eigene Darstellung

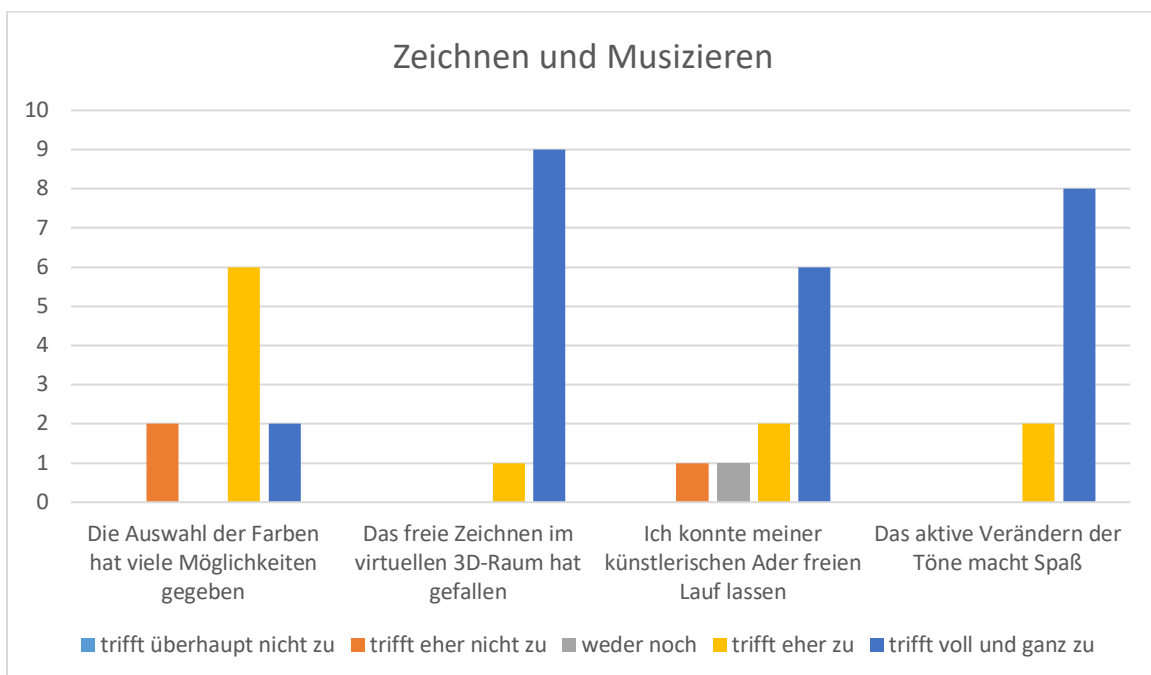


Abbildung 38: Verteilung der Antworten zur Zeichen- und Musizierfunktion
Quelle: Eigene Darstellung

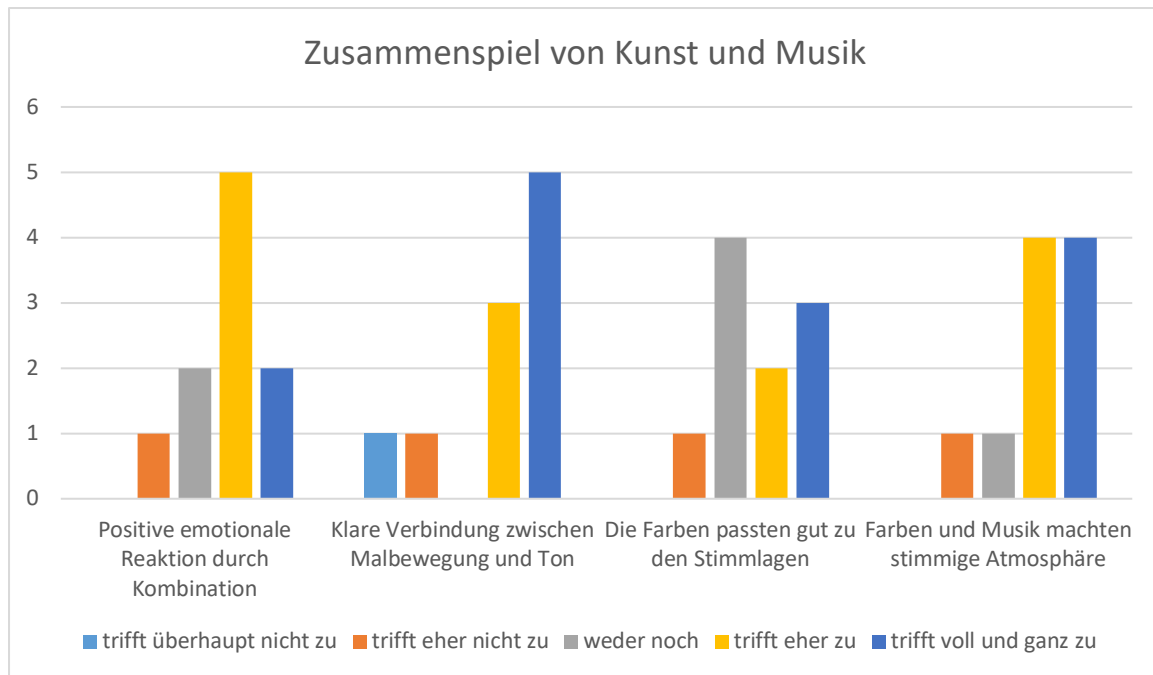


Abbildung 39: Verteilung der Antworten zum Zusammenspiel von Kunst und Musik
Quelle: Eigene Darstellung

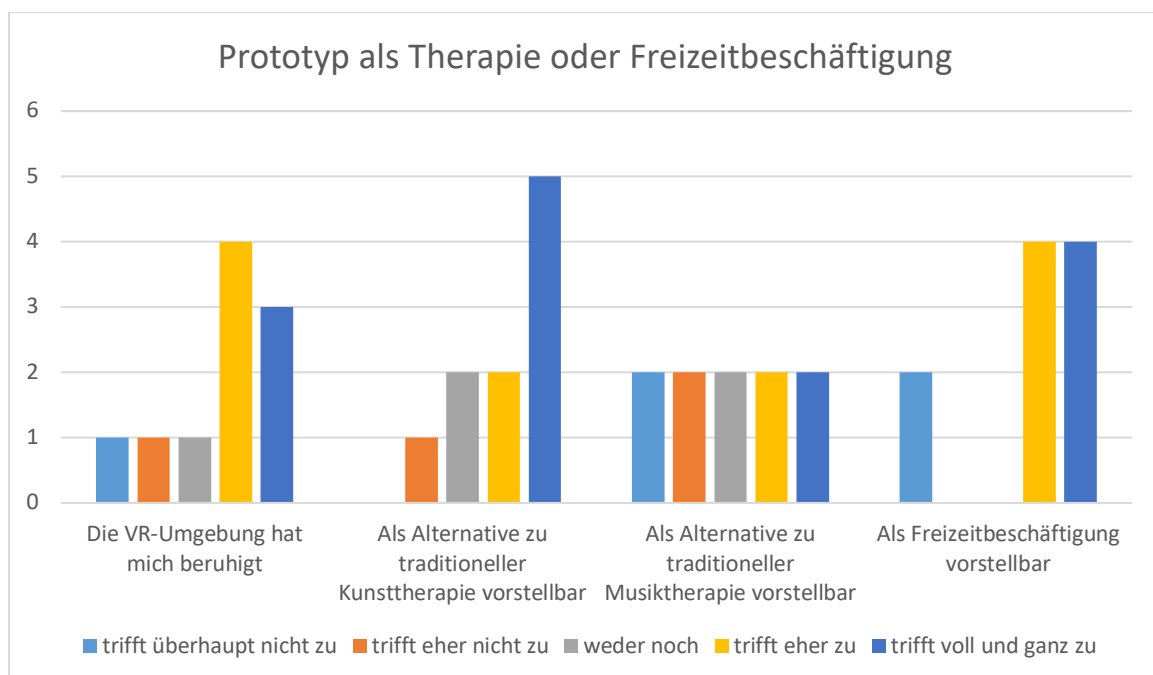


Abbildung 40: Verteilung der Antworten bezüglich des therapeutischen Nutzens des Prototyps
Quelle: Eigene Darstellung

Zum obigen Diagramm muss gesagt werden, dass die Aussagen *Ich kann mir vorstellen, diese Anwendung statt traditioneller Kunsttherapie zu nutzen* und *Ich kann mir vorstellen, diese Anwendung statt traditioneller Musiktherapie zu nutzen* während der Durchführung der Studie mündlich zu *Ich kann mir vorstellen, diese Anwendung eher zu nutzen als klassisches Malen auf Papier* und *Ich kann mir vorstellen, diese Anwendung eher zu nutzen als das klassische Musizieren mit Instrumenten* geändert wurden. Dies war notwendig, da keiner der Probanden angegeben hat, Vorerfahrungen mit Kunst- oder

Musiktherapie zu haben und daher nicht wussten, wie traditionelle Kunst- oder Musiktherapie abläuft. Die Veränderung kam wegen der Nachfrage des ersten Teilnehmers und wurde daraufhin einheitlich kommuniziert und von allen weiteren Teilnehmern verstanden.

Das folgende Diagramm unterscheidet sich von den anderen. Während die vorherigen vier Diagramme die Verteilung der Antworten gezeigt haben, zeigt das nächste Diagramm Mittelwerte der Antworten. Dies hilft der Lesbarkeit des Diagramms, da die folgende Kategorie *Benutzerfreundlichkeit und Interaktivität* aus mehreren Einzelkategorien besteht, die jeweils auch unterteilt sind. Beispielsweise gehören zur Kategorie *Benutzerfreundlichkeit*: *Malen* Aussagen zur Einfachheit der gestellten Zeichen-Aufgaben und die Steuerung beim Wechseln der Farben. Die Ergebnisse werden daher zusammengefasst dargestellt.

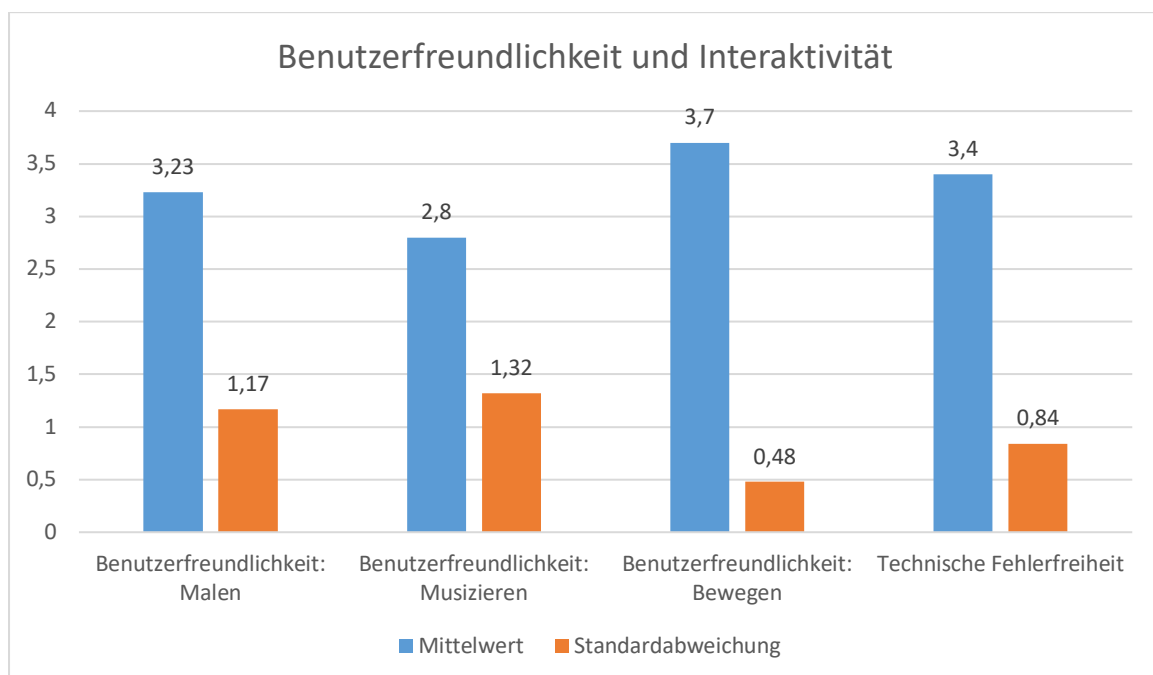


Abbildung 41: Mittelwerte und Standardabweichung bei Antworten zur Benutzerfreundlichkeit und Interaktivität
Quelle: Eigene Darstellung

5.4.1 Auswertung der Ergebnisse zum Prototyp

Die Auswertung der Ergebnisse bezüglich Benutzerfreundlichkeit und Interaktivität erfolgt anhand der acht definierten Anforderungen (A1 bis A8). Dabei werden die oben abgebildeten Diagramme entsprechend den Anforderungen geordnet und interpretiert. Da bei dieser Auswertung keine Mittelwerte miteinander verglichen werden, gilt für diese Evaluation: Werte über 2,5 werden als positiv gewertet, da sie über der neutralen Mitte (2) liegen. Werte über 3 gelten als sehr positiv, da sie nahe am Maximum (4) sind.

Die ersten beiden Kategorien des obigen Diagramms in Abbildung 37 beziehen sich auf **A1: Die KI-generierten Stimmen sollen die Nutzererfahrung nicht beeinträchtigen**. Die folgende Tabelle zeigt die Mittelwerte und Standardabweichungen der Antworten zu den beiden Aussagen.

Tabelle 4: Mittelwerte und Standardabweichung der Fragen bezüglich der KI-generierten Stimmen
Quelle: Eigene Darstellung

Aussage	Mittelwert	Standardabweichung
Die KI-generierten Stimmen klangen angenehm	2,6	1,08
Die KI-generierten Stimmen klangen natürlich	1,6	1,27

Diese Werte zeigen, dass die generierten Stimmen zwar nicht als besonders natürlich empfunden wurden, aber dennoch als angenehm wahrgenommen wurden. Dies deutet darauf hin, dass die geringe Natürlichkeit der Stimmen nicht zwingend zu einer zu einer negativen Nutzerfahrung geführt haben.

Dadurch kann gesagt werden, dass Anforderung A1 grundsätzlich erfüllt wurde, da die Stimmen insgesamt eher positiv bewertet wurden. Dennoch besteht Verbesserungspotenzial, besonders was die Natürlichkeit der Stimmen angeht.

Der zweite Teil des Diagramms in Abbildung 37 bezieht sich auf die erzeugten Akkorde und die damit verbundene Anforderung **A2: Die gesungenen Akkorde im Misch-Modus sollen harmonisch klingen und passen**. Auch diese Anforderung kann als erfüllt betrachtet werden, da der Mittelwert der Antworten (dargestellt in nachfolgender Tabelle) sehr hoch und die Standardabweichung dabei relativ gering ist. Die Aussage, die sich direkt auf die Anforderung bezieht, wurde im Mittel mit 3,2 (SD = 0,42) bewertet, was einer sehr positiven Bewertung entspricht.

Tabelle 5: Mittelwerte und Standardabweichung der Fragen bezüglich der erzeugten Akkorde
Quelle: Eigene Darstellung

Aussage	Mittelwert	Standardabweichung
Die erzeugten Akkorde klingen harmonisch und passend	3,2	0,42
Die erzeugten Akkorde haben eine beruhigende Wirkung auf mich gehabt	2,7	0,82

Das obige Diagramm in Abbildung 38 zeigt, dass sowohl die dritte Anforderung **A3: Es soll genug Möglichkeiten und Motivation geben, im dreidimensionalen Raum nach Belieben zu malen** als auch die vierte Anforderung **A4: Es soll motivieren, frei zu musizieren** beide als erfüllt betrachtet werden können. Die Aussagen zum freien Zeichnen im virtuellen Raum und dem aktiven Verändern der Musik sind mit 3,9 und 3,8 von 4 im Mittelwert sogar die am höchsten bewerteten Aussagen. Auch Anforderung **A5: Die Kombination aus Zeichnen und Musizieren soll eine stimmige Atmosphäre schaffen** ist als erfüllt, da die Aussagen zu dieser Kategorie ebenfalls überwiegend positiv bewertet wurden.

Die nachfolgende Tabelle zeigt die Mittelwerte und Standardabweichungen zu den Aussagen, die sich auf A3, A4 und A5 beziehen.

Tabelle 6: Mittelwerte und Standardabweichungen zum Zeichnen und Malen und der Kombination aus beiden
Quelle: Eigene Darstellung

Aussage	Mittelwert	Standardabweichung
Die Auswahl der Farben hat mir viele gestalterische Möglichkeiten gegeben	2,8	1,03
Das freie Zeichnen im virtuellen 3D-Raum hat mir gut gefallen	3,9	0,32
Ich konnte meiner künstlerischen Ader freien Lauf lassen	3,3	1,06
Das aktive Verändern der Töne hat mir Spaß gemacht	3,8	0,42
Die Kombination aus Klängen und Farben hat eine positive emotionale Reaktion in mir ausgelöst	2,8	0,92
Ich konnte eine klare Verbindung zwischen Malbewegung und Klängen erkennen	3	1,41
Die Farben passten gut zu den Stimmlagen	2,7	1,06
Die Farben und Musik haben eine stimmige Atmosphäre geschaffen	3,1	0,99

Anforderung A6: Nutzer sollen sich die Anwendung als Alternative zur traditionellen Kunst- und Musiktherapie oder als Freizeitbeschäftigung vorstellen können kann aufgrund der veränderten Aussagen in dieser Form nicht als erfüllt betrachtet werden. Selbst dann, wenn die Anforderung an die neuen Aussagen angepasst wird, zeigen die Ergebnisse nur eine teilweise Erfüllung der Anforderung.

Die geänderten Aussagen *Ich kann mir vorstellen, diese Anwendung eher zu nutzen als klassisches Malen auf Papier* und *Ich kann mir vorstellen, diese Anwendung in meiner Freizeit zur Entspannung zu nutzen* wurden sehr positiv ($M = 3,1$ $SD = 1,10$) und positiv ($M = 2,8$ $SD = 1,55$) bewertet. Die Aussage den Prototyp als Alternative zu traditionellen Instrumenten wurde jedoch neutral ($M = 2$ $SD = 1,49$) bewertet. Dabei ist zu beobachten, dass sich die Antworten innerhalb der Gruppen stark unterscheiden. Gruppe 1 mit den Orchestermusikern hat den Prototyp eher als Alternative zu traditionellen Instrumenten gesehen als Gruppe 2 mit den Büromitarbeitern. Der Mittelwert für Gruppe 1 zu dieser

Aussage war 2,6 (SD = 1,67), während Gruppe 2 einen Mittelwert von 1,4 (SD = 1,14) hatte. Diese Ergebnisse legen nahe, dass erfahrene Musiker den Prototyp eher als eine zusätzliche Möglichkeit zur musikalischen Entfaltung betrachten, während Nicht-Musiker eher bei klassischen Instrumenten bleiben möchten.

Tabelle 7: Mittelwerte und Standardabweichungen zu Aussagen zum Prototyp als Alternative
Quelle: Eigene Darstellung

Aussage	Mittelwert	Standardabweichung
Ich kann mir vorstellen, diese Anwendung eher zu nutzen als klassisches Malen auf Papier	3,1	1,10
Ich kann mir vorstellen, diese Anwendung eher zu nutzen als das klassische Musizieren mit Instrumenten	2,8	1,55
Ich kann mir vorstellen, diese Anwendung in meiner Freizeit zur Entspannung zu nutzen	2	1,49

Zuletzt können die beiden Anforderungen **A7: Die Steuerung soll intuitiv und nicht verwirrend sein** und **A8: Die Interaktion soll ohne wahrnehmbare technische Verzögerungen oder Störungen erfolgen** wiederum als erfüllt angesehen werden. Aus dem Diagramm in Abbildung 41 kann entnommen werden, dass der niedrigste Mittelwert bei den Aussagen zur Benutzerfreundlichkeit 2,8 (SD = 1,32) ist, während alle anderen Aussagen mit über 3 bewertet wurden. Am schlechtesten, aber trotzdem positiv, wurde die Steuerung im Mischmodus bewertet (M = 2,7 SD = 1,34). Die Probanden brauchten die meiste Zeit, um sich mit dieser vertraut zu machen.

Die Aussage zu Anforderung A8 wurde mit 3,4 (SD = 0,84) bewertet. Es kam bei keinem der Probanden zu irgendwelchen technischen Problemen, sodass das VR-Erlebnis verzögerungs- und störungsfrei verlief.

Die folgende Tabelle zeigt die Mittelwerte und Standardabweichungen aller einzelnen Aussagen, die im Diagramm aus Abbildung 41 in die Kategorien *Benutzerfreundlichkeit: Malen*, *Benutzerfreundlichkeit: Musizieren* und *Benutzerfreundlichkeit: Bewegen* zusammengefasst wurden.

Tabelle 8: Mittelwerte und Standardabweichung aller Einzelaussagen zur Kategorie Steuerung
 Quelle: Eigene Darstellung

Aussage	Mittelwert	Standardabweichung
Die Aufgabe, die vorgegebene Melodie zu spielen, war einfach und intuitiv	2,8	1,32
Die Aufgabe, das vorgegebene Objekt zu malen, war einfach und intuitiv	3,7	0,68
Die Steuerung des Pinsels war einfach und intuitiv	3,2	1,23
Das Wechseln der Farben war einfach und intuitiv	3,3	1,25
Das Mischen der Farben war einfach und intuitiv	2,7	1,34
Die Bewegung im Raum war angenehm und nicht verwirrend	3,7	0,48

5.5 Auswertung der Kommentare und Beobachtungen

Der Prototyp hat von allen Teilnehmern positives Feedback bekommen. Besonders das erste freie Malen im virtuellen Raum und das erste Erklingen eines vierstimmigen Akkordes haben die Probanden positiv überrascht. Einige Teilnehmer der Gruppe 1 (Orchester) haben die zusätzliche Experimentierzeit am Ende des Tests dafür genutzt, um Stücke zu musizieren, die gerade in der Probe behandelt wurden. Teilnehmer der Gruppe 2 nutzten dabei eher die Zeit, um weitere Dinge zu malen.

Nach dem Abnehmen des HMDs waren viele Teilnehmer enttäuscht wieder in den Alltag zu müssen und bedankten sich für die – wie ein Teilnehmer schrieb – „tolle neue Erfahrung“ (Teilnehmer 5, Gruppe 1).

Teilnehmer 5 der Gruppe 2 nahm als letzter an der Studie teil und nutzte beinahe 15 Minuten, um mit dem Prototyp weiter zu experimentieren. Als anschließender Kommentar wurde folgendes hinterlassen:

„Das VR-Erlebnis war insgesamt sehr beruhigend und inspirierend. Ich hatte das Gefühl, dass mein inneres Kind dadurch zum Vorschein gekommen ist, weil ich durch das Malen im freien Raum eine Assoziation mit einem Film aus meiner Kindheit hatte. Das hat in dem Moment viel Freude in mir ausgelöst. Es war ein besonderes Gefühl rund um sich herum in der Luft malen zu können, da man das im Alltag so nicht praktiziert. Was ich noch sehr schön gefunden hätte, wäre, wenn man zusätzlich zum Farbbänderungs-Modus noch eine Musikänderungs-Option gehabt hätte. Beispielsweise wäre es

sicherlich auch entspannend, wenn eine Farbe für ein einzelnes Instrument stehen würde (z. B. Grün für eine Gitarre, Blau für ein Klavier etc.) und man dadurch die Instrumente beim Malen kombinieren könnte.“ (Teilnehmer 5, Gruppe 2).

Andere Teilnehmer gaben ebenfalls an, dass ihnen die Anwendung sehr gefallen hat, und sie, wenn sie mehr Zeit hätten, gerne weiter experimentiert hätten. Auch der Teilnehmer mit einem negativen Ergebnis schrieb trotz eines unruhigen Studienverlaufs einen positiven Kommentar: „Die unruhige Situation bei der Erklärung und Ausführung des Ganzen hat das Empfinden ein wenig verändert. Sonst ist das Ganze eine sehr interessante Erfahrung“ (Teilnehmer 2, Gruppe 1).

Diese positiven Beobachtungen und Kommentare deuten darauf hin, dass der Prototyp seinen Zweck weitestgehend erfüllt hat, indem eine beruhigende und entspannende, fast schon spielerische Umgebung geschaffen wurde, in der sich die Teilnehmer abgekoppelt von der physischen Realität künstlerisch und musikalisch austoben konnten.



Abbildung 42: Bearbeitung der Aufgabe Sonne und Blume zeichnen mit ergänzendem Fluss und Wiese von Teilnehmer 5 der Gruppe 2 (Screenshot innerhalb des HMDs)
Quelle: Eigene Darstellung



Abbildung 43: Musikalische Experimente von Teilnehmer 3 der Gruppe 1 (Screenshot vom Livestream auf Smartphone)
Quelle: Eigene Darstellung

Da jedoch auch explizit nach Verbesserungsvorschlägen gefragt wurde, bestanden die geschriebenen Kommentare nicht nur aus positiven Danksagungen, sondern auch aus Kritik und Anregungen für neue Features oder Verbesserung bestehender Funktionen.

Neben der von Teilnehmer 5 (Gruppe 2) gewünschten Musikänderungs-Option, die neben Gesangsstimmen auch Instrumentalmusik ermöglichen soll, wurden sich etwas natürlichere Stimmen gewünscht, eine Option zum Löschen des Gemalten und virtuelle Wände, die den Raum, in dem man sich befindet, abbilden sollen.

Da es für die allermeisten Teilnehmer die erste VR-Erfahrung war, gaben viele an, dass sie von der Umgebung und der Immersion positiv überrascht waren. Nur ein Teilnehmer berichtete von einem leichten Schwindelgefühl nach dem Erlebnis.

5.6 Limitationen

Die Studie liefert zwar größtenteils positive Ergebnisse, dennoch gibt es einige Limitationen, auf die Rücksicht zu nehmen ist. Obwohl im Vorfeld darauf geachtet wurde, eine breite Altersspanne (19 bis 64 Jahre) sowie eine ausgewogene Geschlechterverteilung (60 % Frauen, 40 % Männer) zu berücksichtigen, bleibt die Stichprobengröße dennoch gering. Da die Stichprobe mit nur 10 Teilnehmern relativ klein ist, sind die Ergebnisse nur eingeschränkt verallgemeinerbar. Eine größere Anzahl an Probanden könnte ein aussagekräftigeres Ergebnis liefern und die Effektivität des VR-Prototyps in unterschiedlichen Kontexten genauer bewerten.

Die Teilnehmer der Gruppe 2 (Büro) führten die Studie in einem nahezu leeren Raum auf einem fast ungenutzten Stockwerk durch, um ein möglichst isoliertes Umfeld zu schaffen, das einer echten Therapiesitzung ähnelt. Im Gegensatz dazu befanden sich die Teilnehmer der Gruppe 1 (Orchester) zwar ebenfalls in einem separaten Raum, jedoch direkt neben dem Orchesterprobenraum. Dadurch kam es gelegentlich zu Störungen und Zeitdruck, da die Studie in einer zeitlich limitierten Pause stattfand. Zudem kommt hinzu, dass es sich bei dem Durchführenden der Studie nicht um einen ausgebildeten Kunst- oder Musiktherapeuten handelt. Dadurch kann trotz der beobachtbaren positiven Effekte in der Stressbewältigung und Entspannung nicht entschieden werden, ob der Prototyp sich letztendlich für den Therapieeinsatz eignet. Für eine fundiertere Einschätzung wäre eine Untersuchung unter Anleitung ausgebildeter Kunst- oder Musiktherapeuten erforderlich.

Diese Limitationen zeigen auf, dass dieser Prototyp aufgrund seiner vielversprechenden Ergebnisse zwar das Potenzial eines Konzepts der KI-gestützten Kunst- und Musiktherapie bietet, jedoch weitere Untersuchungen und Verbesserungen erforderlich sind, um seine tatsächliche Eignung und Wirksamkeit in therapeutischen Anwendungen zu beurteilen.

6 Fazit

Zusammenfassend zeigen die Ergebnisse dieser Studie, dass der entwickelte VR-Prototyp das subjektive Entspannungsgefühl der meisten Teilnehmer deutlich steigern konnte. Bei 80% der Probanden war nach der Nutzung eine Zunahme der Entspannung zu beobachten. Bei einem Probanden blieb das Entspannungsgefühl auf demselben bereits sehr hohen Niveau und nur bei einem Probanden sank es. Letzteres kann jedoch auf externe Störfaktoren während der Durchführung zurückgeführt werden. Durch die positiven Ergebnisse des FEEs und des RSQs kann Hypothese **H1: Der VR-Prototyp hat einen positiven Einfluss auf das subjektive Entspannungsgefühl unmittelbar nach der Nutzung** weitgehend bestätigt werden. Die Mehrheit der Teilnehmer berichtete von einer Steigerung der Entspannung, einer Verringerung der Muskelspannung und kardiovaskulärer Aktivität sowie einer allgemeinen Beruhigung.

Auch Hypothese **H2: Der VR-Prototyp erfüllt die im Konzept definierten Anforderungen an Benutzerfreundlichkeit und Interaktivität** konnte in großen Teilen bestätigt werden. Die Mehrheit der Anforderungen, darunter eine intuitive Steuerung, die Möglichkeit zur kreativen Gestaltung und eine zusammenpassende audiovisuelle Erfahrung, wurden von den Teilnehmern als erfüllt wahrgenommen. Insbesondere das Zeichnen im dreidimensionalen Raum, welches vom Konzept des bewährten Programms *Tilt Brush* inspiriert ist, wurde als angenehm und inspirierend beschrieben und die Kombination aus visuellen und auditiven Elementen als stimmig empfunden.

Allerdings gab es einige Einschränkungen. So wurde die Einbindung der generierten KI-Stimmen teils als unnatürlich empfunden, was sich negativ auf die Immersion und das Erlebnis auswirken könnte.

Zudem musste die Anforderung, ob der Prototyp als Alternative zur traditionellen Kunst- oder Musiktherapie angesehen wird, verändert werden. Aufgrund fehlender Erfahrung mit diesen Therapieformen konnten die Teilnehmer dies nicht einschätzen. Die umformulierte Frage, ob der Prototyp eine Alternative zum analogen Malen oder Musizieren darstellen kann, wurde auch nur teilweise bejaht. Da sich jedoch die meisten Teilnehmer vorstellen können, den Prototyp in ihrer Freizeit zur Entspannung zu nutzen, kann der VR-Prototyp durchaus als sinnvolle Ergänzung gesehen werden, und nicht als Ersatz oder Alternative für traditionelle Kunst- oder Musiktherapie.

Insgesamt zeigen die Ergebnisse, dass der entwickelte VR-Prototyp großes Potenzial zur Erhöhung des Entspannungsgefühls bietet und von Nutzern als bereichernde Erfahrung wahrgenommen wurde. Während es Optimierungspotenziale gibt, wie die Qualität der KI-Stimmen und der Steuerung im Mischmodus, bestätigen die positiven Rückmeldungen die grundlegende Wirksamkeit und Benutzerfreundlichkeit des Prototyps.

Allerdings wäre für eine allgemeinere Bewertung wichtig, eine größere Stichprobe einzubeziehen und die Anwendung von Fachleuten aus dem Bereich Kunst- und Musiktherapie prüfen zu lassen. Eine

professionelle Einschätzung könnte dabei helfen, die langfristige Wirksamkeit sowie weitere mögliche Einsatzbereiche in anderen therapeutischen Kontexten zu untersuchen.

6.1 Ausblick

Aufbauend auf den positiven Ergebnissen dieser Studie sind weitere Optimierungen und Untersuchungen geplant. Zukünftig sollen alle identifizierten Verbesserungsvorschläge in eine Weiterentwicklung des Prototyps fließen. Insbesondere die Verbesserung der KI-Stimmen, eine bessere Steuerungsmöglichkeit innerhalb des Mischmodus und einer eventuellen Einbindung von weiteren musikalischen Instrumenten als auswählbare Option.

Ein weiterer wichtiger Schritt ist die Einbindung von Fachleuten aus dem Bereich Kunst- und Musiktherapie, um eine professionelle Einschätzung der Wirksamkeit und des therapeutischen Potenzials des VR-Prototyps zu erhalten. Falls sich der Prototyp in zukünftigen Tests eben als ausreichend ausgereift und wirksam erweist, könnte er längerfristig vielleicht als unterstützende Therapieoption eingesetzt werden. Es wäre durchaus denkbar, dass der ausgereifte Prototyp besonders Menschen mit eingeschränktem Zugang zu traditionellen Therapieangeboten eine sinnvolle Ergänzung bietet. Menschen, die keine regelmäßigen Therapiestunden vor Ort wahrnehmen können, könnten in ihrer gewohnten Umgebung mithilfe eines eigenen oder gestellten HMDs ein kreatives und entspannungsförderndes Erlebnis erhalten.

Diese Perspektiven eröffnen spannende Möglichkeiten für die zukünftige Forschung und Entwicklung. Die positiven Ergebnisse der ersten Evaluation unterstreichen das Potenzial des VR-Prototyps, sodass eine weiterführende Optimierung und umfassendere Untersuchung einen vielversprechenden nächsten Schritt darstellen.

Literaturverzeichnis

- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-94463-0>
- Alexanian, S. (2021). *Virtual Reality and music therapy's impact on people's psychological well-being*. American Psychological Association. (2021). *COVID-19 Practitioner Survey*. Abgerufen 22. März 2025 von <https://www.apa.org/pubs/reports/practitioner/covid-19-2021>
- Audacity. (o. J.). Audacity ® | Frequently Asked Questions. Abgerufen 17. März 2025, von <https://www.audacityteam.org/faq/>
- Benett, G., & Rodet, X. (1989). Synthesis of the singing voice. In *Current Directions in Computer music research* (S. 19–44). MIT Press.
- Bernatzky, G., & Kreutz, G. (Hrsg.). (2015). *Musik und Medizin: Chancen für Therapie, Prävention und Bildung*. Springer Vienna. <https://doi.org/10.1007/978-3-7091-1599-2>
- Bishop, C. M. (1994). *Mixture density networks*. Aston University
- Blob Opera. (o. J.). Google Arts & Culture. Abgerufen 13. März 2025, von <https://artsandculture.google.com/experiment/blob-opera/AAHWrq360NcGbw>
- Bradt, J., Dileo, C., & Potvin, N. (2013). Music for stress and anxiety reduction in coronary heart disease patients. *Cochrane Database of Systematic Reviews*, 2021(9). <https://doi.org/10.1002/14651858.CD006577.pub3>
- Brill, M. (2023). *Virtual Reality kompakt: Entwicklung von immersiver Software*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-41245-6>
- Brungardt, A., Wibben, A., Tompkins, A. F., Shanbhag, P., Coats, H., LaGasse, A. B., Boeldt, D., Youngwerth, J., Kutner, J. S., & Lum, H. D. (2021). Virtual Reality-Based Music Therapy in Palliative Care: A Pilot Implementation Trial. *Journal of Palliative Medicine*, 24(5), 736–742. <https://doi.org/10.1089/jpm.2020.0403>
- Bruscia, K. E. (1998). *An introduction to music psychotherapy*.
- Bullerjahn, C. (2004). *Musik im Alltag (unabridged and unedited)*.
- Burianek, S. (2021). *Klirrende Klänge*. OPE[R]NTHEK. Abgerufen 22. März 2025, von <https://opern.news/thek/kapitel/opern.news/thek/kapitel/1689>

- Cabezas, C. A., Arcos, A. R., Carrillo-Medina, J. L., & Arias-Almeida, G. I. (2022). Effects of Virtual Reality and Music Therapy on Academic Stress Reduction Using a Mobile Application. In M. Botto-Tobar, O. S. Gómez, R. Rosero Miranda, A. Díaz Cadena, S. Montes León, & W. Luna-Encalada (Hrsg.), *Trends in Artificial Intelligence and Computer Engineering* (Bd. 407, S. 45–59). Springer International Publishing. https://doi.org/10.1007/978-3-030-96147-3_4
- Conn, C., Lanier, J., Minsky, M., Fisher, S., & Druin, A. (1989). Virtual environments and interactivity: Windows to the future. *ACM SIGGRAPH Computer Graphics*, 23(5), 7–18. <https://doi.org/10.1145/77277.77278>
- Darewych, O. H., Carlton, N. R., & Farrugie, K. W. (2015). Digital Technology Use in Art Therapy with Adults with Developmental Disabilities. *Journal on Developmental Disabilities; Toronto*, 21(2), 95–102.
- De Witte, M., Pinho, A. da S., Stams, G.-J., Moonen, X., Bos, A. E. R., & van Hooren, S. (2022). Music therapy for stress reduction: A systematic review and meta-analysis. *Health Psychology Review*, 16(1), 134–159. <https://doi.org/10.1080/17437199.2020.1846580>
- De Witte, M., Spruit, A., Van Hooren, S., Moonen, X., & Stams, G.-J. (2020). Effects of music interventions on stress-related outcomes: A systematic review and two meta-analyses. *Health Psychology Review*, 14(2), 294–324. <https://doi.org/10.1080/17437199.2019.1627897>
- Der vierstimmige Satz. (o. J.). Abgerufen 15. März 2025, von <http://www.lehrklaenge.de/PHP/Harmonielehre1/VierstimmigerSatz.php>
- DMtG. (o. J.). Was ist Musiktherapie? *Deutsche Musiktherapeutische Gesellschaft*. Abgerufen 4. März 2025, von <https://www.musiktherapie.de/musiktherapie/was-ist-musiktherapie/>
- FEE. (o. J.). Stress-Fragebogen. Abgerufen 26. März 2025, von http://prof-loew.de/data/documents/STRESS4_FEE.pdf
- Blender Foundation. (o. J.). *Blender*. Blender.Org - Home of the Blender Project - Free and Open 3D Creation Software. Abgerufen 19. März 2025, von <https://www.blender.org/>
- Garner, R. (2016). *Digital Art Therapy: Material, Methods, and Applications*. Jessica Kingsley Publishers.

- Gold, C., Solli, H. P., Krüger, V., & Lie, S. A. (2009). Dose–response relationship in music therapy for people with serious mental disorders: Systematic review and meta-analysis. *Clinical Psychology Review*, 29(3), 193–207. <https://doi.org/10.1016/j.cpr.2009.01.001>
- Goodfellow, I., Bengio, Y., & Courville, A. (2018). *Deep Learning: Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze* (G. Lenz, Übers.; 1. Auflage). mitp.
- Google Formulare. (o. J.). Abgerufen 24. März 2025, von <https://docs.google.com/forms/u/0/>
- Heinrichs, B., Heinrichs, J.-H., & Rütger, M. (2022). *Künstliche Intelligenz*. De Gruyter.
- Heinrichs, M., Stächele, T., & Domes, G. (2015). *Stress und Stressbewältigung*. Hogrefe.
- Hono, Y., Hashimoto, K., Oura, K., Nankaku, Y., & Tokuda, K. (2021). Sinsy: A Deep Neural Network-Based Singing Voice Synthesis System. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 2803–2815. <https://doi.org/10.1109/TASLP.2021.3104165>
- Hono, Y., Takaki, S., Hashimoto, K., Oura, K., Nankaku, Y., & Tokuda, K. (2021). Periodnet: A Non-Autoregressive Waveform Generation Model with a Structure Separating Periodic and Aperiodic Components. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6049–6053. <https://doi.org/10.1109/ICASSP39728.2021.9414401>
- HTML RGB and RGBA Colors. (o. J.). Abgerufen 16. März 2025, von https://www.w3schools.com/html/html_colors_rgb.asp
- IBMNN. (o. J.). What Is a Neural Network? | IBM. Abgerufen 13. März 2025, von <https://www.ibm.com/think/topics/neural-networks>
- Isdale, J. (1993). *What is Virtual Reality*. Abgerufen 26. März 2025, von <https://www.columbia.edu/~rk35/vr/vr.html>
- Jones, P., Osborne, T., Sullivan-Drage, C., Keen, N., & Gadsby, E. (2022). *Virtual Reality Methods*. Policy Press. <https://doi.org/10.47674/9781447360773>
- Kaimal, G., Ray, K., & Muniz, J. (2016). Reduction of Cortisol Levels and Participants’ Responses Following Art Making. *Art Therapy*, 33(2), 74–80. <https://doi.org/10.1080/07421656.2016.1166832>

- Kasseler Thesen zur Musiktherapie.* (o. J.). Abgerufen 4. März 2025, von <https://www.musiktherapie.de/wp-content/uploads/2019/07/Kasseler-Thesen-zur-Musiktherapie.pdf>
- Kenmochi, H., & Ohshita, H. (2007). *VOCALOID - commercial singing synthesizer based on sample concatenation*. 4009–4010. Abgerufen 26. März 2025, von https://www.isca-archive.org/interspeech_2007/kenmochi07_interspeech.html#
- Kowal-Summek, L. (2016). *Musiktherapie und Autismus*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-13187-6>
- LaVelle, S. M. (2023). *Virtual Reality*. Cambridge University Press.
- Li, D. (2020, Dezember). *Blob Opera*. Blob Opera by Artist in Residence at Google Arts & Culture: David Li - Experiments with Google. Abgerufen 26. März 2025, von <https://experiments.withgoogle.com/blob-opera>
- Lin, X., Mahmud, S., Jones, E., Shaker, A., Miskinis, A., Kanan, S., & Kim, J.-H. (2020). Virtual Reality-Based Musical Therapy for Mental Health Management. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 0948–0952. <https://doi.org/10.1109/CCWC47524.2020.9031157>
- LoewDownloads.* (o. J.). Downloads: psychologische Tests, Bücher und Artikel (gratis) - www.prof-loew.de. Abgerufen 2. März 2025, von <http://www.prof-loew.de/downloads-psychologische-tests-buecher-und-artikel-gratis/>
- Lu, P., Wu, J., Luan, J., Tan, X., & Zhou, L. (2020). *XiaoiceSing: A High-Quality and Integrated Singing Voice Synthesis System* (No. arXiv:2006.06261). arXiv. <https://doi.org/10.48550/arXiv.2006.06261>
- Macon, M. W. (1997). *Concatenation-based MIDI-to-Singing Voice Synthesis*.
- Maples-Keller, J. L., Bunnell, B. E., Kim, S.-J., & Rothbaum, B. O. (2017). The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. *Harvard Review of Psychiatry*, 25(3), 103–113. <https://doi.org/10.1097/HRP.0000000000000138>
- MathworksLSTM.* (o. J.). Was ist ein Long Short-Term Memory (LSTM)? Abgerufen 14. März 2025, von <https://de.mathworks.com/discovery/lstm.html>

- Menzen, K.-H. (2023). *Grundlagen der Kunsttherapie*. UTB.
- MetaQuest3. (o. J.). Meta Quest 3: Mixed-Reality-VR-Headset – Jetzt kaufen. Abgerufen 15. März 2025, von <https://www.meta.com/de/quest/quest-3/>
- Moser, H. J. (1968). *Allgemeine Musiklehre*: DE GRUYTER. <https://doi.org/10.1515/9783110832983>
- MuseScore. (o. J.). Gratis Musikkompositionen und -notations Software | MuseScore. Abgerufen 17. März 2025, von <https://musescore.org/de>
- Obermeier, P. (2019). *Kunsttherapie für die Praxis*. Elsevier. <https://doi.org/10.1016/C2017-0-03727-8>
- OpenXR. (o. J.). OpenXR Plugin | OpenXR Plugin | 1.14.1. Abgerufen 16. März 2025, von <https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.14/manual/index.html>
- Oura, K., Mase, A., Yamada, T., Muto, S., Nankaku, Y., & Tokuda, K. (2010). *Recent development of the HMM-based singing voice synthesis system—Sinsy*. 211–216. Verfügbar unter: https://www.isca-archive.org/ssw_2010/oura10_ssw.html#
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. <https://doi.org/10.1109/5.18626>
- Reger, G. M., Holloway, K. M., Candy, C., Rothbaum, B. O., Difede, J., Rizzo, A. A., & Gahm, G. A. (2011). Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of Traumatic Stress*, 24(1), 93–96. <https://doi.org/10.1002/jts.20574>
- Riggs, L., & Karwoski, T. (1934). Synaesthesia. *British Journal of Psychology. General Section; London, etc.*, 25(1), 29–42.
- Rodet, X. (2002). *Synthesis and processing of the singing voice*.
- Schuster, M. (2014). *Kunsttherapie in der psychologischen Praxis*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-44923-9>
- Scirea, M., & Brown, J. (2015). *Evolving Four Part Harmony Using a Multiple World Model*. <https://doi.org/10.5220/0005595202200227>
- Shamri Zeevi, L. (2021). Making Art Therapy Virtual: Integrating Virtual Reality Into Art Therapy With Adolescents. *Frontiers in Psychology*, 12, 584943. <https://doi.org/10.3389/fpsyg.2021.584943>

- Shenoy, A., & Wang, Y. (2005). Key, Chord, and Rhythm Tracking of Popular Music Recordings. *Computer Music Journal*, 29(3), 75–86.
- Sinsy. (o. J.). Sinsy - HMM/DNN-based Singing Voice Synthesis System. Abgerufen 17. März 2025, von <https://www.sinsy.jp/>
- SinsyReference. (2021). Sinsy Reference Guide. Abgerufen 17. März 2025, von <https://sinsy.sp.nitech.ac.jp/reference.pdf>
- Steghaus, S., & Poth, C. H. (2022). Assessing momentary relaxation using the Relaxation State Questionnaire (RSQ). *Scientific Reports*, 12(1), 16341. <https://doi.org/10.1038/s41598-022-20524-w>
- Strümke, I., & Dalen, H. P. (2024). *Künstliche Intelligenz: Wie sie funktioniert und was sie für uns bedeutet* (C. Hildebrandt, Übers.; 1. Auflage). Rheinwerk Verlag.
- Sutherland, I. E. (1965). The Ultimate Display. *Proceedings of IFIP Congress*, 506–508.
- Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I on - AFIPS '68 (Fall, Part I)*, 757. <https://doi.org/10.1145/1476589.1476686>
- Syed Abdullah, S., Awang Rambli, D., Sulaiman, S., Alyan, E., Merienne, F., & Mohd Muhaiyuddin, N. D. (2021). The Impact of Virtual Nature Therapy on Stress Responses: A Systematic Qualitative Review. *Forests*, 12(12), 1776. <https://doi.org/10.3390/f12121776>
- Tanghatar, R. (2012). *Stress*. Centaurus Verlag & Media. <https://doi.org/10.1007/978-3-86226-842-9>
- Techniker Krankenkasse. (2021). *Entspann dich, Deutschland—TK-Stressstudie 2021*. Abgerufen 26. März 2025, von <https://www.tk.de/resource/blob/2033600/dabd321631964c329be93cf716020397/entspann-dich-deutschland-data.pdf>
- Theis, S., Pfendler, C., Alexander, T., Mertens, A., Brandl, C., & Schlick, C. (2016). *Head-Mounted Displays – Bedingungen des sicheren und beanspruchungsoptimalen Einsatzes*.
- Tilt Brush. (o. J.). *Tilt Brush – Zentrum für didaktische Computerspielforschung*. Abgerufen 26. März 2025, von <https://zfdc.ph-freiburg.de/tilt-brush/>

- Unity Technologies. (o. J.-a). *Unity - Manual: Rendering lines*. Abgerufen 16. März 2025, von <https://docs.unity3d.com/6000.0/Documentation/Manual/line-rendering-introduction.html>
- Unity Technologies. (o. J.-b). *Unity - Manual: Rigidbody component reference*. Abgerufen 21. März 2025, von <https://docs.unity3d.com/6000.0/Documentation/Manual/class-Rigidbody.html>
- Unity Technologies. (o. J.-c). *Unity - Scripting API: Collider*. Abgerufen 21. März 2025, von <https://docs.unity3d.com/6000.0/Documentation/ScriptReference/Collider.html>
- Unity Technologies. (o. J.-d). *UnityAudio*. Unity - Manual: Audio Overview. Abgerufen 16. März 2025, von <https://docs.unity3d.com/6000.0/Documentation/Manual/AudioOverview.html>
- Unity. (o. J.). Unity. Abgerufen 16. März 2025, von <https://unity.com>
- UnityXRIT. (o. J.). XR Interaction Toolkit | XR Interaction Toolkit | 3.0.7. Abgerufen 16. März 2025, von <https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0/manual/index.html>
- Unreal Engine. (o. J.). Unreal Engine. Abgerufen 16. März 2025, von <https://www.unrealengine.com/de/home>
- Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-Colour Synaesthesia: To What Extent Does it Use Cross-Modal Mechanisms Common to us All? *Cortex*, 42(2), 264–280. [https://doi.org/10.1016/S0010-9452\(08\)70352-6](https://doi.org/10.1016/S0010-9452(08)70352-6)
- Wittpahl, V. (Hrsg.). (2019). *Künstliche Intelligenz: Technologie | Anwendung | Gesellschaft*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-58042-4>
- Wright, S. (2018). *A Guide to Color | New Mexico State University—BE BOLD. Shape the Future*. https://pubs.nmsu.edu/_c/C316/
- Wu, S., Li, J., Sousa, M., & Grossman, T. (2023). Investigating Guardian Awareness Techniques to Promote Safety in Virtual Reality. *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, 631–640. <https://doi.org/10.1109/VR55154.2023.00078>
- Ying-Chun, L., & Chwen-Liang, C. (2018). The Application of Virtual Reality Technology in Art Therapy: A Case of Tilt Brush. *2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, 47–50. <https://doi.org/10.1109/ICKII.2018.8569081>

Zheng, L., Li, W., Song, S., Xiao, X., Low, S. R., Zhang, Y., Yu, X., & Peng, Y. (2024). Effectiveness of Mindfulness-Based Virtual Reality Training on Stress, Anxiety, and Depression among Chinese University Students. *Mindfulness*, 15(4), 899–913. <https://doi.org/10.1007/s12671-024-02335-4>

Anhang

Fragebogen

Nutzerstudie KI-gestützte Kunst- und Musiktherapie in VR

Moin! Danke, dass Du Dich bereit erklärt hast, mir bei meiner Studie zu helfen! Diese Untersuchung

besteht aus drei Teilen:

1. Einigen Prä-Test-Fragen, um Deinen aktuellen Zustand festzuhalten.
2. Dem VR-Erlebnis
3. Einigen Post-Test-Fragen, um Deine Erlebnisse und Deinen Zustand nach dem Erlebnis festzuhalten

Bitte beantworte die Fragen so ehrlich wie möglich - es gibt aber kein Richtig oder Falsch!

Du brauchst keinerlei Erfahrungen mit VR - ich gebe Dir eine kleine Einführung und zeige Dir wie alles geht!

Vor dem Test

Diese Fragen sind vor dem Test auszufüllen und bieten mir sowohl demographische Daten als auch Daten,

die mir helfen die Wirkung der VR-Anwendung zu analysieren.

Bei Fragen mit Kästchen sind mehrere Antworten möglich

1.

Wie alt bist Du? (freiwillige Angabe)

2.

Geschlecht? (freiwillige Angabe)

Männlich

Weiblich

Divers

3.

Hast du schon Erfahrungen mit VR gemacht? *

Ja

Nein

4.

Nutzt Du Musik zur Entspannung? *

Ja, rezeptiv (ich höre gerne Musik zur Entspannung)

Ja, aktiv (ich spiele gerne Musik zur Entspannung)

Nein

5.

Nutzt du Kunst zur Entspannung? *

Ja, ich zeichne oder male

Nein

6.

Hast Du schon Erfahrung mit Kunst- oder Musiktherapie gemacht? *

Ja, mit Kunsttherapie

Ja, mit Musiktherapie

Nein, weder noch

7.

Bitte markiere, wie entspannt du dich gerade fühlst

Gar nicht entspannt 0 Sehr Entspannt 10

1 2 3 4 5 6 7 8 9 10

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

8.

Im Folgenden sind eine Reihe von Feststellungen. Bitte lies sie durch und wähle aus den fünf Antworten eine aus, die auf dein Befinden **jetzt in diesem Moment** zutrifft.

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Mein Atem ist schneller als gewöhnlich	☐	☐	☐	☐	☐
Mein Herz schlägt schneller als sonst	☐	☐	☐	☐	☐
Meine Muskeln fühlen sich angespannt an (Hand zur Faust geballt; Kiefer angespannt; gerunzelte Stirn)	☐	☐	☐	☐	☐
Meine Muskeln fühlen sich entspannt an	☐	☐	☐	☐	☐
Meine Muskeln fühlen sich locker an	☐	☐	☐	☐	☐
Ich fühle mich sehr entspannt	☐	☐	☐	☐	☐
Ich bin gerade vollkommen ruhig	☐	☐	☐	☐	☐
Ich fühle mich schläfrig und müde	☐	☐	☐	☐	☐
Ich bin kurz davor einzunicken	☐	☐	☐	☐	☐
Ich fühle mich erfrischt und wach	☐	☐	☐	☐	☐

VR-Erlebnis

Sag mir Bescheid, wenn du es bis hierhin geschafft hast - dann starten wir gemeinsam das VR-Erlebnis.

Das VR-Erlebnis besteht aus einer kleinen Einführung, zwei kleinen Aufgaben und dann etwas Zeit zur freien Erkundung. Ich bin dabei immer in Deiner Nähe und kann alles auf meinem Smartphone sehen.

Nach dem Test

Diese Fragen sind unmittelbar nach dem Test zu beantworten und helfen mir die Wirkung der VR-Anwendung besser zu analysieren.

9.

Bitte markiere, wie entspannt du dich gerade fühlst

Gar nicht entspannt 0 Sehr Entspannt 10

1 2 3 4 5 6 7 8 9 10

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

10.

Im Folgenden sind eine Reihe von Feststellungen. Bitte lies sie durch und wähle aus den fünf Antworten eine aus, die auf dein Befinden **jetzt in diesem Moment** zutrifft.

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Mein Atem ist schneller als gewöhnlich	☐	☐	☐	☐	☐
Mein Herz schlägt schneller als sonst	☐	☐	☐	☐	☐
Meine Muskeln fühlen sich angespannt an (Hand zur Faust geballt; Kiefer angespannt; gerunzelte Stirn)	☐	☐	☐	☐	☐
Meine Muskeln fühlen sich entspannt an	☐	☐	☐	☐	☐
Meine Muskeln fühlen sich locker an	☐	☐	☐	☐	☐
Ich fühle mich sehr entspannt	☐	☐	☐	☐	☐
Ich bin gerade vollkommen ruhig	☐	☐	☐	☐	☐
Ich fühle mich schläfrig und müde	☐	☐	☐	☐	☐
Ich bin kurz davor einzunicken	☐	☐	☐	☐	☐
Ich fühle mich erfrischt und wach	☐	☐	☐	☐	☐

11.

Die folgenden Aussagen beziehen sich auf die Aufgaben, die du erhalten hast

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Die Aufgabe, die vorgegebene Melodie zu spielen, war einfach und intuitiv	☐	☐	☐	☐	☐
Die Aufgabe, das vorgegebene Objekt zu malen war einfach und intuitiv	☐	☐	☐	☐	☐
Eine Aufgabe zu haben empfand ich als angenehm	☐	☐	☐	☐	☐
Ich hätte es besser gefunden, komplett alleine und frei machen zu können, was ich will	☐	☐	☐	☐	☐
Die Kombination aus freiem und geleitetem Erlebnis fand ich gut	☐	☐	☐	☐	☐

12.

Die folgenden Aussagen beziehen sich auf die gehörten Stimmen

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Die KI-generierten Stimmen klangen angenehm	☐	☐	☐	☐	☐
Die KI-generierten Stimmen klangen natürlich	☐	☐	☐	☐	☐
Die erzeugten Akkorde klangen harmonisch und passend	☐	☐	☐	☐	☐
Die erzeugten Akkorde haben eine beruhigende Auswirkung auf mich gehabt	☐	☐	☐	☐	☐
Das aktive Verändern der Töne hat mir Spaß gemacht	☐	☐	☐	☐	☐

13.

Die folgenden Aussagen beziehen sich auf die gemalten Linien

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Die Auswahl der Farben hat mir viele gestalterische Möglichkeiten gegeben	☐	☐	☐	☐	☐
Das freie Zeichnen im virtuellen 3D-Raum hat mir gut gefallen	☐	☐	☐	☐	☐
Ich konnte meiner künstlerischen Ader freien Lauf lassen	☐	☐	☐	☐	☐

14.

Die folgenden Aussagen beziehen sich auf das Zusammenspiel von Kunst und Musik

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Die Kombination aus Klängen und Farben hat eine positive emotionale Reaktion in mir ausgelöst	☐	☐	☐	☐	☐
Ich konnte eine klare Verbindung zwischen Malbewegung und Klängen erkennen	☐	☐	☐	☐	☐
Die Farben passten gut zu den Stimmlagen	☐	☐	☐	☐	☐
Die Farben und Musik haben eine stimmige Atmosphäre geschaffen	☐	☐	☐	☐	☐

15.

Die folgenden Aussagen beziehen sich auf die therapeutische Wirkung in VR

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Die VR-Umgebung hat mich beruhigt, da ich abgekoppelt von der realen Welt war	☐	☐	☐	☐	☐
Ich verspürte körperliches Unwohlsein in VR (Schwindel etc.)	☐	☐	☐	☐	☐
Ich kann mir vorstellen diese Anwendung statt traditioneller Kunsttherapie zu nutzen	☐	☐	☐	☐	☐
Ich kann mir vorstellen diese Anwendung statt traditioneller Musiktherapie zu nutzen	☐	☐	☐	☐	☐
Ich kann mir vorstellen diese Anwendung in meiner Freizeit zur Entspannung zu nutzen	☐	☐	☐	☐	☐

16.

Die folgenden Aussagen beziehen sich auf die VR-Anwendung an sich

	Trifft überhaupt. nicht zu	Trifft eher. zu	Weder. noch	Trifft eher zu	trifft voll und ganz zu
Die Steuerung des Pinsels war einfach und intuitiv	☐	☐	☐	☐	☐
Das Wechseln der Farben war einfach und intuitiv	☐	☐	☐	☐	☐
Das Mischen der Farben war einfach und intuitiv	☐	☐	☐	☐	☐
Die Bewegung im Raum war angenehm und nicht verwirrend	☐	☐	☐	☐	☐
Das schlichte, graue Design hat mich angespornt, den Raum mehr Farbe zu verleihen	☐	☐	☐	☐	☐
Ich hatte keine technischen Probleme oder Verzögerungen bei der Interaktion	☐	☐	☐	☐	☐

17.

Hast Du noch irgendwelche abschließenden Anmerkungen zum VR-Erlebnis oder Verbesserungsvorschläge?

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel:

KI-gestützte Kunst- und Musiktherapie in VR

selbständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Datum

Unterschrift