

BACHELORARBEIT

zur Erlangung des akademischen Grades B.Sc.

KI-unterstützte Musikkomposition im Test - Eine empirische Untersuchung zu Usability und Akzeptanz der Multi- Track Music Machine in Cubase

vorgelegt am 22. Januar 2024

Hannah Körber XXXXXXXXXX

Erstprüferin: Prof. Dr. Tessa Taefi
Zweitprüferin: Maryam Safi

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**
Department Medientechnik
Finkenau 35
20081 Hamburg

Zusammenfassung

Die Anwendungsfelder im Bereich der künstlichen Intelligenz (KI) für co-kreative Arbeitsprozesse wachsen stetig. Auch in die Musikproduktion haben intelligente Tools ihren Weg gefunden und werden dort nun von den Nutzer:innen digitaler Audio Workstations (DAWs) angewendet. Diese Bachelorarbeit untersucht die Nutzungserfahrungen von unerfahrenen Komponist:innen mit der Multi-Track Music Machine (MMM), einem generativen System für KI-assistierte Musikproduktion. Für die Untersuchung wird MMM in die DAW Cubase 13 integriert. Unter Berücksichtigung internationaler Standards und Empfehlungen von Expert:innen wird eine fragenbogenbasierte empirische Mixed Method Studie konzipiert, die auf bestehende Metriken für die Quantifizierung von Usability und Technology Acceptance zurückgreift. Zum Einsatz kommen qualitative und quantitative Messinstrumente. Die Ergebnisse zeigen eine grundsätzlich positive Bewertung der Usability, während bei der Akzeptanz Uneinigkeiten zu bemerken waren. Diese werden anhand der Wünsche und Anregungen der Teilnehmenden zum System analysiert. Daraus werden Strategien zur Entwicklung effektiverer Anpassungsmöglichkeiten diskutiert, die die Akzeptanz des Systems erhöhen können, um zukünftige Wege aufzuzeigen.

Abstract

Artificial intelligence (AI) applications for co-creative work processes are expanding. Intelligent tools have also found their way into music production, where they are now being explored by users of Digital Audio Workstations (DAWs). This bachelor thesis examines the user experience of novice composers with the Multi-Track Music Machine (MMM), a generative system for AI-assisted music production. For the study, MMM is integrated into the DAW Cubase 13. Taking into account international standards and expert recommendations, a questionnaire-based empirical mixed-method study is designed, drawing on existing metrics for quantifying usability and technology acceptance. Qualitative and quantitative measurement instruments are used. The results show a generally positive assessment of usability, while disagreements were noticeable in terms of acceptance. These are analysed in the light of the participants' wishes and suggestions for the system. From this, strategies for developing more effective customisation options that can increase the acceptance of the system are discussed in order to identify future paths.

Danksagung

Ich möchte mich bei allen bedanken, die mich bei der Erstellung meiner Bachelorarbeit unterstützt und motiviert haben.

An erster Stelle bedanke ich mich ganz herzlich bei Prof. Dr. Tessa Taefi von der HAW Hamburg und Maryam Safi von Steinberg für die hervorragende Betreuung, für die fachliche Unterstützung und die konstruktive Kritik, die stark zur Qualität meiner Forschungsarbeit beigetragen haben.

Ein weiterer Dank geht an das Research Team von Steinberg, insbesondere an Jean-Baptiste Rolland. Vielen Dank für die Hilfsbereitschaft und die zuverlässige Unterstützung während des gesamten Forschungsprozesses. Mein Dank gilt diesbezüglich auch Philippe Pasquier, Jeff Ens und Renaud Bougueng Tchameube von der Simon Fraser University.

Ganz herzlich möchte ich mich außerdem bei Tino Degen von Steinberg bedanken, der mir für die Abschlussarbeit die Zusammenarbeit mit Steinberg ermöglicht hat und der mir auch während meiner Zeit als Werkstudentin in der Firma unzählige Male mit seiner Zeit und seinem Fachwissen zur Seite stand.

Dem Tonlabor der HAW Hamburg danke ich für die Unterstützung während der Versuchsdurchführung. Besonders bei Tobias Falke bedanke ich mich für das Wegweisen zu Beginn der Bachelorarbeit und für die Unterstützung im Tonlabor, die mir viele schnelle Anpassungen ohne große Umwege ermöglicht hat.

Auch bei den Teilnehmenden meiner Studie möchte ich mich an dieser Stelle dafür bedanken, dass sie sich die Zeit für die Teilnahme genommen haben.

Vielen Dank an M, der mich mit viel Geduld, Interesse und konstruktiven Korrekturen während dieser Arbeit sicher begleitet und ermutigt hat. Vielen Dank an C, durch deren Unterstützung ich die Möglichkeit bekommen habe Medientechnik zu studieren.

Inhaltsverzeichnis

Inhaltsverzeichnis.....	III
Abkürzungsverzeichnis	V
Abbildungsverzeichnis	VI
Tabellenverzeichnis.....	VII
1 Einleitung	2
1.1 Forschungsfragen und Zielsetzung.....	3
1.2 Aufbau.....	5
1.3 Beteiligte Parteien	6
2 Theoretischer Kontext	7
2.1 KI-assistierte Musikkomposition	7
2.1.1 Symbolische Musikdarstellung	8
2.1.2 KI-Modelle für Musikgenerierung.....	9
2.1.3 MMM-Modell	11
2.2 User Experience Testing	13
2.2.1 User Experience, Usability und Akzeptanz.....	13
2.2.2 Menschzentrierte Gestaltung nach DIN EN ISO 9241-210	14
2.2.3 Evaluation in der ersten MMM-Iteration durch Tchemeube u. a.	17
2.2.4 Iterative Entwicklung am Beispiel von MMM	18
3 Vorbereitung der User Study.....	19
3.1 Prototyp	20
3.2 Szenario	22
3.3 Untersuchungsform	25
3.4 Erhebungsinstrument.....	25
3.4.1 Task-Based Usability Testing	26
3.4.2 Demographie und Erfahrung.....	28
3.4.3 Quantitative Metriken	28
3.4.4 Qualitatives Feedback (Kommentare).....	35
3.4.5 Fragebögen und Likert-Skala	36
3.4.6 Übersetzungen	37
3.5 Stichprobenverfahren: Beschreibung der User Persona.....	39
3.6 Pretest	40
4 Durchführung und Auswertung der User Study.....	40

4.1	Onboarding und Versuchsablauf.....	40
4.2	Versuchsaufbau	42
4.3	Datenauswertung.....	44
4.4	Ergebnisse	45
4.4.1	Teilnehmende	45
4.4.2	Ergebnis: System Usability Scale	48
4.4.3	Ergebnis: Controllability	48
4.4.4	Ergebnis: Creativity Support Index.....	49
4.4.5	Ergebnis: Technology Acceptance.....	50
4.4.6	Ergebnis: After Scenario Questionnaire.....	52
4.4.7	Ergebnis: Qualitative Befragung (Kommentare)	53
4.5	Diskussion	55
5	Fazit.....	58
	Literaturverzeichnis.....	60
	Anhang	65
	Anschreiben.....	65
	Übersetzungen.....	66
	Checkliste	67
	SUS-Ergebnisse.....	69
	CSI- und Controllabilityergebnisse	70
	TAM-Ergebnisse	71
	ASQ-Ergebnisse	72
	Qualitatives Feedback	72
	Inhalt des beigelegten Datenträgers	73
	Eigenständigkeitserklärung	74

Abkürzungsverzeichnis

AI	Artificial Intelligence
ASIO	Audio Stream Input/Output
ASQ	After Scenario Questionnaire
Convolutional Neural Networks	Convolutional Neural Networks
CSI	System Usability Scale
DAW	Digital Audio Workstations
DIN	Deutsches Institut für Normung
DL	Deep Learning
EN	Europäische Norm
GAN	Generative Adversarial Networks
GmbH	Gesellschaft mit beschränkter Haftung
GPT	Generative Pretrained Transformer
HAW	Hochschule für Angewandte Wissenschaften
HCP	Human-Centred Processes
ISO	International Organization for Standardization
KI	Künstliche Intelligenz
LAN	Local Area Network
LSTM	Long Short-Term Memory
MacOS	Macintosh Operating System
MGP	Music Generation Panel
MIDI	Musical Instrument Digital Interface
ML	Machine Learning
MMM	Multitrack Music Machine
mp3	Moving Picture Experts Group-1 Audio Layer 3
NDA	Non-Disclosure-Agreement
RNN	Recurrent Neural Networks
SUS	System Usability Scale
TAM	Technology Acceptance Model
TN	Teilnehmende/Teilnehmer/Teilnehmerin
UX	User Experience
VST	Virtual Studio Technology
WAV	Waveform Audio File Format
WLAN	Wireless Local Area Network

Abbildungsverzeichnis

Abbildung 1: Pianoroll-Ansicht von MIDI-Daten in Cubase 13	9
Abbildung 2: Multitrack- und Bar-Fill-Repräsentation	12
Abbildung 3: User Experience, Usability und Akzeptanz	14
Abbildung 4: Einflüsse, Auswirkungen, Ziele und Mittel menschenzentrierter Gestaltung	15
Abbildung 5: Iterative Entwicklung im Kontext menschenzentrierter Gestaltung.....	18
Abbildung 6: Zeitplan für User Study	19
Abbildung 7: Music Generation Panel Interface.....	21
Abbildung 8: Dimensionen von Prototypen in Anlehnung an Nielsen	22
Abbildung 9: Cubase Template für Rock.....	24
Abbildung 10: Kanal- und RoutingEinstellungen für Rock-Template	24
Abbildung 11: Ablauf einer Durchführung mit Methodenübersicht.....	26
Abbildung 12: TAM nach Davis	33
Abbildung 13: Ablauf einer Durchführung	41
Abbildung 14: Übersicht: Technisches Setup	43
Abbildung 15: Raumplan R2 und R3	43
Abbildung 16: Fotos des Setups in R2 (links) und R3 (rechts).....	44
Abbildung 17: Altersgruppen der Teilnehmenden	45
Abbildung 18: Gender-Identität der Teilnehmenden	45
Abbildung 19: DAW-Nutzungskontext der Teilnehmenden	46
Abbildung 20: DAW-Nutzungsprofile der Teilnehmenden.....	46
Abbildung 21: DAW-Erfahrungslevel pro Teilnehmer:in	47
Abbildung 22: Kompositionserfahrung pro Teilnehmer:in.....	47
Abbildung 23: SUS-Scores	48
Abbildung 24: Bedienfreundlichkeit insgesamt	48
Abbildung 25: Controllability-Scores	49
Abbildung 26: CSI-Scores	50
Abbildung 27: TAM-Scores.....	51
Abbildung 28: Voraussage über die Nutzungswahrscheinlichkeit	52
Abbildung 29: Nutzungswahrscheinlichkeit in Abhängigkeit vom CSI-Score/TN.....	52
Abbildung 30: ASQ-Scores.....	53

Tabellenverzeichnis

Tabelle 1: Parameter des Music Generation Panel 1.5.0.188.....	21
Tabelle 2: Arbeitsauftrag für die Task	27
Tabelle 3: Variablen des Fragebogens zu Demographie und Erfahrung.....	28
Tabelle 4: SUS-Variablen	29
Tabelle 5: CSI-Variablen	31
Tabelle 6: Controllability-Variablen	32
Tabelle 7: TAM-Variablen.....	34
Tabelle 8: ASQ-Variablen.....	35
Tabelle 9: Prompts für qualitatives Feedback	35

Einleitung

Im Bereich der künstlichen Intelligenz (KI) entwickeln sich breite, neue Anwendungsfelder. Ein großer Teil davon sind co-kreative Arbeitsprozesse zwischen Mensch und Maschine im künstlerischen Bereich: die Generierung von Bild- (Lu u. a. 2023), Video- (Sontakke u. a. 2023), Musik- (Tchemeube u. a. 2023) und Sprachmaterial (Xi u. a. 2023). Im Feld der Musik-Komposition bietet der Stand der Forschung bereits mehrere Systeme, die Deep-Learning Prozesse anwenden, um Musik mit variierender Komplexität zu generieren (Liang u. a. 2017; Roberts u. a. 2018; Donahue u. a. 2019; Oore u. a. 2020; Noel-Hirst und Bryan-Kinns 2023). Auch in die Welt der Musikproduktion haben intelligente Tools ihren Weg gefunden und werden dort nun von den Nutzer:innen Digitaler Audio Workstations (DAWs) angewendet.

Ein Beispiel dafür ist die Multi-Track Music Machine (MMM), ein generatives System zur symbolischen Musikgenerierung, basierend auf dem Transformer-Modell. Durch MMM bekommen DAW-Nutzer:innen die Möglichkeit, ihren Arbeitsaufwand beim Komponieren erheblich zu verringern. Sie sind jederzeit nur einen Klick von der nächsten Mensch-Maschinen-Kollaboration entfernt: Der Mensch stellt die Parameter ein und die Maschine generiert automatisch die Melodien, Harmonien und Rhythmen dazu. Das System ist in der Lage, sich am musikalischen Kontext zu orientieren. MMM lässt sich flexibel in DAWs integrieren und kann somit schnell seinen Weg in die Studios und Wohnzimmer der Musiker:innen finden.

Doch wie wird die neue Technologie von den Benutzer:innen aufgenommen? Würden sie MMM zum Komponieren verwenden? Welche Schwierigkeiten haben sie bei der Benutzung und wie muss das System gestaltet sein, damit sie die bestmögliche Benutzungserfahrung haben? Dies sind typische Fragen der Benutzungserfahrung, auch: User Experience. Um die Wahrnehmung der Benutzer:innen von MMM zu erforschen, wurde bereits eine dreiteilige Mixed-Method-Studie von Tchemeube u. a. (2023) durchgeführt. MMM wurde dafür als Ein-Parameter-Plugin „Music Generation Panel“ in Steinbergs DAW Cubase integriert und einer Gruppe von Expert:innen zur Verfügung gestellt. Die Expert:innen bearbeiteten Kompositionsaufgaben (Tasks) mit MMM und beantworteten Fragebögen zur User Experience, Gebrauchstauglichkeit (Usability) und Akzeptanz. Die Ergebnisse bestimmten weitere Entwicklungsschritte.

Das Einholen von Expert:innenmeinungen ist ein wichtiger erster Schritt in der Softwareentwicklung. Im nächsten Schritt wird der Fokus auf eine andere Nutzer:innengruppe gelegt: Personen mit wenig oder ganz ohne DAW- und Kompositionserfahrung. Steinberg sammelte zwischen 2022 und 2023 Daten von über 12.000 Cubase 12 Nutzer:innen und fragte unter anderem,

ob sie Cubase professionell oder als Hobby nutzen. Die Ergebnisse zeigen, dass die Hobby-Produzent:innen den größten Teil der Benutzer:innen ausmachen. Darunter befinden sich Anfänger:innen und wenig erfahrene Komponist:innen. Durch diese Umfrage liegen Informationen über die bereits registrierten Steinberg-Kund:innen vor. Aus Marketingsicht ist es für eine Softwarefirma auch wichtig, neue Kund:innen zu gewinnen, zum Beispiel durch das Hinzufügen neuer Features. Gerade ein Tool wie MMM, das einen niedrighschwelligigen Einstieg in kreative Arbeitsweisen ermöglicht, die sonst ein hohes Maß an Kompetenz erfordern, sind für Neueinsteiger:innen attraktiv.

Aus diesen Gründen wurde in der zweiten MMM-Iteration eine Benutzer:innenstudie (User Study) mit unerfahrenen Komponist:innen durchgeführt. Diese Studie ist Gegenstand der vorliegenden Arbeit. Dazu wurde ein neuer Prototyp des Music Generation Panels mit sechs einstellbaren Parametern bereitgestellt. Bei der Planung der Untersuchung wurden die internationalen Standards aus der DIN EN ISO 9241-Reihe zur Ergonomie der Mensch-System-Interaktion und Empfehlungen von Expert:innen berücksichtigt.

1.1 Forschungsfragen und Zielsetzung

Für die User Study zur MMM-Nutzungserfahrung von unerfahrenen Komponist:innen wird eine Hauptforschungsfrage gestellt, der sich weitere Fragestellungen unterordnen.

Wie kann die Nutzungserfahrung von unerfahrenen Komponist:innen mit MMM in Cubase optimiert werden?

0. Wie muss eine Task gestaltet sein, damit Benutzer:innen mit wenig DAW- und Kompositionserfahrung die Möglichkeit haben, das MMM- System zu erkunden?
1. Wie wird die Usability von MMM von unerfahrenen Komponist:innen bewertet?
2. Wie akzeptieren unerfahrene Komponist:innen MMM?
3. Welche Aspekte des MMM-Systems können die Bedienung für unerfahrene Komponist:innen erschweren?
4. Welche Verbesserungsempfehlungen ergeben sich aus der Evaluation der Usability und Akzeptanz für eine verbesserte Nutzungserfahrung von MMM, aus Sicht unerfahrener Komponist:innen?

Aus den Forschungsfragen lassen sich die Forschungsziele für diese Arbeit ableiten. Jedes Ziel gilt nur unter bestimmten Kriterien als erfüllt. Als Grundsatz für die gesamte Arbeit und für jedes Einzelziel gilt, dass die Objektivität gewahrt bleiben muss.

Hauptziel: Diese Arbeit soll dazu beitragen, die Nutzungserfahrung von unerfahrenen Komponist:innen mit MMM in Cubase zu optimieren.

Ziel 0: Konzeption einer Task

Kriterien:

- Die Task ist repräsentativ für die spätere Verwendung von MMM.
- Die Task kann von Benutzer:innen mit wenig oder ganz ohne DAW-Erfahrung bearbeitet werden.
- Es werden Hilfsmittel für die Task bereitgestellt, um Unterschieden im Vorwissen der unerfahrenen Komponist:innen entgegenzukommen.
- Die Task ist weder zu umfangreich noch zu trivial für die zur Verfügung stehende Bearbeitungszeit.
- Das Erkunden von MMM steht in der Task im Vordergrund und die Benutzer:innen müssen nicht viel Zeit damit verbringen, sich in Cubase zurechtzufinden.
- Das Ziel der Task ist für die Teilnehmenden klar.
- Die Task ist durch die Durchführung eines Pretests optimiert.
- Die Zufriedenheit der Studienteilnehmer:innen mit der Task wird mit Hilfe eines Fragebogens quantifiziert.

Ziel 1: Usabilitybewertung von MMM durch unerfahrene Komponist:innen

Kriterien:

- Aus der Bewertung durch die Benutzer:innen geht hervor, ob MMM von unerfahrenen Komponist:innen als effektiv, effizient und zufriedenstellend wahrgenommen wird, um kreative Ideen umzusetzen.
- Es werden drei geeignete Usability-Messinstrumenten angewendet, um die Usability zu quantifizieren und zusätzlich ein qualitatives Feedback eingeholt.

Ziel 2: Bewertung der Akzeptanz von MMM bei unerfahrenen Komponist:innen

Kriterien:

- Aus der Bewertung durch die Benutzer:innen geht hervor, ob die Benutzung von MMM leicht und nützlich für sie ist.
- Aus der Bewertung durch die Benutzer:innen geht ihre Bereitschaft zur Nutzung von MMM hervor.
- Es wird ein geeignetes Akzeptanz-Messinstrument angewendet, um diese Akzeptanz zu quantifizieren und zusätzlich ein qualitatives Feedback eingeholt.

Ziel 3: Identifikation von Mängeln

Kriterien:

- Mängel in Bezug auf die Usability von MMM werden anhand der Interpretation der quantitativen Daten zu Usability und Akzeptanz und anhand des qualitativen Feedbacks identifiziert.
- Der Schweregrad der aufgetretenen Probleme wird aus Benutzer:innenperspektive eingeschätzt.

Ziel 4: Ableiten von Verbesserungsempfehlungen

Kriterien:

- Konkrete Verbesserungsempfehlungen für die Usability von MMM werden aus der Interpretation der quantitativen Daten zu Usability und Akzeptanz und aus dem qualitativen Feedback abgeleitet.

1.2 Aufbau

Das erste Kapitel dieser Arbeit stellt die bisher vorliegende Forschung vor. Die beteiligten Parteien werden genannt, um einen Überblick über die wichtigsten Akteur:innen im Forschungskontext zu geben.

Im zweiten Kapitel wird der Forschungsgegenstand in einen theoretischen Kontext eingebettet. Dieses Kapitel gliedert sich in zwei Kernbereiche, in denen das Thema verankert ist. Zunächst wird ein Überblick über das Feld der KI-assistierten Musikkomposition gegeben. Möglichkeiten der symbolischen Musikdarstellung werden aufgeführt, die Entwicklung generativer Modelle in der Musik wird kurz zusammengefasst und das Modell hinter MMM erläutert.

Darauf folgt die theoretische Einbettung des Forschungsthemas in den Bereich User Experience Testing. Grundlegende Begriffe werden erläutert und das Konzept der menschenzentrierten Gestaltung nach DIN EN ISO 9241-210 wird vorgestellt. Anschließend wird der Bogen von der Theorie zur Praxis gespannt. Die Studie von Tchemeube u. a. (2023) aus der ersten MMM-Iteration wird näher beleuchtet und das Prinzip der iterativen Entwicklung wird am Beispiel von MMM praxisnah dargestellt.

Das dritte Kapitel der Arbeit dokumentiert die Vorbereitung der User Study. Der Prototyp wird vorgestellt, das TestszENARIO festgelegt, Untersuchungsform, Erhebungsinstrument und Stichprobenverfahren werden detailliert beschrieben. Das Kapitel schließt mit dem Bericht über den durchgeführten Pretest.

Das vierte Kapitel dokumentiert alle Arbeitsschritte und Ergebnisse der Studiendurchführung. Das Onboarding der Teilnehmenden, der Versuchsablauf und der Versuchsaufbau werden beschrieben und die Methoden der Datenauswertung werden vorgestellt. Darauf folgt eine vollständige Dokumentation, Interpretation und Diskussion der Ergebnisse.

Im fünften Kapitel wird die Fragestellung abschließend beantwortet. Die wichtigsten Empfehlungen für die Usability von MMM aus der Sicht unerfahrener Komponist:innen werden zusammengefasst und ein Ausblick auf zukünftige Arbeitsschritte gegeben.

1.3 Beteiligte Parteien

Die Integration des Systems MMM in Cubase ist eine Zusammenarbeit der Steinberg Media Technologies GmbH und der Simon Fraser University in Burnaby, Kanada.

Steinberg Media Technologies GmbH

Die Hamburger Firma Steinberg Media Technologies besteht seit 1984. Sie entwickelt Software- und Hardwarelösungen für Musiker:innen und Audioingenieur:innen. Angefangen mit dem ersten MIDI Multi-Track Sequencer Pro-16 hat Steinberg sich mit einer Reihe von kreativen Anwendungen im Bereich Audio-Engineering breit aufgestellt. Unter den Anwendungen sind die DAWs Cubase (für Musikproduktion und -aufnahme) und Nuendo (mit speziellen Features für Audio-Postproduktion, Filmkomposition und Game-Audio-Design), die Audio-Editing- und Masteringsoftware WaveLab, die Notationssoftware Dorico und der virtuelle Sampler HALion. Darüber hinaus hat Steinberg durch die Entwicklung von VST (Virtual Studio Technology) und ASIO (Audio Stream Input/Output) bedeutende Beiträge im Bereich der digitalen Audioproduktion geleistet.

Gegenstand der Zusammenarbeit mit Steinberg für diese Arbeit ist die Evaluation des Systems MMM in Steinbergs DAW Cubase, integriert als bisher unveröffentlichtes Plugin Music Generation Panel (im Folgenden abgekürzt mit MGP).

Simon Fraser University

Das von Seiten Steinbergs bereitgestellte MGP dient als Interface (Schnittstelle) zwischen Cubase und MMM, einem System, das von Angehörigen der Simon Fraser University entwickelt wurde (Ens und Pasquier 2020). Eine erste Evaluation des Systems erfolgte ebenfalls durch die Simon Fraser University. Die dazu durchgeführte Studie von Tchemeube u. a. (2023) wird in Abschnitt 2.2.3 dieser Arbeit vorgestellt.

2 Theoretischer Kontext

Es besteht eine theoretische Verankerung des Forschungsthemas in zwei Schlüsselbereichen: der KI-assistierten Musikkomposition und dem User Experience Testing. Im folgenden Kapitel erfolgt die theoretische Einbettung der Arbeit in diese beiden Kernbereiche.

2.1 KI-assistierte Musikkomposition

„Walzer oder Schleifer mit zwei Würfeln zu componieren ohne musikalisch zu seyn noch von der Composition etwas zu verstehen“

So lautet der Titel einer interaktiven Komposition von Wolfgang Amadeus Mozart aus dem 18. Jahrhundert, welche als sehr frühes Modell für algorithmische¹ Musikkomposition angesehen werden kann (vgl.: Briot 2021; MUSOPEN [ohne Datum]). Musikgenerierung nach System ist also keineswegs ein neues Phänomen. Sie findet sich in unterschiedlichen Ausprägungen an vielen Stellen in der Musikgeschichte. In Mozarts Würfelspiel entsteht die Partitur erst dadurch, dass ein:e Komponist:in 16 von Mozart vorkomponierte Takte durch Würfeln neu zusammensetzt. Das Ergebnis ist eine einzigartige, zufallsgesteuerte Komposition, die verschiedene vordefinierte Teile neu kombiniert.

Die Prämisse für das Würfelspiel und die in dieser Arbeit dokumentierten User Study ist dieselbe: Eine Person ohne kompositorische Vorkenntnisse erstellt mit Hilfe des ihr vorliegenden Systems eine Komposition. Hinter den beiden Systemen stecken jedoch sehr unterschiedliche Modelle: Die Teilnehmenden der Studie benutzen zum Komponieren keine Würfel, sondern ein KI-System. Dieses System liefert keine präkomponierten Musikabschnitte, sondern kann selbstständig neue Musik erzeugen. Das System basiert nicht auf Regeln, sondern auf Daten.

KI ist der übergeordnete Begriff für Maschinen, die in der Lage sind, menschenähnliche Problemlösungsfähigkeiten nachzubilden. KI-Systeme werden seit Jahrzehnten dafür genutzt Prozesse zu automatisieren. Das Anwendungsspektrum ist mittlerweile sehr breit. Es reicht von Berechnungen auf der Grundlage von mathematischen Regeln (zum Beispiel durch einen Schachcomputer) bis hin zur Fähigkeit auf Basis tiefer, mehrschichtiger Netzwerke komplizierte Konzepte zu erlernen und aus Erfahrung zu lernen (vgl.: Goodfellow u. a. 2016: 1ff.). Letzteres Konzept wird als Deep Learning (DL) bezeichnet. Es ermöglicht Maschinen, intuitive menschliche Fähigkeiten zu erlernen, wie zum Beispiel das Erkennen von Emotionen in

¹ Als Algorithmus kann ein „Lösungsverfahren in Form einer Verfahrensanweisung, die in einer wohldefinierten Abfolge von Schritten zur Problemlösung führt“ bezeichnet werden (Lackes u. a. 2018).

Gesichtern. Deep Learning ist eine spezielle Form des Machine Learnings (ML), also der Fähigkeit von Systemen, nicht nur auf Wissen und Regeln zuzugreifen, sondern darauf aufbauend Muster zu erkennen und Schlüsse zu ziehen, die das „Wissen“ dieser Systeme erweitern.

Welches Modell hinter dem System MMM steckt und was dieses Modell von anderen generativen Systemen unterscheidet, wird in den folgenden Abschnitten beschrieben.

2.1.1 Symbolische Musikdarstellung

Mozarts Würfelspiel ist ein Beispiel aus einer Zeit lange vor den ersten Computern. Die ersten computergenerierten Kompositionen entstanden in den 1950er Jahren (Briot 2021: 40). Musik kann in symbolischer und damit für Computer verarbeitbarer Form dargestellt werden. Zum einen natürlich als diskretisierte Schallwelle². Aber auch die Darstellung als Matrix oder als Tokensequenz ist möglich. Die symbolische Repräsentation enthält alle musikalischen Elemente, Strukturen und Informationen und kann als ein dekonstruiertes Musikstück betrachtet werden. Um dieses Musikstück abspielen zu können, muss die Information erst in Klang umgewandelt werden, etwa durch das Zuweisen eines virtuellen Instruments in einer DAW.

Bei der symbolischen Darstellung von Musik als Matrix (Pianoroll-Darstellung) wird auf der vertikalen Achse die Tonhöhe (Pitch) und auf der horizontalen Achse die Zeit dargestellt. Die Tokendarstellung hingegen erfolgt in einer eindimensionalen Sequenz. Die Sequenz enthält alle musikalischen Ereignisse, dargestellt als Token (Symbol für eine Informationseinheit). Eine Übersetzung von der einen Darstellungsform in die andere ist möglich und beispielsweise für die Kommunikation von Systemen und für das Hör- und Sichtbarmachen der musikalischen Daten notwendig.

Ein Kommunikationsprotokoll zur Übertragung musikalischer Informationen ist das MIDI-Protokoll. Die Abkürzung MIDI steht für Musical Instrument Digital Interface (digitale Schnittstelle für Musikinstrumente). Das Protokoll ermöglicht die Weiterleitung der Informationen zwischen virtuellen Instrumenten, Computern und anderen MIDI-fähigen Geräten. Diese Informationen können vom empfangenden Gerät interpretiert und akustisch oder grafisch (etwa als MIDI-Noten in der MIDI-Pianoroll-Ansicht) wiedergegeben werden. In Abbildung 1 ist die Pianoroll-Ansicht von MIDI-Daten in Cubase 13 zu sehen. Am oberen Rand des Bildes sind

² Schallwellen (Audiosignale) können diskretisiert werden. Das analoge Audiosignal hat eine kontinuierliche Wellenform. Dieses Signal wird in regelmäßigen Intervallen (Abtastfrequenz) abgetastet und auf eine begrenzte Anzahl von Werten (Quantisierungsstufen) gerundet. So entsteht eine Folge diskreter Werte, die im Anschluss codiert und gespeichert werden kann (zum Beispiel im WAV-Format).

einige Metainformationen der ersten MIDI-Note zu sehen: Start- und Endzeitpunkt, Spieldauer, Pitch, Velocity (Anschlagsgeschwindigkeit), Channel (Kanal in Cubase), Off-Velocity (Loslassgeschwindigkeit), Articulations (Nuancen in der Spielweise) und Release Length (Loslassdauer). Die darunter stehenden Zeitangaben erfolgen im Format Bars/Beats (Takte/Schläge)³.

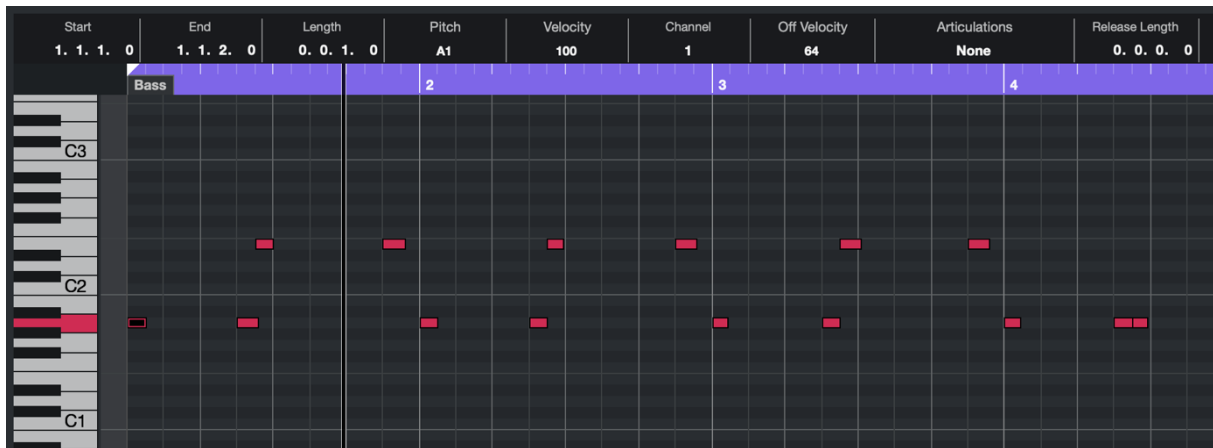


Abbildung 1: Pianoroll-Ansicht von MIDI-Daten in Cubase 13
Quelle: Screenshot aus Cubase 13.0.10.123 auf MacOS 10.15.7

2.1.2 KI-Modelle für Musikgenerierung

Frühe Modelle der Musikgenerierung verwenden Markov-Ketten. Markov-Ketten sind stochastische Modelle, die auf der Annahme beruhen, dass das nächste Ereignis nur vom aktuellen Zustand und nicht von vorherigen Ereignissen abhängt (vgl.: Ens 2023: 27). In der Musikkomposition kann ein Zustand beispielsweise die Tonhöhe sein. Angenommen, der Ausgangszustand ist die Note C. Das Modell kennt für jeden Zustand die Übergangswahrscheinlichkeiten zu anderen Zuständen, weiß also zum Beispiel, dass nach dem C mit einer Wahrscheinlichkeit von 40 Prozent ein E kommt, mit einer Wahrscheinlichkeit von 20 Prozent ein D, und so weiter. Der Algorithmus verwendet die definierten Übergangswahrscheinlichkeiten, um nach und nach eine musikalische Sequenz zu generieren. In jedem Schritt wird ein neuer Zustand basierend auf den vorherigen Zuständen ausgewählt (vgl.: Bell 2011: 101). Kommen versteckte, nicht direkt beobachtbare Zustände hinzu, spricht man von Hidden-Markov-Modellen (vgl.: Ens 2023: 37). Ein versteckter Zustand kann zum Beispiel die Musikstimmung sein (melancholisch/fröhlich/lebhaft/etc.).

Deep-Learning-Modelle verwenden sogenannte neuronale Netzwerke. Diese entstehen durch die Analyse riesiger Datenmengen, die Modelle werden also mit Datensätzen trainiert. Die

³ Die vier Ziffern des Bars/Beats-Formats beziehen sich auf Takt, Schlag, Sechzehntel (eines Schlags) und Tick. Ein Sechzehntel hat 120 Ticks.

Daten werden vorwärts durch das Netzwerk geleitet (Feedforward-Netzwerke), um Vorhersagen zu treffen (vgl.: Goodfellow u. a. 2016: 163). Ein Beispiel sind Convolutional Neural Networks (CNNs). Der Name leitet sich von der mathematischen Faltung ab, die die Modelle für die Datenverarbeitung nutzen. CNNs werden hauptsächlich für Daten verwendet, die als Gitter strukturiert werden können, wie die Pixel eines Bildes. Es gibt auch Anwendungen für musikalische Modelle, in denen die musikalischen Informationen als Pianoroll-Matrix dargestellt werden, zum Beispiel in MusicVae (Roberts u. a. 2018).

Werden in einem neuronalen Netzwerk Outputs durch Feedbackschleifen in das Modell zurückgeführt, spricht man von Recurrent Neural Networks (RNNs) (vgl.: Goodfellow u. a. 2016: 163). Durch die Rückführung haben RNNs ein „Gedächtnis“. Sie können die aktuelle Information als zusätzlichen Input für den nächsten Schritt benutzen und nach und nach komplexere Muster lernen (vgl.: ebd. 364ff.). RNNs werden für Daten verwendet, die sequenziell angeordnet sind, beispielsweise eine Folge von Noten, also eine Melodie. Eine Weiterentwicklung davon sind Gated RNNs, die durch eingebaute „Tore“ entscheiden, welche Informationen behalten und welche vergessen werden sollen, um die Verarbeitung von langfristigen Abhängigkeiten in Sequenzen zu verbessern (vgl.: ebd. 2016). Auf diese Weise funktionieren zum Beispiel Long Short-Term Memory Modelle (LSTMs). Eine Anwendung für LSTMs wäre etwa das Fortführen einer Musiksequenzen im gleichen Stil (vgl.: Eck und Schmidhuber 2002).

Bei Generative Adversarial Networks (GANs) treten zwei Netzwerke gegeneinander an. Ein Generatornetzwerk erstellt Replikat von Daten aus einem Datenset und ein Diskriminatornetzwerk versucht, generierte Daten von Originaldaten zu unterscheiden (vgl.: Goodfellow u. a. 2016: 690ff). Wenn keine Unterscheidung mehr möglich ist, liegen perfekte Replikat vor. Ein bekanntes Modell ist DrumGAN zur synthetischen Erzeugung von Schlagzeugklängen (vgl.: Nistal u. a. 2020).

In Transformer-Netzen können alle Positionen der Eingangssequenz parallel verarbeitet werden, was die Effizienz der Modelle etwa im Vergleich zu RNNs erhöht. Durch einen Selbstaufmerksamkeitsmechanismus sind sie in der Lage, sich auf bestimmte Teilmengen ihres Inputs zu konzentrieren (vgl.: Vaswani u. a. 2017). Dadurch können die Systeme bei jedem Generierungsschritt auf einen beliebigen Teil der zuvor generierten Ausgabe zugreifen, auch wenn dieser weit entfernt ist (vgl.: Huang u. a. 2018: 1). Im Transformermodell gibt es einen Sequenzencoder und einen Sequenzdecoder, die jeweils aus mehreren gestapelten Selbstaufmerksamkeits-schichten bestehen. Diese Systeme werden vor allem zur Sprachgenerierung eingesetzt. Die generativen Musik-Systeme MMM (Ens und Pasquier 2020), MusicTransformer (Huang u. a.

2018), VampNet (Garcia u. a. 2023), Composer’s Assistant (Malandro 2023) und MuseNet (Payne 2019) basieren alle auf dem Transformer-Modell, haben aber unterschiedliche spezifische Architekturen.

2.1.3 MMM-Modell

Die Multi-Track Music Machine (MMM) wurde von Ens und Pasquier (2020) entwickelt, um einen neuen Ansatz für Transformer-basierte Musikgenerierung zu bieten. Ein GPT-2-Modell (generativer, vortrainierter Transformer) wurde trainiert. Dabei entstand ein System mit einem Aufmerksamkeitsfenster von 16 Takten. Die Trainingsdaten stammten aus dem MetaMIDI Dataset (MMD), das rund eine halbe Millionen MIDI- und Metadateien enthält.

„Multi-Track“ bezieht sich auf die Fähigkeit des Systems, für zwei oder mehr Instrumentenspurten gleichzeitig Musik zu generieren. Das System kann vorhandenes Musikmaterial fortführen oder ganz von vorne beginnen. Bei Musik-Transformern sind die musikalischen Ereignisse verschiedener Tracks, die als Token dargestellt werden, klassischerweise alle in einer Sequenz miteinander verschachtelt (interleaved). Bei MMM hingegen wird zunächst für jeden Track eine eigene, zeitlich geordnete Sequenz erstellt. Erst danach werden die Track-Sequenzen zu einer Gesamtsequenz verkettet. Durch diese Änderung der Darstellung ist es sowohl möglich, auf die Attribute jedes einzelnen Tracks zuzugreifen als auch einzelne Noten durch das System hinzufügen zu lassen (Inpainting). Diese beiden Funktionen sind bei vielen anderen Modellen nicht vorhanden. Ausnahmen sind Composer’s Assistant (Malandro 2023) und VampNet (Garcia u. a. 2023).

Die symbolische Darstellung der Musik erfolgt in MMM mit Hilfe von Tokensequenzen. In anderen Systemen, wie zum Beispiel MuseNet (Payne 2019), bekommt jeder Pitch auf jedem Track ein eigenes Token. Ein C auf Track 1 hat einen anderen Namen als ein C auf Track 2. Dadurch entstehen schnell große Mengen von Token. Da die Größe des Token-Vokabulars begrenzt ist (vgl.: Ens 2023: 127), führt dies letztlich zu einer Limitierung der möglichen Anzahl von Tracks. In MMM sind die Tracks getrennt. Dadurch können in jedem Track die gleichen NOTE_ON- und NOTE_OFF-Token für die Noten verwendet werden, und die Anzahl der unterschiedlichen Token wird deutlich reduziert, was die Anzahl der möglichen Tracks wiederum erhöht. Gleiches gilt für die Instrumente. MMM erlaubt die Verwendung aller 128 Standard-MIDI-Instrumente. Der Composer’s Assistant (Malandro 2023) verfügt ebenfalls über 128 Instrumente, während zum Beispiel MuseNet (Payne 2019) auf zehn Instrumente beschränkt ist.

Auf der Taktebene gibt es bei MMM 128 NOTE_ON- und 128 NOTE_OFF-Token (einen für jeden möglichen Pitch) und 48 TIME_SHIFT-Token⁴. BAR_START- und BAR_END-Token umrahmen jeden Takt. In der Sequenz sind die Tracks verkettet. Sie beginnen jeweils mit einem TRACK_START-Token und enden mit einem TRACK_END-Token. Weitere Meta-Token, wie beispielsweise das ausgewählte MIDI-Instrument, befinden sich ebenfalls auf Track-Ebene. Das Musikstück beginnt mit einem PIECE_START-Token. Das Ende des Stücks wird bei n Spuren durch das n-te TRACK_END-Token signalisiert. Diese Darstellungsform wird Multitrack-Repräsentation genannt.

MMM erlaubt auch die Generierung einzelner Takte in einem Track. Dies erfordert eine etwas andere Repräsentationsform, die dem System Zugriff auf jeden einzelnen Takt erlaubt (Bar-Fill-Repräsentation). In dieser Darstellungsform werden die zu ersetzenden Takte aus der Sequenz entfernt und durch ein FILL_PLACEHOLDER-Token ersetzt. Nach dem letzten TRACK_END-Token werden dann die neuen Takte angehängt, jeweils umrahmt von einem FILL_START- und einem FILL-END-Token.

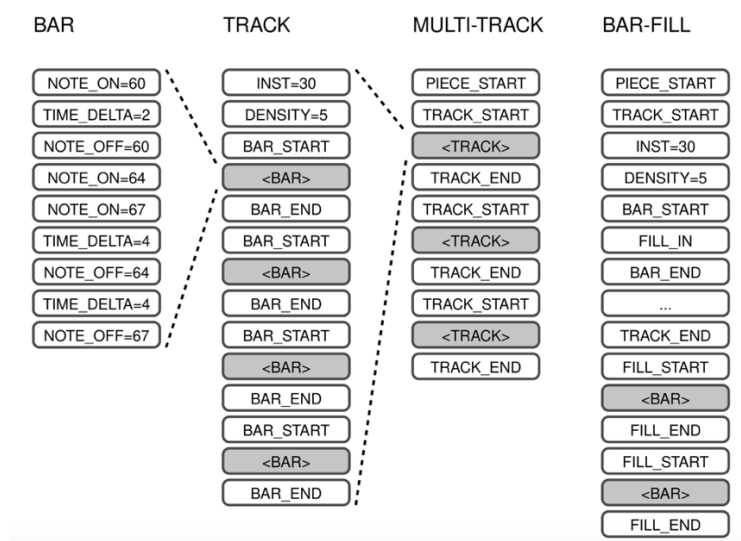


Abbildung 2: Multitrack- und Bar-Fill-Repräsentation
Quelle: Ens (2023: 128)

Zur Veranschaulichung der Darstellungsformen dient die Übersicht in Abbildung 2. Für das Training von MMM wurden zwei Modelle separat trainiert. Ein Modell in der Multitrack-Repräsentation, zum Füllen einzelner Tracks, und eines in der Bar-Fill-Repräsentation, zum Füllen einzelner Takte. Für beide Modelle wurde eine Version bereitgestellt, die vier Takte mit

⁴ Ein Beat wird in zwölf Zeitabschnitte quantisiert. Bei vier Beats pro Takt ergeben sich daraus 48 Token, mit denen alle möglichen Rhythmen abgebildet werden.

maximal zwölf Instrumententracks generieren kann und eine, die acht Takte mit maximal sechs Instrumententracks generieren kann.

2.2 User Experience Testing

User Experience Testing wird im Kontext dieser Arbeit als Oberbegriff für alle Evaluationsmethoden verwendet, die sich mit der Nutzer:innenerfahrung beschäftigen. Dazu gehören auch die Usability- und Technology Acceptance-Tests, die im Rahmen der vorliegenden Untersuchung durchgeführt wurden. Die Idee solcher Erhebungen ist, dass reale Nutzer:innen mit dem untersuchten System interagieren und die Beobachtenden daraus Schlüsse über konkrete Anwendungsprobleme ziehen können (vgl.: Nielsen 1993: 165).

In diesem Abschnitt werden zunächst die Begriffe User Experience, Usability und Akzeptanz kurz erläutert. Anschließend wird der vom Deutschen Institut für Normung erarbeitete Standard der DIN EN ISO 9241-210 zur menschenzentrierten Gestaltung interaktiver Systeme vorgestellt. Die Ziele und Prozesse dieses Ansatzes werden erläutert und am Beispiel der benutzer:innen-zentrierten Evaluation wird eine mögliche Integration der Normen in die Praxis aufgezeigt. Anschließend werden die Evaluationsergebnisse der ersten MMM-Iteration aus der Studie von Tchameube u. a. (2023) vorgestellt und das Prinzip der iterativen Entwicklung am konkreten Beispiel von MMM veranschaulicht.

2.2.1 User Experience, Usability und Akzeptanz

Der Begriff User Experience, kurz UX, setzt sich aus den englischen Wörtern „user“ (Benutzer:in) und „experience“ (Erfahrung) zusammen. Er bezeichnet die Wahrnehmungen und Reaktionen von Menschen bei der Nutzung eines Produkts, einer Dienstleistung oder eines Systems. Dazu gehören auch alle Gefühle, die eine Person vor, während und nach der Nutzung empfindet, ebenso wie die Erwartungen, Erfahrungen und das Wissen, das sie bereits mitbringt. UX umfasst das Markenimage, die Attraktivität, die Usability und Nützlichkeit des Produkts (vgl.: Jacobsen und Meyer 2017: 36). Die ganzheitliche Betrachtung der Prozesse und Zusammenhänge zwischen diesen Faktoren ist der Kern des UX-Ansatzes.

Usability (Gebrauchstauglichkeit) sagt nach ISO 9241-210 aus, wie leicht ein System, ein Produkt oder eine Dienstleistung von bestimmten Anwender:innen in einer bestimmten Situation genutzt werden kann, um bestimmte Ziele gut und schnell zu erreichen. Entscheidend ist also der Nutzungskontext - wer nutzt was, wie und wofür? Die Usability wird immer nur für den spezifischen Nutzungskontext bewertet. Für die Bewertung können verschiedene Aspekte

während der Nutzung betrachtet werden: Effektivität (Kann ich mit dem System mein Ziel erreichen?), Effizienz (Kann ich mit dem System zeit- und ressourcenschonend arbeiten?) und Zufriedenstellung (Entspricht mein Benutzungserlebnis mit dem System meinen Erwartungen?). Je nach System und Nutzungskontext können sich diese Aspekte durch unterschiedliche Eigenschaften auszeichnen. Ein co-kreatives System wie MMM beispielsweise wird als effektiv wahrgenommen, wenn es gut kontrollierbar ist, als effizient, wenn sich die Benutzer:innen schnell auf der Bedienoberfläche zurechtfinden und als zufriedenstellend, wenn sie sich mit Hilfe des Systems kreativ entfalten können.

Die Akzeptanz im Zusammenhang mit UX bezieht sich darauf, inwieweit Benutzer:innen ein Produkt, ein System oder eine Dienstleistung annehmen, mögen oder bereit sind, es zu verwenden. Es ist entscheidend, dass ein Produkt nicht nur funktionell ist, sondern auch positiv wahrgenommen wird. In der Entwicklungsphase können so Prognosen über den Erfolg bei der Markteinführung gemacht werden. Die Akzeptanz kann durch die Usability und andere UX-Aspekte beeinflusst werden. Handelt es sich um die Akzeptanz technischer Systeme, wird auch der Begriff Technology Acceptance verwendet.

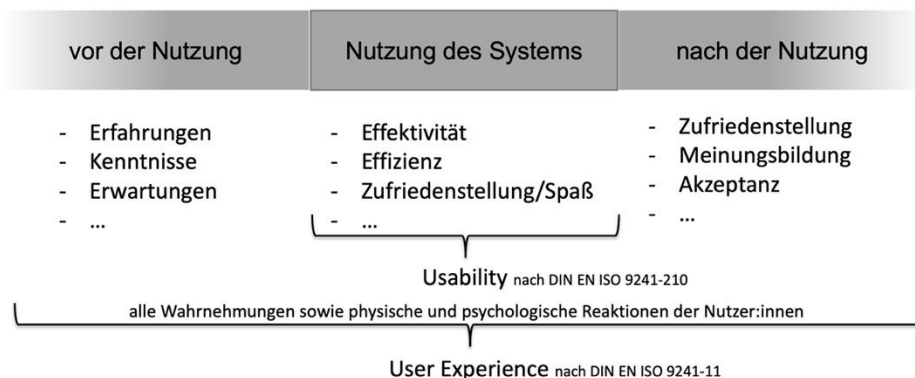


Abbildung 3: User Experience, Usability und Akzeptanz

Quelle: Eigene Darstellung, in Anlehnung an Jacobsen und Meyer 2017:60 und ISO 9241-11:2020: 33

2.2.2 Menschzentrierte Gestaltung nach DIN EN ISO 9241-210

Nielsen (1993), Rosenzweig (2015) und Jacobsen und Meyer (2017) beziehen sich für die Begründung von User Experience Testing auf die Normen zur „Ergonomie der Mensch-System-Interaktion“. Darunter sind international gültige Standards zusammengefasst, die sich mit dem Verständnis der Interaktionen zwischen menschlichen und sonstigen Komponenten eines Systems befassen. Die Normen sind für private Unternehmen nicht verpflichtend, sondern als Anregung zu verstehen.

Für diese Arbeit ist ein Teil der Normenreihe von besonderem Interesse: ISO 9241-210 stellt mit dem Konzept der menschenzentrierten Gestaltung (auch: Human-Centred Design) einen konkreten Ansatz bereit, der in die Systemgestaltung mit einfließen kann. In der Norm werden vier menschenzentrierte Gestaltungsaktivitäten, auch „Human-Centred Processes“ (HCPs) genannt, vorgestellt. Ergänzend dazu werden in ISO 9241-220 detaillierte Prozessbeschreibungen für die Umsetzung der Aktivitäten gegeben. Zum Verständnis der grundlegenden Begriffe und Konzepte der Usability wird an einigen Stellen auf ISO 9241-11 verwiesen.

Ein System gilt als „interaktiv“, wenn die Benutzer:innen für dessen Bedienung mit Software, Hardware, Dienstleistungen oder anderen Menschen interagieren. „Menschzentriert“ bedeutet, dass alle Akteur:innen des Systems einbezogen werden. Es geht also nicht nur um die Benutzer:innen, sondern auch alle Stakeholder:innen werden mit ihren Bedürfnissen berücksichtigt und sind von den Auswirkungen betroffen.

Die Anwendung des menschenzentrierten Ansatzes wird in der Norm empfohlen, da sie erhebliche positive Auswirkungen für alle Akteur:innen des Systems haben kann: „In hohem Maße gebrauchstaugliche Systeme und Produkte sind sowohl in technischer als auch in kommerzieller Hinsicht tendenziell erfolgreicher“ (ISO 9241-210:2020: 12). Es entstehen dadurch ökonomische und soziale Vorteile für Benutzer:innen, Arbeitgeber:innen und weitere Stakeholder:innen. Im ganzheitlichen Ergonomieansatz der Normenreihe wird darüber hinaus von einer Steigerung der Nachhaltigkeit und sozialen Verantwortung gesprochen.

Abbildung 4 gibt einen Überblick über die Einflüsse und Auswirkungen menschenzentrierter Gestaltung. Als Beispiel für eine Gestaltungsaktivität ist die benutzer:innenzentrierte Evaluierung dargestellt, bei der ein Wechsel von der human-centred Ebene zur user-centred Ebene erfolgt.

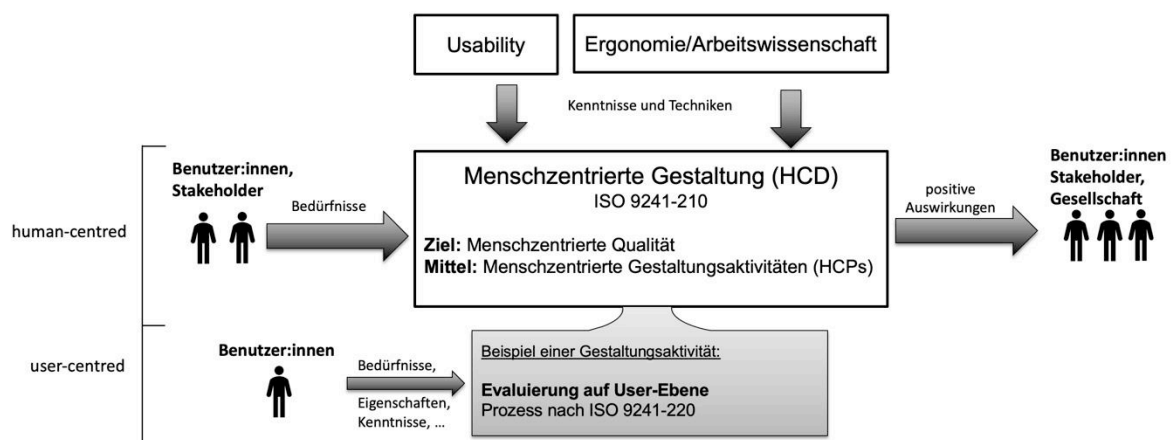


Abbildung 4: Einflüsse, Auswirkungen, Ziele und Mittel menschenzentrierter Gestaltung
Quelle: Eigene Darstellung

Menschzentrierte Qualität

Der menschenzentrierte Ansatz wendet Kenntnisse und Techniken aus den Bereichen der Ergonomie/Arbeitswissenschaft und der Usability an. Ziel ist es, bestimmte Qualitätsstandards für das Gesamtsystem zu erreichen. Dieser Qualitätsbegriff bezieht sich auf die Komponenten Usability, Barrierefreiheit, allgemeine Nutzungserfahrung (UX) und die Vermeidung nutzungsbedingter Schäden. Zwischen den einzelnen Komponenten bestehen wechselseitige Beziehungen. Die Zuordnung von Systemanforderungen zu einer bestimmten Komponente ist nicht immer eindeutig, aber auch nicht zwingend, solange die menschenzentrierte Qualität als Gesamtziel im Vordergrund steht.

Menschzentrierte Gestaltungsaktivitäten

Um das Gesamtziel menschenzentrierter Qualität zu erreichen, werden in der ISO 9241-210 vier Aktivitäten (HCPs) vorgestellt. Die erste Aktivität ist das Verstehen und Beschreiben des Nutzungskontextes. Dafür werden zunächst die verschiedenen Benutzer:innen und Stakeholder:innen identifiziert. Benutzer:innen können bei Bedarf in Typen eingeteilt werden. Für jeden Typ werden spezifische Merkmale aufgezählt (Kenntnisse, Kompetenzen, Erfahrung, Ausbildung, Übung, physische Merkmale, Gewohnheiten, Vorlieben und Fertigkeiten). Danach werden die Ziele und Aufgaben der Benutzer:innen und die organisatorische, technische und physische Umgebung identifiziert.

Die zweite Aktivität ist das Festlegen der Nutzungsanforderungen, die sich aus dem Nutzungskontext ergeben. Dazu gehören beispielsweise auch konkrete Anforderungen an die Usability. Daraufhin werden als dritte Aktivität anwendbare Gestaltungsregeln erarbeitet. Gestaltet werden müssen die Aufgaben (für Benutzer:innen oder für die Technik), die Interaktionen (zwischen Menschen und Technik) und das Interface. Es kann zum Beispiel zunächst ein Prototyp hergestellt werden. Bestimmte Grundsätze, wie Controllability (Steuerbarkeit), werden dabei schon bei der Gestaltung beachtet.

Die vierte Aktivität zur Erreichung menschenzentrierter Qualität ist die benutzer:innenzentrierte Evaluierung der Gestaltung. Durch regelmäßiges Testen in unterschiedlichen Phasen (beispielsweise mit Prototypen) ergeben sich neue Informationen für alle anderen Aktivitäten. ISO 9241-220 listet typische prozessbezogenen Aktivitäten für die Planung und Durchführung von Evaluationen auf.

Einbindung in die Praxis

Menschzentrierte Gestaltung nach ISO 9241-210 ist ein umfassendes Rahmenwerk, das als inspirierende Leitlinie für Unternehmen zur Verfügung steht. Den Unternehmen steht es in jeder Phase der Entwicklung frei, diejenigen Elemente der menschzentrierten Gestaltungsaktivitäten zu übernehmen, die ihr bestehendes Gestaltungskonzept, etwa Wasserfall, objektorientierte oder agile Entwicklung ergänzen. Das ermöglicht eine flexible Anpassung an die individuellen Gegebenheiten und Prozesse der Unternehmen.

Steinberg arbeitet mit der agilen Projektmanagementmethode Scrum. Diese zeichnet sich wie die menschzentrierte Gestaltung durch iterative Arbeitszyklen aus. Bei beiden Ansätzen stehen die Bedürfnisse der Benutzer:innen im Vordergrund. Ein Hinzuziehen der Norm für einzelne Arbeitsschritte in der agilen Entwicklung bietet sich an.

Ein Beispiel hierfür ist die regelmäßige Durchführung von benutzer:innenzentrierter Evaluation im Entwicklungsprozess von MMM. Bei dieser iterativen Methode findet ein Wechsel von der human-centred Ebene zur user-centred Ebene statt. Gesellschaftliche und soziale Auswirkungen werden ausgeklammert und der Fokus liegt auf dem, was die Benutzer:innen mitbringen, was sie brauchen, wo ihre Bedürfnisse noch nicht erfüllt sind und was verändert werden muss, damit sich ihr Nutzungserlebnis verbessert.

2.2.3 Evaluation in der ersten MMM-Iteration durch Tchemeube u. a.

In der Studie von Tchemeube u. a. (2023) wurde die Multi-Track Music Machine (MMM) als ko-kreatives KI-Tool in Cubase für Musikkomposition mit einem Mixed-Method-Ansatz (qualitativ und quantitativ) evaluiert. Es wurden Usability, User Experience und Technology Acceptance evaluiert. Für die Integration von MMM in Cubase stand ein Prototyp zur Verfügung. Das Interface bestand lediglich aus einem einstellbaren Parameter („Temperature“) und einem „Generate“-Button.

18 Personen nahmen an der Studie teil. Alle verfügten über Kompositionserfahrung auf Expert:innenniveau. Die Studie erstreckte sich über fünf Wochen und bestand aus drei Phasen, in denen die Teilnehmenden mit Hilfe des Plugins je eine Kompositionsaufgabe von zu Hause aus bearbeiteten. In jeder Phase füllten sie quantitative Fragebögen zu Usability und Technology Acceptance aus und beantworteten qualitative Fragen zu ihren Nutzungserfahrungen.

Die Ergebnisse für Usability und Technology Acceptance waren insgesamt positiv. Es konnte jedoch festgestellt werden, dass das Ein-Parameter-Design für die Expert:innen nicht

zufriedenstellend war. Für zukünftige Gestaltungsüberlegungen für das Tool wurde von den Forschenden empfohlen, die Erfahrungen und Kenntnisse der Benutzer:innen zu berücksichtigen. Sie wiesen darauf hin, dass das System bereits die Möglichkeit bietet, weitere Parameter zu integrieren. Der Zugriff auf diese Parameter müsse lediglich im Interface implementiert werden. Mehr Einfluss auf die Musikgenerierungen des Systems durch mehr Parameter könnte den Benutzer:innen das Gefühl vermitteln, mehr Kontrolle und Autonomie zu haben und die Zufriedenheit während der Benutzung steigern.

2.2.4 Iterative Entwicklung am Beispiel von MMM

Menschzentrierte Gestaltungsaktivitäten können als Kreislauf mit wechselseitigen Abhängigkeiten zwischen den Aktivitäten dargestellt werden. Der Entwicklungsprozess kann als eine Abfolge von Schritten ohne festen Anfangspunkt verstanden werden. Die einzelnen Aktivitäten werden ständig durch neue Informationen aus den anderen Schritten beeinflusst. Die Wiederholung der Schritte wird als Iteration bezeichnet. Somit entsteht ein Entwicklungsfluss, der sich stark an den Benutzer:innen orientiert.

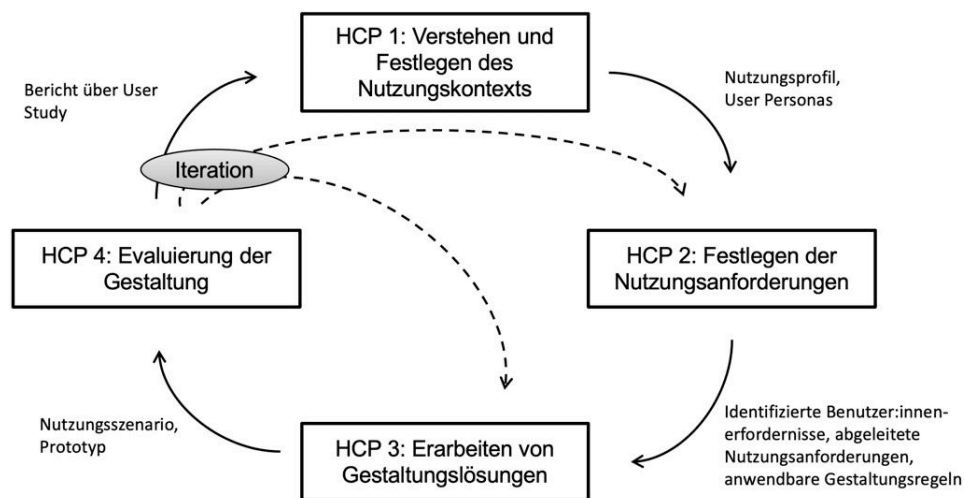


Abbildung 5: Iterative Entwicklung im Kontext menschzentrierter Gestaltung

Quelle: Eigene Darstellung, in Anlehnung an ISO 9241-210:2020: 21

Am Beispiel von MMM lässt sich dieser Fluss nachvollziehen, siehe Abbildung 5. In der ersten Iteration wurden Usability, UX und Akzeptanz von MMM durch Tchemeube u. a. (2023) evaluiert (HCP 4). Die untersuchte Benutzer:innengruppe waren Expert:innen. Das Ergebnis der Evaluation zeigte, dass die Expert:innen sich mehr Gestaltungseinfluss auf die Komposition

wünschten, was zu einer Erweiterung der Nutzungsanforderungen (HCP 2) und der Erarbeitung einer neuen Gestaltungslösung (HCP 3) führte.

Für die zweite Iteration wurde also ein neuer Prototyp gestaltet. Basierend auf den Ergebnissen aus der ersten Evaluation wurden diesem Prototypen mehr einstellbare Parameter für mehr Gestaltungseinfluss hinzugefügt. Auch in der zweiten Iteration sollte eine Evaluation durchgeführt werden, dieses Mal mit Nutzer:innen mit wenig Kompositionserfahrung. Der Nutzungskontext wurde entsprechend angepasst (HCP 1). Daraus ergaben sich neue Nutzungsanforderungen (HCP 3): Das Interface musste für Anfänger:innen intuitiv verständlich sein, das betraf zum Beispiel die Beschriftungen und Funktionsbeschreibungen auf der Bedienoberfläche (wenig Fachwörter, intuitiv verständliche Benennung der Parameter und Skalen). Dadurch sollte das System auch für Anfänger:innen schnell erlernbar sein, ohne dass der Funktionsumfang sie überforderte. Diese Überlegungen flossen auch in die Gestaltung des neuen Prototypen mit ein. Genau so stand er dann für eine erneute Evaluation zur Verfügung (HCP 4).

3 Vorbereitung der User Study

Für die Studie wurde der in Abbildung 6 dargestellte Zeitplan mit fünf Phasen erstellt. Die Phasen wurden nach den fünf Hauptphasen für empirische Studien von Diekmann (2010) eingeteilt. Dabei wurde darauf geachtet, dass die für diesen spezifischen Anwendungsfall relevanten Empfehlungen zum Vorgehen bei benutzer:innenzentrierter Evaluation aus ISO 9241-220 berücksichtigt werden. An den entsprechenden Stellen wurden die Vorgaben von Diekmann ergänzt oder verfeinert.

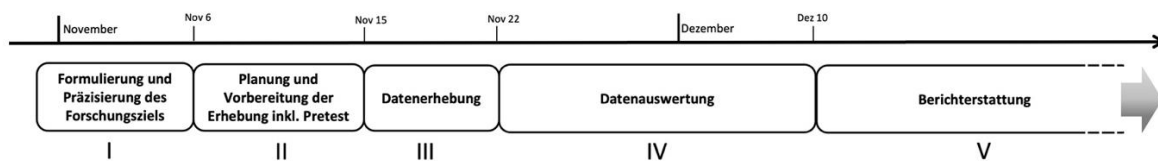


Abbildung 6: Zeitplan für User Study

Quelle: Eigene Darstellung, Phasen nach Diekmann (2010: 187)

In Phase I wurde die Fragestellung formuliert, eine Literaturrecherche gemacht, ein praktisches Problem formuliert, die angestrebten Ziele festgelegt und die beteiligten Akteur:innen genannt. Die Dokumentation von Phase I ist in Kapitel 0 nachzulesen.

In Phase II wurde die Erhebung geplant. Die Dokumentation der Planung erfolgt im folgenden Kapitel. Zunächst musste sichergestellt werden, dass das zu untersuchende System (in diesem Fall als Prototyp) zur Verfügung stand, siehe dazu Abschnitt 3.1. Außerdem musste ein

realitätsnaher Nutzungskontext für die Erhebung geschaffen werden. Dafür wurde ein Testszenario definiert, dokumentiert in Abschnitt 3.2. Die Untersuchungsform und das Erhebungsinstrument sowie weitere Überlegungen und Vorbereitungen dazu wurden ebenfalls in der Planungsphase festgelegt, siehe Abschnitte 3.3 und 3.4. Auch Größe und Typ der zu untersuchenden Benutzer:innengruppe wurden festgelegt, siehe Abschnitt 3.5. Zum Schluss wurde die Erhebungsform in einem Pretest erprobt, siehe Abschnitt 3.6.

In Phase III wurde das Erhebungsinstrument eingesetzt. Darauf schloss sich in Phase IV die Auswertung der Daten an. Die genaue Vorgehensweise ist in Kapitel 4 beschrieben. Phase V, die Berichterstattung, ist Gegenstand der vorliegenden Arbeit.

3.1 Prototyp

Als Prototyp wurde das Plugin Music Generation Panel (MGP) 1.5.0.188 entwickelt, mit dem das MMM-System in Cubase integriert ist. Abbildung 7 zeigt das Interface in den Standardeinstellungen. In der für die Erhebung verwendeten Version des Plugins (2. MMM-Iteration) gibt es sechs einstellbare Parameter: Genre, Mode, Duration, Density, Repetition und Crazyness. Diese Parameter wurden sorgfältig unter der Berücksichtigung der Zielgruppe der User Study ausgewählt. Tabelle 1 beschreibt diese Parameter genauer.

Das Plugin generiert MIDI-Noten für das aktuell ausgewählte Event im Cubaseprojekt. Als Event wird ein (Zeit-)Abschnitt auf einer Instrumentenspur bezeichnet. Einer angelegten Instrumentenspur wird zunächst ein Instrument zugewiesen (zum Beispiel HALion Sonic von Steinberg/Kontakt von Native Instruments/...). Wichtig dabei ist, dass die Spur richtig benannt wird (zum Beispiel „Piano“ für Klavier/„Guitar“ für Gitarre/...), damit MMM das Instrument erkennen kann, für das Musik generiert werden soll. In der Instrumentenspur wird dann ein Event erstellt. Wenn dieses ausgewählt ist, können vom MGP entweder neue MIDI-Noten dafür generiert werden (Replace Selection) oder weitere MIDI-Noten zum Event hinzugefügt werden (Add to Selection).

Tabelle 1: Parameter des Music Generation Panel 1.5.0.188

Parametername	Ausprägungen	Beschreibung
Genre	Alternative Rock, Ambient, Blues, Break-Beat, Classical, Classic Rock, Contemporary Rock, Country, Drum'n'Bass, Folk, Gospel, Hardcore Punk, Hardcore Techno, Heavy Metal, HipHop, House, Industrial, Jazz, Latin, Pop, Punk, Punk Rock, R&B, Reggae, Rock'n'Roll, Techno, Trance, Utility, World, None	Parameter für die Auswahl der Musikrichtung.
Mode	Auto*, Polyphonic, Monophonic, Drums	Parameter für die Auswahl des Modus', bezogen auf die Satztechnik der Musik. Das Instrument kann entweder mehrere Noten gleichzeitig spielen (Polyphonie) oder nur eine Note zur Zeit (Monophonie). Für die Generierung einer Schlagzeugspur gibt es den Drums-Modus.
Duration	Whole, Half, Quarter, Eighth, Sixteenth, Thirty-Second	Parameter für die Einstellung der erlaubten Notenlängen: ganze, halbe, Achtel-, Sechzehntel- und Zweiunddreißigstel-Noten. Mehrfachauswahl ist hier möglich.
Density	Auto*, Low, Medium, High	Parameter für die Einstellung der Notendichte (hoch, mittel, niedrig).
Repetition	Auto*, Low, Medium, High	Parameter für die Einstellung der Noten- bzw. Melodiemotivwiederholungen (hoch, mittel, niedrig).
Crazyness (ehemals Temperature)	1, 2, 3, 4, 5	Parameter für die Einstellung der „Verrücktheit“, wobei 1 eher konservative und 5 eher experimentelle Generierungen hervorbringt.

* Wenn die Auto Option gewählt ist, wird diese Einstellung vom System automatisch vorgenommen.

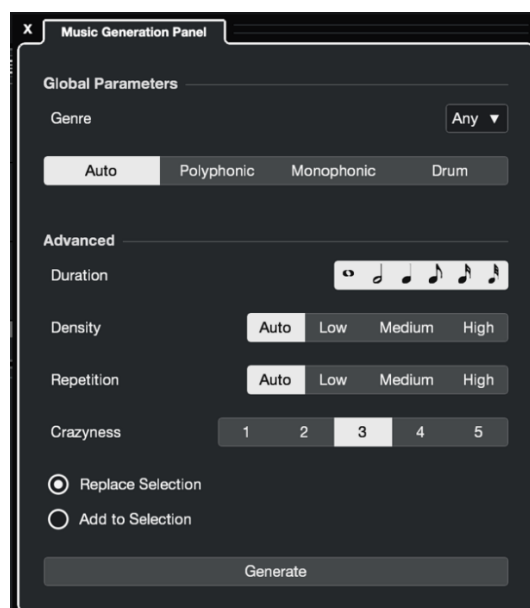


Abbildung 7: Music Generation Panel Interface

Quelle: Screenshot aus dem Music Generation Panel 1.5.0.188 in Cubase 13.0.10.123 auf MacOS 10.15.7

Der Prototyp war zum Zeitpunkt der Erhebung noch nicht voll funktionsfähig. Es kann nach Nielsen (1993: 93ff.) weder von einem horizontalen Prototypen (alle Funktionen vorhanden, eingeschränkte Funktionalität), noch von einem vertikalen Prototypen (eingeschränkter Funktionsumfang mit voller Funktionalität) gesprochen werden. In einem solchen Fall wird ein Szenario festgelegt. Der Prototyp funktioniert dann unter bestimmten Bedingungen fehlerfrei, die vorher von der Versuchsleitung festgelegt werden (s. Abbildung 8).

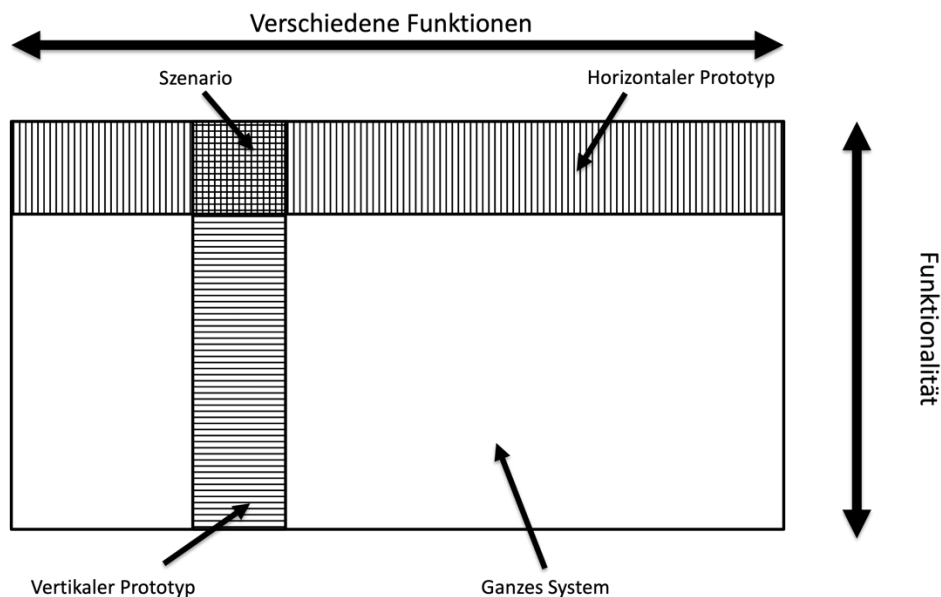


Abbildung 8: Dimensionen von Prototypen in Anlehnung an Nielsen
Quelle: Nielsen (1993: 94)

3.2 Szenario

Es war ein nicht behobener Bug⁵ des Systems bekannt. Dieser trat in der Planungsphase der Studie sehr oft auf. Durch einen Fehler in der Implementierung konnte das System nicht zuverlässig für Events mit weniger als acht Takten Musik generieren, genauer gesagt wurden dann die Parameter nicht zuverlässig berücksichtigt. Dadurch war beispielsweise Inpainting nicht möglich. Durch den begrenzten Zeitrahmen für die Anfertigung der Forschungsarbeit gab es nicht die Möglichkeit diesen Fehler vor Studienstart zu beheben. Es wurde beschlossen, das Auftreten des Bugs durch die Szenarioentwicklung so gut es geht zu umgehen. Bei genau acht Takten trat dieser Bug fast nie auf, weswegen für die Task festgelegt wurde, dass nur für achttaktige Events generiert werden soll. Es war nicht vollständig ausgeschlossen, dass die

⁵ Als Bug wird ein Softwarefehler bezeichnet, durch den sich das Programm unerwünscht verhält.

Teilnehmenden auf den Bug trafen, deswegen wurde dieser Umstand in der Ergebnisauswertung berücksichtigt.

Das Szenario für die Teilnehmenden wurde in Form eines voreingestellten Testprojekts (Template) erstellt. Vor Beginn der Task wurden sie von der Versuchsleitung auf die Rahmenbedingungen hingewiesen. In einem Tutorialvideo lernten sie die Projektstruktur kennen. Die Templates wurden auch deshalb bereitgestellt, weil die TN unterschiedliche Kenntnisse in speziell Cubase und DAWs im Allgemeinen mitbrachten. Auf diesem Weg mussten sie sich nicht mit Projekteinstellungen beschäftigen, sondern konnten ihre Zeit vollständig zum Komponieren nutzen. Alle wichtigen Voreinstellungen waren also bereits vorgenommen.

Es gab vier verschiedene Templates, die sich durch das Genre unterschieden (HipHop, Jazz, Pop, Rock). Jedes Genre hatte eine Genre-typische Instrumentierung und jeweils ein leeres, acht-taktiges Event pro Instrumentenspur, wobei die acht Takte der Drums-Spur bereits mit MIDI-Content gefüllt waren. Ein Stereo-Bus wurde für jedes Projekt angelegt. Das Computer-interne Audiogerät war als Audiosystem voreingestellt und die externen Kopfhörer als Output festgelegt. Für jedes Genre gab es bereits vier Instrumentenspuren: Drums, Bass, ein Lead- und ein Rhythmusinstrument. Alle Instrumente stammten aus der Library des Steinberg Instruments HALion Sonic SE. Hinzu kamen zwei bis drei Send-Effekte (Steinbergs RoomWorks SE, Ping-PongDelay und MonoDelay) sowie Insert-Effekte für einzelne Instrumente (zum Beispiel Steinbergs virtueller Röhrenverstärker DaTube für einen E-Bass). Die vier Instrumente wurden in jedem Template grob abgemischt, räumlich auf der Stereo-Basis verteilt (Panning) und teilweise mit Insert-Effekten versehen und/oder an Send-Effekte geschickt. Ein Loop-Cycle wurde für acht Takte eingeschaltet, sodass beim Abspielen immer wieder nach dem achten Takt zum ersten Takt übergegangen wird. In Abbildung 9 ist als Beispiel ein Rock-Template abgebildet. Die dazugehörigen Kanal- und Routing-Einstellungen sind in Abbildung 10 zu sehen.

Weitere Details des Szenarios sind in den Abschnitten zu Task-Based Usability Testing (3.4.1) und zum Versuchsaufbau (4.2) dokumentiert.

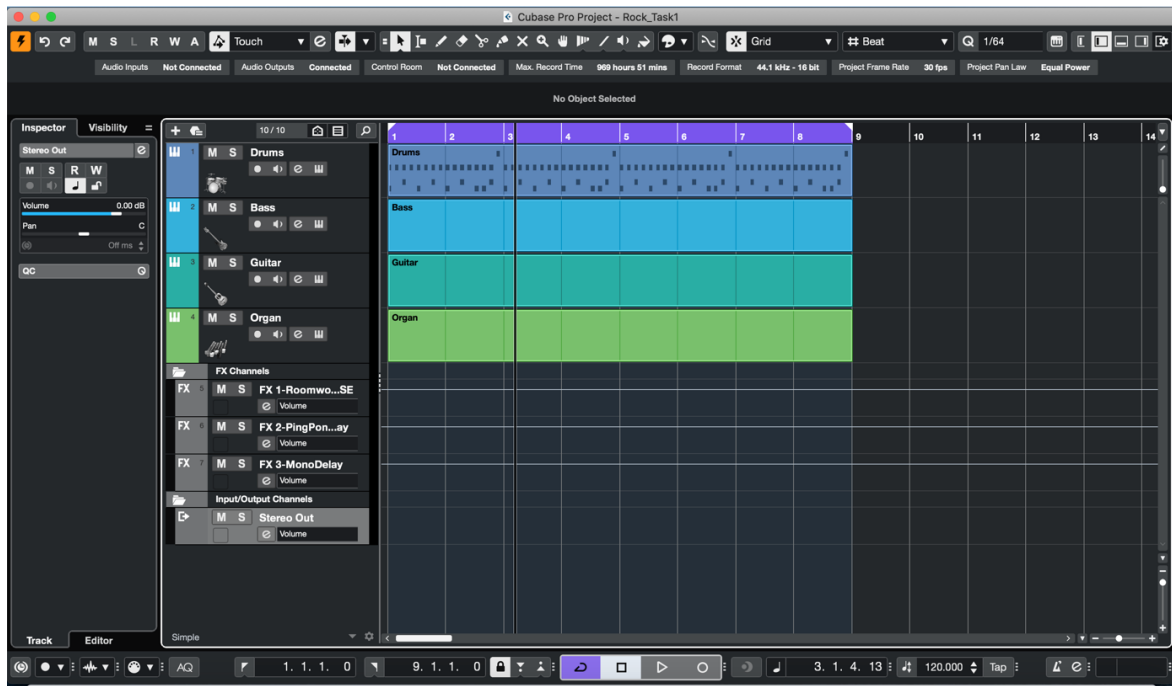


Abbildung 9: Cubase Template für Rock
 Quelle: Screenshot aus Cubase 13.0.10.123 auf MacOS 10.15.7



Abbildung 10: Kanal- und Routing Einstellungen für Rock-Template
 Quelle: Screenshot aus Cubase 13.0.10.123 auf MacOS 10.15.7

3.3 Untersuchungsform

Die Untersuchungsebene der User Study betrachtet die Interaktionen und Erfahrungen der Benutzer:innen mit dem MMM-System in Cubase, wobei sie spezifisch auf die Anwender:innen und deren Wechselwirkung mit dem Plugin ausgerichtet ist. Es handelt sich um eine iterative Studie, bei der nach der Bewertung einer vorangegangenen Iteration Änderungen vorgenommen wurden und nun ein neuer Test durchgeführt wird. Die für die Beantwortung der vorliegenden Forschungsfrage relevanten Elemente der ersten Studie wurden in identischer Form mit in die zweite übernommen. Daher kann man von einem Trenddesign sprechen. Beim Trenddesign werden Daten zu den gleichen Variablen zu unterschiedlichen Zeiten mit unterschiedlichen Stichproben erhoben (vgl.: Diekmann 2010: 305). Es gibt keine Vergleichs- oder Kontrollgruppe, deshalb handelt es sich um ein nicht-experimentelles Design (vgl.: ebd. 194).

3.4 Erhebungsinstrument

Die für die Beantwortung der vorliegenden Forschungsfrage wichtigen Messinstrumente wurden aus der Studie der ersten MMM-Iteration übernommen. Tchemeube u. a. haben ihrerseits gängige psychometrische⁶ Tests aus dem Bereich der UX-Forschung eingesetzt. Der Vorteil dabei ist, dass diese Fragebögen bereits auf die Gütekriterien Objektivität, Reliabilität und Validität überprüft wurden.

Durch die Objektivität eines Messinstruments wird ausgedrückt, „in welchem Ausmaß die Ergebnisse unabhängig sind von der jeweiligen Person, die das Messinstrument anwendet“ (Diekmann 2010: 249). Wenn das Messinstrument von einer anderen Person in gleicher Weise angewendet wird und dabei die gleichen Ergebnisse herauskommen, ist die Objektivität maximal. Für den Nachweis der Objektivität wird meist der Korrelationskoeffizient zwischen zwei Messungen berechnet, wobei ein Wert zwischen minus eins (minimale Objektivität) und eins (maximale Objektivität) herauskommt.

Die Reliabilität eines Messinstruments ist „ein Maß für die Reproduzierbarkeit von Messergebnissen“ (vgl.: ebd. 250). Reliabilität kann ebenfalls mit Hilfe eines Korrelationskoeffizienten nachgewiesen werden. Meistens werden dazu zwei parallele Messungen mit zwei

⁶ Psychometrische Tests sind „Befragungsverfahren der empirischen Psychologie, die psychologische Variablen quantifizieren. Es kann sich um Messungen der Reaktion auf physikalische Reize, Leistungstests, Intelligenztests, Einstellungsmessungen, Messung von Persönlichkeitseigenschaften u.a. handeln“ (Spektrum Lexikon der Psychologie 2000a)

vergleichbaren Messinstrumenten oder zwei aufeinanderfolgende Messungen mit gleichem Messinstrument verglichen.

Die Validität eines Messinstruments gibt an, in welchem Maß durch das Instrument auch wirklich das gemessen wird, was gemessen werden soll (vgl.: ebd. 257ff). Je besser die ausgewählten Items die zu messende Eigenschaft repräsentieren, desto höher ist die Inhaltsvalidität. Je stärker die Resultate mit anderen relevanten Merkmalen korrelieren, desto höher die Kriteriumsvalidität. Je mehr andere Variablen mit dem vom Messinstrument erfassten Konstrukt in theoretisch begründbaren Zusammenhängen steht, desto höher die Konstruktvalidität.

In Abbildung 11 sind im Zentrum des Zeitstrahls die drei Abschnitte der Testdurchführung „Demographie&Erfahrung“, „Task“ und „Usability&Acceptance“ zu sehen, jeweils mit einer Übersicht der darin verwendeten Methoden. Für die Usability-Untersuchung wurden die quantitativen Metriken System Usability Scale (SUS), Creativity Support Index (CSI) und Controllability verwendet, für die Technology Acceptance Untersuchung das Technology Acceptance Model (TAM). Es gab zwei Möglichkeiten für die Teilnehmenden ein qualitatives Feedback zu geben. Das geschah schriftlich durch zwei freie Kommentarfelder. Zur Bewertung der Zufriedenheit mit der Task wurde der quantitative Fragebogen After Scenario Questionnaire (ASQ) verwendet.

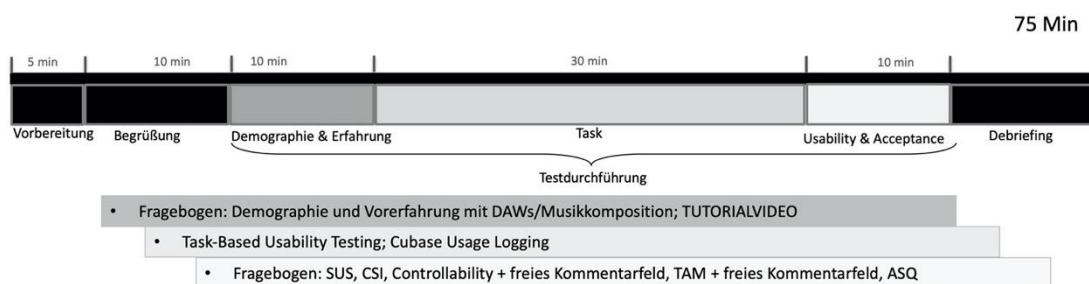


Abbildung 11: Ablauf einer Durchführung mit Methodenübersicht

Quelle: Eigene Darstellung, Abschnitte in Anlehnung an Nielsen (1993: 187ff.)

Im Folgenden werden die Messinstrumente näher erläutert. Es wird darauf hingewiesen, an welchen Stellen begründete Änderungen an den Originalfragebögen von Tchemeube u.a. vorgenommen wurden.

3.4.1 Task-Based Usability Testing

Für iterative Entwicklung im Sinne der DIN EN ISO 9241-210 wird vorausgesetzt, dass Beschreibungen, Spezifikationen und Prototypen regelmäßig überarbeitet und verbessert werden, um auf neue Informationen zu reagieren. Dadurch soll sichergestellt werden, dass das in der

Entwicklung befindliche System den Anforderungen der Benutzer:innen genügt. Die Rückmeldung der Anwender:innen über den Prototyp ist Teil des Prozesses. Eine einfache Demonstration reicht, um ein umfassendes Feedback zu erhalten. Effektiver ist es die Anwender:innen aufzufordern, Aufgaben (Tasks) mit dem Prototypen zu erledigen. Task-Based Usability Testing ist nach Rosenzweig (2015: 135f.) ein gutes Instrument, um im geleiteten Rahmen einer vorher festgelegten Aufgabe festzustellen, ob ein System in einem realistischen Szenario gut nutzbar ist.

Nielsen (1993: 185ff.) empfiehlt eine Task zu kreieren, die möglichst repräsentativ für den späteren Einsatz des Systems ist, wobei auch im Interface schon die wichtigsten Funktionen abgedeckt sein sollten. Weiterhin soll die Task weder zu umfangreich noch zu trivial für die angesetzte Zeit sein. Für die TN muss klar formuliert sein, was das Ziel der Task ist. Der Arbeitsauftrag soll möglichst in Schriftform an die TN gegeben werden, wobei es vor allem darum ginge, dass er allen TN auf dieselbe Weise mitgeteilt wird und sie später darauf zurückgreifen können. Aus diesen Empfehlungen wurde die Task für den vorliegenden Versuch konzipiert.

Tabelle 2: Arbeitsauftrag für die Task

Task Arrangement (30 Min)
Verwenden Sie das Plug-in "Music Generation Panel", um Musik für 3 neue Instrumentenspuren in Cubase zu komponieren. Bitte behalten Sie die vordefinierte Projektstruktur bei und folgen Sie den Anweisungen im bereitgestellten Tutorialvideo.

Dazu wurde den TN ein speziell auf das Szenario zugeschnittenes Tutorialvideo zur Verfügung gestellt, das sie sich bei Bedarf jederzeit wieder anschauen konnten. Darin wird zuerst der Aufbau des Projekts mit Drums, Bass, Lead- und Rhythmusinstrument sowie Send-Effekten erklärt. Es wird gezeigt, wie die Send-Effekte benutzt werden können und wie der Mixer funktioniert. Das MGP wird geöffnet und am Beispiel einer Bass-Spur wird gezeigt, wie es zur MIDI-Generierung benutzt werden kann. Zuletzt wird das Editieren der Noten im MIDI-Editor demonstriert.

Den TN wurde dazu gesagt, dass das Ziel der Task sei, das Plugin so gut es geht zu erkunden. Sie wurden gebeten Zwischenversionen abzuspeichern, wenn sie etwas kreiert haben, das ihnen gefällt. Diese Zwischenversionen mussten nicht alle Instrumente enthalten. Es konnte auch nur eine Bassline zu den Drums sein. Die TN konnten nach dem Zwischenspeichern auch von vorne beginnen, ohne dass ihre vorher komponierte Musik verloren ging. Dadurch bestand weder die Gefahr, dass einer Testperson langweilig wird, noch dass eine Person, die tendenziell länger braucht, am Ende ohne Ergebnis dasteht.

3.4.2 Demographie und Erfahrung

In diesem Fragebogen wurden Single-Choice und Multiple-Choice Fragen, so wie Fragen mit semantischer Differentialskala verwendet.

Tabelle 3: Variablen des Fragebogens zu Demographie und Erfahrung

Skala	Variable	Item
Demographie	DEM-01	Welche Geschlechtsidentität haben Sie? [Weiblich; Männlich; Transgender; Nicht-Binär; Möchte Ich nicht sagen; (Sonstige)]
	DEM-02	Welcher Altersgruppe gehören Sie an? [<20 Jahre; 20-29 Jahre; 30-39 Jahre; 40-49 Jahre; 50-59 Jahre; ≥60 Jahre]
Erfahrung	ERF-01	In welchem Kontext haben Sie bisher eine Digitale Audio Workstation (DAW) genutzt? (Mehrfachnennung möglich) [Ich habe eine DAW in meiner Freizeit genutzt; Ich habe eine DAW in meinem Beruf genutzt; Ich habe eine DAW in meiner Ausbildung/meinem Studium genutzt; (Sonstiges)]
	ERF-02	Wie hoch schätzen Sie Ihre Erfahrung mit Digitalen Audio Workstations ein? [0: Anfänger:in – 10: Expert:in]
	ERF-03	Welchem Nutzungsprofil würden Sie sich selbst in der angefügten Grafik zuordnen? [Entdecker:in; Enthusiast:in; Anfänger:in; Expert:in]
	ERF-04	Haben Sie schon einmal eine eigene Musikkomposition in einer Digital Audio Workstation produziert? [Ja (→ERF-05); Nein (→ERF-07)]
	ERF-05	Wie hoch schätzen Sie Ihre Erfahrung mit digitaler Musikkomposition ein? [0: Anfänger:in – 10: Expert:in]
	ERF-06	Was war Ihnen bei der Musikkomposition am wichtigsten? (Mehrfachnennung möglich)? [Kreativität ausleben; Innovative Klänge erzeugen; Neue Technologien ausprobieren; Eigene Persönlichkeit ausdrücken; Aufgabe / Ziel / Zweck erfüllen; Eigene Idee umsetzen; Software kennenlernen; Bekannte Klänge / Produktionstechniken imitieren; (Sonstige)]
	ERF-07	Haben Sie Erfahrung mit KI-unterstützter Musikkomposition? [Ja; Nein]

3.4.3 Quantitative Metriken

System Usability Scale

Die System Usability Scale (SUS) wurde von Brooke (1996) entwickelt. Die Intention war, einen einfachen, kurzen psychometrischen Test zu gestalten, der einen Überblick über die subjektive Bewertung der Bedienfreundlichkeit bietet (vgl.: Brooke 1996: 191). Brookes Test besteht aus zehn Zustimmungsaussagen mit fünfstufiger Likert-Skala. Die Aussagen sind abwechselnd positiv und negativ formuliert. Dadurch sollen Verzerrungen vermieden werden, da die Befragten über jede Aussage im Einzelnen nachdenken müssen (vgl.: ebd.: 191).

Der SUS-Fragebogen ist ein häufig angewendetes Instrument im Usability Testing. Es werden damit viele verschiedene Systeme evaluiert, ob Buchungssystem (Gustafsson 2019), App für

den Digitalen Mutterpass (Nissen u. a. 2023) oder Augmented Reality System für virtuellen Gitarrenunterricht (Martin-Gutierrez u. a. 2020).

In der Studie von Tchemeube u. a. (2023) und in dieser Erhebung wurde das Instrument übernommen, um die Usability von MMM in Cubase zu bewerten. Es wurde im Anschluss zu den zehn Statements noch ein Item zur Wahrnehmung der Usability insgesamt auf einer Skala von eins („nicht bedienfreundlich“) bis zehn („bedienfreundlich“) hinzugefügt.

Tabelle 4: SUS-Variablen

Skala	Variable	Item
SUS <i>Likert-Skala (5)</i>	SUS-01	Ich kann mir gut vorstellen, das Plugin regelmäßig zu nutzen.
	SUS-02	Ich empfand das Plugin als unnötig komplex.
	SUS-03	Ich empfand das Plugin als einfach zu nutzen.
	SUS-04	Um das Plugin zu nutzen bräuchte ich technischen Support.
	SUS-05	Ich fand, dass die verschiedenen Funktionen im Plugin gut integriert waren.
	SUS-06	Ich fand, dass es im Plugin zu viele Inkonsistenzen gab.
	SUS-07	Ich kann mir vorstellen, dass die meisten Leute schnell lernen, das Plugin zu benutzen.
	SUS-08	Ich empfand die Bedienung des Plugins als sehr umständlich.
	SUS-09	Ich habe mich sehr sicher im Umgang mit dem Plugin gefühlt.
	SUS-10	Ich musste sehr viele Dinge lernen, bevor ich das Plugin benutzen konnte.
Usability insgesamt <i>Skala: von 1: nicht bedienfreundlich bis 7: bedienfreundlich</i>	USA	Ich würde die Bedienfreundlichkeit dieses Produkts insgesamt wie folgt bewerten:

Für die Auswertung wird ein Gesamtscore der zehn SUS-Items zwischen 0 und 100 berechnet. Zuerst werden die Antworten der TN, die auf einer Skalenposition zwischen eins und fünf liegen, in eine Punktzahl umgewandelt. Das geschieht für SUS-01, -03, -05, -07 und -09 durch die Skalenposition minus 1. Für SUS-02, -04, -06, -08 und -10 wird 5 minus die Skalenposition gerechnet (Formel 1). Nun liegen die Punktzahlen zwischen null und vier. Danach wird für jede:n TN die Summe der Punkte gebildet, die dann mit 2,5 multipliziert wird, um einen Score zwischen 0 und 100 zu erreichen (Formel 2). Aus allen Scores wird zum Schluss ein Mittelwert gebildet (Formel 3).

Wenn $X_{i,j}$ die Skalenposition der i -ten Person auf das j -te SUS-Item repräsentiert:

$$Punktzahl_{i,j} = \begin{cases} X_{i,j} - 1 & \text{für } j \in \{01, 03, 05, 07, 09\} \\ 5 - X_{i,j} & \text{für } j \in \{02, 04, 06, 08, 10\} \end{cases} \quad (1)$$

$$Score_i = 2,5 * \sum_j Punktzahl_{i,j} \quad (2)$$

$$Score_{SUS} = \frac{1}{N} \sum_i Score_i \quad (3)$$

Für die Bewertung des Scores wurde von Tchemeube u. a. (2023: 5772) festgelegt, dass ein Score unter 50 inakzeptabel, zwischen 50 und 70 marginal und über 70 akzeptabel sei. Das entspricht auch den Bewertungskriterien von Bangor u. a. (2008).

Ein akzeptabler SUS-Score spräche dafür, dass unerfahrene Komponist:innen ihre Kompositionen durch die Verwendung von MMM effizient und zufriedenstellend kreieren können, während ein marginaler oder inakzeptabler Score auf Mängel und Hindernisse im Interface oder Gesamtsystem hinweisen könnte.

Creativity Support Index

Der Creativity Support Index (CSI) von Cherry und Latulipe (2014) ist ein psychometrischer Fragebogen, um zu bewerten, wie gut ein kreatives Tool dafür geeignet ist, Benutzer:innen bei kreativer Arbeit zu unterstützen. Dabei werden sechs Dimensionen bewertet: „Erkundung“, „Vereinnahmung“, „Ausdruckskraft“, „Spaß“, „Ergebnisse lohnen Aufwand“ und „Kollaboration“. Jede Dimension wird über zwei Zustimmungsaussagen bewertet. Als „Creativity Support Tool“ gelten nach Cherry und Latulipe (2014: 2) alle Werkzeuge, mit denen jemand in einem unbegrenzten Prozess neue Artefakte schaffen kann, egal ob digital oder analog. Zusätzlich werden die Dimensionen im paarweisen Vergleich gegeneinander aufgewogen (zum Beispiel „Ausdruckskraft“ gegen „Spaß“: „Bei dieser Aufgabe ist mir besonders wichtig...“ [„dass ich in der Lage bin, während der Aktivität sehr ausdrucksstark und kreativ zu sein“; „dass ich Spaß daran habe das Tool zu benutzen“]). Dadurch erfolgt eine Gewichtung, die bei der Auswertung berücksichtigt wird.

Der CSI kann auch als Vergleichsinstrument benutzt werden. So gab es in einer Studie von Sano und Yamada (2022) zum Beispiel eine Gruppe, die ein in der Studie untersuchtes KI-Tool benutzte und eine Kontrollgruppe, die ein beliebiges Tool ihrer Wahl (oder kein Tool) benutzten, um die Ästhetik von Designkonzepten zu benennen. Beide Gruppen beantworteten im Anschluss den CSI-Fragebogen. Die CSI-Ergebnisse zeigten, dass die Testpersonen mit KI-Tool viel mehr das Gefühl hatten, während der Task kreativ unterstützt zu werden, besonders was die Dimension Erkundung betraf.

Tchemeube u. a. benutzten für ihre User Study in der ersten MMM-Iteration eine stark vereinfachte Version des CSI. Eine Gewichtung der Dimensionen durch paarweisen Vergleich wurde nicht vorgenommen. Eine weitere Änderung war die Reduzierung von zwei

Zustimmungsstatements pro Bewertungsdimension auf nur eines. Dadurch wird auf eine höhere statistische Aussagekraft über einen Reliabilitätstest zwischen zwei Variablen zur gleichen Dimension verzichtet. Von den sechs Bewertungsdimensionen wurde außerdem von Tchemeube u. a. die Dimension „Kollaboration“ weggelassen, weil bei der Nutzung von MMM im Cubase Plugin keine Kollaboration vorgesehen ist (Tchemeube u. a. 2023: 5772). In anderen Studien gab es die gleiche Situation. Sano und Yamada (2022) haben sich dafür entschieden, die Dimension zu behalten aber die Option „Not applicable“ („nicht zutreffend“) hinzuzufügen. Cabral und Correia (2017) entschieden sich dafür die Statements zur Kollaboration wegzulassen.

Tabelle 5: CSI-Variablen

Skala	Variable	Item
CSI <i>Likert-Skala (5)</i>	CSI-01	Erkundung: Es war einfach für mich, viele verschiedene Ideen, Optionen, Entwürfe oder Ergebnisse mit dem Plugin zu erkunden
	CSI-02	Vereinnahmung: Ich war sehr eingenommen von dieser Aktivität - ich habe es genossen und würde es wieder tun.
	CSI-03	Ausdruckskraft: Ich war in der Lage, während der Aktivität sehr ausdrucksstark und kreativ zu sein.
	CSI-04	Spaß: Ich hatte Spaß daran das Plugin zu benutzen.
	CSI-05	Ergebnisse lohnen Aufwand: Was ich erreichen konnte, war die Mühe wert, die ich dafür aufbringen musste

Bei Cherry und Latulipe (2014) wird eine zehnstufige Likert-Skala benutzt und am Ende mit Hilfe einer CSI-Applikation aus allen Daten ein Gesamtscore zwischen 0 und 100 berechnet. Bei Tchemeube u. a. (2023) und in dieser Untersuchung werden die Ergebnisse für die Dimensionen im Einzelnen betrachtet und nicht hochskaliert. Die TN bewerten die fünf Aussagen auf einer fünfstufigen Likert-Skala. Für die Auswertung wird pro Bewertungsdimension ein Mittelwert für den Score zwischen eins und fünf berechnet.

In Anlehnung an die Score-Bewertung von Cherry und Latulipe (2014:8), wird für die Auswertung festgelegt, dass ein Score ab 4,5 sehr gut ist, das Tool also in der Task für diese Dimension des kreativen Arbeitens außerordentlich hilfreich war. Ein Score von 3,7 ist ein zufriedenstellender Mittelwert. Ein Score unter 2,5 zeigt an, dass das Tool in dieser Bewertungsdimension nicht unterstützend ist.

Hohe CSI-Scores würden darauf hinweisen, dass MMM auf effektive und effiziente Weise den kreativen Prozess der Musikproduktion bei unerfahrenen Komponist:innen fördert und dadurch zu ihrer Zufriedenheit mit dem Tool beiträgt.

Controllability

Im Kontext eines co-kreativen KI-Tools bezieht sich Controllability auf die Möglichkeit, die Handlungen und Entscheidungen einer KI zu steuern oder zu beeinflussen. Hier wird auch von Steerability gesprochen: die Fähigkeit der Benutzer:innen, ihrem KI-Tool ihre künstlerischen Absichten mitzuteilen, sodass dessen algorithmisches Verhalten sich danach ausrichtet (Chung 2023: 58). Dies sei gerade bei KI-Anwendungen wichtig, weil die Modelle dahinter für Anwender:innen so unvorhersehbar seien.

Für eine Einschätzung der Controllability von MMM gab es in dieser Untersuchung zwei von Tchemeube u. a. verfasste Zustimmungsaussagen und ein freies Kommentarfeld (s. qualitatives Feedback, 3.4.4).

Tabelle 6: Controllability-Variablen

Skala	Variable	Item
Controllability	CON-01	Ich empfinde das Plugin als leicht kontrollierbar.
Skala 1: überhaupt nicht bis 10: voll und ganz	CON-02	Ich wünsche mir mehr Kontrolle über den Output des Plugins.

Die TN konnten auf einer Skala von eins („überhaupt nicht“) bis zehn („voll und ganz“) zustimmen. Bei CON-01 bedeutet ein hoher Score eine positive Bewertung der Controllability des Plugins. Bei CON-02 bedeutet ein hoher Score, dass das Bedürfnis nach Kontrolle nicht befriedigt wurde. Für die Bewertung von CON-01 wurde hier festgelegt, dass ein Score unter fünf für ungenügende und ein Score größer gleich acht für sehr gute Steuerbarkeit spricht. Für CON-02 gilt das entsprechend andersherum.

Eine positive Controllabilitybewertung spräche dafür, dass die musikalischen Ideen der unerfahrenen Komponist:innen von MMM auf eine effektive, effiziente Weise in die Musikgenerierung eingebunden werden, wodurch ihre Zufriedenheit mit dem System gesteigert wird. Eine negative Bewertung spräche dafür, dass es entweder zu wenig Einflussmöglichkeiten für sie auf den Output gibt oder das System technische Probleme mit der Umsetzung ihrer Eingaben hat. Das qualitative Feedback soll darüber weiteren Aufschluss geben.

Technology Acceptance Model

Das Technology Acceptance Model (TAM) ist ein psychometrisches Instrument zur Bewertung der Akzeptanz eines Systems mit Mensch-Maschine-Schnittstelle. Nach Kohnke (2015) ist es „das bisher am weitesten verbreitete und empirisch sehr gut überprüfte Akzeptanzmodell für

(neue) Informationssysteme". Das Messinstrument wurde von Davis (1989) konzipiert und später weiterentwickelt (Davis und Venkatesh 1996). Seine Hypothese war, dass die beiden von externen Anreizen beeinflussten Variablen für die „Wahrgenommene Nützlichkeit“ des Systems und für die „Wahrgenommene Leichtigkeit der Systemnutzung“ die Verhaltensabsicht der Benutzer:innen und somit ihre „Tatsächliche Systemnutzung“ bestimmen (s. Abbildung 12). Er entwickelte durch Pretesting Skalen-Items für beide und überprüfte diese auf Reliabilität und Konstruktvalidität. Er wies Korrelationen zwischen der Wahrgenommenen Nützlichkeit des Systems und den Angaben der TN über ihr aktuelles Nutzungsverhalten und ihre Vorhersage über eine zukünftige Nutzung nach. Auch für die Wahrgenommene Leichtigkeit der Systemnutzung wurde diese Korrelation festgestellt, wenn auch schwächer. Davis vermerkt mit Hilfe von Regressionsanalysen⁷, dass die Leichtigkeit eine kausale Vorbedingung für die Nützlichkeit sein könnte.

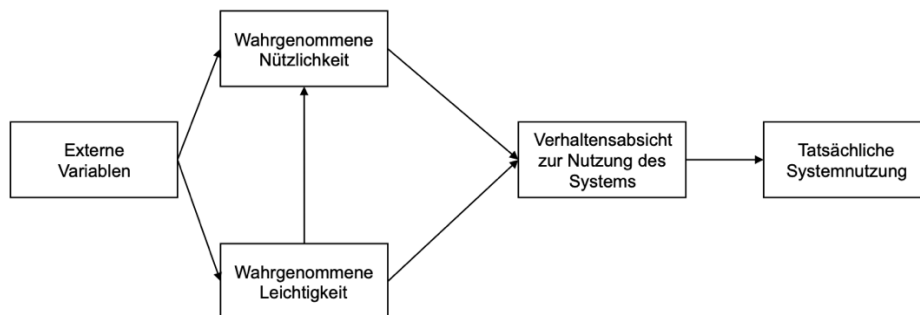


Abbildung 12: TAM nach Davis
Quelle: Davis (1996: 20)

TAM lässt sich durch seine Einfachheit und Flexibilität sehr gut erweitern, anpassen oder kombinieren. So nutzte etwa Kohnke (2015) das Modell von Davis für die Entwicklung eines Modells zur Anwender:innenakzeptanz unternehmensweiter Software. Ein Anwendungsfall, bei dem die Benutzer:innen nicht freiwillig eine neue Technologie ausprobieren, wie es die Ausgangslage bei Davis ist, sondern durch ihre Arbeit dazu gezwungen sind. Ein Unternehmen könnte mit einem solchen Modell testen, ob sich die Investition in eine neue Standardsoftware lohnt. Jiang u. a. (2016) kombinierten in einer Task-basierten Studie TAM mit einem anschließenden semi-strukturierten Interview, das in Bezug auf die Konzepte des TAM-Rahmens codiert und analysiert wurde. Ambareta u. a. (2022) fügten dem Modell in einer Studie zwei weitere Variablen „Trust“ und „Perceived Risk“ als Vorbedingung für die Wahrgenommenen

⁷ Regressionsanalysen sind eine „Gruppe statistischer Verfahren, mit denen die stochastische Abhängigkeit einer Kriteriumsvariablen von einer oder mehreren unabhängigen Variablen bestimmt werden kann“ (Spektrum Lexikon der Psychologie 2000b).

Nützlichkeit hinzu und passten die Items zu den einzelnen Variablen nach ihren Bedürfnissen an.

Bei Tchemeube u. a. (2023) und in dieser Erhebung wurden die jeweils sechs Zustimmungstatements für die Wahrgenommene Nützlichkeit und Leichtigkeit der Systemnutzung mit einer fünfstufigen Likertskala verwendet. Zusätzlich gab es ein Item zur Wahrscheinlichkeit der zukünftigen Nutzung des Plugins, mit einer semantischen Differentialskala von eins („unwahrscheinlich“) bis zehn („wahrscheinlich“). Da das Plugin noch nicht veröffentlicht wurde, konnten durch die TN keine Aussage zur bisherigen Nutzung gemacht werden.

Tabelle 7: TAM-Variablen

Skala	Variable	Item
Wahrgenommene Nützlichkeit des Systems <i>Likert-Skala (5)</i>	TAM-01	Der Einsatz des Plugins ermöglicht es mir, Kompositionen schneller zu erstellen.
	TAM-02	Der Einsatz des Plugins verbessert meine Arbeitsleistung beim Komponieren.
	TAM-03	Der Einsatz des Plugins verbessert meine Produktivität beim Komponieren.
	TAM-04	Der Einsatz des Plugins verbessert meine Effektivität beim Komponieren.
	TAM-05	Der Einsatz des Plugins macht es mir leichter Kompositionen anzufertigen.
	TAM-06	Ich finde den Einsatz des Plugins nützlich beim Komponieren.
Wahrgenommene Leichtigkeit der Systemnutzung <i>Likert-Skala (5)</i>	TAM-07	Es ist leicht für mich, den Umgang mit dem Plugin zu lernen.
	TAM-08	Es ist leicht für mich, das Plugin das machen zu lassen, was ich möchte.
	TAM-09	Ich empfinde die Interaktion mit dem Plugin als klar und verständlich.
	TAM-10	Ich empfinde die Interaktion mit dem Plugin als flexibel.
	TAM-11	Es ist leicht für mich, gute Kompetenzen im Umgang mit dem Plugin zu entwickeln.
	TAM-12	Ich schätze das Plugin als einfach zu nutzen ein.
Zukünftige Nutzung <i>Skala 1: unwahrscheinlich bis 10: wahrscheinlich</i>	TAM-13	Wie wahrscheinlich wäre es, dass Sie das Plugin in Zukunft benutzen?

Für die Bewertung der Scores für die Wahrgenommene Leichtigkeit und Nützlichkeit wird festgelegt, dass bei einem Ergebnis unter 3,5 dringender Handlungsbedarf besteht.

Eine positive TAM-Bewertung würde darauf hinweisen, dass unerfahrene Komponist:innen eine positive Haltung gegenüber MMM haben und dazu geneigt sind, das System zu akzeptieren und in Zukunft zum Komponieren zu verwenden.

After Scenario Questionnaire (ASQ)

Der After Scenario Questionnaire wurde von Lewis (1991) für Szenario-basierte Usability Studies entwickelt, um die Benutzer:innenzufriedenheit während der Teilnahme zu bewerten. Die psychometrische Befragung besteht aus drei Zustimmungsaussagen. Im Original ASQ werden sie auf einer siebenstufigen Skala von eins („strongly agree“) bis sieben („strongly disagree“) beantwortet. Es gibt darin auch die Möglichkeit, „not applicable“ (nicht zutreffend) anzukreuzen und zu jedem Statement eine Kommentaroption.

In dieser Umfrage wird eine fünfstufige Likertskala verwendet, um die Kontinuität mit den in den anderen Fragebögen verwendeten Skalen zu gewährleisten. Die Option "nicht zutreffend" wird nicht angeboten, da jede der drei Fragen auf alle Teilnehmenden zutrifft.

Tabelle 8: ASQ-Variablen

Skala	Variable	Item
ASQ <i>Likert-Skala (5)</i>	ASQ-01	Insgesamt bin ich zufrieden mit dem Schwierigkeitsgrad der Task.
	ASQ-02	Insgesamt bin ich zufrieden mit der Zeit, die ich zum Erfüllen der Task benötigt habe.
	ASQ-03	Insgesamt bin ich zufrieden mit den unterstützenden Informationen zum Erledigen der Task.

Die Items werden für die Auswertung zu einem Wert zusammengefasst. Es entsteht ein Score zwischen eins und fünf, der aussagt, wie zufrieden die TN mit dem Testszenario waren. Für die Bewertung des Scores wird festgelegt, dass ein Ergebnis unter 3,5 eine Überarbeitung der Task erfordert und ein Score über vier für eine angemessene Task steht.

3.4.4 Qualitatives Feedback (Kommentare)

Die TN bekamen zusätzlich zu den quantitativen Fragen auch die Möglichkeit, ihre Meinungen und Erfahrungen qualitativ mitzuteilen. Dafür gab es im Fragebogen in den Abschnitten Controllability und Akzeptanz die Möglichkeit, einen Kommentar zu schreiben.

Tabelle 9: Prompts für qualitatives Feedback

Skala	Variable	Item
Controllability	CON-03	Kommentar zur Controllability des Systems:
TAM	TAM-14	Geben Sie bei Bedarf weiteres Feedback:

Die Kommentare wurden für die Auswertung durch induktive Kategorienbildung geordnet und zusammengefasst.

3.4.5 Fragebögen und Likert-Skala

Es wurden zwei Fragebögen in Microsoft Forms erstellt. Der erste Fragebogen wurde nach der Begrüßung ausgefüllt. Im ersten Abschnitt wurden Fragen zu Demographie und Erfahrung der TN gestellt. Im zweiten Abschnitt wurde die Kurzformulierung der Task mit Hinweis auf das Tutorialvideo angezeigt, das sich die TN direkt im Anschluss selbstständig anschauen sollten. Die Bearbeitungszeit für den ersten Fragebogen betrug maximal fünf Minuten.

Der zweite Fragebogen wurde im Anschluss zur Task ausgefüllt und bestand aus vier Abschnitten: Usability (SUS und Usability insgesamt), Kreativität (CSI und Controllability), Akzeptanz (TAM) und Zufriedenheit (ASQ). Für den Fragebogen wurden zehn Minuten Bearbeitungszeit angesetzt.

Für die quantitativen Fragebögen wurde als Instrument die Likert-Bewertungsskala ausgewählt. Nur die Fragen zur Einschätzung der Usability insgesamt, Controllability und zukünftigen Nutzung haben eine andere Bewertungsskala bekommen. Auf einer Likert-Skala hat die befragte Person die Möglichkeit, eine Frage aus einer bestimmten Anzahl von Antwortkategorien durch Ankreuzen zu beantworten. Alternativ kann auch eine Aussage am Anfang stehen, zu der die befragte Person ihre Zustimmung bewerten soll. Meist bleiben die Antwortkategorien für mehrere Fragen/Aussagen (Items) dieselben, was eine übersichtliche Darstellung als Itemmatrix ermöglicht (vgl.: Menold und Bogner 2015: 8). Für die Usability- und Akzeptanzbefragungen der von Tchemeube u. a. (2023) durchgeführten Studie, aus der die Fragebögen für diese Befragung übernommen wurden, wurde eine fünfstufige, bipolare Skala gewählt, das heißt die Antwortmöglichkeiten bilden zwei gegensätzliche Pole ab. Das entspricht der gängigen Praxis die Anzahl der Antwortkategorien zwischen fünf und sieben zu wählen (vgl.: Menold und Bogner 2015: 2). Es gibt eine Mittelkategorie und die Stufen wurden symmetrisch, ausbalanciert, allgemein verständlich und universell beschriftet (vgl.: ebd. 3): [„Stimme voll und ganz zu“, „Stimme zu“, „Weder noch“, „Stimme nicht zu“, „Stimme überhaupt nicht zu“].

Gegen ein Design mit Mittelkategorie hätte argumentiert werden können, dass die Bereitschaft zur mentalen Anstrengung bei Befragten einer Studie mit der Zeit nachlässt. Sie neigen dann dazu, einen einfacheren Weg zu nehmen und wählen beispielsweise die Antwort, die ihnen erwünscht oder angemessen erscheint („Satisficing“, Krosnick 1991:214-15). Bei einer bipolaren Likert-Skala könnte in dem Fall der einfachste Weg sein, die Mittelkategorie anzukreuzen oder sie als Äquivalent für „Weiß nicht“ zu benutzen (vgl.: Sturgis u. a. 2014: 20). Andererseits könnten aber auch Befragte, die wirklich eine neutrale Position haben, durch die fehlende Mittelkategorie gezwungen werden, ihre Antwort in die eine oder andere Richtung zu

lenken, obwohl dies nicht ihrer wahren Einstellung entspricht (vgl.: ebd. 33). Es ist durchaus möglich, dass die Befragten nach dem Ausprobieren des ihnen völlig neuen Plugins zu bestimmten Fragen eine neutrale Meinung haben. Es ist ihnen zuzumuten, dass sie für die kurze Dauer der Befragung konzentriert bleiben und authentisch antworten. Durch die einheitliche Gestaltung der Fragebögen zu SUS, CSI, TAM und ASQ und durch übersichtliche fünf Antwortkategorien, ist es ihnen möglich, sich schnell und leicht durch die Befragung durchzuarbeiten.

3.4.6 Übersetzungen

Die Fragebögen der Studie von Tchemeube u.a. lagen auf Englisch vor. Für diese Erhebung sollten sie ins Deutsche übersetzt werden. Dafür wurde sich an dem Leitfaden für Fragebogenübersetzungen von Behr (2023) orientiert.

Die zwei wichtigsten Punkte beim Übersetzen sind erstens die Bedeutung und die Messeigenschaften beizubehalten und zweitens sowohl die allgemeinen Anforderungen an Fragebogendesigns als auch die grammatikalischen, idiomatischen und pragmatischen Anforderungen der Zielkultur zu berücksichtigen. Nicht immer ist es möglich, alle Richtlinien einzuhalten, dann ist eine Abwägung nötig.

Die Art der Gestaltung in Microsoft Forms wurde aus den Originalfragebögen beibehalten, also die Reihenfolge der Fragen in den übernommenen Abschnitten und Textformatierungen, etwa fettgedruckte Wörter.

Beim Übersetzen ist es wichtig, auf Konsistenz zu achten. Beispielsweise wurde in der deutschen Fassung durchgehend „das Plugin“ für „system“ oder „tool“ benutzt. Die TN kannten diese Bezeichnung schon aus der Einladung und der Begrüßung zur Studie.

Oft mussten grammatikalische, idiomatische und pragmatische Anpassungen durchgeführt werden. So wäre beispielsweise für SUS-01 „I think that I would like to use this system frequently“ die direkte Übersetzung „Ich denke, dass ich das Plugin gerne häufig nutzen würde“ sehr fremd zur gängigen Ausdrucksweise und dazu schwer zu verstehen. Stattdessen wurde das Item übersetzt mit „Ich kann mir gut vorstellen, das Plugin regelmäßig zu nutzen“. Für SUS-09 wurde für „I felt very confident using the system“ ebenfalls eine sinngemäße statt direkter Übersetzung gewählt: „Ich habe mich sehr sicher im Umgang mit dem Plugin gefühlt“. Für die Bewertungskategorie „Enjoyment“ in CSI-04 wurde statt „Vergnügen“ oder „Genuss“ das Wort „Spaß“ als passend erachtet. Für ASQ-01 wurde mit der Übersetzung von „Overall, I am satisfied with the ease of completing the tasks in this scenario“ zu „Insgesamt bin ich zufrieden mit dem

Schwierigkeitsgrad der Task“ der positiv konnotierte Begriff „ease“ in den eher negativ konnotierten Begriff „Schwierigkeitsgrad“ umgewandelt, weil der deutsche Begriff „Einfachheit“ oder „Leichtigkeit“ für „ease“ hier im Satzzusammenhang umständlich formuliert klang. Es musste abgewogen werden. Für SUS-08 „I found the system very awkward to use“ wurde eine passende Übersetzung gesucht, weil „awkward“ mit „peinlich“ oder „unangenehm“ nicht sinngemäß schien. Hier half ein Blick in die Originalquelle für die System Usability Scale von Brooke (1996), in der „cumbersome“, statt „awkward“ benutzt wurde. Daraus folgte die Übersetzung zu „Ich empfand die Bedienung des Plugins als sehr umständlich“.

Sehr selten gab es bereits in deutscher Literatur eine Übersetzung, die übernommen werden konnte. So übersetzt Kohnke (2015) für das Technology Acceptance Model die Begriffe „Perceived Usefulness“ zu „Wahrgenommene Nützlichkeit des Systems“ und „Perceived Ease of Use“ zu „Wahrgenommene Leichtigkeit der Systemnutzung“.

Es wurde außerdem auf Adäquatheit der Sprache für die Zielgruppe und die Einhaltung kultureller Sprachkonventionen geachtet. Zum Beispiel darauf, auch im Deutschen genderneutrale Sprache zu verwenden, damit sich alle TN angesprochen fühlen. So wurde für das Item zur Usability insgesamt der Begriff „user-friendliness“ übersetzt mit „Bedienfreundlichkeit“, eine Formulierung, die sich auf das Verb „bedienen“ bezieht und nicht den „Benutzer“ oder die „Benutzerin“. Es wurde entschieden die TN im Fragebogen zu Siezen und nicht zu Duzen, was im Englischen nicht unterschieden wird. Dadurch sollte Professionalität und Höflichkeit zum Ausdruck gebracht werden. Außerdem wurde in TAM-03 der Zusatz „in my job“ einfach weggelassen, vor dem Hintergrund, dass die Zielgruppe der zweiten Studie keine professionellen Komponist:innen oder Produzent:innen waren.

Für die Beschriftung der fünfstufigen Likert-Skala war darauf zu achten, dass die Zustimmungsggrade und die Abstände dazwischen ihre Wertigkeit und Symmetrie beibehalten, die Polarität (in diesem Fall bipolar) nicht verändert wird und die Antwortkategorien getrennt blieben. Es musste eine Übersetzung für Adverbien gefunden werden, die im Deutschen idiomatisch passend schien. So wurde aus „Strongly agree“ „Stimme voll und ganz zu“ und aus „Strongly disagree“ „Stimme überhaupt nicht zu“. Für die Items mit anderen Skalen waren jeweils die Extrempunkte beschriftet, wofür eine direkte Übersetzung möglich war (zum Beispiel „improbable“/„probable“ zu „unwahrscheinlich“/„wahrscheinlich“).

3.5 Stichprobenverfahren: Beschreibung der User Persona

Durch die Studie sollte ein bestimmter Teil der Benutzer:innen repräsentiert werden. Schon die Einladung zur Teilnahme sollte speziell diese Zielgruppe ansprechen. Die genaue Kenntnis der Anwender:innen ist das oberste Gebot und der Grundstein von Usability Engineering (vgl.: Nielsen 1993: 73). Bei der Einteilung der Benutzer:innen in Profile wird von User Personas gesprochen. Nach Rosenzweig (2015) gehören zur Beschreibung einer User Persona demographische Angaben (Alter, Einkommen und Bildung), Ziele, Motivationen und Frustrationen der Gruppe sowie ihre typischen Verhaltensweisen, Aufgaben, und Einschränkungen (vgl.: Rosenzweig 2015: 47).

In dieser Studie soll es um unerfahrene Komponist:innen gehen, die im Zuge ihres Audio- oder Medien-bezogenen Studiums oder einer Ausbildung angefangen haben, sich mit DAWs zu beschäftigen, aber in diesem Bereich noch nicht im Berufsleben stehen. Diese Personen sind durch ihren Bildungskontext und ihre Peers über die Entwicklungen im Bereich KI informiert und haben Interesse daran, neue Tools auszuprobieren. Sie sind technikaffin und haben bereits erste Erfahrungen im Umgang mit DAWs gesammelt. Entweder durch Projekte im Zuge ihrer Ausbildung, durch erste Jobs oder aus eigenem Interesse in ihrer Freizeit. Die Beschäftigung mit DAWs ist eines von vielen Interessen und Projekten, für die sie ihre Zeit aufteilen müssen. Diejenigen mit mehr DAW-Erfahrung haben sich nicht primär mit Musikkomposition, sondern mit anderen Nutzungsmöglichkeiten beschäftigt, zum Beispiel mit Podcastproduktion, Tonschnitt, Recording oder Mastering.

Die Benutzer:innen haben entweder schon einmal eine eigene Komposition in einer DAW angefertigt oder sie wünschen sich, den Einstieg zu finden. Sie sind weit davon entfernt Kompositionsprofis zu sein. Es fällt ihnen schwer, etwas aus dem Nichts entstehen zu lassen, und sie durchschauen nicht alle künstlerischen und technischen Möglichkeiten. Umfangreiche DAWs überfordern sie dabei eher. Sie würden gerne schnell(er) Ergebnisse erzielen und sich einen neuen Skill aneignen. Sie haben keine musiktheoretischen Kenntnisse und kennen somit nicht alle Grundbegriffe. Ihnen ist dadurch gegebenenfalls nicht unmittelbar verständlich, was die jeweilige Parametereinstellung im MGP bewirkt. Sie finden es durch wiederholtes Ausprobieren und Vergleichen heraus.

Aufgrund der Empfehlungen von Rosenzweig (2015: 135) und in Orientierung an ähnlichen User Studies (Gustafsson 2019; Malandro 2023; Tchemeube u. a. 2023) wurde eine TN-Zahl von 18 bis 25 Personen angestrebt.

3.6 Pretest

Die ersten fünf Versuchsdurchführungen wurden als Pretest angesehen. Sie dienten dazu, das Erhebungsinstrument und die Task zu testen, aber auch den Ablauf und die Funktionalität der Technik. Während dieser Durchläufe wurde beispielsweise die Checkliste für die Durchführung ergänzt. Es wurde klar, dass der Arbeitsauftrag an die TN nicht deutlich genug formuliert war. Sie trafen dadurch oft auf einen bestimmten Bug und waren deswegen in der Task frustriert. Das Testszenario wurde daraufhin angepasst. Die Bearbeitungszeiten für die Fragebögen konnten nach dem Pretest realistischer eingeschätzt werden und es wurde für die Task festgestellt, dass 30 Minuten eine angebrachte Zeitspanne sind. Nachdem die Formulierung des Arbeitsauftrags überarbeitet wurde, war sie so formuliert, dass für die TN klar ersichtlich war, was das Ziel der Task ist und was von ihnen erwartet wird.

4 Durchführung und Auswertung der User Study

Ein Blick zurück auf die fünf Hauptphasen für empirische Studien (Abbildung 6) zeigt, dass die nächsten Schritte des Forschungsprozesses die Datenerhebung (Phase IV) und -auswertung (Phase V) sind. Im folgenden Kapitel wird das Vorgehen während dieser Phasen dokumentiert. Als Teil der Berichterstattung in Phase VI werden dann die Ergebnisse der Erhebung präsentiert und diskutiert.

Zunächst zur Datenerhebung: In Abschnitt 4.1 wird erläutert, wie das Onboarding der TN umgesetzt wurde und wie eine Versuchsdurchführung ablief. Der genaue Versuchsaufbau wird in Abschnitt 4.2 beschrieben.

Für die Datenauswertung wurden Berechnungen mit den quantitativen Daten durchgeführt. Die Vorgehensweise wird in Abschnitt 4.3 erläutert. In den Abschnitten 4.4 und 4.5 werden die Ergebnisse präsentiert und diskutiert.

4.1 Onboarding und Versuchsablauf

Die Einladung zur Studienteilnahme wurde auf dem Mediacampus der Hochschule für Angewandte Wissenschaften Hamburg händisch verteilt und zusätzlich über die Lernplattform Moodle per Mail an alle Studierenden der Fakultät Design, Medien und Information der HAW geschickt. Dadurch wurde direkt ein großer Pool von Menschen angesprochen, die mit großer Wahrscheinlichkeit der gesuchten User Persona entsprachen. Es wurde bei der Rekrutierung auch auf Genderdiversität geachtet. In der Einladung wurde den TN als Dankeschön eine Steinberg Software Lizenz für Cubase Elements, Dorico Elements oder Wavelab Elements

angeboten. Ein solcher Anreiz (Incentive) sollte die Zahl der Anmeldungen erhöhen, wie es beispielsweise für E-Mail-Umfragen (Church 1993), postalische Erhebungen (Edwards u. a. 2002) und Web Studies (Göritz 2006) nachgewiesen wurde. Interessierte konnten sich online im Buchungskalender mit Namen und E-Mail-Adresse für einen Termin eintragen und bekamen dann eine automatische Terminbestätigung zugeschickt. Sie konnten ihren Termin bei Bedarf selbstständig im Buchungssystem stornieren. Der Zeitraum für die Studie war der 13. bis 22. November 2023 (s. Abbildung 6). Davon waren zwei Tage für den Pretest vorgesehen und sechs für den Versuch. Jeden Tag wurden ein bis vier Zeitslots angeboten. Jeder Zeitslot wurde mit 90 Minuten angesetzt (75 Minuten pro Durchlauf, 15 Minuten Zeitpuffer) und konnte von zwei TN gleichzeitig gebucht werden. Der Ort der Durchführung waren zwei Tonregionen (R2 und R3) des Tonlabors der HAW Hamburg in der Finkenau 35.

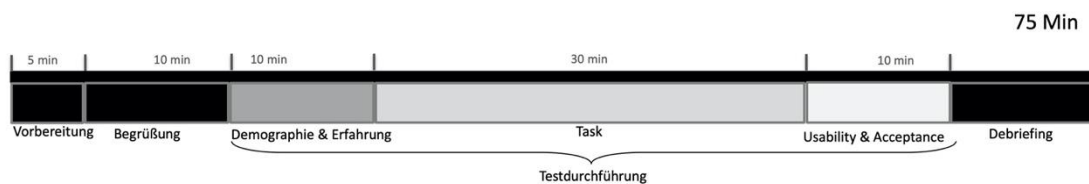


Abbildung 13: Ablauf einer Durchführung

Quelle: Eigene Darstellung, Abschnitte in Anlehnung an Nielsen (1993: 187ff.)

Abbildung 13 bildet die Abschnitte eines Versuchsdurchlaufs ab. Die Einteilung lehnt an die „Stages of a Test“ von Nielsen (1993: 187ff.). Die Empfehlungen daraus wurden berücksichtigt. Für die Versuchsdurchführung wurde eine Checkliste angefertigt, damit der Ablauf immer derselbe ist und kein Schritt vergessen wird.

Vorbereitung

In der Vorbereitung wurde ein Techniktest durchgeführt. Zuerst wurde getestet, ob Maus, Tastatur, Kopfhörer und Bildschirm richtig angeschlossen sind. Die Internetverbindung über LAN wurde überprüft, WLAN und Bluetooth wurden ausgestellt. Im Browser, in dem später die Fragebögen in Microsoft Forms bearbeitet werden würden, wurde die Chronik gelöscht. Es wurde sichergestellt, dass alle unwichtigen Applikationen geschlossen, Benachrichtigungen ausgestellt und Konten ausgeloggt sind. Das Tutorialvideo war bereits geöffnet. Für jede teilnehmende Person wurde ein Ordner nach dem Benennungsschema YYYY-MM-DD-hh-Nr vorbereitet. Im Steinberg Activation Manager wurde überprüft, ob die Lizenzen für Cubase 13 und das MGP aktiv sind. In Cubase wurde ein Soundcheck über die Kopfhörer gemacht und sichergestellt, dass das MGP in den Default-Einstellungen ist.

Begrüßung

Die Versuchsteilnehmer:innen kamen selbstständig zum Tonlabor. Dort wurden sie in Empfang genommen und zu den Regien R2 und R3 geführt. Zunächst wurden ihnen der Kontext und das Ziel der Studie erklärt. Dabei wurden die Beteiligten Steinberg und Simon Fraser University genannt. Dann wurden ihnen der Versuchsablauf und das Szenario erklärt. Die TN wurden gebeten ein Non-Disclosure-Agreement (NDA) zu unterschreiben. Es wurde notiert, welche Software Lizenz von Steinberg sie als Dankeschön für ihre Teilnahme erhalten wollen und ob sie im Falle einer weiteren Studie kontaktiert werden dürfen.

Testdurchführung

Die TN wurden nach der Begrüßung angewiesen, den ersten Fragebogen zu Demographie und Erfahrung auszufüllen und sich danach das bereits geöffnete Tutorialvideo anzuschauen. Währenddessen verließ die Versuchsleitung den Raum. Zur Einleitung der Task sollten sich die TN eine Musikrichtung aussuchen, und die Versuchsleitung legte daraufhin das entsprechende Projekt im TN-Ordner ab. Den TN wurde noch gezeigt, wie sie eine Zwischenversion abspeichern können und dann begannen ihre 30 Minuten Bearbeitungszeit. Die Versuchsleitung kam zum Ende der Zeit wieder in den Raum, um den zweiten Fragebogen zu öffnen.

Debriefing

Nach Ausfüllen des letzten Fragebogens fand ein kurzes Debriefing statt, in dem die TN meistens Redebedarf hatten und von ihrem Erlebnis berichteten. Sie wurden anschließend verabschiedet und ihr Projekt gespeichert. Es wurde von jeder Zwischenversion ein Stereo-Mixdown erstellt und ein Backup des Cubase Projekts gemacht.

4.2 Versuchsaufbau

Für den Versuch wurden von Steinberg zwei Mac Book Pro (2020) mit MacOS 10.15.7 bereitgestellt, auf denen das Music Generation Panel 1.5.0.188 in Cubase 13.0.10.123 installiert war. In R2 gab es zwei zusätzliche Bildschirme, in R3 nur einen. Außerdem gab es in jedem Setup eine externe Maus und Tastatur. Die TN arbeiteten mit Kopfhörern. Fotos vom Setup sind in Abbildung 16 zu sehen. Übersichten des Setups und der Räume werden in Abbildung 14 und Abbildung 15 gezeigt.

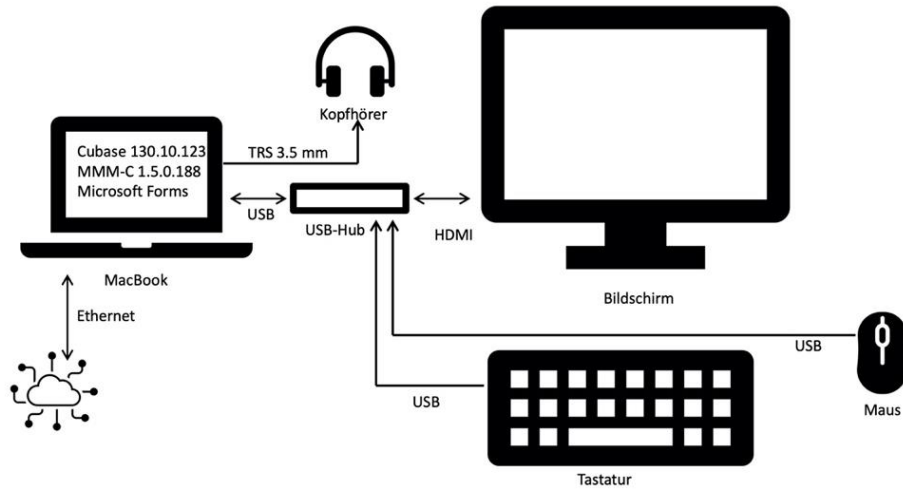


Abbildung 14: Übersicht: Technisches Setup
Quelle: Eigene Darstellung

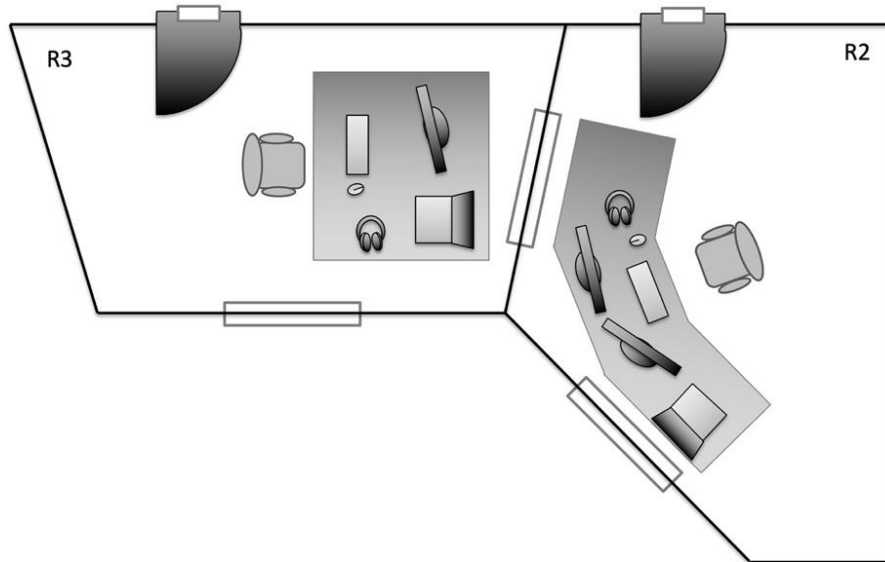


Abbildung 15: Raumplan R2 und R3
Quelle: Eigene Darstellung



Abbildung 16: Fotos des Setups in R2 (links) und R3 (rechts)
Quelle: Eigene Darstellung

4.3 Datenauswertung

Zur Auswertung der Ergebnisse wurden Quartile für die jeweiligen Datensätze errechnet und Boxplotdiagramme erstellt. Für alle Boxplotdiagramme gilt: Die untere Grenze der Box markiert das erste Quartil ($Q_{0,25}$), das heißt 25 Prozent der Daten liegen unter dieser Grenze. Die obere Grenze markiert das dritte Quartil ($Q_{0,75}$), das heißt 75 Prozent der Daten liegen darunter. Der Bereich zwischen $Q_{0,25}$ und $Q_{0,75}$ spannt den Interquartilsabstand (IQA) auf und wird als farbige Box dargestellt. Die eingezogene Linie innerhalb der Box zeigt das zweite Quartil an ($Q_{0,5}$, auch: Median), das heißt 50 Prozent der Daten liegen darunter. Die Quartile werden inklusive Median berechnet.

$$IQA = Q_{0,75} - Q_{0,25} \quad (4)$$

Für die Berechnung der Quartile ist es wichtig, dass der Datensatz zunächst aufsteigend geordnet wird. Es entsteht eine sortierte Zahlenreihe S mit Elementen x_i . Nun wird für die drei Quartile jeweils eine Stelle x_p berechnet. Wenn der Index p genau einem der Indizes i entspricht, dann ist Q_j einfach das entsprechende Element x_i . Falls p zwischen zwei Indizes i und $i+1$ liegt, dann ist Q_j der Mittelwert der beiden Elemente.

$$S = \{x_{i=1}, x_{i=2}, \dots, x_n\} \quad (5)$$

$$Q_j = x_p = x_{(j(n-1)+1)} \quad \text{für } j = \{0,25, 0,5, 0,75\} \quad (6)$$

Das Kreuz innerhalb der Box im Boxplotdiagramm markiert das arithmetische Mittel \bar{x} der Daten, auch genannt Mittel- oder Durchschnittswert.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

Die Antennen an der Box zeigen den Minimal- und Maximalwert an. Extreme Ausreißer (Werte über respektive unter dem Ausreißerlimit) jedoch werden außerhalb der Antennen mit einem Kreis dargestellt.

$$\text{Oberes Ausreißerlimit: } Q_3 + 1,5 * IQA \quad (8)$$

$$\text{Unteres Ausreißerlimit: } Q_1 - 1,5 * IQA \quad (9)$$

Die Standardabweichung (SD) bezeichnet die Streubreite um den Mittelwert \bar{x} . Dafür wird berechnet, wie weit die Werte durchschnittlich vom Mittelwert abweichen.

$$SD = \sqrt{\frac{\sum_{i=1}^n x_i - \bar{x}}{n}} \quad (10)$$

Zusätzlich wird das prozentuale Verhältnis von SD zu \bar{x} als Variationskoeffizient (CV) mit angegeben.

$$CV = \frac{SD}{\bar{x}} * 100\% \quad (11)$$

4.4 Ergebnisse

An der Studie nahmen 19 Personen teil. Alle TN haben die Task und die Befragung vollständig abgeschlossen. Im Folgenden werden die Ergebnisse präsentiert.

4.4.1 Teilnehmende

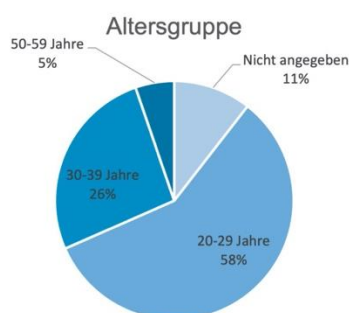


Abbildung 17: Altersgruppen der Teilnehmenden
Quelle: Eigene Darstellung

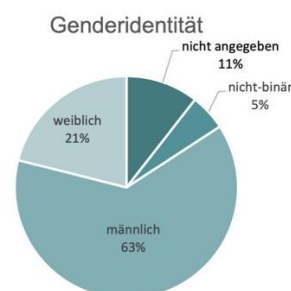


Abbildung 18: Gender-Identität der Teilnehmenden
Quelle: Eigene Darstellung

Von 19 TN waren vier weiblich, zwölf männlich, eine:r nicht-binär und zwei blieben ohne Angabe (s. Abbildung 18). Die Gruppe im Alter zwischen 20 und 29 Jahren war mit elf Personen die größte. Fünf Personen befanden sich in der Altersgruppe 30-39, eine in der Gruppe 50-59

und zwei machten keine Angabe (s. Abbildung 17). Diese Altersgruppenverteilung war zu erwarten, weil sie der gesuchten User Persona entspricht.

Die TN wurden gebeten anzugeben, in welchem Kontext sie bisher eine DAW genutzt haben. Es konnten mehrere Antworten gewählt werden. Das Ergebnis zeigt, dass die meisten sowohl in ihrer Freizeit als auch im Zuge des Studiums oder der Ausbildung eine DAW genutzt haben. Der berufliche Kontext wurde am seltensten selektiert (s. Abbildung 19).

DAW-Nutzungskontext

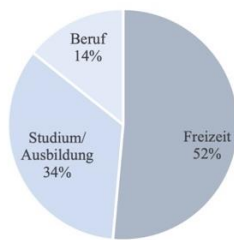


Abbildung 19: DAW-Nutzungskontext der Teilnehmenden
Quelle: Eigene Darstellung

Nutzungsprofile: Digital Audio Workstations

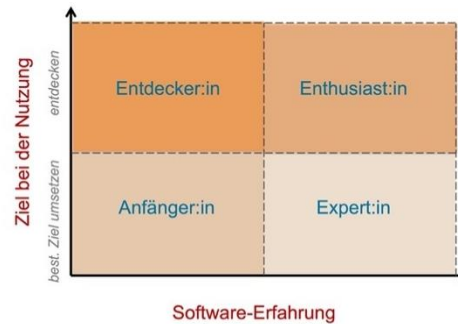


Abbildung 20: DAW-Nutzungsprofile der Teilnehmenden
Quelle: Eigene Darstellung

Bis auf eine Ausnahme schätzten alle TN ihr Erfahrungslevel bei der DAW-Nutzung als niedrig oder mittel ein: Acht Personen wiesen sich ein Level zwischen null und drei zu, zehn Personen ordneten sich zwischen vier und sieben ein und eine Person bei acht. Damit liegt der Durchschnittswert für die DAW-Erfahrung bei 3,89 (s. Abbildung 21). Die meistselektierte DAW-Nutzungsgruppe war „Entdecker:in“, der sich zehn Personen zugehörig fühlten. Fünf Personen sahen sich als „Enthusiast:in“, drei als „Anfänger:in“ und eine als „Expert:in“ (s. Abbildung 20). Daraus lässt sich ableiten, dass das Ziel der TN bei der DAW-Nutzung in den meisten Fällen (15/19) das Erkunden der Möglichkeiten war und sie kein spezielles Ziel vor Augen hatten.

Die Studie richtete sich an Personen ohne professionellen Hintergrund im Feld digitaler Musikkomposition. Die Ergebnisse bestätigen, dass diese Bedingung erfüllt ist. 9 von 19 TN gaben an schon einmal eine eigene Komposition in einer DAW erstellt zu haben. Von diesen neun schätzten drei ihr Erfahrungslevel zwischen null und drei ein (niedrig) und sechs zwischen vier und sieben (mittel). Der durchschnittliche Wert für die Kompositionserfahrung liegt bei 4,22 (s. Abbildung 22). Keine:r der TN hat vorher schon einmal mit musikbezogener KI gearbeitet.

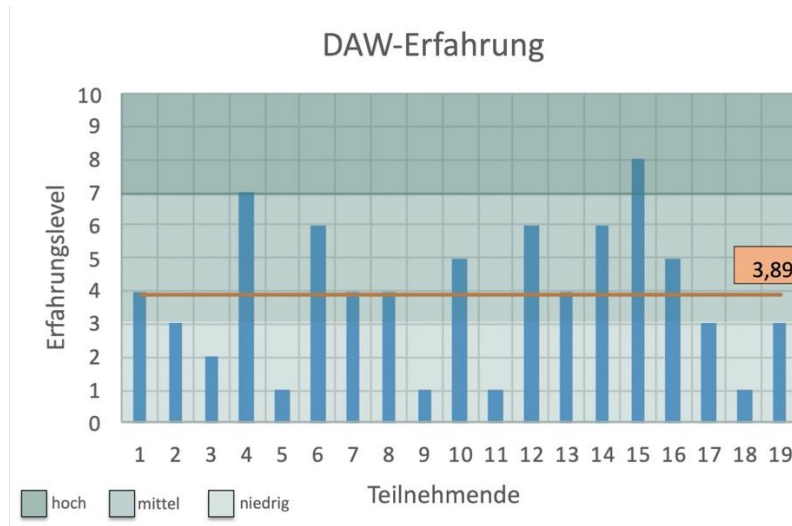


Abbildung 21: DAW-Erfahrungsniveau pro Teilnehmer:in
Quelle: Eigene Darstellung

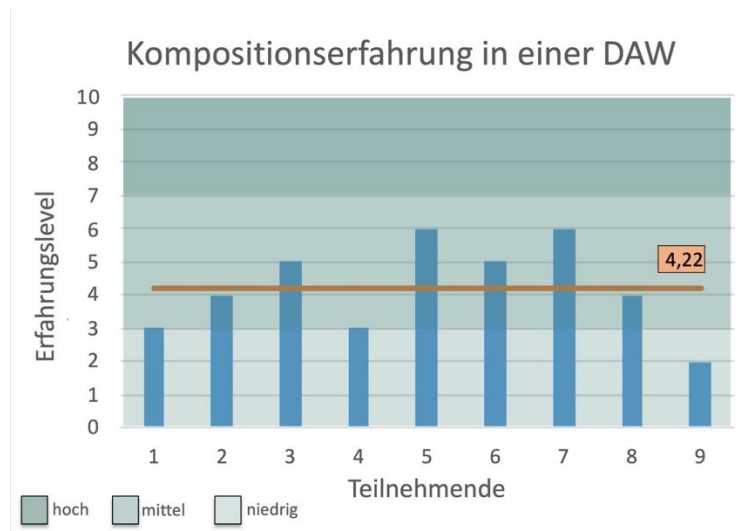


Abbildung 22: Kompositionserfahrung pro Teilnehmer:in
Quelle: Eigene Darstellung

Die TN mit Kompositionserfahrung gaben zusätzlich an, was ihnen dabei wichtig war. Dafür konnten sie mehrere Items wählen. Das Ergebnis zeigt, dass das Wichtigste für sie die Selbstentfaltung war, also eigene Ideen umsetzen (8 Mal selektiert) und der eigenen Kreativität (6) und Persönlichkeit (4) Ausdruck verleihen. Einige bekundeten ihr Interesse daran, innovative Sounds herzustellen (4). Wenige gaben an, beim Komponieren neue Technologien auszuprobieren (2) oder die Software kennenlernen zu wollen (1). Ebenso wenige hatten das Ziel, eine bestimmte Aufgabe zu erfüllen (2) oder bekannte Klänge/Produktionstechniken zu imitieren (2). Obwohl die TN also generell in der DAW bisher viel ausprobieren und entdecken wollten, starteten sie mit klaren Vorstellungen und eigenen Ideen in das Komponieren.

4.4.2 Ergebnis: System Usability Scale

Das Plugin wurde insgesamt von den TN als sehr einfach zu nutzen bewertet. Der Score über alle TN liegt mit einer erreichten Punktzahl von 73,16/100 (SD = 14,19) im akzeptablen Bereich über 70. Für die Hälfte der TN wurde ein Score zwischen 66,25 und 77,5 berechnet. Unter der Grenze zum akzeptablen Bereich lagen nur 6 von 19 Bewertungen. Ein extremer Ausreißer macht sich bemerkbar. Die Antworten dieses einen TN weichen stark von den restlichen ab. Er empfand das Plugin als nicht einfach zu nutzen. Genaueren Aufschluss darüber gibt es nicht, weil der TN nicht die Möglichkeit für ein qualitatives Feedback genutzt hat. Eine Darstellung der SUS-Ergebnisse ist in Abbildung 23 zu sehen.

Die Bewertung für die insgesamt Bedienerfreundlichkeit fiel mit einem Durchschnittswert von 5,53/7 (SD = 0,96) sehr gut aus (s. Abbildung 24).

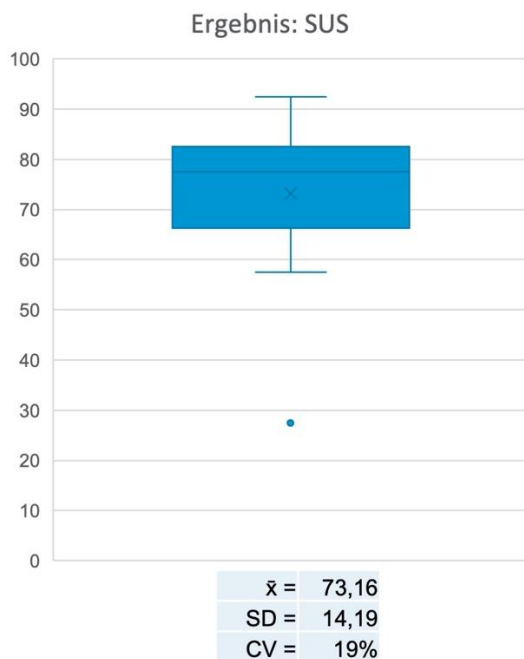


Abbildung 23: SUS-Scores
Quelle: Eigene Darstellung

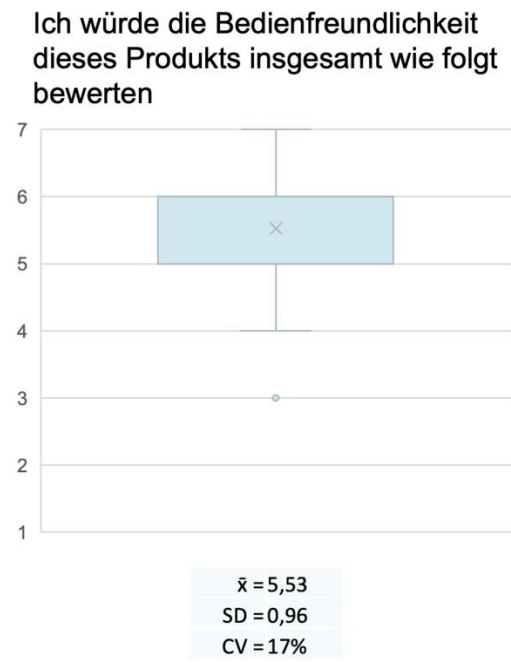


Abbildung 24: Bedienerfreundlichkeit insgesamt
Quelle: Eigene Darstellung

4.4.3 Ergebnis: Controllability

Die Controllability wurde quantitativ mit zwei Metriken CON-01 und CON-02 bewertet (s. Abbildung 25). CON-01 sagt aus, wie leicht kontrollierbar die TN das Plugin fanden. Die durchschnittliche Bewertung lag hier mit 5,42/10 (SD = 2,09) nur knapp über der kritischen

Grenze. Es fällt auf, dass die Bewertungen zu CON-01 bei den TN weit auseinander lagen und sich fast über die gesamte Bewertungsskala erstrecken, dass sie also in beide Richtungen vom Mittelwert abweichen (erkennbar an den Antennen des Boxplots). Die Controllability wurde also von unterschiedlichen TN sehr unterschiedlich wahrgenommen. Angesichts dessen hat der Mittelwert für CON-01 nicht viel Aussagekraft.

CON-02 drückt aus, zu welchem Grad sich die TN mehr Kontrolle über den Output wünschen. Der Mittelwert zeigt hier mit 8,11/10 (SD = 1,76) deutlich an, dass die TN das Bedürfnis nach mehr Kontrolle hatten. Das qualitative Feedback gibt dazu weiteren Aufschluss (s. Ergebnis der qualitativen Befragung, 4.4.7).

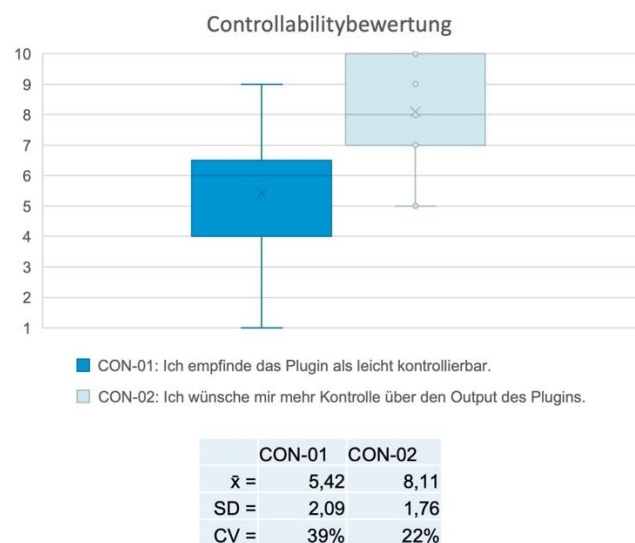


Abbildung 25: Controllability-Scores
Quelle: Eigene Darstellung

4.4.4 Ergebnis: Creativity Support Index

Für die Bewertung der Kreativitätsförderung durch das Plugin wurden fünf Bewertungsdimensionen ausgewertet. Die Anwender:innen waren zufrieden mit den Möglichkeiten durch das Plugin eigene Ideen, Optionen, Entwürfe und Ergebnisse für ihre Kompositionen zu erkunden. Im Durchschnitt wurde für die Kategorie „Erkundung“ ein Score von 3,74/5 vergeben. Der Score von 3,47 Punkten in der Kategorie „Ergebnisse lohnen Aufwand“ weist darauf hin, dass die TN durch den Output, den das Plugin ihnen generiert hat, tendenziell noch nicht zufrieden gestellt waren. Für die am schwächsten bewertete Kategorie „Ausdruckskraft“ lag der Mittelwert bei 2,79 Punkten. Das zeigt, dass es Raum für Verbesserungen gibt, um die kreative Ausdruckskraft der Benutzer:innen weiter zu stärken und bestätigt die Tendenzen aus dem Controllability Feedback. Besser schneidet das Plugin bei der Bewertung für die Kategorie

„Vereinnahmung“ ab. Hier wurde ein Score von 3,89 vergeben, was auf eine gute Bedienfreundlichkeit und Integration in den kreativen Workflow hinweist. Der höchste Score wurde für die Kategorie „Spaß“ vergeben. Mit 4,47 als Durchschnittswert wird hier deutlich, dass das Plugin den wenig erfahrenen Benutzer:innen Freude bereitet und als unterhaltsam wahrgenommen wird. Dieser positive Effekt trug zur allgemeinen Zufriedenheit bei, trotz Schwierigkeiten im Bereich Controllability. In Abbildung 26 sind die CSI-Scores zu sehen.

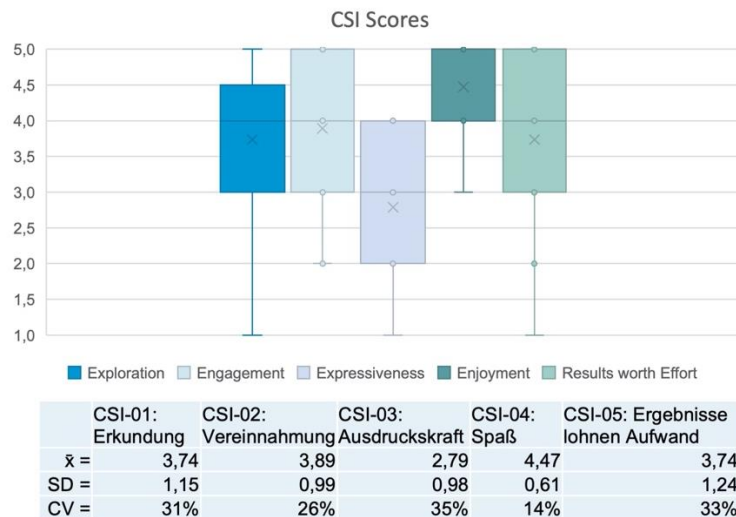


Abbildung 26: CSI-Scores
Quelle: Eigene Darstellung

4.4.5 Ergebnis: Technology Acceptance

Die durchschnittliche Bewertung für die Wahrgenommene Nützlichkeit lag bei 3,32/5 (SD = 0,30). Dies deutet darauf hin, dass die TN eine moderate Wahrnehmung der Nützlichkeit des Systems haben, wobei die Meinungen relativ konsistent sind. In Bezug auf die Wahrgenommene Einfachheit der Systemnutzung lag der Mittelwert bei 3,74 (SD = 0,76). Dies legt nahe, dass die TN das System im Allgemeinen als zufriedenstellend einfach in der Handhabung betrachten, jedoch mit einer gewissen Varianz in den individuellen Bewertungen. Abbildung 27 zeigt die TAM-Scores für diese zwei Metriken.

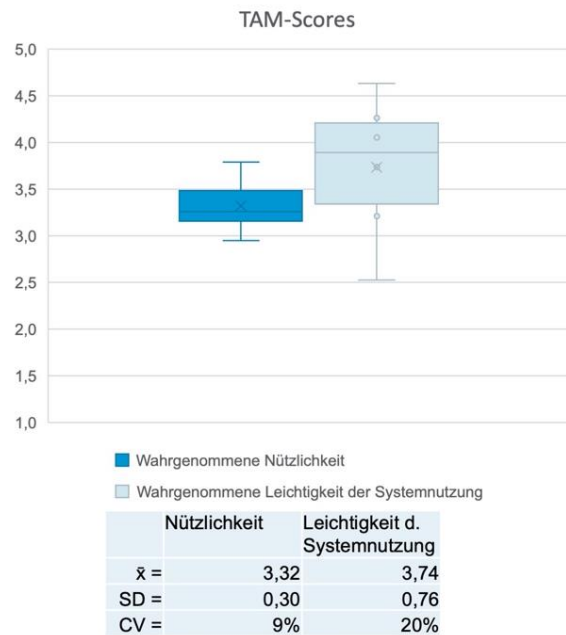


Abbildung 27: TAM-Scores
Quelle: Eigene Darstellung

Hinsichtlich der Wahrscheinlichkeit der zukünftigen Nutzung des Plugins lag eine hohe Variabilität vor. Hier gingen die Aussagen der TN sehr auseinander. Durchschnittlich antworteten sie mit 4,26/7 (SD = 1,79). Diese moderate Prognose spiegelt die Ergebnisse für die Wahrgenommene Leichtigkeit und Nützlichkeit des Systems wider. Dazu ist jedoch zu sagen, dass 50 Prozent der Antworten in der breiten Spanne zwischen zwei und sechs liegen. Der Variationskoeffizient beträgt damit 42 Prozent und der Mittelwert hat keine starke Aussagekraft. Es haben beispielsweise ebenso viele TN auf die Frage „Wie wahrscheinlich wäre es, dass Sie das Plugin in Zukunft benutzen?“ mit 2/7 geantwortet, wie TN mit 6/7 geantwortet haben. In Abbildung 28 ist die breite Streuung im Boxplotdiagramm klar zu erkennen. Das qualitative Feedback gibt genaueren Aufschluss darüber, welche unterschiedlichen Meinungen es diesbezüglich gab.

Um einen möglichen Zusammenhang zwischen der Kreativitätsförderung und der Voraussage über eine zukünftige Nutzung zu untersuchen, wurde aus den Ergebnissen aus der CSI-Befragung für jede:n TN ein Mittelwert über die fünf Bewertungskategorien berechnet und dieser mit ihrer jeweiligen Nutzungsvoraussage in einem Diagramm dargestellt (s. Abbildung 29). Es kann zumindest eine Tendenz wahrgenommen werden: Testpersonen mit einer guten Erfahrung im Bereich Kreativitätsförderung während der Task gaben eher an, in Zukunft das Plugin nutzen wollen.

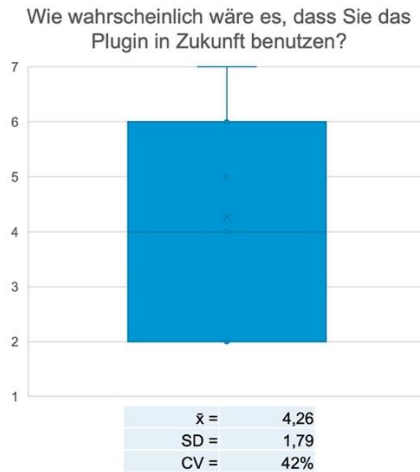


Abbildung 28: Voraussage über die Nutzungswahrscheinlichkeit
Quelle: Eigene Darstellung

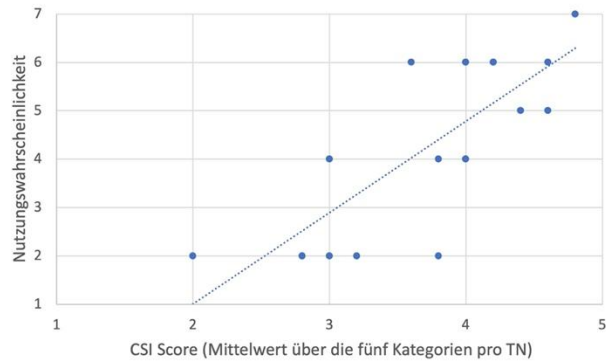


Abbildung 29: Nutzungswahrscheinlichkeit in Abhängigkeit vom CSI-Score/TN
Quelle: Eigene Darstellung

4.4.6 Ergebnis: After Scenario Questionnaire

Die Zufriedenheit der TN mit der Task war äußerst hoch. Mit einem ASQ-Score von 4,58/5 (SD = 0,19) liegt die Punktzahl im sehr guten Bereich, nahe am Höchstwert. Dieses Ergebnis deutet darauf hin, dass die TN trotz unterschiedlicher Vorkenntnisse mit der angesetzten Zeit, den bereitgestellten Hilfsmitteln und dem Schwierigkeitsgrad der Task zufrieden waren.

In Abbildung 30 wird der ASQ-Score mit den drei Statements zu Zeit, Schwierigkeitsgrad und Hilfsmitteln für die Task aufgeschlüsselt. Bei der Bewertung waren sich die TN sehr einig. Ein paar wenige gaben am Ende der Task das mündliche Feedback, dass sie gerne noch mehr Zeit gehabt hätten, um das Plugin weiter auszuprobieren. Das Testszenario war überschaubar und hat die TN weder über- noch unterfordert. Zwar lernten alle eine für sie neue Anwendung kennen, doch waren sie in der Lage, sich mit Hilfe des Tutorialvideos und des Cubasetemplates schnell zurechtzufinden.

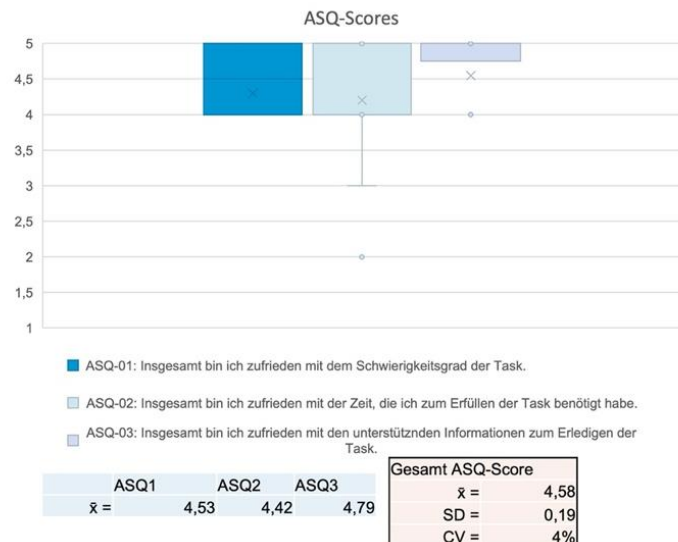


Abbildung 30: ASQ-Scores
 Quelle: Eigene Darstellung

4.4.7 Ergebnis: Qualitative Befragung (Kommentare)

In der Befragung gab es zwei Möglichkeiten für die TN, qualitatives Feedback über ein Kommentarfeld zu geben. Die Auswertung im Folgenden bezieht sich auf beide Kommentarfelder. Zunächst wurde für die Analyse das Feedback zur Controllability aus den Kommentaren herausgefiltert. In der Gesamtheit der Kommentare wurde für jeden Parameter mindestens ein Vorfall gemeldet, bei dem die gewünschte Einstellung nicht umgesetzt wurde. Diesen Meldungen wurden aufgrund der Implementierungsschwierigkeiten erwartet. Der auftretende Bug wurde in einer neuen Version des Plugins inzwischen behoben und wird die Controllabilitywahrnehmung in zukünftigen Tests nicht mehr behindern. Das Feedback dazu wird deswegen in diesem Abschnitt ausgeklammert.

Für das restliche Feedback wurde induktiv vorgegangen. Es bezog sich vor allem auf Wünsche für die Gestaltung des Tools. Diese Wünsche wurden kategorisiert als Parameter-, Feature- oder Systemwünsche. Es gab aber auch einige Kommentare zum individuellen Erlebnis der TN oder ihrer Wahrnehmung des Systems.

Kommentar

Jede:r TN hatte eigene Präferenzen und einige teilten diese in den Kommentaren mit. So empfand beispielsweise TN13 einige Melodie-Generierungen als „zu gradlinig, d.h. Betonung und Noten zu sehr auf der 1. Note“. Daraus entstand ein System-Wunsch (weiter unten aufgeführt). TN6 fand die Skala für den Parameter „Density“ (low/middle/high) „nicht wirklich intuitiv“.

TN11 empfand die Generierung im monophonen Modus „meist schlüssiger und spannender“ als im polyphonen Modus. TN16 lobte, dass „[m]it gleichen Einstellungen vielfältige mögliche Ergebnisse“ generiert werden können. Während TN11 das Plugin in Zukunft höchstens „zum Spaß mal ausprobieren“ will, hat TN16 „Lust, mit dem Plugin mehr auszuprobieren“. TN23 findet das Plugin „[g]ut als Startpunkt oder Ideengenerierung“, an der dann „noch viel geschraubt werden“ kann. TN15 schließt sich dieser Meinung an. Auch in Unterhaltungen während des Debriefings haben viele TN die Unterstützung bei der Ideenfindung und den erleichterten Zugang zum Komponieren durch das Plugin wertgeschätzt.

Parameterwunsch

Als Parameterwunsch wurde mehrfach die Tonart genannt. Während es für erfahrene Komponist:innen in der DAW leicht ist, die Tonart einer generierten Spur im MIDI-Editor manuell zu ändern, wäre es für Anfänger:innen komfortabler, die Einstellung im Interface des Plugins integriert zu sehen. Außerdem wurde nach einem Parameter für Emotion und Pausen gefragt. So schrieb etwa TN19 „Tonleiter manuell einstellen wär cool oder eine Option für die Emotion, so etwas wie: fröhlich/traurig/düster...“ und TN20 „Mehr Steuerparametern wie zB. Pausen oder Tonart wären gut“. TN15 wünschte sich einen Parameter für die Art der Generierung, also ob sie „mehr eine Fläche, Melodie etc.“ sein sollte. Dies kann bereits durch die Parameter „Duration“, „Density“ und „Monophonic/Polyphonic“ beeinflusst werden, was aber nicht unmittelbar verständlich für alle Benutzer:innen war. Deswegen könnte ein abstrakterer Parameter für alles in Einem hilfreich sein. Es gab noch mehr Stimmen, die für mehr Parameter im Allgemeinen sprachen, ohne konkrete Beispiele zu nennen.

Featurewunsch

Als Featurewunsch wurde genannt, dass es einen Button geben könnte, der eine Variation der vorherigen Generierung erstellt, „z.B. so ähnlich, wie die letzte Generation nur schneller“ (TN8). Ein weiterer Vorschlag war ein „Reduce from Selection“ Button, der in einem generierten Event die MIDI-Noten reduziert, als Pendant zu „Add to Selection“.

Systemwunsch

Von TN13 wurde der Wunsch geäußert, dass das System mehr rhythmisch und melodisch variante Outputs produzieren sollte („Synkopen wären abwechslungsreicher“). Dies ist einerseits eine Frage der Trainingsdaten, andererseits auch der Implementation und Anwendung, da es

bereits den Parameter „Crazyness“ gibt, der Einfluss darauf hat, wie gradlinig oder außergewöhnlich das generierte Event ist.

Es wurde von mehreren TN angemerkt, dass der „Genre“ Parameter manchmal im Zusammenspiel mit den restlichen Parametern nicht zufriedenstellend funktioniert. Daraus entstand der Wunsch, diesen Parameter zu überdenken. So schreibt zum Beispiel TN6: „einige Genres ergeben mit bestimmten Settings unbenutzbare, seltsame Ergebnisse, sodass ich ein anderes Zusammenspiel zwischen Genre und den restlichen Einstellungen begrüßen würde“. Genres zeichnen sich durch typische Rhythmen, Harmonien, Tempi, und so weiter aus, weswegen die Restriktion von Notendauer und -dichte und anderen Parametern dazu führen kann, dass das Genre nicht mehr zu erkennen ist und das Ergebnis womöglich falsch klingt. Hier könnte also überlegt werden, wie die Hierarchie dieser Parameter gestaltet werden sollte, an welcher Stelle eine Parametereinstellung eine Vorschrift oder nur ein Vorschlag ist und wie das mit der UI an die Anwender:innen kommuniziert werden kann.

Die TN äußerten auch den Wunsch nach einer Generierung für kürzere Zeitabschnitte. Damit war entweder das taktbasierte Generieren gemeint (TN8: „Ansonsten fehlte mir eine Option nur Teile der Spur zu generieren. (z.B. nur einen Takt“), oder das Inpainting einzelner MIDI-Noten (TN18: „Ich hätte mir gewünscht es gäbe die Möglichkeit präziser einen gewissen [Teil] der Spur zu ersetzen“). Diese beiden Funktionen war im festgelegten Testszenario aufgrund des damit einhergehenden Bugs der Version 1.5.0.188 nicht vorgesehen. Das System ist aber dazu fähig (s. Abschnitt 2.1.3) und in der gefixten Version des Plugins sind diese Optionen vorhanden.

TN22 schlug vor, das System so zu gestalten, dass es möglich ist, eine eigene Referenzmusik als Input bereitzustellen: „Wenn man einen Input liefern könnte, an dem sich der Output orientiert, wäre das sicher eine coole Sache (wie bei KI-generierten Bildern ein Bild hochladen, an dem sich der Content orientiert)“. Beispielsweise könnte das die .wav- oder .mp3-Datei eines bereits existierenden Songs sein.

4.5 Diskussion

Die Ergebnisse liefern einen umfangreichen Einblick in die Wahrnehmung von MMM durch unerfahrene Komponist:innen. Zunächst einmal muss erwähnt werden, dass es im Anfänger:innenspektrum sehr unterschiedliche Ausprägungen von Benutzer:innen gibt. „Unerfahren“ reicht in diesem Fall von Personen ohne jegliche DAW- und Kompositionserfahrung bis hin zu

Amateur-Komponist:innen mit langer DAW-Erfahrung. Diese Unterschiede im Kenntnisstand führen mitunter zu einer großen Ergebnisvarianz.

Die Vorbereitung der Task enthielt die Entwicklung eines geeigneten Prototypen, die Produktion eines Tutorialvideos zur Benutzung des Prototypen und die Definition eines Testszenarios mit Hilfe von voreingestellten Cubasetemplates. Die Task, die entwickelt wurde, ist repräsentativ für die spätere Verwendung von MMM. Das System wurde so entwickelt, dass es flexibel in DAWs integriert werden kann. Mit einem Plugin wie dem Music Generation Panel könnte das System auch in andere DAWs implementiert werden. Alle TN konnten die Task bearbeiten, unabhängig davon, ob sie schon einmal eine DAW benutzt hatten oder nicht. Das als Hilfsmittel bereitgestellte Tutorialvideo und die Verwendung von voreingestellten Templates kamen den unterschiedlichen DAW-Kenntnissen der Benutzer:innen entgegen, sodass alle die Möglichkeit hatten, die Zeit ganz zur Erkundung des Plugins zu nutzen. Dies wird auch in den Ergebnissen der ASQ-Befragung deutlich. Die Benutzer:innen empfanden die Task für die vorgesehenen 30 Minuten weder als zu umfangreich noch als zu trivial. Die Länge des Zeitslots wurde im Pretest ermittelt und die Ergebnisse der ASQ-Befragung bestätigen, dass die TN damit zufrieden waren.

Die positiven SUS-Ergebnisse weisen auf eine allgemeine Zufriedenheit unerfahrener Komponist:innen mit der Bedienung von MMM hin. Das Plugin ist somit auch mit den zusätzlich einstellbaren Parametern als bedienfreundlich und einfach zu erlernen zu bewerten, selbst wenn es von unerfahrenen Benutzer:innen verwendet wird. Vergleicht man das Ergebnis mit den SUS-Scores der ersten Studie von Tchemeube u. a., zeigt sich, dass die Ergebnisse beinahe identisch sind. Das spricht für eine gute Integration der neuen Parameter in das Interfacedesign und für eine intuitive, gut verständliche Bedienung trotz Zunahme der Komplexität.

Bei der Interpretation der CSI-Ergebnisse ist zu beachten, dass die Standardabweichungen der Scores relativ hoch sind, wie der Vergleich der Variationskoeffizienten zeigt. Das Erleben scheint bei den TN sehr individuell zu sein. Die schwächer bewerteten Dimensionen „Erkundung“, „Ausdruckskraft“ und „Ergebnisse lohnen Aufwand“ weisen auch eine höhere Variation auf, das heißt die Aussagekraft der Mittelwerte ist geschwächt. Bei den am höchsten bewerteten Dimensionen „Spaß“ und „Vereinnahmung“ waren sich die TN im Vergleich deutlich einiger. Auch in der ersten Studie von Tchemeube u. a. (2023) war die Varianz bei den CSI-Scores etwas höher als beispielsweise bei den SUS- oder TAM-Scores. Die Bewertung der Dimension „Spaß“ schnitt bei den Expert:innen ebenfalls am besten und die der Dimension „Ausdruckskraft“ am schwächsten ab. Das Bedürfnis nach kreativem Ausdruck wurde also sowohl bei der

Gruppe mit dem Ein-Parameter-Plugin als auch bei der Gruppe mit dem Sechs-Parametern noch nicht ausreichend befriedigt. Durch die Frage ERF-06 „Was war Ihnen bei der Musikkomposition am wichtigsten?“ aus dem Fragebogen zu Demographie und Erfahrung ist bekannt, dass für die TN der zweiten Studie das Ausleben der eigenen Kreativität und der Ausdruck der Persönlichkeit beim Komponieren besonders wichtig ist, was die Bedeutung der Kreativitätsförderung unterstreicht.

Die größere Streuung bei der Wahrgenommenen Leichtigkeit der Systemnutzung in den Ergebnissen der TAM-Befragung könnte auf unterschiedliche Erfahrungslevel und individuelle Präferenzen bei der Nutzung des Systems zurückzuführen sein. Auch bei der Voraussage über die zukünftige Nutzung gab es große Uneinigkeit. Das freie Kommentarfeld im Anschluss an die Frage gab den TN die Möglichkeit, diese Einschätzung näher zu erläutern. Diese Möglichkeit wurde jedoch nicht von allen genutzt. Insbesondere bei den TN, die das Plugin voraussichtlich eher nicht nutzen wollen, fehlt ein qualitatives Feedback. Nur eine Person (TN11) schrieb: "Ich würde es wahrscheinlich zum Spaß mal ausprobieren. Diese halbe Stunde Nutzung hat mich allerdings nicht so sehr begeistert, dass ich es unbedingt noch einmal nutzen möchte“. TN11 kreuzte 2/6 an. Er war ein DAW-Enthusiast (DAW-Erfahrung 6/10), der vorher bereits komponiert hatte (Kompositionserfahrung 5/10), wobei es ihm ausschließlich um die Umsetzung eigener Ideen ging. Die festgestellte Tendenz, dass Benutzer:innen, die sich in ihrer Kreativität gefördert fühlen, das Tool auch gerne weiter nutzen wollen, erscheint schlüssig und unterstreicht einmal mehr, wie wichtig die Kreativitätsförderung für die Akzeptanz des Systems ist.

Die Ergebnisse der Controllabilityfragen zeigen vor dem Hintergrund des vorab bekannten Bugs der MGP-Version 1.5.0.188 erwartungsgemäß ein mittelmäßiges bis unzureichendes Ergebnis. Je nachdem, ob die Befragten mit dem Bug konfrontiert waren oder nicht, fiel die Einschätzung der Controllability grundsätzlich unterschiedlich aus. Durch die bereits erfolgte Behebung des Bugs und durch die Berücksichtigung der konkreten Vorschläge der Testpersonen für weitere Parameter und Funktionen kann die Controllability von MMM erheblich verbessert werden.

Das qualitative Feedback der TN ist sehr interessant, da es zeigt, dass der Wunsch nach mehr Kontrolle auch mit zusätzlichen Parametern in der zweiten Version des Prototypen bestehen bleibt. In der ersten Prototypversion gab es nur einen Parameter. Die Ergebnisse der User Study mit dem Ein-Parameter-Prototyp zeigten, dass die Expert:innen das Bedürfnis nach mehr Kontrolle hatten. Durch die zweite Studie kann nun festgestellt werden, dass sich auch die Anfänger:innen viel kreativen Einfluss auf die KI-Komposition wünschen. Die Erwartung der

Benutzer:innen an das System scheint zunächst zu sein, mit konkreten Ideen in den Prozess zu starten, und diese dem System so genau wie möglich mitzuteilen. Viele sind jedoch offen dafür, sich auf eine andere Art der Interaktion einzulassen, als sie zu Beginn erwartet hatten. Sie begrüßen die Möglichkeit, sich durch die vielseitigen musikalischen Kreationen von MMM inspirieren zu lassen und einen einfachen Zugang zum Komponieren zu erhalten.

Aus dem qualitativen Feedback lassen sich folgende Empfehlungen für die Usability von MMM ableiten: Bei der Gestaltung des Systems sollte aus Sicht unerfahrener Benutzer:innen darauf geachtet werden, dass das Verständnis der Parameter und Skalen keine fortgeschrittenen Musiktheoriekenntnisse erfordert. Abstraktere Parameter, wie „Stimmung“ oder „Musikalische Struktur“ sind hier hilfreich. Unerfahrene Benutzer:innen schätzen aufgrund ihrer begrenzten Kenntnisse im MIDI-Editing und anderen Arbeitsschritten der Musikproduktion die Möglichkeit, direkt in der Systemoberfläche viel Einfluss auf die Generierung nehmen zu können. Eine große Vielfalt an Parametern ermöglicht ihnen einen effizienteren und effektiveren Workflow beim Komponieren. Darüber hinaus wird dadurch die Kreativitätsförderung erhöht, was zu einer hohen Akzeptanz von MMM führt.

5 Fazit

Die vorliegende Arbeit untersucht die Nutzungserfahrungen unerfahrener Komponist:innen mit dem KI-Kompositionstool MMM. Dafür wurde unter Berücksichtigung internationaler Standards und Empfehlungen von Expert:innen eine Task-basierte User Study durchgeführt, die die Wahrnehmung der Usability und die Akzeptanz von MMM aus Sicht dieser Benutzer:innen-gruppe beleuchtet. Ziel der Befragung war es einzuschätzen, ob das System von unerfahrenen Komponist:innen als effizient, effektiv und zufriedenstellend wahrgenommen wird, um eigene kreative Ideen umzusetzen, und ob sie die Benutzung als leicht und nützlich empfinden. Als Erhebungsinstrument wurden Fragebögen mit qualitativen und quantitativen Anteilen verwendet.

Im Zuge der Arbeit wurde eine Task konzipiert, die für den Einsatz in einer Task-basierten Usability- und Akzeptanzbefragung mit Anfänger:innen mit variierenden Vorkenntnissen geeignet ist und für weitere Tests angewendet werden kann.

Die Ergebnisse zeigen eine positive Bewertung der Bedienung von MMM, die weitestgehend als benutzer:innenfreundlich und leicht erlernbar für unerfahrene Komponist:innen zu beurteilen ist. Ein weiterer Ausbau der kreativen Einflussmöglichkeiten auf das System wird empfohlen, um die Akzeptanz des Systems zu erhöhen und die Nutzungserfahrung von Anfänger:innen

zu optimieren. Um das zu erreichen, wird vorgeschlagen, weitere abstrakte, intuitiv verständliche Parameter in das Interface zu integrieren.

Als nächster Schritt in der Entwicklung könnte die User Study mit der gleichen Stichprobe unter Verwendung eines voll funktionsfähigen Prototypen mit zwei bis drei zusätzlichen Parametern wiederholt werden. Der Vergleich der Testergebnisse, insbesondere im Hinblick auf die Controllability und die Kreativitätsförderung, wären hierbei interessant. Es kann vermutet werden, dass die Akzeptanz des Systems durch diese Veränderungen bei den gleichen Personen deutlich erhöht wird.

Eine weitere zukünftige Option wäre die Erweiterung des Blicks auf die Entwicklung der KI-assistierten Musikkomposition aus einer menschenzentrierten Perspektive. Wichtige Fragen des KI-Aufschwungs in der Musik, die von Expert:innen bereits diskutiert werden, sind unter anderem rechtliche Fragen in Bezug auf musikalische Trainingsdaten und gesellschaftliche Fragen in Bezug auf die Entwicklung der Musikproduktionsbranche und eine befürchtete Übernahme der kreativen Berufe durch Maschinen.

Die feste Integration benutzer:innenzentrierter Tests in den Entwicklungsprozess von MMM und die Berücksichtigung weitergehender menschenzentrierter Überlegungen können den Erfolg des Produkts maßgeblich steigern.

Literaturverzeichnis

- Ambareta, Y., Rosana und Lusianah. 2022. Development of Digital Society: Factors Affecting the Usage of Mobile Payment in Jakarta Metropolitan Area. In: *Proceedings of the 2022 IEEE International Conference of Computer Science and Information Technology: Boundary Free: Preparing Indonesia for Metaverse Society, ICOSNIKOM 2022*. Laguboti, North Sumatra, Indonesia: Institute of Electrical and Electronics Engineers Inc., S. 421–427. doi: 10.1109/ICOSNIKOM56551.2022.10034910.
- Bangor, A., Kortum, P.T. und Miller, J.T. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24(6), S. 574–594.
- Behr, D. 2023. *What to consider and look out for in questionnaire translation*. Mannheim: ESIS – Leibniz-Institute for the Social Sciences (GESIS – Survey Guidelines). doi: 10.15465/gesis-sg_en_043.
- Bell, C. 2011. Algorithmic Music Composition Using Dynamic Markov Chains and Genetic Algorithms. *The Journal of Computing Sciences in Colleges* 27(2), S. 99–107.
- Briot, J.-P. 2021. From artificial neural networks to deep learning for music generation: history, concepts and trends. *Neural Computing and Applications* 33, S. 39–65. Verfügbar unter: <https://doi.org/10.1007/s00521-020-05399-0> [Zugegriffen: 4 Januar 2024].
- Brooke, J. 1996. Sus: A 'quick and dirty' usability scale. In: *Usability Evaluation In Industry*. London: Taylor & Francis Ltd, S. 189–204.
- Cabral, D. und Correia, N. 2017. Video editing with pen-based technology. *Multimedia Tools and Applications* 76(5), S. 6889–6914. doi: 10.1007/S11042-016-3329-Y.
- Cherry, E. und Latulipe, C. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21(4). Verfügbar unter: <https://dl.acm.org/doi/abs/10.1145/2617588> [Zugegriffen: 19 September 2023].
- Chung, J. 2023. *Steerable AI-powered Art-making Tools*. Michigan: University of Michigan. Verfügbar unter: <http://deepblue.lib.umich.edu/handle/2027.42/176627> [Zugegriffen: 30 Dezember 2023].
- Church, A.H. 1993. Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly* 57, S. 62–79. Verfügbar unter: <https://academic.oup.com/poq/article/57/1/62/1833464> [Zugegriffen: 29 Dezember 2023].

- Davis, F.D. und Venkatesh, V. 1996. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International Journal of Human-Computer Studies* 45(1), S. 19–45. doi: 10.1006/IJHC.1996.0040.
- Diekmann, A. 2010. *Empirische Sozialforschung*. Reinbek: Rowohlt Taschenbuch Verlag.
- DIN EN ISO. 2018. *Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte* DIN EN ISO 9241-11:2018.
- DIN EN ISO. 2019. *Ergonomie der Mensch-System-Interaktion - Teil 220: Prozesse zur Ermöglichung, Durchführung und Bewertung menschenzentrierter Gestaltung für interaktive Systeme in Hersteller- und Betreiberorganisationen*.
- DIN EN ISO. 2020. *Ergonomie der Mensch-System Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme* DIN EN ISO 9241-210:2019.
- Donahue, C., Mao, H.H., Li, Y.E., Cottrell, G.W. und McAuley, J. 2019. LakhNES: Improving multi-instrumental music generation with cross-domain pre-training. *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019, November 4-8, 2019, Delft, The Netherlands*, S. 685–692.
- Eck, D. und Schmidhuber, J. 2002. A First Look at Music Composition using LSTM Recurrent Neural Networks. Technical Report, IDSIA/USI-SUPSI. *IDSIA* 07(02). Verfügbar unter: <https://people.idsia.ch/~juergen/blues/IDSIA-07-02.pdf> [Zugegriffen: 11 Januar 2024].
- Edwards, P., Roberts, I., Clarke, M., Diguseppi, C., Pratap, S., Wentz, R. und Kwan, I. 2002. Increasing response rates to postal questionnaires: systematic review. *BMJ* 324, S. 1183–1192. doi: 10.1136/bmj.324.7347.1183.
- Ens, J. 2023. Corpus-based Symbolic Music Generation: Data, Representation, Models, Evaluation. *Simon Fraser University*. Verfügbar unter: https://summit.sfu.ca/_flysystem/fedora/2023-03/etd22360.pdf [Zugegriffen: 18 September 2023].
- Ens, J. und Pasquier, P. 2020. MMM : Exploring Conditional Multi-Track Music Generation with the Transformer. Verfügbar unter: <https://arxiv.org/abs/2008.06048v2> [Zugegriffen: 18 September 2023].
- Garcia, H.F., Seetharaman, P., Kumar, R. und Pardo, B. 2023. VampNet: Music Generation via Masked Acoustic Token Modeling. In: *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR 2023*. Milano, S. 359–366. Verfügbar unter: <https://arxiv.org/abs/2307.04686v2> [Zugegriffen: 22 November 2023].

- Goodfellow, I., Bengio, Y. und Courville, A. 2016. *Deep Learning*. Massachusetts: Massachusetts Institute of Technology.
- Göritz, A.S. 2006. Incentives in Web Studies: Methodological Issues and a Review. *International Journal of Internet Science* 1(1), S. 58–70.
- Gustafsson, F. 2019. *An Explorative Design Study of a Booking System Evaluating the Usability and Experience of a User Interface for Novice Admin Users*. Stockholm: KTH Royal Institute of Technology. Verfügbar unter: <http://www.diva-portal.org/smash/get/diva2:1305553/FULLTEXT01.pdf> [Zugegriffen: 2 Januar 2024].
- Huang, C.-Z.A. u. a. 2018. Music Transformer: Generating Music with Long-Term Structure (Preprint). Verfügbar unter: <https://arxiv.org/abs/1809.04281v3> [Zugegriffen: 18 September 2023].
- Jacobsen, J. und Meyer, L. 2017. *Praxisbuch Usability & UX*. Bonn: Rheinwerk Verlag GmbH.
- Jiang, Y., Liu, X., Liu, F., Wu, D. und Anumba, C.J. 2016. An Analysis of BIM Web Service Requirements and Design to Support Energy Efficient Building Lifecycle. *Buildings* 2016, Vol. 6, Page 20 6(2), S. 20–43. doi: 10.3390/BUILDINGS6020020.
- Kohnke, O. 2015. *Anwenderakzeptanz unternehmensweiter Standardsoftware*. Wiesbaden: Springer Fachmedien. doi: 10.1007/978-3-658-08206-2.
- Krosnick, J.A. 1991. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *APPLIED COGNITIVE PSYCHOLOGY* 5, S. 213–236. doi: 10.1002/acp.2350050305.
- Lackes, R., Siepermann, M. und Lübbecke, M. 2018. *Algorithmus*. Verfügbar unter: <https://wirtschaftslexikon.gabler.de/definition/algorithmus-27106> [Zugegriffen: 5 Januar 2024].
- Liang, F., Gotham, M., Johnson, M. und Microsoft, J.S. 2017. Automatic Stylistic Composition of Bach Chorales With Deep LSTM. In: *Proceedings of the 18th ISMIR Conference, October 23-27, 2017, Suzhou, China*. S. 449–456.
- Lu, Z., Huang, D., Bai, L., Qu, J., Wu, C., Liu, X. und Ouyang, W. 2023. Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images. In: *Conference Submission for the 37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks*. doi: 10.48550.

- Malandro, M.E. 2023. Composer's Assistant: An interactive transformer for multi-track MIDI infilling. In: *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR 2023*. Milano, S. 327–334.
- Martin-Gutierrez, J., Del Rio Guerra, M.S., Lopez-Chao, V., Gastelum, R.H.S. und Bojórquez, J.F.V. 2020. Augmented Reality to Facilitate Learning of the Acoustic Guitar. *Applied Sciences* 2020, Vol. 10, Page 2425 10(7), S. 2425–2438. doi: 10.3390/APP10072425.
- Menold, N. und Bogner, K. 2015. *Gestaltung von Ratingskalen in Fragebögen (Version 1.1)*. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). Verfügbar unter: https://doi.org/10.15465/gesis-sg_015 [Zugegriffen: 31 Oktober 2023].
- MUSOPEN. [ohne Datum]. *Mozart Musikalisches Würfelspiel, KV 516f*. Verfügbar unter: <https://musopen.org/de/music/2914-musikalisches-wurfelspiel-k-516f/> [Zugegriffen: 4 Januar 2024].
- Nielsen, J. 1993. *Usability Engineering*. San Francisco: Morgan Kaufmann.
- Nissen, M. u. a. 2023. Usability and Perception of a Wearable-Integrated Digital Maternity Record App in Germany: User Study. *JMIR Pediatr Parent* 2023;6:e50765 <https://pediatrics.jmir.org/2023/1/e50765> 6(1). Verfügbar unter: <https://pediatrics.jmir.org/2023/1/e50765> [Zugegriffen: 2 Januar 2024].
- Nistal, J., Lattner, S. und Richard, G. 2020. DrumGAN: Synthesis of Drum Sounds With Timbral Feature Conditioning Using Generative Adversarial Networks. Verfügbar unter: <https://arxiv.org/abs/2008.12073v2> [Zugegriffen: 15 Januar 2024].
- Noel-Hirst, A. und Bryan-Kinns, N. 2023. *An Autoethnographic Exploration of XAI in Algorithmic Composition*. Verfügbar unter: <https://xaixarts.github.io/> [Zugegriffen: 10 Oktober 2023].
- Oore, S., Simon, I., Dieleman, S., Eck, D. und Simonyan, K. 2020. This time with feeling: learning expressive musical performance. *Neural Computing and Applications* 32(4), S. 955–967.
- Payne, C. 2019. *MuseNet*. Verfügbar unter: <https://openai.com/research/musenet> [Zugegriffen: 15 Januar 2024].
- Roberts, A., Engel, J., Raffel, C., Hawthorne, C. und Eck, D. 2018. A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. In: *Proceedings of the 35th*

International Conference on Machine Learning, PMLR 80, July 10-15, 2023, Stockholm, Sweden. PMLR, S. 4364–4373.

Rosenzweig, E. 2015. *Successful User Experience. Strategies and Roadmaps.* Amsterdam Boston Heidelberg: Morgan Kaufmann (Elsevier).

Sano, S. und Yamada, S. 2022. AI-Assisted Design Concept Exploration Through Character Space Construction. *Frontiers in Psychology* 12. doi: 10.3389/FPSYG.2021.819237/BIBTEX.

Sontakke, N., Utekar, S., Rastogi, S. und Sonawane, S. 2023. Comparative Analysis of Deep-Fake Algorithms. *International Journal of Computer Science Trends and Technology* 11. Verfügbar unter: <https://arxiv.org/abs/2309.03295> [Zugegriffen: 18 Januar 2024].

Spektrum Lexikon der Psychologie. 2000a. *psychometrische Tests.* Verfügbar unter: <https://www.spektrum.de/lexikon/psychologie/psychometrische-tests/12121> [Zugegriffen: 31 Dezember 2023].

Spektrum Lexikon der Psychologie. 2000b. *Regressionsanalyse.* Verfügbar unter: <https://www.spektrum.de/lexikon/psychologie/regressionsanalyse/12688> [Zugegriffen: 2 Januar 2024].

Sturgis, P., Roberts, C. und Smith, P. 2014. Middle Alternatives Revisited: How the neither/nor Response Acts as a Way of Saying „I Don’t Know“? *Sociological Methods and Research* 43(1), S. 15–38. doi: 10.1177/0049124112452527/FORMAT/EPUB.

Tchemeube, R.B., Ens, J., Plut, C., Pasquier, P., Safi, M., Grabit, Y. und Rolland, J.-B. 2023. Evaluating Human-AI Interaction via Usability, User Experience and Acceptance Measures for MMM-C: A Creative AI System for Music Composition. In: *IJCAI-23: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, August 19-25, 2023, Macao, SAR.* S. 5769–5778.

Vaswani, A. u. a. 2017. Attention Is All You Need. In: *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017).* Long Beach, CA. Verfügbar unter: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html [Zugegriffen: 18 Januar 2024].

Xi, Z. u. a. 2023. *The Rise and Potential of Large Language Model Based Agents: A Survey.* Verfügbar unter: arXiv:2309.07864 [Zugegriffen: 10 Oktober 2023].

Anhang

Anschreiben

STUDIE ZU KI-UNTERSTÜTZTER MUSIKKOMPOSITION

Teilnehmende gesucht

Liebe Musik-Interessierte,

für meine Bachelorarbeit führe ich in Zusammenarbeit mit der Firma Steinberg eine User Experience Studie durch. Hast du Interesse ein kreatives KI-System für Musikkomposition zu testen?

Als Teilnehmende:r bekommst du die Möglichkeit, ein Machine Learning System mit einem hochmodernem Algorithmus auszuprobieren, das dich als Plugin in Steinbergs Digitaler Audio Workstation (DAW) Cubase darin unterstützt, eine mehrspurige Komposition mit spannenden Melodien, Rhythmen und Harmonien zu produzieren.

Für den Test wirst du gebeten, das Plugin durch das Bearbeiten einer vorgegebenen Aufgabe zu erkunden und anschließend einen Fragebogen über dein Nutzungserlebnis auszufüllen. Du solltest insgesamt 90 Minuten Zeit mitbringen.

Als Dankeschön darfst du dir eine Softwarelizenz für Cubase Elements 13, Dorico Elements 5 oder WaveLab Elements 11 als Geschenk aussuchen!

Voraussetzung für die Teilnahme ist, dass du mit der grundsätzlichen Bedienung von Cubase (oder anderen DAWs) vertraut bist. Vorerfahrung mit digitaler Musikkomposition ist keine Voraussetzung. Gesucht werden explizit Anfänger:innen und Amateurkomponist:innen. Ausgeschlossen sind professionelle Komponist:innen.

Ich freue mich über jede Teilnahme!

Zur Terminvergabe für die Teilnahme folge dem Link oder scanne den QR-Code.



Ort: Tonlabor des Medien-campus der HAW Hamburg (Finkenau 35, 22081 Hamburg, Gebäude B, 3. Stock)

<https://terminplaner6.dfn.de/b/ec5dc66c163f79ace635228e592f82fa-458923>

Hannah Körber (Medientechnik, HAW) [REDACTED]



Übersetzungen

Skala	Variable	Englisch (aus Studie von Tchemeube u. a. 2023)	Deutsche Interpretation
SUS	SUS-01	I think that I would like to use this system frequently	Ich kann mir gut vorstellen, das Plugin regelmäßig zu nutzen.
	SUS-02	I found the system unnecessarily complex	Ich empfand das Plugin als unnötig komplex.
	SUS-03	I thought the system was easy to use	Ich empfand das Plugin als einfach zu nutzen.
	SUS-04	I think that I would need the support of a technical person to be able to use this system	Um das Plugin zu nutzen bräuchte ich technischen Support.
	SUS-05	I found the various functions in this system were well integrated	Ich fand, dass die verschiedenen Funktionen im Plugin gut integriert waren.
	SUS-06	I thought there was too much inconsistency in this system	Ich fand, dass es im Plugin zu viele Inkonsistenzen gab.
	SUS-07	I would imagine that most people would learn to use this system very quickly	Ich kann mir vorstellen, dass die meisten Leute schnell lernen, das Plugin zu benutzen.
	SUS-08	I found the system very awkward to use	Ich empfand die Bedienung des Plugins als sehr umständlich.
	SUS-09	I felt very confident using the system	Ich habe mich sehr sicher im Umgang mit dem Plugin gefühlt.
	SUS-10	I needed to learn a lot of things before I could get going with this system	Ich musste sehr viele Dinge lernen, bevor ich das Plugin benutzen konnte
Usability insgesamt	USA	Overall, I would rate the user-friendliness of this product as	Ich würde die Bedienfreundlichkeit dieses Produkts insgesamt wie folgt bewerten:
CSI	CSI-01	Exploration: It was easy for me to explore many different ideas, options, designs, or outcomes, using this plugin.	Erkundung: Es war einfach für mich, viele verschiedene Ideen, Möglichkeiten und Outcomes mit dem Plugin zu erkunden
	CSI-02	Engagement: I was very absorbed/engaged in this activity - I enjoyed it and would do it again.	Vereinnahmung: Ich war sehr eingenommen von dieser Aktivität - ich habe es genossen und würde es wieder tun.
	CSI-03	Expressiveness: I was able to be very creative while doing the activity inside this system or tool.	Ausdrucksstärke: Ich war in der Lage, während der Aktivität sehr ausdrucksstark und kreativ zu sein.
	CSI-04	Enjoyment: I enjoyed using the system or tool.	Spaß: Ich hatte Spaß daran das Plugin zu benutzen.
	CSI-05	Results Worth Effort: What I was able to produce was worth the effort I had to exert to produce it.	Ergebnisse lohnen Aufwand: Was ich erreichen konnte, war die Mühe wert, die ich dafür aufbringen musste.
Controllability	CON-01	Controllability 1: I find the system easy to control	Controllability 1: Ich empfinde das Plugin als leicht kontrollierbar.
	CON-02	Controllability 2: I would like more control over the system generation	Controllability 2: Ich wünsche mir mehr Kontrolle über den Output des Plugins.
	CON-03	Comment more on the controllability of the system	Kommentar zur Controllability des Systems:
TAM		Perceived Usefulness	Wahrgenommene Nützlichkeit des Systems
	TAM-01	Using the plugin would enable me to accomplish composition tasks more quickly	Der Einsatz des Plugins ermöglicht es mir, Kompositionen schneller zu erstellen.
	TAM-02	Using the plugin would improve my performance when composing	Der Einsatz des Plugins verbessert meine Arbeitsleistung beim Komponieren.
	TAM-03	Using the plugin in my job would increase my productivity when composing	Der Einsatz des Plugins verbessert meine Produktivität beim Komponieren.
	TAM-04	Using the plugin would enhance my effectiveness when composing	Der Einsatz des Plugins verbessert meine Effektivität beim Komponieren.
	TAM-05	Using the plugin would make it easier to create compositions	Der Einsatz des Plugins macht es mir leichter Kompositionen anzufertigen.
	TAM-06	I would find the plugin useful when composing	Ich finde den Einsatz des Plugins nützlich beim Komponieren.
		Perceived Ease of Use	Wahrgenommene Leichtigkeit der Systemnutzung
	TAM-07	Learning to operate the plugin would be easy for me	Es ist leicht für mich den Umgang mit dem Plugin zu lernen.
	TAM-08	I find it easy to get the plugin to do what I want to do	Es ist leicht für mich das Plugin das machen zu lassen, was ich möchte.
	TAM-09	My interaction with the plugin would be clear and understandable	Ich empfinde die Interaktion mit dem Plugin als klar und verständlich.
	TAM-10	I would find the plugin flexible to interact with	Ich empfinde die Interaktion mit dem Plugin als flexibel.
	TAM-11	It would be easy for me to become skillful at using the plugin	Es ist leicht für mich gute Kompetenzen im Umgang mit dem Plugin zu entwickeln.
	TAM-12	I would find the plugin easy to use	Ich schätze das Plugin als einfach zu nutzen ein.
	TAM-13	Assuming TOOL would be available, I predict that I will use it on a regular basis in the future.	Wie wahrscheinlich wäre es, dass Sie das Plugin in Zukunft benutzen?
	TAM-14	Provide additional feedback if necessary	Geben Sie bei Bedarf weiteres Feedback:

Skala	Variable	Englisch (aus Studie von Lewis 1991)	Deutsche Interpretation
After Scenario Questionnaire	ASQ-01	Overall, I am satisfied with the ease of completing the tasks in this scenario.	Insgesamt bin ich zufrieden mit dem Schwierigkeitsgrad der Task.
	ASQ-02	Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.	Insgesamt bin ich zufrieden mit der Zeit, die ich zum Erfüllen der Task benötigt habe.
	ASQ-03	Overall, I am satisfied with the support information when completing the tasks?	Insgesamt bin ich zufrieden mit den unterstützenden Informationen zum Erledigen der Task.

Checkliste

1. Davor

- Neuer Ordner YYYY-MM-DD-hh-Nr
- Fragebogenlinks überprüfen → Internetverbindung überprüfen
- Ausloggen aus Microsoft Forms, Chronik löschen
- SAM: Lizenzen aktiv?
- Setup prüfen: Bildschirm (Fensteranordnung), Maus, Tastatur
- Soundcheck
- Tutorialvideo öffnen
- Cubase-Version und MMM-Version checken
- MGP in Defaulteinstellungen zurückstellen
- Bluetooth und WLAN ausstellen

1. Einführung

- Herzlich Willkommen
- Beschreibung der Studie
 - Bachelorarbeit in Zusammenarbeit mit Steinberg und SFU
 - Usability eines Plugins untersuchen
 - Ungefähre Dauer: 60-90 Minuten
- Notieren: Fragen, ob TN für ergänzende Tests kontaktiert werden darf
- NDA unterzeichnen, ggf. Fragen klären
- Ablauf der Studie erklären
- Notieren: Welche Lizenz als Dankeschön? E-Mail-Adresse des Steinberg Accounts

2. Fragebogen zum Einstieg

- Ca. 3-5 Minuten
- Ziel: Demographie, bisherige DAW- und Kompositionskenntnisse erfassen
- Notieren: Teilnahmenummer bei Umfrage

3. Task

- 30 Minuten
- Setup zeigen: Laptop, Bildschirm, Maus, Tastatur, Kopfhörer
- Erwähnen, dass die anderen Geräte des Studios nicht angeschlossen sind
- Zeigen, wo man den Systemsound lauter und leiser stellt
- Hinweisen, dass auf der Tastatur Cubase Shortcuts zu sehen sind
- Zeigen, wo das Video zu finden ist (Quicktime, Desktop)

- Genre wählen lassen
- Entsprechendes Projekt in den TN-Ordner legen und öffnen

- Aufbau des Projekts zeigen: Drums-Spur, Bass, Lead Instrument, Rhythm Instrument, Effekte
- Zeigen, wo das Midi Generator Panel zu finden ist
- Zeigen, wie „Neue Version speichern“ funktioniert
- Ziel: Erkundung des Plugins. Die TN können eine neue Version speichern und von vorne beginnen.

4. Fragebogen nach der Task

- Ca. 10 Minuten
- Ziel: Das Nutzungserlebnis bewerten
- Notieren: Teilnahmenummer bei Umfrage

5. Danach

- Projekt speichern
- Mixdowns erstellen
- TN-Ordner Backup erstellen
- TN in Tabelle eintragen
- Tonlabor Kabel zurückstecken

SUS-Ergebnisse

TN ID	Ich kann mir gut vorstellen, das Plugin regelmäßig zu nutzen.	Ich empfand das Plugin als unnötig komplex.*	Ich empfand das Plugin als einfach zu nutzen.	Um das Plugin zu nutzen bräuchte ich technischen Support.*	Ich fand, dass die verschiedenen Funktionen im Plugin gut integriert waren.	Ich fand, dass es im Plugin zu viele Inkonsistenzen gab.*	Ich kann mir vorstellen, dass die meisten Leute schnell lernen, das Plugin zu benutzen.	Ich empfand die Bedienung des Plugins als sehr umständlich.*	Ich habe mich sehr sicher im Umgang mit dem Plugin gefühlt.	Ich musste sehr viele Dinge lernen, bevor ich das Plugin benutzen konnte.*
6	3	1	4	1	2	5	4	2	3	2
7	1	2	2	3	3	5	1	4	1	3
8	2	2	4	2	4	4	4	2	2	1
9	4	1	5	1	5	1	4	1	4	1
10	4	2	4	3	5	2	4	1	3	2
11	2	1	4	1	4	5	5	1	5	1
12	3	1	4	1	4	3	5	1	4	1
13	4	2	5	1	4	2	4	2	4	1
14	4	3	3	2	4	5	5	3	3	3
15	4	2	4	1	2	3	4	2	3	2
16	5	1	4	3	4	1	4	2	3	3
17	3	1	4	1	3	3	5	1	5	1
18	4	1	5	1	4	3	4	2	4	1
19	3	1	4	1	4	3	5	1	3	1
20	2	1	5	1	2	5	3	1	3	1
21	5	2	4	1	4	3	5	4	5	2
22	4	2	4	1	4	3	4	1	4	1
23	4	1	5	1	3	2	5	1	5	4
24	5	3	4	1	4	3	4	2	4	1

* Im Fragebogen wechseln sich positive und negative Aussagen ab. Die Antworten auf die mit Stern markierten Fragen werden vor dem nächsten Schritt für die Punktvergabe daher invertiert. Für die mit Stern markierten Aussagen wird dann also eine 5 zur 1, eine 4 zur 2, usw. sodass positives Feedback zum Plugin eine hohe Punktzahl bekommt und negatives eine niedrige. Anhand der Farbcodierung lässt sich die Bedeutung der Antworten für den SUS-Score erkennen.

Umrechnung in Punkte

TN ID	Ich kann mir gut vorstellen, das Plugin regelmäßig zu nutzen.	Ich empfand das Plugin als unnötig komplex.*	Ich empfand das Plugin als einfach zu nutzen.	Um das Plugin zu nutzen bräuchte ich technischen Support.*	Ich fand, dass die verschiedenen Funktionen im Plugin gut integriert	Ich fand, dass es im Plugin zu viele Inkonsistenzen gab.*	Ich kann mir vorstellen, dass die meisten Leute schnell lernen, das Plugin zu	Ich empfand die Bedienung des Plugins als sehr umständlich.*	Ich habe mich sehr sicher im Umgang mit dem Plugin gefühlt.	Ich musste sehr viele Dinge lernen, bevor ich das Plugin benutzen konnte.*	Summe	SUS Score = Summe *2,5
6	2	4	3	4	1	0	3	3	2	3	25	62,5
7	0	3	1	2	2	0	0	1	0	2	11	27,5
8	1	3	3	3	3	1	3	3	1	4	25	62,5
9	3	4	4	4	4	4	3	4	3	4	37	92,5
10	3	3	3	2	4	3	3	4	2	3	30	75
11	1	4	3	4	3	0	4	4	4	4	31	77,5
12	2	4	3	4	3	2	4	4	3	4	33	82,5
13	3	3	4	4	3	3	3	3	3	4	33	82,5
14	3	2	2	3	3	0	4	2	2	2	23	57,5
15	3	3	3	4	1	2	3	3	2	3	27	67,5
16	4	4	3	2	3	4	3	3	2	2	30	75
17	2	4	3	4	2	2	4	4	4	4	33	82,5
18	3	4	4	4	3	2	3	3	3	4	33	82,5
19	2	4	3	4	3	2	4	4	2	4	32	80
20	1	4	4	4	1	0	2	4	2	4	26	65
21	4	3	3	4	3	2	4	1	4	3	31	77,5
22	3	3	3	4	3	2	3	4	3	4	32	80
23	3	4	4	4	2	3	4	4	4	1	33	82,5
24	4	2	3	4	3	2	3	3	3	4	31	77,5

TN ID	Ich würde die Bedienfreundlichkeit dieses Produkts insgesamt wie folgt bewerten:
6	5
7	3
8	5
9	6
10	5
11	7
12	4
13	6
14	5
15	5
16	6
17	6
18	6
19	6
20	7
21	6
22	5
23	6
24	6

CSI- und Controllabilityergebnisse

CSI:

TN ID	Erkundung: Es war einfach für mich, viele verschiedene Ideen, Möglichkeiten und Outcomes mit dem Plugin zu erkunden	Vereinnahmung: Ich war sehr eingenommen von dieser Aktivität - ich habe es genossen und würde es wieder tun.	Ausdruckskraft: Ich war in der Lage, während der Aktivität sehr ausdrucksstark und kreativ zu sein.	Spaß: Ich hatte Spaß daran das Plugin zu benutzen.	Ergebnisse lohnen Aufwand: Was ich erreichen konnte, war die Mühe wert, die ich dafür aufbringen musste.
6	4	3	4	5	4
7	1	3	2	5	3
8	3	3	2	4	4
9	5	4	4	5	4
10	5	5	4	5	5
11	5	2	2	4	1
12	5	4	2	4	5
13	4	5	3	5	3
14	5	4	3	5	2
15	3	3	2	4	3
16	4	5	4	5	5
17	2	4	2	4	3
18	4	5	2	5	3
19	4	5	4	5	5
20	2	2	1	3	2
21	3	5	2	4	4
22	4	4	4	4	5
23	4	4	3	5	5
24	4	4	3	4	5

Controllability:

TN ID	CON-01: Ich empfinde das Plugin als leicht kontrollierbar.	CON-02: Ich wünsche mir mehr Kontrolle über den Output des Plugins.
6	3	7
7	1	10
8	6	8
9	7	5
10	6	5
11	4	8
12	4	8
13	6	9
14	4	9
15	3	10
16	9	5
17	8	10
18	8	7
19	6	10
20	4	10
21	6	10
22	8	7
23	4	8
24	6	8

TAM-Ergebnisse

Wahrgenommene Nützlichkeit des Systems

TN ID	Der Einsatz des Plugins ermöglicht es mir, Kompositionen schneller zu erstellen.	Der Einsatz des Plugins verbessert meine Arbeitsleistung beim Komponieren.	Der Einsatz des Plugins verbessert meine Produktivität beim Komponieren.	Der Einsatz des Plugins verbessert meine Effektivität beim Komponieren.	Der Einsatz des Plugins macht es mir leichter Kompositionen anzufertigen.	Ich finde den Einsatz des Plugins nützlich beim Komponieren.
6	4	2	2	3	5	3
7	2	1	1	1	1	1
8	4	3	4	3	4	2
9	4	4	4	3	4	5
10	5	5	3	4	5	4
11	1	1	1	1	1	1
12	3	3	4	3	3	4
13	4	4	3	3	5	4
14	3	3	3	2	4	2
15	4	3	4	3	4	4
16	5	5	5	5	5	5
17	4	2	2	3	4	1
18	2	4	2	4	4	4
19	4	5	4	4	4	4
20	2	2	3	3	3	4
21	4	3	4	4	4	4
22	4	3	4	3	4	4
23	5	4	4	2	4	4
24	3	3	3	2	4	4

Wahrgenommene Leichtigkeit der Systemnutzung

TN ID	Es ist leicht für mich den Umgang mit dem Plugin zu lernen.	Es ist leicht für mich das Plugin das machen zu lassen, was ich möchte.	Ich empfinde die Interaktion mit dem Plugin als klar und verständlich.	Ich empfinde die Interaktion mit dem Plugin als flexibel.	Es ist leicht für mich gute Kompetenzen im Umgang mit dem Plugin zu entwickeln.	Ich schätze das Plugin als einfach zu nutzen ein.
6	5	2	4	3	4	5
7	3	1	2	2	2	3
8	4	2	3	3	3	4
9	5	3	5	5	5	4
10	4	4	4	4	5	4
11	5	2	5	3	3	2
12	4	2	4	3	3	5
13	5	3	4	4	3	5
14	5	1	4	3	3	4
15	4	2	3	3	3	3
16	5	3	5	4	5	5
17	5	2	5	3	4	5
18	5	3	4	4	5	4
19	5	3	4	3	5	5
20	5	2	4	3	3	5
21	5	3	4	3	4	4
22	5	4	4	4	4	4
23	5	3	4	2	3	5
24	4	3	5	2	4	5

Zukünftige Nutzung

TN ID	Wie wahrscheinlich wäre es, dass Sie das Plugin in Zukunft benutzen?
6	4
7	2
8	2
9	5
10	7
11	2
12	4
13	6
14	2
15	4
16	6
17	2
18	4
19	5
20	2
21	6
22	6
23	6
24	6

ASQ-Ergebnisse

TN ID	Insgesamt bin ich zufrieden mit dem Schwierigkeitsgrad der Task.	Insgesamt bin ich zufrieden mit der Zeit, die ich zum Erfüllen der Task benötige habe.	Insgesamt bin ich zufrieden mit den unterstützenden Informationen zum Erledigen der Task.
6	4	5	5
7	4	4	5
8	4	4	4
9	5	5	5
10	5	5	5
11	5	5	5
12	5	5	5
13	5	5	5
14	4	4	5
15	5	5	5
16	5	5	5
17	4	4	4
18	4	4	5
19	4	2	5
20	5	5	5
21	4	4	4
22	5	5	5
23	5	5	4
24	4	3	5

Qualitatives Feedback

TN ID	Qualitatives Feedback (schriftlich) [Controllability] Prompt: "Kommentar zur Controllability des Systems:"	Interpretation
6	Die Funktion Density fiel mir als besonders willkürlich auf, die Skala low-middle-high war nicht wirklich intuitiv	Controllability: Density willkürlich Kommentar: Density Skala "low, middle, high" nicht intuitiv
7		
8	Es fehlen Parameter wie Tonhöhe und Bezug zu vorherigen Generationen. (z.B. so ähnlich, wie die letzte Generation nur schneller) Ansonsten fehlte mir eine Option nur Teile der Spur zu generieren. (z.B. nur einen Takt))	Wunsch: Parameter für Tonart Wunsch: Bezug zur letzten Generierung Wunsch: Generieren für kleinere Zeitabschnitte, Inpainting
9		
10	Die Crazyness funktion wirkte auf mich nicht, als würde ich da viel kontrollieren, bei den meisten Generierungen.	Controllability: Crazyness willkürlich
11	Insbesondere im monophonen Modus wurden auch immer wieder Töne übereinander gelegt und das Ergebnis somit nicht mehr rein monophon. Trotzdem waren die Ergebnisse für mich meist schlüssiger und spannender als der polyphone Modus. Wenn man im Monophonen Modus noch einen Overlap-Regler hätte, um zu bestimmen, wie stark Töne sich überschneiden dürfen, wäre das cool.	Controllability: monophon/polyphon Kommentar: Generierungen im monophonen Modus schlüssiger und spannender Wunsch: Overlap-Regler für Monophonie
12	Eine Einstellung für die verwendete "Scale" wäre hilfreich	Wunsch: Parameter für Tonart (scale)
13	Ich bin mir nicht sicher, ob die gewählten Parameter z.B. Tonlänge immer korrekt umgesetzt wurden.	Controllability: Duration und andere Parameter willkürlich
14	Den Sound den ich erreichen wollte konnte ich wenig kontrollieren. Nach erneuter Generierung mit gleichen Parametern war der Sound ein ganz anderer.	Controllability: allgemein (alle Parameter) Wunsch: Generierung mit gleichen Parametern führt zu Variation statt ganz neuer Version.
15	Es wäre gut dem Plugin sagen zu können ob es mehr eine Fläche, Melodie etc. erzeugen soll und nicht anhand der länge der Noten zu sagen wie die Melodie erstellt werden soll.	Wunsch: Parameter für Art der Generierung (z.B. Fläche/Pattern/Melodie) statt Duration
16	Mit gleichen Einstellungen vielfältige mögliche Ergebnisse. Ich kann mir vorstellen, dass es zu wenig Optionen gibt, wenn man speziellere Vorstellungen hat.	Kommentar: Vielfältige Ergebnisse trotz gleicher Einstellung (Lob) Wunsch: mehr Parameter für spezifische Vorstellungen
17	Genre-Auswahl wirkt nicht effektiv	Controllability: Genre willkürlich
18	Ich hätte mir gewünscht es gäbe die Möglichkeit präziser einen gewissen [Teil] der Spur zu ersetzen	Wunsch: Generieren für kleinere Zeitabschnitte, Inpainting
19	Tonleiter manuell einstellen wär cool oder eine Option für die Emotion, so etwas wie: fröhlich/traurig/düster...	Wunsch: Parameter für Tonart , Parameter für Emotion
20	Mehr Steuerparametern wie z.B. Pausen oder Tonart wären gut.	Wunsch: Parameter für Tonart , Parameter für Pausen
21	Nicht nur eine Add to Selection, sondern vielleicht Reduce from Selection?	Wunsch: Option Reduce from Selection
22	Wenn man einen Input liefern könnte, an dem sich der Output orientiert, wäre das sicher eine coole Sache (wie bei KI-generierten Bildern ein Bild hochladen, an dem sich der Content orientiert).	Wunsch: Referenzmusik "füttern"
23	Tonarten auszuwählen wäre noch super. War sonst so sehr vereinzelt und chaotisch (was natürlich auch gut sein kann)	Wunsch: Parameter für Tonart Controllability: Generierte Noten sehr vereinzelt und chaotisch (Density)
24		

TN ID	Qualitatives Feedback (schriftlich) [Technology Acceptance] Prompt: "Geben Sie gegebenenfalls zusätzliches Feedback"	Interpretation
6	einige Genres ergeben mit bestimmten Settings unbenutzbare, seltsame Ergebnisse, sodass ich ein anderes Zusammenspiel zwischen Genre und den restlichen Einstellungen begrüßen würde	<u>Controllability</u> : best. Kombinationen von Parametereinstellungen und Genrewahl ergeben schlechte Ergebnisse Wunsch : Genres trennen von anderen Parametern / anders lösen
7		
8		
9		
10		
11	Ich würde es wahrscheinlich zum Spaß mal ausprobieren. Diese halbe Stunde Nutzung hat mich allerdings nicht so sehr begeistert, dass ich es unbedingt noch einmal nutzen möchte.	<u>Kommentar</u> : Würde Plugin wahrscheinlich noch mal nutzen, war während der Task enttäuscht.
12		
13	Einige Melodievorschläge waren m.E. zu gradlinig, d.h. Betonung und Noten zu sehr auf der 1. Note. Synkopen wären abwechslungsreicher.	<u>Kommentar</u> : Generierte Melodien teilweise zu gradlinig Wunsch : Mehr Synkopen, mehr rhythmische Abwechslung
14		
15	Bisher ist es ein guter Anfang um, eine Idee zu bekommen, man kann sich damit quasi "vorschläge" Generieren, die man dann bearbeitet. Ich finde es noch ein wenig zu undurchsichtig was das plugin da generiert, man wählt irgendwas aus und dann macht es irgendwas. Besser wären mehr einstellungen um ihm mehr zu sagen wo es hin gehen soll.	<u>Kommentar</u> : Plugin-Nutzung für Inspiration/als Startpunkt. <u>Controllability</u> : Zu hohe Willkür. Wunsch : mehr Parameter
16	Ich habe zwar wenig Vorerfahrung mit DAWs und Musikkomposition, aber ich hätte Lust, mit dem Plugin mehr auszuprobieren.	<u>Kommentar</u> : Möchte mehr mit dem Plugin ausprobieren
17		
18		
19		
20		
21	Noch ein paar Parameter mehr wären Interessant	Wunsch : Mehr Parameter
22		
23	Gut als Startpunkt oder Ideengenerierung, aber danach musste noch viel geschraubt werden	<u>Kommentar</u> : Plugin-Nutzung für Inspiration/als Startpunkt. <u>Controllability</u> : Viel Nachjustieren nötig.
24		

Inhalt des beigelegten Datenträgers

1. Digitale Fassung der Bachelorarbeit
2. Tutorialvideo für MGP im vorgegebenen Szenario
3. Cubase-Templates für das vorgegebene Szenario
4. Mixdowns der durch die Teilnehmenden erstellten Musik

Eigenständigkeitserklärung

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

KI-unterstützte Musikkomposition im Test – Eine empirische Untersuchung zu Usability und Akzeptanz der Multi-Track Music Machine in Cubase

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.

Hamburg, 22.01.2024

Datum

Unterschrift