Bachelorarbeit im Studiengang Audiovisuelle Medien

Sounddesign für virtuelle Räume in Wwise am Beispiel "Nanna"

vorgelegt von Elisabeth Drobny
an der Hochschule der Medien Stuttgart
am 28. Februar 2021
zur Erlangung des akademischen Grades eines Bachelor of Engineering.

Erstprüfer: Prof. Oliver Curdt
Zweitprüfer: Prof. Dr. Johannes Schaugg

Hochschule der Medien Stuttgart
Februar 2021

Ehrenwörtliche Erklärung

Hiermit versichere ich, Elisabeth Drobny, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Sounddesign für virtuelle Räume in Wwise am Beispiel “Nanna”“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden. Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Stuttgart, den 28. Februar 2021

Kurzfassung

Der interaktive Inhalt einer Virtual-Reality-Anwendung erfordert bestimmte Vorgehensweisen, die sich vom Sounddesign in linearen Medien unterscheiden. Mithilfe einer Middleware wie Wwise, lassen sich diese Methoden implementieren und auf Anforderungen des jeweiligen Projekts anpassen. Virtual Reality ist eine simulierte Umgebung, die das Ziel hat, bei Nutzer*innen eine Immersion zu erzeugen. Dies geschieht über das Ansprechen der verschiedenen Sinne. Ein wichtiger Faktor für die Immersion ist das Sounddesign. Der gehörte Raum wird durch Wahrnehmungseffekte, wie die interaurale Zeitdifferenz und kontextuelle Informationen gebildet.
Die Infrarot-Technik eines VR-Headsets und der restlichen Raum-Sensorik ermöglicht es, Bewegungen und Handlungen in Echtzeit zu registrieren und zu verarbeiten, um ein Feedback auf die Interaktion in Echtzeit zu erzeugen.
Bei der Vorbereitung eines Projekts werden aufgrund der Ideen die Eigenschaften festgelegt und das Level der Interaktionsmöglichkeiten bestimmt. Da virtuelle Anwendungen üblicherweise auf Basis einer Game Engine programmiert werden, existiert zu Beginn kein Originalton, der verwendet werden kann. Das Sounddesign kann auf Basis von Foleyaufnahmen gestaltet werden.
Wwise lässt sich in die Game Engine Unity implementieren. Die erstellten Ordnerstrukturen und Eigenschaften der Audio-Clips werden mithilfe von Soundbanks aus der Middleware importiert. Alle Elemente der Soundbanks können in der Game Engine an bestimmte Objekte gebunden werden.
Die Mischung des fertigen Sounddesigns erfolgt im Team. Zu zweit kann der Aufgabenbereich so aufgeteilt werden, dass eine Person innerhalb der Anwendung Höreindrücke überprüft, während die andere Person die vorgeschlagenen Änderungen implementiert.
Die vorliegende Arbeit vermittelt ein Grundverständnis für die Arbeit in Virtual Reality aus der Perspektive der Tongestaltung. Die Durchführung wird anhand eines Projekts beispielhaft erläutert. Bei der Arbeit mit einem interaktiven Medium spielen nicht-lineare Abläufe eine große Rolle. Die Verfahren zur Konzeption müssen der Interaktivität angepasst werden.

Abstract

The interactive content of a virtual reality application requires specific sounddesign methods, that are different from methods in linear media. Using a middleware, like Wwise, enables the implementation of these methods. Wwise enables to adjust the sound for different projects respectively.

Virtual Reality is a simulated environment with the goal of creating immersion for its users. This effect triggers the different senses. A central point of immersion is the sounddesign. The auditory room is shaped with perceptual effects, such as interaural time difference or contextual evidence.

A VR-Headset and the interconnected room sensors are able to register movement and actions in real time through infrared light. This produces real time feedback for all interactions.

The features of a project are based on the ideas, that also determine the possibilities of interaction in preparation for a project. As virtual applications are usually developed in a game engine, there is no sound from location recordings available. This enables the use of foley recordings in the developement of the sounddesign.

Wwise can be implemented in the game engine Unity. The audio clips, organized and structured in file paths with defined properties, are imported from the middleware through soundbanks. Every soundbank element can be linked to a specific game object inside the game engine.

Mixing the final sounddesign is team work. A team of two allows to separate tasks and enable one person to review auditory implementations. The other person is able to change settings according to the perception of the person wearing the VR-Headset. This thesis conveys a basic understanding of working with virtual reality from the perspective of sounddesign. The sound implementation is explained using a specific examplary project. Non linear processes play a big role in working with interactive media. The used methods to develop a concept have to be adapted to the goal of being interactive.

Inhaltsverzeichnis

1 Einführung **5**

1.1 Aufbau der Arbeit 6

1.2 Studioproduktion NANNA 7

2 Grundlagen Virtual Reality **11**

2.1 Was ist Virtual Reality? 11

2.2 Häufig verwendete Begriffe in der Fachliteratur 13

2.3 Sensorik 14

2.3.1 Raumwahrnehmung 15

2.3.2 Individuelles Hören 16

2.4 Technik 17

3 Durchführung der Tongestaltung für das Projekt **19**

3.1 Sounddesign 19

3.2 Konzeption des Projekts 19

3.2.1 Ideen und Entwicklung 20

3.2.2 Vergleich mit linearen Medien 23

3.3 Vorbereitungen 26

3.3.1 Eigenschaften und Verwendungszwecke einer Anwendung 26

3.3.2 Layout des virtuellen Raums 28

3.3.3 Informationsaustausch mit den Programmierer*innen 28

3.3.4 Welche Aufnahmen werden benötigt 29

3.4 Auswahl der Middleware 30

3.5 Foleyaufnahmen 31

3.5.1 Beispiele 31

3.6 Aufbereitung der Soundfiles für Verwendung in Wwise 32

3.7 Skizzierung der Atmo 33

4 Implementierung des Sounddesigns in Wwise **34**

4.1 Allgemeine Funktionen der Software 34

4.2 Verbindung mit Unity 35

4.2.1 Programmierseitig 35

4.2.2 Integrierung eines bestehenden Unity-Projekts mit Wwise 36

4.3 Arbeits-Strukturen anlegen 37

4.3.1 Aufbau einer Session 38

4.3.2 Leistung und CPU-Auslastung 42

4.3.3 Soundbanks 43

4.3.4 Events 44

4.4 Spezielle Gestaltungsmöglichkeiten 45

4.4.1 Orientierung durch Geräusche 45

4.4.2 Atmo 46
4.4.3 Raumbezüge herstellen 47
4.4.4 Interaktionsgeräusche 48
4.5 Mischung 49

5 Ausblick **55**
5.1 Was hat nicht geklappt 55
5.2 Mögliche Erweiterungen in zukünftigen Anwendungen 56

Literaturverzeichnis **59**

Abbildungsverzeichnis **61**

1 Einführung

Durch die stetige Weiterentwicklung und das steigende Interesse an der Technik, die nicht mehr nur Forschungsinstitutionen einen Zugang zu einer simulierten, virtuellen Realität (VR) ermöglicht, eröffnen sich auch neue Möglichkeiten in der Gestaltung. Eine Zunahme der Inhalte, die für dieses Medium produziert werden, steigert auch die Anforderungen an die Qualität und Optionen der Verfahren, die im Sounddesign angewendet werden.

Die Platzierung einzelner Geräusche in einer räumlichen Wiedergabe kann durch Panning in Stereo- oder Mehrkanalsystemen erfolgen. Für die interaktive und reaktive Nutzung und Wiedergabe im virtuellen Raum kann eine objektorientierte Platzierung verwendet werden. Diese ist nicht kanalgebunden und jede Geräuschquelle hat eine Position im virtuellen Raum, die unabhängig von der Kopfbewegung bestehen bleibt.

Eine Software wie Wwise ermöglicht eine Verarbeitung dieser Sound-Objekte, sowie ihrer Platzierung, Eigenschaften und unter welchen Umständen sie zu hören sind.
Die Erläuterung zum Sounddesign in der folgenden Arbeit bezieht sich neben der Tongestaltung, der zu hörenden Elemente, auch auf deren Einfluss auf ihre Umgebung, sowie die Aktion oder den Umstand, der zum Abspielen des Clips führt. Die Tongestaltung im virtuellen Raum ist nicht zeitgebunden und eine vorhersehbare Abstimmung zwischen den Gestaltungselementen ist nicht möglich. Anhand des Beispiels wird eine Gestaltung im interaktiven Format erläutert und die Planung und Vorgehensweise nicht linearer Medien dargestellt.
Das Projekt Nanna, das als Beispiel herangezogen wird, dient der Verdeutlichung allgemeiner Erklärungen und liefert Beispiele von möglichen Vorgehensweisen.

Das Ziel der Projekt-Arbeit im virtuellen Raum wird passend beschrieben von Schütze:

> As the audience decides where they will place their focus, the world responds to create a unique perspective based on how the audience chooses to interact and position themselves in the world (Schütze & Irwin-Schütze, 2018).

Denn bei all der technischen Arbeit geht es letztendlich um die Perspektive der einzelnen Person, die mit der Anwendung selbstbestimmt interagieren kann und sich dadurch in dieser Welt positioniert.
Der Fokus dieser Arbeit liegt auf der Durchführung eines VR-Projekts, doch die freie Wahl der Nutzer*innen wurde in jede Entscheidung eingebunden und bestimmte die Umsetzung der kreativen und der technischen Möglichkeiten.

Eine Anforderung, die an die vorliegende Arbeit gestellte wurde, war es, die Schritte des Projekts soweit zu dokumentieren, dass man mithilfe dieser “Anleitung” ein vergleichbares Projekt durchführen kann. Hierbei wurde davon ausgegangen, dass einzelne Knöpfe und Einstellungen selbst gefunden werden müssen, aber diese detailliert benannt werden, um einzelne Schritte möglichst nachvollziehbar zu machen.

Außerdem werden alle Schritte und damit verbundene Entscheidungen soweit erklärt, dass diese verständlich Nachzuvollziehen sind und Hintergründe klar werden. Gleichzeitig wurde darauf geachtet, dass dies den Umfang der Thesis nicht sprengt.

Im Anhang sind Links zu der Software-Dokumentation der Entwicklerfirma (*Wwise Dokumentation*, 2021). In dieser Dokumentation finden sich Informationen, die maßgeblich zum Projekt beigetragen haben. In den folgenden Kapiteln werden sie nicht wörtlich zitiert.
Da seit Beginn des Projekts mit der englischen Original-Version der Software gearbeitet wurde, werden im folgenden Text die spezifischen englischen Begriffe verwendet, wie sie auch in der Software zu finden sind. Audiospezifische Begriffe, werden soweit möglich übersetzt oder um eine Erklärung ergänzt.

1.1 Aufbau der Arbeit

Das erste Kapitel wird für die Leser*innen die gleiche Grundlage schaffen, die zu Beginn des Projektes Nanna erarbeitet werden musste. Der Aufbau und das Anwendungsprinzip des Projekts wird im folgenden Abschnitt vorgestellt. Die Rahmenbedingungen und die Ideenfindung werden hier beschrieben.
Darauf folgt eine Einführung in Virtual Reality als Medium, es wird das grundsätzliche Funktionsprinzip vorgestellt.
Es werden häufig auftretende Begriffe in der Literatur rund um Virtual Reality erläutert. Hier wird die dahinterstehende Theorie der Wahrnehmung behandelt.
Da sich der weitere Teil der Arbeit vorrangig auf die Durchführung eines Projekts bezieht, schaffen die Hintergründe und Begrifflichkeiten Verständnis für getroffene Entscheidungen im Workflow.
Virtual Reality benötigt einen sehr spezifischen Aufbau und spezielle Technik. Diese wird am Beispiel der HTC Vive Pro erklärt. Im Unterschied zu vielen anderen Medien, die auf verschiedenen Geräten konsumiert werden können, sind VR-Anwendungen größtenteils plattformgebunden.

Diese Arbeit beleuchtet den Teil der Tongestaltung, der im Rahmen einer Studioproduktion durchführbar ist. Dieser Rahmen bezieht sich auf einen Zeitrahmen von einem Semester und kleinem Team ohne spezifische Erfahrung bezüglich Virtual Reality.Zusätzlich wird aufgezeigt, welchem Zweck die einzelnen Schritte dienten und welche Möglichkeiten zur Erweiterung denkbar sind.
Der Sound-Workflow wird spezifischer beschrieben. Von der Konzeption eines kompletten Projekts chronologisch bis zum Sounddesign in der fertigen Anwendung, mit besonderem Schwerpunkt auf den Verwendungsmöglichkeiten der genutzten Software und den Eigenheiten der Tongestaltung in einem virtuellen Raum.

1.2 Studioproduktion NANNA

Das Projekt mit dem Namen “Nanna”, das im Rahmen der Studioproduktion 2018 durchgeführt wurde, hatte sich zum Ziel gesetzt, eine virtuelle Umgebung zu schaffen, die Nutzer*innen erforschen können und im Zuge dessen mit der eigenen Wahrnehmung auseinandersetzen können. Aufgeteilt in drei Räume, die durch Teleportationsplatten verbunden wurden, konnten sich die Nutzer*innen frei bewegen, ohne von einer Narration geleitet zu werden.
Eine Studioproduktion umfasst die Arbeit im Team über ein Semester hinweg, bei dem ein gemeinsames Projekt erstellt und präsentiert wird.
Die Anwendung lief auf der Game Engine Unity mit HTC Vive Pro Headset, steuerbar mit Handheld Controllern. Die im Headset eingebauten Lautsprecher wurden durch separate Kopfhörer ersetzt. Durch künstliche Reize wird die Wahrnehmung der Nutzenden getestet, ausgereizt und kann in eigener Geschwindigkeit erforscht werden.

Das Team bestand aus fünf Leuten, die alle in erster Linie mit Virtual Reality arbeiten wollten. Das gemeinsame Interesse an der Art des Projektes führte die Teammitglieder zusammen. Daher entschieden sie sich für ein Projekt im Bereich der Interaktiven Medien. Das Projekt wurde gemeinsam entworfen. Jedes Teammitglied konnte sich für die Arbeit in einem Departement entscheiden. Basierend auf eigenen Interessen und bisheriger Erfahrung wurde der mögliche Aufwand für die jeweiligen Aufgaben grob umrissen.
Es gab zwei Programmierer, die vor allem auf Unity basierte Shader und Funktionen programmieren wollten. Ein weiterer Programmierer, legte seinen Schwerpunkt auf Grafik und Raumgestaltung.
Zwei Teammitglieder entschieden sich für das Departement Ton: neben der Autorin dieser Arbeit, die den Schwerpunkt auf Sounddesign und Integrierung der Geräusche über Wwise wählte und dem Teamkollegen, der sich auf prozeduralen Sound in einer grafischen Audiosoftware (Max MSP) konzentrierte.
Aus dieser Aufteilung ergab sich, dass weniger einzelne Räume zu erstellen waren, da diese auch alle eine visuelle Oberfläche benötigen würden.

Gestaltung der Anwendung

Die Anzahl der virtuellen Räume legten wurde auf drei festgelegt, um eine unterschiedliche Gewichtung auf diese legen zu können. Es ergaben sich drei unterschiedlich umfangreich gestaltete Räume, die als Teil einer gemeinsamen Ausstellung drei voneinander abgetrennte Umgebungen bilden.
Um keine gefühlte Unausgeglichenheit der Gewichtung der Räume zu erzeugen, da eine Balance zwischen zwei Räumen nicht erzielt werden konnte und um die entwickelten Ideen unterbringen zu können, wurde die oben genannte Anzahl gewählt.
So war es möglich, dass der Raum, in dem man beginnt, einen Grafik-Schwerpunkt mit Soundunterstützung hat, ein weiterer sich mehr auf Funktionalität und Interaktion bezieht und der letzte eine ausgewogene Kombination von Geräusch und Bild ergibt. Klanglich

gestaltet wurde von der Autorin der erste und der letztgenannte Raum, im weiteren als *Foyer* und *Basic Emotions*-Raum betitelt.

Das Foyer

Zum Anfang der Anwendung befindet man sich in einem Aufzug, der untermalt von einer Shepard-Skala und klassischem Aufzugs-Jazz, nach einer vordefinierten Zeit im Foyer ankommt. Beim Ankommen öffnet sich der Raum zu allen Seiten gleichzeitig und man kann sich frei im Foyer bewegen.

Abbildung 1: Foyer Raum in Nanna

In diesem Raum befinden sich verschiedene Objekte, mit denen interagiert werden kann. Schwebende Würfel, die einen leichten Sound emittieren und dazu anregen sollen, ihnen näher zu treten. Der Raum wird nicht durch Wände begrenzt, sondern existiert als Plattform in einer undefinierten Umgebung. Diese Plattform ist umgeben von vier großen Stelen. Diese viereckigen Säulen sind auf verschiedene Tonhöhen gestimmt und geben Geräuschflächen wieder. Je nach Position zwischen den vier Eckpunkten mischt sich eine verschiedene Ton-Kombination. Die Gestaltung des Foyers ist vorwiegend monochromatisch und ohne Schattierungen. Die Anforderung war es, Formen und Dimensionen durch Musterung und Tiefenstaffelung der sich überlagernden Objekte zu schaffen. Von der Hauptfläche des Foyers schwebten auf beiden Seiten kleinere Flächen. Diese heben sich farblich ab. Die Loslösung mit der Spalte zwischen den Platten sollte Nutzer*innen zur Nutzung des Teleports bringen. Die Funktion des Foyers war vor allem, als “tutorial” zu fungieren. Durch die simple Positionierung der Objekte und die direkte Reaktion auf Kollision mit der eigenen Position sollte ein Gefühl für die Interaktion geschaffen werden.

Rhythmus-Raum

Die Plattform auf der rechten Seite teleportierte Nutzer*innen in den *Rhythmus-Raum".

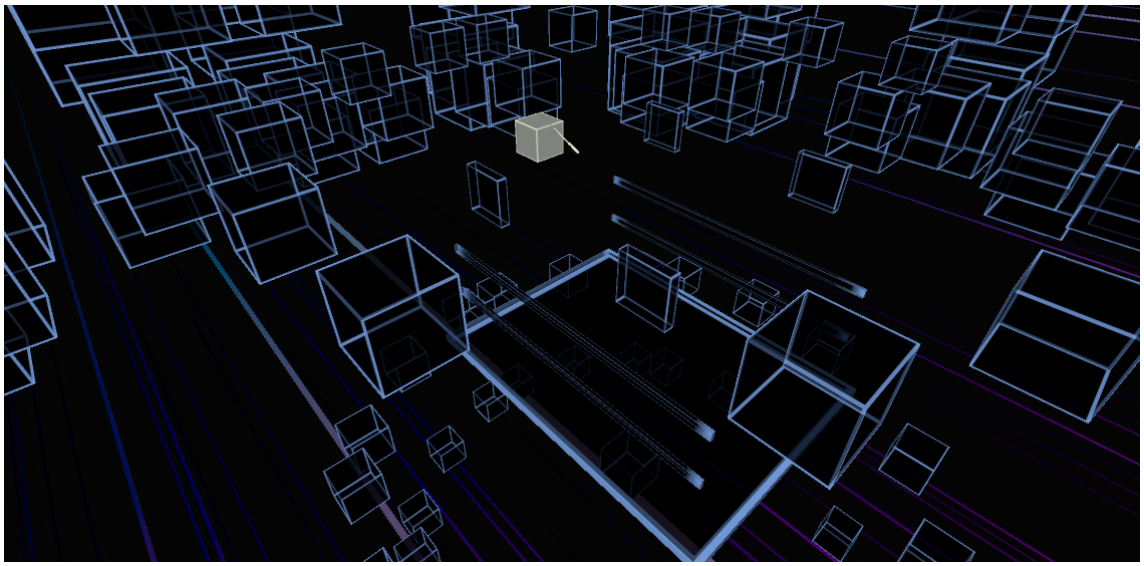

Abbildung 2: Rhythmus-Raum in Nanna

Auch hier wiederholt sich das Motiv der schwebenden Plattform, die die Bewegungsfreiheit definiert. Diese kleine Plattform ist durch den Raum steuerbar und wird umgeben von langen Lichtbahnen die eine Tiefe simulieren. Mit einem Dirigierstab lässt sich ein Rhythmus angeben. Stimmt der Rhythmus, der mit dem Stab angegeben wird, mit dem Pulsieren des Raumes überein, so erscheinen in Kreisen angeordnete Würfel, die auf dieselbe Rhythmik reagieren. Dieser Rhythmus wird von prozedural erzeugter Musik unterstützt, die mit jedem erscheinenden Kreis vielschichtiger werden.
Von einer separaten roten Plattform kann man sich wieder zurück ins Foyer teleportieren.

Zwischen den Räumen gibt es nur Verbindungen zum Foyer. Man sollte in diesem Raum anfangen und auch wieder dahin zurückkehren.

Basic-Emotions-Raum

Der Raum, der von der linken Plattform im Foyer aus erreichbar ist, hat seinen Arbeitstitel "Basic-Emotions" behalten. Dieser hat den Fokus auf kreuzmodale Korrelation verschiedener Sinneswahrnehmungen gelegt. Einige wie Nähe und Weite oder Größe sind in VR einfach nachzubilden. Dieser Raum war der Versuch, Sinneswahrnehmungen wie Wärme, Gewicht und Haptik annähernd zu simulieren.
Aufgeteilt in zwei Abschnitte, die mit Membranen von einem mittleren Durchgangsbereich getrennt wurden, sollen die Gegensätze der Effekte durch den gestalterischen Kontrast verdeutlicht werden.

Der gestalterische Kontrast fand sich in der Farbgebung und Oberflächenstruktur wieder, sowie in den räumlichen Dimensionen.
In dem Mittelteil wurden verschiedene Objekte platziert, mit denen verschiedene Interakti-

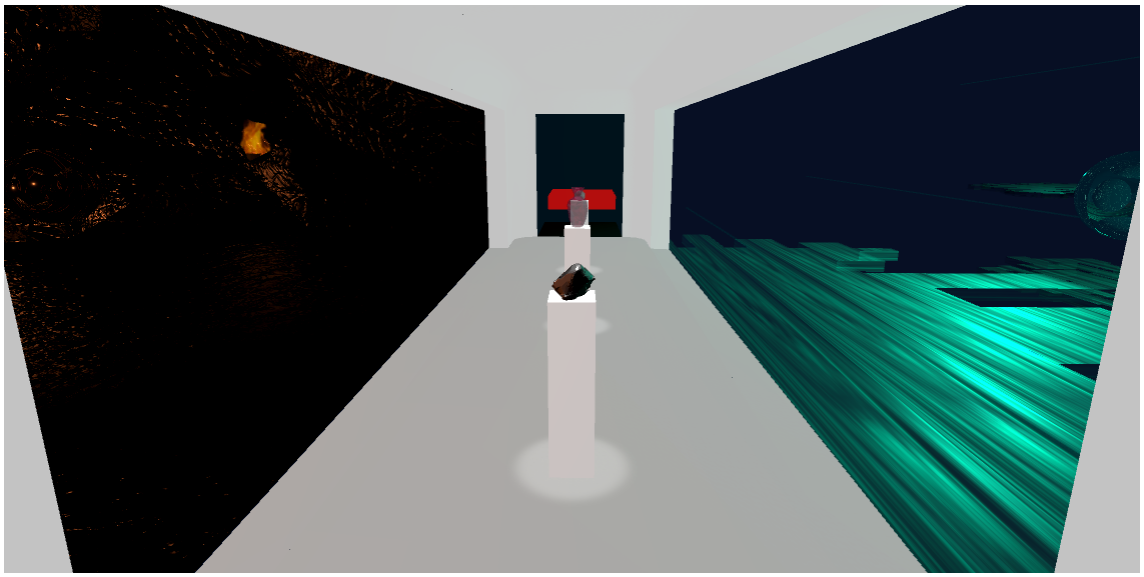

Abbildung 3: Basic-Emotions-Raum in Nanna

onsmöglichkeiten in den beiden angrenzenden Räumen bestehen.

In den beigelegten Dateien findet sich ein kurzer Rundgang in Video-Form. Hierbei ist zu beachten, dass dieses 2D-Video nur als grobe Repräsentation der Anwendung zu sehen ist. Es schafft allerdings einen visuellen Bezug, der durch die Bilder nicht möglich ist.
Der eine Teil ist nach dem Vorbild einer steinernen Höhle gestaltet, der andere ist durch die schwebenden Schollen definiert, die an Glas oder Eis erinnern sollen. In den beigelegten Dateien sind hierzu weitere detaillierte Bilder.

In der *"Konzeption des Projekts"* wird noch genauer ausgeführt, welche Ideen hinter der Raumgestaltung und dem Layout stehen.

2 Grundlagen Virtual Reality

Das Projekt wurde mit dem Fokus auf Virtual Reality erstellt. Die erste Entscheidung, definierte das Ziel, die Arbeit mit Virtual Reality kennenzulernen. Im vorgegebenen Rahmen sollten die Grenzen dieses Mediums getestet werden. Virtual Reality, oder VR, sowohl für das Projektteam, als auch in der Welt der Medien gleichzeitig altbekannt und gleichzeitig noch in seinen Kinderschuhen steckt.
Auf Basis der bestmöglichen Verwendung und Inszenierung von Virtual Reality wurden Entscheidungen getroffen. Zur besseren Nachvollziehbarkeit der Software-Auswahl, technischer Arbeit, sowie kreativen Entscheidungen, wird in diesem Kapitel eine Verständnisgrundlage geschaffen.

Wie sich in der Projektarbeit zeigte, war es für alle Departements wichtig, sich allgemein mit VR als Medium auseinandersetzen. Dies war nötig, um später die eigenen Aufgaben einordnen zu können. Dieser Chronologie wird gefolgt. Den größeren Teil der Arbeit nehmen Tongestaltung und die Arbeit mit der Software ein.

2.1 Was ist Virtual Reality?

Definition
Es gibt verschiedene Definitionen und Versuche, den Begriff der Virtual Reality einzugrenzen und abzugrenzen. Im Folgenden werden Definitionen dargestellt und diskutiert. Es wird ein Einblick gegeben wie VR verwendet wird und wo es angewandt werden kann.
Man kann Virtual Reality klar von konventioneller Computergraphik unterscheiden und so einer Definition näher kommen. Dabei ist einmal die multimodale Präsentation relevant, die mehrere Sinne anspricht und sich nicht auf eine ausschließlich visuelle Wiedergabe beschränkt.
Außerdem charakterisiert sich VR durch die zeitkritische Darstellung, die erfordert, das alle ausgegebene Informationen (visuelle und auditive) in Echtzeit aktualisiert werden, um eine Interaktion zu ermöglichen (Dörner et al., 2019).
Diese Interaktion findet im dreidimensionalen Raum statt, im Vergleich zur zweidimensionalen Darstellung am Bildschirm. Durch die Immersion im Raum wird ermöglicht, mithilfe von Gestik und Bewegungen Eingaben zu tätigen.

Ein weiterer Versuch einer Definition, der sich auf die kognitive Eigenheit von Virtual Reality bezieht, lautet:

> "Virtual Reality is the use of computer technology to create the effect of an interactive three-dimensional world in which the objects have a sense of spatial presence." (Bryson, 2013)

Der Begriff "computer technology" wird von anderen Darstellungen abgegrenzt, die in sogenannter Telepräsenz eine Echtheit auf Bildschirmen imitieren.
Weiterhin wird ausgeführt, dass gezielt von einem Effekt und nicht von einer Illusion die

Rede ist, da es sich um einen Wahrnehmungseffekt handelt und nicht um eine Illusion, deren Ziel es ist, eine Täuschung hervorzurufen. Interaktivität wird im weiteren Verlauf der Arbeit detaillierter brieben. An dieser Stelle soll es vor allem den Gegensatz zu konventionellen 3D-Animation deutlich machen.
Der letzte Teil der Definition hebt hervor, dass die Objekte einen Ort im virtuellen Raum belegen, der sowohl unabhängig von den Nutzer*innen, wie auch vom Display existiert. Dies ermöglicht, dass die Objekte als tatsächlich vorhandene Dinge wahrgenommen werden. Bryson distanziert sich von dem Begriff der Immersion, der in vielen anderen Definitionen vorkommt und auch in dieser Arbeit eine zentrale Rolle einnimmt. Er stellt klar, dass VR im gesamten Umfang nutzbar ist, auch ohne das Gefühl zu haben, präsent zu sein und sich innerhalb der virtuellen Umgebung zu befinden.
Bei den Beispielen, die Bryson für dieses Argument anführt, handelt es sich allerdings um große Displays, die mithilfe von 3D-Brillen, ähnlich wie man sie aus dem Kino kennt, eine räumliche Darstellung abbilden, mit der interagiert werden kann. Es ist wichtig anzumerken, dass diese Definition vor der Jahrtausendwende formuliert wurde und sich der technische Teil stark verändert hat.
Die Brillen, mit denen auf Bildschirmen abgebildeten 3D-Animationen ein räumlicher Eindruck geschaffen werden kann, haben nicht zum Ziel, die äußeren Eindrücke zu ersetzen. Es ist möglich, mit ihnen den realen Raum wahrzunehmen.
Im heutigen Kontext würde dieser letzte Teil der Definition eher in den Bereich der Augmented Reality passen, die auch nur einen Teil der Realität ersetzt und eher Glaubhaftigkeit als volle Immersion zum Ziel hat.
Eine weitere Anmerkung die Bryson macht, ist die Abgrenzung von der Definition, die besagt, dass Virtual Reality zur Aufgabe hat, die echte Welt nachzuempfinden. Es ist zwar sehr verlockend, Nutzer*innen zu dadurch zu überzeugen, alle Details nachzubilden. Allerdings werden dadurch Kreativität und Gestaltungsmöglichkeiten eingegrenzt.
Diese Unterscheidung war bei Nanna sehr präsent, da eben diese Nachbildung der echten Welt und physikalischen Regeln und Gegebenheiten mit neuen Eigenschaften vermischt werden sollten. Es ist eine kreative Frage aber auch eine, die vom Nutzen der Anwendung abhängt: In welcher Balance man die reelle Nachbildung mit der Neuschaffung verbinden möchte.

Anwendungen neben Spielemarkt

Obwohl der Spielemarkt und die Verwendung in privaten Haushalten ein treibender Faktor für die Weiterentwicklung der Technik rund um VR sind, gibt es einige weitere Bereiche, in denen Virtual Reality-Anwendungen verwendet werden:
Forschern dient die Technik als Werkzeug, mehr über die menschliche Wahrnehmung zu lernen. In der Industrie können Produktentwürfe visualisiert und Prototypen realitätsnah vorgestellt werden. So kann sich jemand in ein neues Automodell setzen und dieses von allen Seiten im virtuellen Raum begutachten. Nicht nur zu Präsentationszwecken, sondern auch als Analyse und Planungstool kann VR eingesetzt werden.

Ein virtueller Flugsimulator, in dem kein echtes Flugzeug abheben muss, kann beim Anlernen von Piloten nicht nur die Gefahr minimieren, sondern auch Auswirkungen auf die Umwelt reduzieren.
Die Simulation der Steuerung von Maschinen wird bereits bei Schiffen, Straßenbahnen und Zügen sowie LKWs genutzt.
Ebenso können Ärzte den Ablauf von Operationen üben und wiederholen, ohne Gefahren einzugehen (Dörner et al., 2019).
In vielen Fällen dient die Simulation in VR der Geldersparnis und der Vermeidung von unnötigen Gefahren und Aufwand.
Spricht man zur Zeit von VR, so ist damit meist das Tragen eines Headsets gemeint, das das Sichtfeld durch eine computergenerierte Ansicht ersetzt. An dem Headset sind Kopfhörer angebracht, die alles Hörbare maskieren und durch abgespieltes Audio ersetzen (Schütze & Irwin-Schütze, 2018).
Sitzt das Headset korrekt, ist es möglich das ganze Licht und alle Geräuschquellen zu überdecken. Bei nicht korrektem Sitz ist das Display durch seitlich eindringendes Licht schlechter sichtbar und die Sounds sind schwieriger zu lokalisieren und zu identifizieren, da sie sich mit den Umgebungsgeräuschen mischen. Diese Faktoren können sowohl die Nutzung, als auch das Gefühl der räumlichen Anwesenheit in der virtuellen Welt einschränken.
Allerdings wird der Begriff "Virtual Reality" auch in anderen Anwendungen und Aufbauten verwendet. In diesen Fällen wird VR als allgemeiner Sammelbegriff verwendet. Mit der Bezeichnung "VR" wird, gerade im Consumer-Marketing, oft jede Art von interaktivem, bildgestütztem Content bezeichnet.

2.2 Häufig verwendete Begriffe in der Fachliteratur

Einige Begriffe treten sowohl in Veröffentlichungen zu Games sowie speziell VR-Anwendungen auf. Da viele VR Anwendungen Spiele sind, überschneiden sich die Felder der verwendeten Begriffe.
Folgende Begriffe beziehen sich auf die eigene Anwesenheit in der Anwendung, die bei der Involvierung des ganzen Körpers direkter gegeben ist, als bei der Bedienung eines User-Interfaces und der Informationsübermittlung über einen Bildschirm.

New Realities oder Extended Realities
Unter den Schirm der New Realities oder Extended Realities (XR) fallen alle neuen Realitäten. Dazu gehören:

AR
Augmented oder Alternate Reality ist eine Darstellung virtuellen Inhalts, bei der die reale Umgebung noch zu sehen ist. Durch diese überlagerten Bilder soll die Realität erweitert und abgewandelt werden. Bilder, Text oder andere Inhalte werden auf Smartphones, Brillengläsern oder auf der Frontscheibe eines Fahrzeugs angezeigt (Schütze & Irwin-Schütze, 2018).

MR
Mixed Reality bezieht sich auf alle interaktiven Medien, die weder unter den AR, noch den VR Begriff passen, und dient als eine Art Platzhalter für alle Medienformate, die neue Ansätze der Implementierung in den New Realities bieten.

360 Grad Video
Dieses Format wird auch CVR, Cinematic Virtual Reality, genannt. 360 Grad Video kann ebenfalls auf einem Headset-Display wiedergegeben werden und ermöglicht die freie Wahl der Blickrichtung durch Kopfbewegungen (Rothe et al., 2018).

FOV
Field of View ist der sichtbare Bildausschnitt, der in den 360 Grad Darstellungen durch den Blickwinkel begrenzt wird. Das Field of View wird in Grad gemessen und kann in den gängigen Head Mounted Displays Werte zwischen 80 und 140 Grad annehmen (Bezmalinovic, 2020)

HMD
Das Head Mounted Display ist das Headset einer VR-Ausrüstung.

2.3 Sensorik

Im virtuellen Raum kann die Sensorik des Menschen auf verschiedene Weise angesprochen werden.
Zur Wiedergabe einer virtuellen Szene wird beiden Augen ein stereoskopisches Bild errechnet. Dieses entsteht nach den Prinzipien der Zentralperspektive. Im Head Mounted Display werden die beiden perspektivischen Bilder angezeigt und den Kopfbewegungen angepasst in Echtzeit aktualisiert (Pfisterer, 2019).

Wie die Augen aus zwei Bildern im *Head-Mounted Display* ein Bild berechnen, so wird auch aus den Schallinformationen, die im rechten und linken Ohr ankommen, aus zwei Signalen ein räumlicher Hör-Eindruck gebildet.
Dies ist nur dann möglich, wenn die zwei Signale voneinander unterscheidbar sind. Diese sogenannte *Dekorrelation* macht aus einem in der "Hör-Mitte" platzierten Geräusch ein räumlich platziertes Geräusch.
Dieser Effekt tritt bei der Unterscheidung zwischen zwei Mono-Signalen auf, die identisch aus zwei Treibern kommen, und einem Stereo Signal, das aus zwei dekorrelierten, sich unterscheidenden Signalen besteht.

IATD
Eine Art der Dekorrelation ist die Interaurale Zeit Differenz (oft: IATD - Interaural Time Difference). Diese Differenz beschreibt den zeitlichen Unterschied des jeweiligen Schalleintritts eines Geräusches an den Ohren. Das bedeutet, der Schall aus der Umgebung kommt zeitverzögert am jeweiligen Ohr an.
Wir haben gelernt, dass das näher an der Schallquelle befindliche Ohr den eintreffenden

Schall als erstes empfängt. Unsere individuelle Kopfgröße und der Einfallswinkel bestimmen das Maß der Verzögerung, mit der das weiter entfernte Ohr dasselbe Schallereignis wahrnimmt. Die einzige Ausnahme sind Schallquellen, die auf der Ebene liegen, die zwischen unseren Ohren verläuft: die mediane Ebene, die unser Gesicht halbiert. Hier treffen die ersten Schallwellen gleichzeitig bei beiden Ohren ein.
Das ermöglicht es uns, auf wenige Grad genau zu bestimmen, aus welchen Richtungen Geräusche kommen. Um Schallquellen genauer zu lokalisieren, nutzen wir sogenannte Peilbewegungen. Dabei bewegen wir unseren Kopf, um vermutete Positionen bestätigen zu können. Im Folgenden wird der Vollständigkeit halber ein Einblick in die anatomische Funktionsweise des menschlichen Ohres gegeben:
Die Ohren sind ein paarweises Organ, ein Ohr an beiden Seiten des Kopfes, der das Sinneszentrum des Ohrs enthält, die Hörschnecken. Diese sitzt tief im Schläfenbein eingebettet. Die Aufgabe des Ohres ist es, den Schall nach innen an die Hörschnecke weiterzuleiten, die die Vibrationen umwandelt. Diese Umwandlung findet durch die Haarzellen statt, die einen Nervenimpuls auslösen (Alberti, 2006).

2.3.1 Raumwahrnehmung

Um einen Raum und eine Umgebung einschätzen zu können und sich zurecht zu finden, hilft unser Gehör. Virtual Reality wird oft gleichbedeutend mit “virtueller Welt” oder “virtueller Raum” verwendet. Die Realität, auch die virtuelle, wird eng mit der Existenz einer Umgebung verknüpft.
In dieser Umgebung gibt es Objekte, von denen manche Schall emittieren, andere reflektieren diesen oder dämpfen ihn ab. Den Eigenschaften der Geräuschquelle, wie ihrem räumlichen Bezug entnehmen wir Informationen um uns besser zurechtzufinden. So werden die Geräuschquellen in das Umgebungsgefühl eingefügt, und der Raum kann erschlossen werden.

Um Objekte in einen Raum einordnen zu können, gibt zum einen die Lautstärkedifferenz Informationen über die Entfernung eines Schallereignisses, es kann die Tiefe des Schallereignisses wahrnehmen, die Entfernung.
Weiß man, welche Lautstärke eine Geräuschquelle hat, kann man mit diesem Vergleichwert feststellen, ob sie weiter oder näher ist als zuvor. Bei einigen Geräuschen weiß man kontextuell, wie laut sie zu erwarten sind. Hört man eine Fliege laut summen, weiß jeder, dass sich diese in unmittelbarer Nähe befindet. Hört man ein leises Schiffshorn, muss sich dieses in der Ferne befinden.
Bestimmte Geräusche können so eine Aussage über den Raum machen in dem man sich befindet. Ein lautes und direktes Grillenzirpen verrät einem, dass man sich im Freien befindet, ohne dass man auf Reflektionen an eventuell vorhandenen Wände lauschen müsste. Bei unbekannten oder nicht so gegenständlichen Geräuschen wie den oben genannten Beispielen, ist die Ortung ausschließlich anhand der Lautstärke schwieriger, da meist kein direkter Vergleichspegel der “Normallautstärke” existiert. Deshalb sind für die Wahrneh-

mung der Distanz zu einem akustischen Ereignis die Pegelunterschiede und die zeitliche Latenz zwischen Direktschall, Erstreflexionen und Nachhall relevant. Befindet sich die Quelle des Geräuschs nah am Ohr, so wird der Direktschall ohne Zeitverzögerung und mit nahezu Originallautstärke gehört. Der Direktschall sind die ersten Schallwellen, die sich unbeeinflusst Reflexionen oder Verdeckungen durch den Raum bewegt. Die Verzögerung und Abschwächung der Erstreflexionen im Vergleich zum Nachhall ist in diesem Fall sehr groß.
Ist die Quelle des Geräuschs hingegen weit vom Ohr entfernt, so ist auch die Differenz des Direktschalls zu Erstreflexionen geringer. Außerdem nimmt man zur räumlichen Einordnung auch die Richtung wahr, aus der der Schall eines akustischen Ereignisses unsere Ohren erreicht.

Hört man sehr gedämpfte Stimmen, durchmischt mit Werbemusik, weiß man nicht nur, dass sich in der Nähe Leute unterhalten und ein Fernseher läuft, sondern man kann auch vom Maß der Dämpfung und fehlender Klarheit ableiten, dass sich Wände zwischen meiner Position und der anderer Leute befinden. So etabliert sich ein Raum auch wenn man keine Wände sehen kann oder nur simulierte Raumgrenzen, wie zum Beispiel weiße Flächen vor den Augen hat.

2.3.2 Individuelles Hören

Die Wahrnehmung über das Gehör und das insgesamte Hörerlebnis weicht von Mensch zu Mensch individuell ab. Ein großer Faktor dafür sind anatomische Unterschiede. Ein individuelles Hörerlebnis exakt nachzubilden oder ein allgemein anwendbares Modell zu erstellen, stellt daher eine sehr große Herausforderung dar.
Dieses individuelle Hörerlebnis nachzubilden, oder ein allgemein anwendbares Modell zu erstellen, ist eine Problematik, die nicht so leicht gelöst werden kann. Beim Headset von Vive lassen sich der Augenabstand und damit der “Sweet Spot” an die Anatomie der Nutzer*innen anpassen und für jede Nutzung individuell verändern.
Der Abstand der Ohren kann allerdings nicht so leicht in die Audio-Wiedergabe mit einberechnet werden, obwohl dieser zur räumlichen Wahrnehmung beiträgt.
Noch schwieriger zu berücksichtigen ist die, nicht als Millimeterangabe formulierbare, individuelle Form der Ohrmuscheln, zu berücksichtigen. Diese Form filtert den Schall beim Eintreten unterschiedlich. Hierfür benötigt es persönliche Ausmessungen und das Erstellen einer Head Related Transfer Function (HRTF), die Außenohr-Übertragungsfunktion, die Kopfform und Rumpf miteinbezieht.
Ein weiterer persönlicher Faktor ist die grundsätzliche Wahrnehmung von Schall einer einzelnen Person. Das Hörvermögen wird am häufigsten bei Freizeitbeschäftigungen oder der Ausübung des Berufs durch Lärmbelästigung geschädigt. Dabei kommt es zu Einschränkungen der Wahrnehmung und zu einer erhöhten frequenzabhängigen Hörschwelle (Streppel & Robert-Koch-Institut, 2006).
Jede und jeder potenzielle Nutzer*in geht unterschiedlichen Hobbies nach, die das Ge-

hör mehr oder weniger intakt hinterlassen. Zusätzlich dazu wird der Höreindruck durch die individuelle Geräuschsensitivität beeinflusst. Die Geräuschsensitivität ist neben dem Hörvermögen dafür verantwortlich, welche Geräusche deutlicher wahrgenommen werden (Ellermeier et al., 2001).
Hinzu kommt der Rechenaufwand eines in Echtzeit simulierten Hörereignisses. Nur in Echtzeit und mit möglichst latenzfreier Wiedergabe kann positionsgenau gearbeitet werden und können unabsichtliche Bewegungen wie auch Peilbewegungen des Kopfes miteinzogen werden.

2.4 Technik

Die Illusion, sich in einem virtuellen Raum zu befinden, wird durch das Ersetzen möglichst vieler realer Sinneseindrücke durch künstlich erzeugte verstärkt. Die Anforderungen an die verwendeten Technik ist es, diese Signale möglichst überzeugend zu ersetzen und gleichzeitig die Einflüsse der echten Welt auszublenden.

Die Wahl des Headsets war einfach, da für Projekte der Hochschule der Medien Vive Pro Headsets zur Verfügung stehen, die auch bei der Präsentation des Projekts verwendet werden konnten.
Das Headset verfügt über Sensorlinsen, die von den Basisstationen durch Infrarot-Abtastung verfolgt werden. So kann das Headset im Raum positioniert und diese Position in die Anwendung übertragen werden.
Das Vive Pro HMD, ein Head Mounted Display, wird mit einem Bügel und Gurt am Kopf angepasst und bedeckt die Augenpartie.
Zur Nutzung des Headsets ist es notwendig die IPD (Inter Pupillary Distance), den Pupillenabstand in Millimetern zu messen. An einem Rad kann der Abstand zwischen den beiden Teil-Displays eingestellt werden, sodass sich beide Bilder für das jeweilige Auge in zentrierter Position befinden. Ansonsten wirkt das Bild verschwommen und weniger räumlich.
Der Abstand zwischen Display und Augen kann individuell geändert werden, wobei keine qualitative Verbesserung des Bildes erzielt wird. Vielmehr kann ein vergrößerter Abstand das Tragen angenehmer machen, falls man lange Wimpern hat oder eine Brille trägt. Mit zunehmendem Abstand verkleinert sich allerdings das mögliche Sichtfeld.
Das Sichtfeld des Head Mounted Displays beträgt 110 Grad.

Durch die simulierte Triangulation zwischen den beiden Kopfhörerlautsprechern umfasst unser Gehör auch im virtuellen Raum 360 Grad. Diese Information lässt sich beispielsweise nutzen, um den Blick zu leiten, indem Geräusche außerhalb des Sichtfelds hörbar und lokalisierbar sind und Nutzer*innen ihren Kopf drehen. Dadurch ist ein erster Interaktionsschritt bereits getan und Nutzer*innen können den Raum erfahren. Die Angaben zum verwendeten Vive Pro Headset können auf der vive website (*VIVE Pro | The Professional-Grade VR Headset*, 2021) abgerufen werden.

Die Bauweise des Headsets ermöglichte es, mit wenigen Handgriffen die angebauten Kopfhörer abzuschrauben, ebenso ihre Halter. Dadurch konnten die Kopfhörer durch ein anderes Modell ersetzt werden. Beim Modell der Wahl handelte es sich um beyerdynamic DT 770. Die geschlossene Bauweise dieser Kopfhörer ermöglicht klares Abhören ohne Umgebungsgeräusche. Diese Entscheidung basierte vor allem auf der Präsentationsart. Die Präsentation von Nanna erforderte eine geschlossene Bauart der Kopfhörer, damit das Hörerlebnis möglichst abgekoppelt und isoliert von den Umgebungsgeräusche stattfinden konnte. Die kleineren vive-Kopfhörer schlossen nicht verlässlich ab, da sie über einen Klapp-Mechanismus verfügen und auch über den Ohren schweben können. Die Wahl der zusätzlichen Kopfhörer brachte allerdings auch weitere Kabel mit sich, die versorgt werden mussten.
Eine Schwierigkeit, die sich während der Endphase ergab, waren die Kommunikationsmöglichkeiten mit der Person, die sich gerade in der Anwendung befindet. Die Kommunikation mit Nutzer*innen, die sich in der Anwendung befinden, wird durch die geschlossene Variante erschwert. Da sehr wenig Schall von außen durch die Kopfhörer eintritt, ist eine Kommunikation über Sprache schwer möglich. Fiel es also einer Versuchsperson schwer, voran zu kommen, oder beim Mischen mit Teammitgliedern, war eine Kommunikation mit der Person vonnöten, die das Headset trug. Hier sollte man erfahrungsgemäß vorsichtig sein, da man sich recht schnell erschrecken kann.

Das Setup bestand neben dem Headset und den Kopfhörern, die beide direkt mit dem Rechner verbunden wurden, aus den “Lighthouses”, die die Bewegung registrieren. Lighthouses sind kleine schwarze Boxen, die Headsets und Controller durch Infrarot vertikal und horizontal erfassen. Dadurch kann die Position vom echten Raum in den virtuellen übersetzt werden.
Darüber hinaus benötigt man für die Bedienung der Menüs und für die Interaktion mit Objekten zwei kabellose Handheld Controller.

3 Durchführung der Tongestaltung für das Projekt

3.1 Sounddesign

Sounddesign bezeichnet die Gestaltung des Tons. Der Ursprung des Game Sounddesigns und Bereich, an dem sich in vielen Punkten orientiert wird, ist das Film Sounddesign. Dieses hat immernoch einen starken Einfluss auf die Gestaltung des Tons (Grimshaw, 2011).
Im Film-Ton wird klassischerweise die Musik vom Rest der Tongestaltung getrennt. Bei der modernen Herangehensweise wird bevorzugt, die Unterscheidung zwischen diegetischen und nicht-diegetischen Geräuschen zu priorisieren. Dieser Ansatz erlaubt es, den Geräuschen komplexere Eigenschaften und Beziehungen zuzuweisen (Görne, 2019).
Im Game-Sounddesign ist diese Unterscheidung grundsätzlich auch anwendbar, allerdings gestaltet sich die Frage nach der Diegese im virtuellen Raum schwieriger. Durch die eigene Präsenz nehmen Nutzer*innen oft nicht nur die Rolle der Handelnden ein, sondern auch die Perspektive. Im Kapitel *Orientierung durch Geräusche* wird diese Frage ausführlicher erörtert.
Die Einteilung in Musik/Soundscape oder diegetisch/nicht-diegetisch hilft, die gestalterischen Aufgaben zu definieren, die die einzelnen Elemente übernehmen sollen.

3.2 Konzeption des Projekts

Das Konzept entwickelte sich vorwiegend im Verlauf der Studioproduktion, da viele Entscheidungen auf praktischen Erfahrungen mit Umsetzbarkeit und Technik beruhten, die im Laufe der Arbeit mit den neuen Softwares gemacht wurden. Im Folgenden wird auf die ersten Ideen eingegangen und es werden die Kriterien vorgestellt, nach denen die Ideen bewertet und aussortiert wurden.
Da es für alle Teammitglieder das erste große interaktive Projekt war, wird besonders Wert auf die Unterscheidung von linearen und interaktiven, nicht-linearen Medien gelegt. Diese Unterscheidung war bei einigen Entscheidungen maßgebend.

Das Projektteam setzte sich zum Ziel, eine Umgebung zu erstellen, die es Nutzer*innen ermöglicht, mit den Einschränkungen der Realität zu spielen. Seien es Erwartungen, Gravitation, Richtungen, Distanz, Temperatur, Geschwindigkeit, und Bewegung. Der Anspruch hierbei wurde schnell klar: technisch muss es den Anforderungen genügen, um den Nutzer weit genug von der Einhaltung physikalischer Regeln zu überzeugen und eine Erwartungshaltung aufzubauen, die dann erfüllt oder umgekehrt werden sollte. Weder Bild noch Ton sollten durch offensichtliche Fehler, Latenzen oder physikalische Ungereimtheiten Aufmerksamkeit auf sich ziehen.
Alles Wahrgenommene sollte zu einem Gefühl der Echtheit des virtuellen Raumes beitragen. Nutzer*innen sollten zum Interagieren angeregt werden; jedoch bewusst ohne eine Narration, um ein geleitetes Durchlaufen der Simulation zu vermeiden. Diese Entscheidung basierte auf der Befürchtung, dass die Leitung oder Erzählung zu sehr von der eigenen Wahrnehmung

ablenkt und zu viel Aufmerksamkeit auf andere Charaktere gelegt wird. Die eigene Präsenz steht im Mittelpunkt.

Als konzeptionelle Grundlage erarbeitete das Projektteam zu Beginn des Projekts folgende Punkte:
Die Nutzer*innen werden durch emotionalen Bezug zur Interaktion angeregt werden. Um dies zu erreichen, galt es Neugier, Unbehagen oder Genugtuung bei der Nutzer*in zu wecken.
Den Nutzer*innen wurde keine spezifische Aufgabe gestellt, die in eine Narration eingebettet ist oder auf einem Schild zu lesen wäre. Es ging um die Möglichkeit, sich auf eigene Faust umzusehen, Orientierung zu gewinnen und Interaktionen zu finden.
Dies bietet den Vorteil, sowohl geübte Computerspieler*innen, als auch Neulinge im VR-/Games-Bereich weder zu unter- noch überfordern. Es soll dadurch erreicht werden, jederzeit ohne Frustrationsgefühl aufhören zu können.
Frustration, so die Befürchtung während der Konzeption, würde Nutzer*innen daran hindern, die Anwendung mehrmals zu starten und Neues herauszufinden.

Als Alternative zum Stellen einer festgelegten Quest, wurde ein neues Regelwerk in der Realität des Raumes aufgestellt, das es zu erforschen gibt.
Gewohnte Regeln, wie Gravitation, wurden aufgehoben oder verstärkt. Auch, dass ein Raum eine sichtbare Grenze haben muss, wurde entkräftet. Die unendliche Bewegungsfreiheit wurde durch das Platzieren der einzelnen Inseln/Glasplatten beschränkt. So konnte eine Weite suggeriert werden, ohne dass Nutzer*innen auf die weiter entfernten Flächen springen konnten, und sich so zu weit vom Startpunkt wegzubewegen.
Als Ziel des Spiels wurde kein Erkenntnisgewinn oder eine moralische Mitteilung gesetzt, sondern das Erforschen des eigenen Verhaltens. So konnte auf spezifisches Game-Design und Balancieren von Spielemechaniken oder KI verzichtet werden. Der Fokus lag auf Räumlichkeit und Gestaltung.
Im Abschnitt *Konzeption* werden diese Überlegungen erneut aufgegriffen und ihre Auswirkungen beleuchtet.

3.2.1 Ideen und Entwicklung

Um sich dem Prinzip von VR-Anwendungen anzunähern, wurden als Vorbereitung zu Beginn der Projektarbeit zunächst eigene Eindrücke gesammelt.
Zum Testen anderer VR-Anwendungen wurden der Verfügbarkeit und Bekanntheit nach primär Spiele genutzt.
Zunächst richtete sich der Fokus auf die Frage, welche Auslöser zu welchen Reaktionen führen können. Hierbei fiel auf, dass das Warten in einem Virtuellen Raum der Wahrnehmung der Qualität und Glaubhaftigkeit des Ganzen unzuträglich ist. Das Ziel wurde gesetzt, eine Balance zu finden. Den Nutzer*innen sollte einerseits die notwendige Zeit zur Verfügung stehen, sich in Ruhe umschauen können.

Zur gleichen Zeit sollte aber vermieden werden, dass die Nutzer*innen zu lange am selben Ort stehen und etwaige Wiederholungen im Ton wahrnehmen. Grafische Details könnten auffallend werden, in denen sich Muster wiederholen.
Die Nutzer*innen sollten weiterhin nicht zu oft in der Lage sein, technische Grenzen zu erfahren und so den Raum als digitale Simulation entlarven zu können. Das beinhaltet auch, zu vermeiden, dass sie mit den realen Begrenzungen des Versuchsraumes, in dem das Headset montiert ist, in Berührung kommen.
Um diese Problematik nicht aufkommen zu lassen, wurde im Konzept auf verschiedene Aspekte geachtet. Diese Aspekte beinhalten einerseits die Platzierung von Objekten, die durch Ton und Interaktionsmöglichkeiten die Aufmerksamkeit der Nutzer*innen auf sich ziehen. Auch durch Auswahl und Bearbeitung der Sound Files wird vermieden, dass Muster auffällig werden. Die angewandten Lösungen werden im folgenden Kapitel “Layout des virtuellen Raums und Platzierungen der Sound Quellen” beschrieben.

Die Anwendung wurde in der Grundidee an eine Ausstellung angelehnt, bei der mehrere Räume besucht werden können. In diesen Räumen können voneinander abgekoppelte Erfahrungen gemacht werden, um verschiedene Aspekte der Gestaltung im virtuellen Raum zu zeigen.
Nachdem die Nutzer*innen durch einen simulierten Aufzug das Foyer betreten, werden sie in diesem dazu angeregt, sich durch den Raum zu bewegen, mit Gegenständen zu interagieren und ein Gespür für die Position des eigenen Kopfes zu bekommen.
Im Raum positionierte Stelen regen die Nutzer*innen dazu an, sich zu bewegen, da je nach Position der Stele ein anderer Klangeindruck entsteht.
Im Raum verteilt wurden Kugeln platziert, die auf ansonsten scheinbar festem Boden zu schwimmen scheinen. Stoßen die Nutzer*innen diese an, ertönt ein Klang, während die Kugel davon treibt.
Die Wahrnehmung der eigenen Position und der eigenen Ausmaße werden durch schwebende Würfel verstärkt, die im Foyer verteilt sind. Groß genug, um den Kopf hineinzustecken, bergen zwei Würfel jeweils noch einen anderen Raum in sich. Dieser innere Raum sprengt die erwarteten Größenverhältnisse und suggeriert mit Farben und Mustern eine Weite, die Desorientierung auslöst. Das Integrieren von Räumen verschiedener Dimensionen ineinander wäre außerhalb einer Simulation nicht möglich. An dieser Stelle wurde die Entscheidung getroffen, statt einer realistischen Repräsentation ein alternatives Erlebnis zu gestalten. Diese Phänomene, die durch die Interaktion mit den Stelen, Kugeln und Würfeln entstehen und Grundformen, die nur mit Farbabgrenzungen und Mustern auf sich aufmerksam machen, werden von der Tongestaltung unterstützt.

Vom Foyer kommen Nutzer*innen zwei verschiedene Plattformen in jeweils einen weiteren Raum. Der eine besteht aus drei Abschnitten, die durch Membranen voneinander getrennt sind. Es werden Temperaturempfindungen, verschiedene Oberflächenstrukturen, unterschiedlich wirkende Schwerkraft, sowie Größen, Stimmungen und Strukturen simuliert und erfahren.

Der mittlere Bereich wurde bewusst reizarm gestaltet, da er der Präsentation der interaktiven Objekte und der Auswahl der beiden gegensätzlichen Teil-Räume dient.
Dabei wurde sich an einer raumschiffartigen Kapsel orientiert, sowohl das Bild als auch den Ton betreffend. Für die beiden Seiten des “Basic-Emotions-Raumes” wurden Listen von Adjektiven erstellt, die einander gegensätzlich sind. Es wurden die Adjektive ausgewählt, zu denen klar einordenbare Assoziationen existieren. Andere, wie beispielsweise “hart” und “weich”, die ein breiteres Assoziationsspektrum besitzen, wurden ersetzt durch “technisch” und „organisch" oder “rau und glatt”.

- klar - dicht
- glatt - rau
- scharf - bröselig
- hell - dunkel
- dünn - massiv
- weit - nah
- kalt - warm
- leicht - schwer

Durch diese konzeptionellen Begriffe ergaben sich einige Parameter, die technisch umsetzbar sind:

- Frequenz (Hoch-/Tiefpass)
- Hall (weit/kurz)
- “texturiert” (Geräuschqualität)
- Recordings oder Synthese (Herkunft des Audio File)

Insgesamt waren vier Räume inklusive Aufzug zu erstellen und klanglich darzustellen. Da aber der Sound des “Rhythmus-Raums” prozedural erstellt werden sollte, wurde dieser nicht über die Middleware sondern über MaxMSP eingeflochten.

3.2.2 Vergleich mit linearen Medien

Als lineares Medium hat ein Endprodukt immer einen fixen Ablauf, so Walder (Somberg, 2019). Dieser ermöglicht es, Bild und Ton sorgfältig zu verbinden, um einen maximal emotionalen Effekt zu erzielen.
In Anwendungen, bei denen es keinen fixen Ablauf gibt, können Nutzer*innen trotzdem in eine annähernd lineare Erzählung "gezwungen" (nach englischer Übersetzung) werden. Durch beschränkte Interaktionen oder "Cutscenes", kurze filmische Szenen, die eingespielt werden. Lineare Medien haben einen festen Start und Endpunkt, das bedeutet, dass das Material an jedem beliebigen Zeitpunkt immer den gleichen Bildausschnitt und dasselbe Audio-Sample benutzt. Dadurch ergibt sich eine exakte Reproduzierbarkeit, bei der sich weder in der Narration, noch in der Gestaltung einzelner Elemente Unterschiede ergeben. Bei Musik, die im Radio läuft oder von einer CD abgespielt wird, ein Film im Kino, Fernsehen oder On Demand, Shows im Fernsehen, sie alle haben vordefinierte Abläufe. Der Ablauf wird in der Komposition, im Drehbuch, oder im Skript definiert und ist nach der Durchführung reproduzierbar. Bei Aufzeichnungen kann das Material später noch bearbeitet werden. So kann es verschiedene Schnittversionen eines Films aus demselben Material geben, allerdings lässt sich hier keine Interaktivität finden, da die Bearbeitungsschritte losgelöst von einem Publikum durchgeführt werden und nach Veröffentlichung und Verbreitung unveränderlich sind.

Interaktion
Interaktivität ist nur möglich, wenn von der Seite der Ersteller*innen des Mediums ein Angebot besteht, das es dem Publikum ermöglicht, durch eigene Entscheidung den Ablauf des konsumierten Mediums zu ändern. Ansonsten könnte man argumentieren, dass das Drücken der "Pause", "Play" oder Vorspul-Tasten auch eine Möglichkeit der Interaktion darstellt. Da dies aber nicht die oben genannte Reproduzierbarkeit beeinflusst, sind diese Funktionen im vorliegenden Kontext nicht gemeint, wenn von Interaktivität die Rede ist.

Die Interaktivität in der Welt der Medienkunst wird von Pfisterer wie folgt definiert:

> I[nteraktion], wie sie in der interaktiven Medienkunst als technisches Verfahren verwendet wird [...], ist der Computer-Betriebsmodus, der die Steuerung eines Rechenprozesses durch die Benutzer während der Verarbeitung zulässt. Er erlaubt im grundlegenden Schema von Eingabe, Verarbeitung und Ausgabe während der Verarbeitung wieder Nutzereingaben, so dass sich eine Rückkopplungsschleife zwischen Nutzer-Aktion und Rechner-Reaktion schließen kann (Pfisterer, 2019).

Weiterhin beschreibt er, dass die Interaktion mit Technik, als ein Teil der weitergefassten, zwischenmenschlichen Interaktion zu verstehen sei. Es geht um den zwischenmenschlichen Bezug, wenn eine Person auf eine andere reagiert und dadurch wiederum ihr Handeln beeinflusst wird.

Auch bei der technischen Interaktion geht es um den Einfluss von Handlungen und das Beeinflusstwerden in einer gegenseitigen Abhängigkeit.
Aus dieser Definition und dem Erwähnen der Rückkopplung lässt sich schließen, dass es immer ein beidseitiges Verhalten ist. Wird, kein wie oben beschriebenes Angebot gemacht und Nutzer*innen schalten aus eigener Motivation das Bild in den Schwarzweiß-Modus, findet keine Interaktion in diesem Sinne statt.
Dasselbe gilt, wenn es eine Aufforderung zur Nutzung einer Funktion gibt, der nicht gefolgt wird. So besteht eventuell eine Möglichkeit zur Interaktion, aber das Medium wird nicht interaktiv genutzt.
Deshalb kann es bei der Konzeption umso wichtiger sein, durch "Cues" im Sounddesign oder ein bewegendes Narrativ, Nutzer*innen dazu anzuregen, sich bei bestehenden Interaktionsmöglichkeiten zu beteiligen. Die Interaktion stellt eine notwendige Bedingung für Virtualität dar. Die Feedbackschleife muss nicht nur eng sein, sondern jede neue Handlung muss ein unmittelbar wahrnehmbares Feedback auslösen, damit die Nutzer*innen erneut auf dieses mit einer Handlung reagieren können (Pfisterer, 2019).

Der Begriff der multilinearen Medien wird definiert von Bucher, der argumentiert, dass es beispielsweise auf Websiten und in der Tageszeitung die Möglichkeit gibt, Nutzungspfade individuell auszuwählen. Hierzu müssen die Informationseinheiten netzartig verbunden sein, statt wie in den linearen Medien sequenziell angeordnet. Hier unterscheidet er zwischen Hörfunk und Fernsehen, sowie Büchern und den meisten Textmedien, die linear sind und oben genannten Websiten und Tageszeitungen, die in der Nutzung anders gedacht sind, und damit multilinear (Bucher, 2001).
Somit kann festgehalten werden, dass es nicht "Das" lineare Medium gibt, sondern ein Spektrum der Interaktivität die definierenden Begrifflichkeiten bestimmt.
Hier spielt die Intention eine große Rolle. Es kann sich um dieselbe Funktion handeln, wobei ein Medium diese interaktiv nutzt, das nächste jedoch nicht. Dies soll anhand eines Beispiels verdeutlicht werden:
Hat ein Film eine angekündigte Möglichkeit, dass Nutzer*innen in einem Detektivfilm einen Farbfilter an oder ausschalten können, der in der Schwarzweißversion inhaltliche Hinweise zeigt, die in der Farbversion nicht sichtbar sind (oder umgekehrt) ist dies der gleiche Knopf auf der Fernbedienung, wie bei einem anderen Film. Besteht allerdings keine Intention und damit kein Angebot der Interaktion, so ändert zwar die Farbumstellung auf Schwarzweiß bei einem anderen Film das Bild, gibt den Nutzer*innen aber keine Mehrinformation.

Im Gegensatz dazu gibt es *nicht lineare* Medien. Da die Interaktivität in der Unterscheidung zu linearen Medien eine zentrale Rolle spielt, werden diese auch *interaktive Medien* genannt. Schütze beschreibt, wie sich die Zielsetzung der Neuen Realitäten ("New Realities") und der linearen Medien, wie Film und Fernsehen unterscheiden. Die linearen Medien geben Einblick durch ein definiertes Fenster, durch das festgelegte Aspekte der dargestellten Welt sichtbar sind. Es werden die Aspekte gezeigt und in Szene gesetzt, die am wichtigsten für die Erzählung sind.

Die Zuständigen für Drehbuch und Regie entscheiden, welche Elemente wichtig sind, um eine möglichst unterhaltsame Szenerie zu gestalten.
Durch dieses Fenster in die Erzählung ist es wichtig, das Publikum in Sounds einzuhüllen, um die Brücke zu dem abgegrenzten Fenster zu überwinden und das Gefühl der Immersion zu verstärken. Das bedeutet aber auch, dass die spezifische Position einer Geräuschquelle oder die exakte Entfernung zum Publikum nicht stark ins Gewicht fällt, solange dieses durch umgebende Geräusche Einhüllung ins Geschehen erfährt (Schütze & Irwin-Schütze, 2018).

Es ist einfacher festzustellen, welche einzelnen Elemente eines "Medien-Objekts" interaktiv sind, als das ganze Medien-Objekt (den Film, das Spiel etc) als lineares oder interaktives Medium zu definieren oder den Grad der Interaktivität/Linearität festzulegen.
Dennoch kann es hilfreich sein, ein Medium im Ganzen als "interaktives Medium" zu benennen, so wie es bei Nanna der Fall war. Dadurch wird deutlich gemacht, dass sich jede zu treffende Entscheidung und jedes zu wählende Gestaltungselement am Maß der Interaktivität messen lassen muss.
Verschiedene Elemente, die linear oder nicht-linear sein können, sind zum Beispiel:
Eine Narration, die sich als Baumstruktur abhängig von den Entscheidungen der Nutzer*innen verschieden entwickelt und verzweigt. Durch das Setzen der Knotenpunkte wird festgelegt, wie anpassungsfähig der Erzählstrang tatsächlich ist, oder ob sogar nur die Illusion einer Interaktion besteht, und es nur ein oder zwei verschiedene Ausgänge der Geschichte gibt. Durch Anzahl und Anordnung der Knotenpunkte definiert man für das Projekt die Linearität der Erzählung.
Sprache ist ein weiteres Element, das ein Level der oben genannten Multilinearität erfüllt. Nutzer*innen wählen sich aus vorhandenem Material einen Dialog aus und können diesen in selbst gewählter Reihenfolge und beliebiger Sprache abspielen.
Musik kann auch mehr oder weniger linear verwendet werden. Vom festgelegten Soundtrack über situationsabhängige Überblendungen, die auf Fortschritte reagieren, bis hin zu deutlich kleineren Elementen, die sich aus Spieleparametern ergeben.
Auch Bildausschnitte sind Elemente, die verschiedene Interaktionen bieten können. Sie sind entweder festgesetzt oder es besteht eine Auswahl aus zwei oder mehr Perspektiven.Oder eine unbegrenzte Anzahl der Perspektiven, die in den New Realities durch Kopfbewegungen frei wählbar sind und den Bildausschnitt definieren.

All diese Elemente können entweder durch harte Grenzen festgelegt werden, in dem beispielsweise zwei Entscheidungsmöglichkeiten ohne Mittelweg angeboten werden. Oder es wird in Teilen eine komplett lineare Cut-Scene eingefügt, die nicht übersehen werden kann. Weniger lineare Elemente können durch weichere Faktoren der Beeinflussung, wie Geräuschhinweise oder blinkende Lichter den Blick lenken. Auch beispielsweise durch drängende Musik können Nutzer*innen dazu bewogen werden, schneller in einen anderen Raum zu gehen, damit ein anderes Element der Umgebung nicht zu lange beachtet wird. Dabei muss allerdings immer im Blick behalten werden, dass sich Nutzer*innen trotz aller Bemühungen

absichtlich oder unabsichtlich dagegen entscheiden können. Das war eine der größten Herausforderungen bei der Mischung von Nanna. Denn Geräusche können noch so sorgfältig platziert sein, wenn niemand den Kopf in die Nähe des großen Würfels bringen möchte, und keine Hinweise bekommt, dies zu tun, dann wird auch niemand dieses Geräusch hören können.

3.3 Vorbereitungen

Im Folgenden werden die relevanten Vorbereitungen des Projekts erläutert, die zum Gelingen des Projekts essentiell waren. Bei der Betrachtung des Gesamtprojektes ist der Aspekt der Vorbereitung für das Gelingen eines interaktiven Projekts besonders wichtig. Durch den iterativen Workflow, bei dem die verschiedenen Produktionsschritte wiederholt werden, sind die verschiedenen Departements aufeinander während der Produktion aufeinander angewiesen. Viele Schritte können nicht nacheinander und abgegrenzt bearbeitet werden. Aus diesem Grund werden die Vorbereitungen in der Pre-Produktionder als eigener Teil gestellt. Dieser Schritt befindet sich zwischen dem der Konzeptionserstellung und der Durchführung. Die Vorbereitungen lassen sich nicht, wie bei Film- oder Musikaufnahmen, eindeutig in Aufnahme und Bearbeitung unterteilen, da bestimmte Schritte der Vorbereitung immer wieder aufgegriffen werden müssen. Das Anfangsmaterial wird von anderen Departements übernommen, deren Material daran angepasst. Im Anschluss kommt das Arbeitspaket wieder zurück, für weitere Anfügungen. Diese Schritte wiederholen sich. Die Grundlagen müssen zwischen den Departements Bild, Ton und Programmierung gut abgesprochen werden, ansonsten kommt es zu Schwierigkeiten im weiteren Verlauf der Produktion. Anders verhält es sich bei den typischen linearen Medien, wie beispielsweise beim Film, bei dem auch die Produktion linear verläuft und geplant werden kann. Das bedeutet, es gibt die Möglichkeit, die Departements dann dazu zu holen, wenn die vorherigen Schritte abgeschlossen sind und das Material übergeben werden kann. Im Film nennt man einen Zeitpunkt der Materialübergabe Picture-Lock, der Produktions-Zeitpunkt, zu demgewährleistet werden kann, dass am Bildmaterial kein einziges Frame mehr geändert wird. Dann ist der Schnitt und die Dramaturgie festegelegt und die Tongestaltung kann beginnen, unter anderem Soundeffekte zu setzen und den Dialog zu bearbeiten.

3.3.1 Eigenschaften und Verwendungszwecke einer Anwendung

Einige Eigenschaften einer Anwendung sollten vor der Produktion festgelegt werden. Welche Grenzen ein Projekt hat und mit welchen Mitteln man zum Ziel kommen möchte,legen die Entscheidungen zwischen Formaten oder Auswahl der Präsentation fest: Das Medium steht zu diesem Zeitpunkt schon fest, da dies durch die Besetzung des Teams und die organisatorischen Gegebenheiten festgelegt werden muss. Im Fall des Projekts Nanna handelte es sich um eine interaktive VR Anwendung.
Das Genre ist eine allgemeinere Definition, denn innerhalb eines Genres können Anforderungen technisch sehr unterschiedlich gelöst werden. Das können zum Beispiel Spiele sein,

wie Shooter, Echtzeit Strategie (RTS) oder Adventure Games, aber auch Serious Games, bei denen es sich um Spiele oder Anwendungen handelt, die den Zweck der Vermittlung eines Lerninhalts haben (Bopp, 2009).
Oder es sind Anwendungen mit speziellem technischen Ziel oder die Übertragung einer Software in den virtuellen Raum.
Die Entscheidungen, auf wie viele Nutzer*innen die Anwendung ausgelegt wird, ob es Mehrspieler-Modi geben wird und ob es eine Online-Version der Anwendung geben wird, wird entweder zu Anfang oder auch erst später getroffen werden. Hier kann aber tonseitig auf jeden Nutzer und jede Nutzerin das selbe etablierte System angewendet werden, wenn alle Nutzer*innen die gleichen Fähigkeiten haben. Durch die Dopplung entsteht in der Gestaltung grundsätzlich kein Mehraufwand.
Allerdings besteht ein deutlicher Gestaltungsunterschied bei der Verwendung verschiedener *Point of View* (POVs). Der POV bestimmt, ob die Kamerasicht mit der Perspektive der Nutzer*innen übereinstimmt oder über einem spielbaren Charakter schwebt und eine Draufsicht bietet. Ein anderes Blickfeld und eine eventuell deutlich erhöhte Perspektive lässt die Nutzer*in eine andere Klangumgebung erwarten.
Ebenfalls können Umgebungsgeräusche in der Vorbereitungsphase definiert werden, um Gestaltungsentscheidungen zu erleichtern. Gibt es ein spezifisches Setting, wie Weltraum, Fantasy-Welt oder reale Großstadt, so ergeben sich einige Elemente der Gestaltung durch Assoziationen und vergleichbare Spiele oder Anwendungen können als Referenz herangezogen werden.

Florian Füsslin beschreibt vier Stufen der Tongestaltung die Projekte je nach Umfang durchlaufen (Somberg, 2019). Bevor ein Projekt beginnt, kann durch Festlegen der erreichbaren Stufe der Umfang des Projekts eingeschätzt werden.
Die erste Stufe ist die, Feedback zu erzeugen. Eine Reaktion auf eine Handlung, die keine realistische Vertonung hat und als Platzhalter genutzt werden kann. Beispielsweise zeigt ein kurzes Piepen eine Kollision an. Dies erfüllt den Zweck, die Aufmerksamkeit der Nutzer*innen zu wecken. Ein vergleichbarer Effekt lässt sich bei Ear-Cons finden, die für alle Arten von Anwendungen Warnsignale oder Nutzgeräusche abgeben.
Sobald das Objekt eine realistischere oder narrative Rolle übernimmt, so Füsslin, müsse auch der Sound angepasst werden.
Um die zweite Stufe zu erreichen, die der Immersion, muss das Geräusch auf die Parameter reagieren, die von der Anwendung ausgegeben werden. Diese Parameter sind: die Geschwindigkeit oder die Richtung aus der das Geräusch erklingt. Aber auch Variationen bei mehreren Interaktions-Instanzen.
In der nächsten, der dritten Stufe werden die Emotionen mit in Betracht gezogen. Hier benötigt man Parameter, die für die Nutzer*innen eine Bedeutung haben und wichtig sind, wie zum Beispiel den Spielstand oder eine Gesundheitsanzeige in einem Spiel. Diese werden durch Sounds repräsentiert und verstärkt.
Auf der vierten Stufe werden die Nutzer*innen durch technische Eigenschaften der Control-

ler (z.B. die Rumble-Funktion) oder das Anheben der Lautstärke direkt miteingebunden. Diese Effekte haben eine direkten Einfluss auf Wohlbefinden der Nutzer*innen. Das haptische oder die reine Lautstärke sind beide in der realen Welt wie auch in der virtuellen Welt wahrzunehmen und bilden so eine Verbindung zwischen beiden.

Bei Füsslin benennt der Begriff Immersion den zweiten Schritt, also Teil des Einstiegs in die Einbindung von Nutzer*innen. Immersion wird nicht als Endziel definiert. Dies lässt sich jedoch problemlos in der Projekt-Konzeption in Einklang bringen, wenn man mit Einführung der dritten Stufe, der Emotionen, die Vorhergegangene nicht als abgeschlossen begreift.

3.3.2 Layout des virtuellen Raums

Wie Cancellaro treffend beschreibt:

> Large spaces do not sound like small spaces. Tiles and hard walls do not sound the same as rugged, dry wall textures (Cancellaro, 2005).

Dass große und kleine Räume verschiedener Materialien unterschiedlich klingen, ist eine Eigenschaft, die auch im Sounddesign genutzt wird.
Weite und Nähe können durch Reverb (Hall-Effekte), Echo und Reflektionen (simuliert durch Filter oder errechnet) dargestellt werden. Je mehr Reverb genutzt wird, desto größer klingt ein Raum grundsätzlich. Cancellaro aber macht deutlich, dass jede Umgebung von Sounddesigner*innen sorgfältig gestaltet werden muss.
Ein großes Büro mit viel Einrichtung und Teppichboden hallt natürlich nicht vergleichbar stark, wie eine Kapelle gleicher Größe aus Sandstein.

3.3.3 Informationsaustausch mit den Programmierer*innen

Im Projektteam Nanna führte jede Person mehrere Aufgaben aus, die sich zum Teil überschnitten. Daher war es besonders wichtig, die Aufteilung der Aufgaben zu besprechen.
Das Modellieren einer zerbrechlichen Vase war beispielsweise eine der Aufgaben der Autorin. Die Simulation der abgebrochenen Teile war dann Aufgabe des Programmierers in Unity. Die Gestaltung des Klangs der zu Boden fallenden Teile fiel wieder in den Aufgabenbereich der Autorin.
Um einen verlässlichen Workflow zu gewährleisten, wurden innerhalb des Teams immer wieder Modelle und Skizzen ausgetauscht.
In einem solch kleinen Team, mit fünf Personen, funktionierte dieser intensive Austausch problemlos und ermöglichte zeiteffektives Arbeiten. Bereits vor der Fertigstellung des Modells konnte mit einer frühen Version festgelegt werden, welche Größe die Bruchstücke haben sollen, wie weit diese fliegen und was danach mit ihnen passiert. Anhand einer beispielhaften Animation, die bei jedem Herunterfallen etwas variierte, aufgrund gleicher Parameter aber ähnlich genug war, ließ sich dann eine Audioskizze erstellen. Diese Audioskizze passte zu jeder Iteration der Splitter. Würde man allerdings immer dieselbe Audioskizze

verwenden, würde dies den Nutzer*innen nach wenigen Wiederholungen auffallen. Dies wurde durch die mehrfache Aufnahme des gleichen Gegenstands beim Foley-Recording verhindert, sowie durch die Abänderungen der Parameter in der Audio-Middleware, die die nötigen Variationen ermöglichten.

3.3.4 Welche Aufnahmen werden benötigt

Sogenannte Foleys sind eine weit verbreitete Gestaltungsmöglichkeit für den/die Tongestalter*in in der Vertonung von Material, bei dem kein Original Tonmaterial vorhanden ist. Die Ton-Aufnahmen, die im Zielmedium verwendet werden, um bestimmte Objekte und Schallereignisse zu vertonen werden meist im Studio aufgenommen. Das ermöglicht eine kontrollierbare Umgebung. Je nach Anforderung und Geschmack können die Foleys auf verschiedenste Weise aufgenommen werden und mit ausgefallenen oder alltäglichen Gegenständen und Materialien erzeugt werden.
Cancellaro priorisiert bei der Repräsentation von Objekten das verwendete Frequenzspektrum um eine Illusion zu erschaffen: Tiefere Frequenzen repräsentieren schwerere und dickere Objekte, höhere Frequenzen dünnere und leichtere Objekte (Cancellaro, 2005). Um im Mix zwischen höherfrequenten und tieferfrequenten Objekten unterscheiden zu können, müssen die Soundclips im EQ angepasst werden. Bei der Auswahl der Foleys kann bereits auf höhere und tiefere Anteile geachtet werden. Geräusche, die mit dem Mikrofon aufgenommen werden, haben grundsätzlich ein breites Frequenzspektrum.

3.4 Auswahl der Middleware

Grundsätzlich kann man bei der Games-Programmierung einfache Audio-Bearbeitungs und Integrierungs-Features nativ in der Game Engine finden.
So hat man in Unity die Möglichkeit, einen Sound abzuspielen. Mit einem einfachen Befehl: audioSource.Play();
Dazu muss der Sounddesigner bereit sein, sich mit dem Programmieren in C# auseinanderzusetzen. Middlewares bieten die Möglichkeit dies auszulagern. Als Software von Drittanbietern bildet sie das Bindeglied zwischen Ton-Verarbeitungs-Software und Game-Engine. Middlewares besitzen eine grafische Bedienoberfläche, sind also ohne Programmiersprache zu bedienen. Nach Wahl finden sich außerdem aus DAWs (Digital Audio Workstation) bekannte Arbeits-Strukturen und Wellenform-Anzeigen, um ähnliche Workflows aus dem Sounddesign zu übernehmen.
Wenn man sich für diese Option entscheidet, bleibt die Notwendigkeit bestehen, einzelne Zeilen in Code zu schreiben. Das Erzeugen der Objekte und Linken mit Events als Schnittstelle ist nur mithilfe der Programmierzeilen möglich. Da das aber die gleichen, wiederverwendbaren Zeilen sind, handelt es sich nicht um großen zeitlichen Aufwand.

Das Projektteam traf zunächst die Entscheidung, die Tongestaltung von der Game Engine zu entkoppeln. Dies erforderte die Nutzung der Middleware. Sollten Probleme innerhalb Unity-Programmierung auftreten, die nicht zu lösen sind, wäre die Tongestaltung davon nicht direkt betroffen. Es geht an dieser Stelle vor allem um die Vermeidung von zeitlicher Abhängigkeit. Problemlösung ist ein unvermeidbarer Teil eines Projekts. Da Nanna auf Unity aufgebaut werden sollte, stellte sich als zweites die Frage, welche Middleware mit Unity am besten geeignet sein würde.

Ein großer Vorteil der engine-unabhängigen Middleware ist, dass sich jedes Team-Mitglied nur den für seine Aufgaben relevanten Teil des Programmierens aneignet. Nicht jeder Einzelne muss die gesamten Programmier-Grundlagen in Unity als Basis für die eigene Arbeit legen.
Durch das Erlernen eines spezifischen Programms ließ sich die dem Einzelnen zur Verfügung stehende Zeit gezielter einsetzen. So war effektiveres Arbeiten möglich, da sich die Skill-Sets nicht doppelten. Dennoch war das Projektteam gelegentlich darauf angewiesen, Aufgaben genauer zu definieren, sich abzusprechen und eng zusammenzuarbeiten.
Bei der Auswahl der Middleware wurde sich auf die Grundlagen und die möglichst problemlose Integrierung mit der Game Engine Unity konzentriert. Es war wichtig, über die Unity Audio Engine Funktionen hinaus weitere Möglichkeiten der Soundbearbeitung zu haben.

3.5 Foleyaufnahmen

Die Foleyaufnahmen wurden im Studio und damit in einem kontrollierten Umfeld erstellt. Hier war das Ziel, möglichst wenig Raumgeräusche auf den Aufnahmen zu haben, um diese neutral verwenden zu können. Wenn die Aufnahmen möglichst viel, bis ausschließlich, Direktschall einfangen, kann später je nach Anforderung das Geräusch im Raum platziert werden. Es können ohne Editieren der Datei beispielsweise Hall-Effekte, Delays, IATDs o.Ä. zur Erweiterung angewandt werden. Den Diffusschall, also den Raum, auf einer Ton-Aufnahme wegzurechnen oder anzupassen, erweist sich als aufwändigere Aufgabe.
Die Entfernung vom Mikrofon richtet sich nach dem jeweiligen Objekt und aus dem Kontext, in dem man das aufgenommene Geräusch normalerweise hören kann.
Durch die Aufnahme von Foleys ergibt sich die Möglichkeit, Geräusche eines Objekts einzeln aufzunehmen, ohne Umgebungsgeräusche

3.5.1 Beispiele

Bei der Auswahl und Aufnahme der Foley-Aufahmen wurde vorwiegend konzeptionell vorgegangen. Einige Geräusche, das Klicken eines Knopfes oder das Splittern einer Vase sind recht klar und eindeutig reproduzierbar. Für die Aufnahmen kann das Klicken eines Knopfes und das Zerbrechen von Glas oder Keramik verwendet werden. Bei beiden Beispielen sind Herkunft und Repräsentation direkt erkennbar und gut einzuordnen.
Andere Geräusche, die kein tatsächliches Pendant in unserer physikalischen Welt haben, wie die der schwebenden Inseln und der knackenden Glas-/Eisschollen oder das Brummen der Höhle, orientieren sich nur grob an realen Geräuschen. Die Anforderung, dass die großen Glas-/Eisplatten eindeutig räumlich zuzuordnen sind, besteht nicht. Diese Geräusche müssen nicht die Anforderung erfüllen, direkt wiederzuerkennen zu sein.
Legt die Konzeption den Fokus auf die Wirkung, lässt sich aus vielen einzelnen Clips das erwünschte Geräusch erstellen. Hierzu werden einige scheinbar unzusammenhängende Geräusche überlagert und in verschiedenen Anteilen zusammengemischt. Diese Geräusche stammen von Gegenständen, die ohne größere Umstände aufgenommen werden können.
Am Beispiel der schwebenden Eisschollen, die zerbrochen werden können, handelte es sich um:

- verschiedene Ketten, die über Glas und Metall gestreift und abgezogen wurden
- ein Blechschild
- verschiedene Spraydosen unter Druck
- Plastik- und Aluminiumfolien
- Knallbrause, auch Knisterzucker genannt, der bei Berührung mit Wasser zischelt und knistert

3.6 Aufbereitung der Soundfiles für Verwendung in Wwise

Nachdem alle Aufnahmen gemacht wurden und ausgewählt wurde, welche Files man verwenden möchte, geht es an die Aufbereitung und Bearbeitung der Aufnahmen. Damit alle Clips ohne Verzögerung abgespielt werden können, sollten die Fades so kurz wie möglich sein. Die Ein- und Ausblendungen müssen am Anfang und Ende jedes Clips gesetzt werden. Alle herausstechenden Charakterisierungen, die einen Clip von den anderen unterscheidbar machen, sollten genau geprüft werden. Sollen die Clips in Typus-Gruppen sortiert werden, sollten sie möglichst ähnlich klingen. Jedes Alleinstellungsmerkmal ist bei wiederholt abgespielten Gruppen erkennbar. Durch die Vermeidung der Verwendung von zu charakteristischen Clips, lassen sich auch kleinere Gruppen erstellen. So ist weniger Material notwendig um einen großen Zeitraum durch wiederholtes Abspielen der Clips abzudecken. Diese Gruppen können wiederum in Kategorien eingeordnet werden. Zur Implementierung in Wwise ist es sehr praktisch, wenn die Arten und ihre Zusammengehörigkeit schon festgelegt sind.
Beispielsweise wurde bei Nanna das Feuer mit Clips unterschiedlicher Typen angesiedelt. Es wurde in die Typen “Lodern”, “lautes Knacken” und “konstantes Knistern” unterteilt. Innerhalb dieser Typen wurde zwischen hochfrequenten und tieffrequenten Clips unterschieden. Diese Unterscheidung konnte durch weiteres Bearbeiten der Frequenzspektren verstärkt werden. Durch Hoch- und Tiefpassfilter konnten die jeweilig charakterisierenden Anteile herausgearbeitet werden.
Damit eröffnet sich beispielsweise die Möglichkeit, das Knistern erst ab einer bestimmten Nähe zum Feuer hörbar zu machen. So wird die Abdämpfung hochfrequenter Geräusche simuliert, ohne andere Anteile des hörbaren Feuers zu beeinflussen. Das “Knacken” kann durch die Gruppierung in größeren zeitlichen Abständen aktiviert werden, während das “Lodern” durchgehend aneinander gereiht wird. Die einzelnen Teile sind zeitlich unabhängig voneinander und können so immer neue Kombinationen gleichzeitig abgespielter Geräusche bilden. So ist kein Loop erkennbar, falls sich Nutzer*innen auf unbestimmt lange Zeit an der gleichen Stelle, in dem Fall direkt neben einem Feuer, aufhalten.
Diese Art der Aneinanderreihung erfordert keine Angleichung der Schnittstellen, wie bei der Erstellung von Loops. Das ermöglicht zu jedem Zeitpunkt Clips zu ändern oder hinzuzufügen, ohne Bestehende abzuändern.
Wwise bietet die Möglichkeit, die Geräusch-Kombinationen probezuhören und zieht bei jeder Instanz die Clips zufällig (wie eingestellt) aus den Ordnern. Die Schwierigkeit ist hier, dass sich die Kombinationen nur für ein einzelnes Objekt abhören lassen, unabhängig der in der Umgebung platzierten Objekte oder Atmos.

3.7 Skizzierung der Atmo

Für die Bildgestaltung war das Erstellen einer Atmo-Skizze äußerst hilfreich. Sie zeigt ein mögliches Bild der fertigen Atmo. Es wird ein erster Eindruck und eine Annäherung daran geschaffen, wie eine Szene in der späteren Anwendung klingen könnte. Eine Atmo-Skizze übermittelt die Idee der Tongestaltung.
Die aufgenommenen Foleys wurden hierzu überlagert und in einer vollständigen Mischung angeordnet. Dies spiegelte die Wirkung und den Ablauf aller gleichzeitig in der Anwendung vorkommenden Geräusche wider.
Solche Skizzen wurden in Protools, einer Digital Audio Workstation erstellt und in Stereo gemischt. Dies hatte vor allem einen zeitlichen Vorteil, da schnell Änderungen durchgeführt und auf Feedback angepasst werden konnte. Durch das Herunterbrechen auf zwei Kanäle ist eine klassische Mischung möglich. Diese ist repräsentativ für eine bewegungslose Szenerie in der Anwendung. Bewegungslos in dem Sinne, dass die Mischung aus einer statischen oder vordefinierten Position oder Bewegung besteht. Vergleichbar mit schon festen Kamerfahrt. Die interaktive Erfahrung der Mischung durch eigene Positionsänderung oder Kopfbewegungen ist dabei noch nicht möglich.
In dieser Mischung können nach dem Editieren der Sound Clips erste Versuche angestellt werden, wie sich alle verwendeten Sounds balancieren lassen. Basierend auf dieser Erkenntnis, kann man Foleys recht stark durch einen EQ in ihrem Frequenzspektrum anpassen, um zu vermeiden, dass bestimmte Frequenzbereiche sich zu sehr überlagern und "ausgelastet" sind. Als "ausgelastet" gelten jene Bereiche, die in mehreren Geräuschen präsent sind und sich überdecken, sodass die einzelnen Sounds schwerer voneinander zu unterscheiden sind. Eine solche Skizze der verschiedenen Räume hat sich als hilfreich gezeigt.
Einerseits um die Bearbeitung der Files zu erleichtern, zum anderen wurde die Kommunikation zwischen Departements verbessert, da die Skizze die richtige Wortwahl den Ton betreffend erleichterte. Im Projekt Nanna hatten alle Beteiligten bereits Erfahrungen im Audio-Bereich gesammelt. Ein Grundverständnis von audio-spezifischen Begriffen war gegeben. Gleichzeitig erleichtert eine solche non-verbale Kommunikation das gegenseitige Verstehen.

Die Skizzen bestanden vor allem aus Flächen und wurden von Objekt-Geräuschen ergänzt. Die Positionierung im Stereofeld repräsentierte die räumliche Anordnung.
Das Panning und der Hall gaben eine Idee wie es sich später im Raum anhören könnte. Als Grundlage für kreative Entscheidungen stellte sich dieses Vorgehen als ausreichend heraus. Um einen Eindruck zu schaffen, ist eine der Audioskizzen im Anhang dieser Arbeit angefügt. Diese ist eine der ersten Repräsentationen des Raums, der die kalte, glatte und sterile Umgebung repräsentiert.

4 Implementierung des Sounddesigns in Wwise

In diesem Kapitel werden ausgewählte Prozessschritte beschrieben, die in dem Authoring Tool stattfinden, um das erstellte Material, unter Berücksichtigung der Konzeptionsgrundlagen zum fertigen Sounddesign zu verarbeiten. Ein Authoring Tool oder Audio Middleware ist eine Software, die als Zwischenschritt zwischen dem klassischen Sounddesign und der Games Programmierung existiert. Somit bietet sie eine Alternative zur Audio-Programmierung, bei der die Verarbeitung in Skriptform, statt auf einer grafischen Benutzeroberfläche stattfindet.

In Wwise können eigene Projekte erstellt werden, aber auch bereits bestehende Unity-Projekte integriert werden. Wwise kann unter anderem dazu verwendet werden, um Events zu erstellen, Geräuscheigenschaften und Quellen zu definieren, die Positionierung und Richtung der Sounds anzugeben, sowie Kanalbelegungen festzulegen, Mischungen vorzunehmen und Soundbanks zu organisieren.

Folgende Elemente bilden die Grundlage des Workflows:

- *Audio Assets* sind Mikrofonaufnahmen oder bereits existierende Dateien, die in einer DAW editiert und bearbeitet und im gewählten Format (bei uns .wav, 48kHz, mono) importiert werden.
- *Events* sind Audio Assets, die in Wwise benannt und eingeordnet und deren Verhalten festgelegt wird.
- *Trigger* werden festgelegt um zu definieren welche Situationen zur Auslösung eines Events führen.
- *Die Übergabe an Unity* erfolgt durch Skripte, die von Unity erkannt werden können um die Zuordnung und den Zugriff auf die Audio Assets zu ermöglichen.

4.1 Allgemeine Funktionen der Software

Alle Informationen zu Geräuschen und Bewegung, die in einer Anwendung vorkommen, werden in einem einzelnen Projekt in Wwise zusammengefasst. Daraus folgt: Wwise ist eine rein projektbasierte Software.
Innerhalb eines Projekts hat man verschiedene Möglichkeiten, die Informationen aus der Game Engine und importierte Dateien zu manipulieren. Dazu gehört das Ordnen von Geräuschen, sowie von Dialog und Musik. Letztere spielen im Projekt Nanna jedoch eine kleinere Rolle. Aufgrund der vielen manipulierbaren Parameter haben wir uns dazu entschieden, die Anwendung auf ein Feld zu begrenzen. Wir konzentrierten uns auf den Bereich der Geräusche.
Die importierten Dateien werden in einem Projekt in Objektkategorien eingepflegt. Dazu muss zuerst festgelegt werden, welche Eigenschaften sie besitzen sollen und wie sie sich verhalten, wenn sie gerufen werden. Das Verhalten beim Abspielen umfasst Reihenfolge und Zufälligkeit. Eigenschaften können Lautstärke, Filter und Delays sein.

Das Definieren der Eigenschaften und des Verhaltens wird “authoring” genannt.
Um Objekte zu rufen kann man Events erstellen. Events sind Ereignisse, die gerufen werden, wenn im Spiel bestimmte Trigger ausgelöst werden. Diese Trigger lösen meist eine Reaktion der Objekte in der Anwendung, sowie synchron des Sounds in Wwise aus. Diese Trigger können von Spielmechaniken oder vom Listener ausgelöst werden.
Alle Objekte können entsprechend ihrer Priorität eingeordnet werden. Sounddesigner*innen legen fest, welche Geräusche für Nutzer*innen relevanter sind. Wwise funktioniert durch Ordner-Strukturen in denen alle Eigenschaften, Variationen und Verknüpfungen der Audio-Files definiert werden können.

4.2 Verbindung mit Unity

Für das Projekt Nanna wurde mit der Version 2019.2.7f2 von Unity gearbeitet.

4.2.1 Programmierseitig

In den Skripten im Unity-Code können alle Objekte, die direkt mit Wwise interagieren, durch die vorangestellte Kennzeichnung “AK” identifiziert werden. Die Abkürzung “AK” ist abgeleitet von Audio$K$inetic, dem Hersteller der Wwise Software.

In jeder Anwendung gibt es sogenannte Audio Emitter, die Geräuschquellen repräsentieren, und Audio Listener, die die Position des gespielten Charakters oder der Kamerasicht einnehmen. Die Geräuschquelle kann durch die Bezeichnung als AkGameObject erkannt werden. Dieses “Audiokinetic Game Object” steht in Verbindung mit allen Events, die sich auf das dazugehörige Geräusch beziehen.
Das “Audiokinetic Audio Listener”- Objekt, also der virtuelle Zuhörer, stellt den Bezug zwischen Zuhörenden und Game Object her. Man kann den Listener als Mikrofon sehen, das alle aktiven Events in Positionsbezug und Richtungsbezug im dreidimensionalen Raum bringt.
Während die Anwendung läuft, werden die Koordinaten der oben genannten Objekte immer wieder verglichen, um eine aktive Umgebung zu reproduzieren. Im Bereich der Programmierung ist besonders wichtig, dass die Informationen, die Wwise von der Position des Listeners hat, immer aktuell übertragen werden können.
Da die Aufgaben der Programmierer*innen und Sounddesigner*innen durch die Middleware klar aufgeteilt werden, gibt es einige Aufgaben, die zur Überbrückung notwendigerweise erfüllt werden müssen. Diese wiederholen sich für alle Events und können somit leicht von Sounddesigner*innen erlernt oder Programmier*innen übernommen werden, ohne jeweils zuviel Zeit in Anspruch zu nehmen.

An dieser Stelle bot sich für unser Projekt eine gute Chance, zeitliche Puffer einzubauen. Da sich die Zuständigkeit für diese Integrierungs-Tasks etwas überschnitten hat, konnte die Integrierung von der Person übernommen werden, die mehr Spielraum im Zeitplan hatte. Gleichzeitig konnten aber beide Departements weitgehend zeitunabhängig voneinander

weiterarbeiten, da die Schnittstelle jederzeit hergestellt werden konnte und Wwise und Unity unabhängig voneinander ihre jeweiligen Arbeitsvorgänge vervollständigen konnte.

4.2.2 Integrierung eines bestehenden Unity-Projekts mit Wwise

Einen Großteil des Setups nimmt einem der Wwise Launcher ab. Sobald dieser heruntergeladen ist, kann man mit ihm die richtige Wwise Version installieren. Die kostenlose Lizenz ist mit wenigen Klicks und ohne Wartezeit beantragt, und kann dann direkt verwendet werden, wenn angegeben werden kann, dass kein Gewinn aus dem Projekt hervorgehen soll. In der nicht-kommerziellen Version ist man auf 200 Assets pro Soundbank beschränkt, bei Nanna haben wir uns dieser Zahl trotz Vollversion nicht angenähert. Der Launcher schlägt dann diverse Plugins vor, darunter Wwise-interne, sowie Plugins von Partnern wie izotope. Mit der Basisversion ohne Plugins lässt sich alles in dieser Arbeit Beschriebene durchführen. Die Software bietet die Möglichkeit über die Wwise API eine Integrierung in allen existierenden und selbst-gecodeten Game Engines zu schaffen. Unity und Unreal Engine lassen sich ohne API-Lösung integrieren.

Bei der Integration wird ein Wwise Ordner im Unity Projekt angelegt. Während der Installation bietet sich die Möglichkeit das notwendige *WwiseGlobal Object*, sowie den Listener, der direkt auf der Hauptkamera seine Default-Position bezieht, im Unity-Ordner automatisch anlegen zu lassen. Ist die Integration erfolgreich, kann man in Unity über *Windows* den *Wwise Picker* rufen und hat damit Zugriff auf die Ordnerstrukturen in Wwise. Es besteht die Möglichkeit, die Events und deren Namen zu finden und schneller Objekte dafür zu erstellen. Wenn im Unity *Project Tab* die Szene geöffnet wird, für die Sound erstellt werden soll, kann man im automatisch erzeugten *WwiseGlobal Tab* Einstellungen zu den von Wwise genutzten Ressourcen (Memory Pool) vornehmen.

4.3 Arbeits-Strukturen anlegen

Wwise legt Files in einer hierarchischen "Parent-Children"-Struktur an. Das kann man sich als eine Art Baum-Diagramm vorstellen, bei dem im untersten "Child" die Audio Files liegen. Immer wenn ein neues "Child" erstellt wird, ist dieses einem bestimmten "Parent" zugeordnet. Bei vielen Einstellungen hat man die Möglichkeit diese global, auf alle Objekte anzuwenden, oder für das dazugehörige *Parent*-Objekt und damit auch alle gleichgestellten *Child*-Elemente des gleichen Typs zu übernehmen. Ebenfalls besteht die Möglichkeit, Einstellungen nur für ein oder mehrere *Child*-Elemente festzulegen. Diese Hierarchie kann man als Möglichkeit nutzen, Übersichtlichkeit zu schaffen, indem alle Geräusche des gleichen Typs im gleichen *Parent* zu finden sind. Ein Beispiel für eine Einteilung kann sein:

```
> Umgebungsgeräusche
> Charakter
    > Dialog
    > Bewegungen
        > Kampfgeräusche
        > Werkzeug
        > Schritte
            > Sumpf
                > rechts
                > links
            > Stein
                > rechts
                > links
            > Holz
                > rechts
                > links
```

Die Schritte werden verschiedenen Untergrund-Typen zugeteilt. Unter diesen gibt es jeweils Rechte und Linke und seltenere, schwere und viele, leichte Tritte. Das ermöglicht es, die Auswahl für sequentiell Abgespieltes, wie Schritte, genau zu definieren:
Es kann zum Beispiel definiert werden, mit welcher Häufigkeit und Reihenfolge die Schritte abgespielt werden und unter welchen Umständen die Geräusche des gleichen Typus wiedergegeben werden sollen.
Auch bei weniger differenzierten Unterteilungen kann es sich lohnen, Foleys beispielsweise nach Frequenzspektrum einzuteilen. So lässt sich gewährleisten, dass die Frequenzanteile ausgewogen sind, auch wenn zufällig ausgewählt wird. So kann ein endlos fortlaufendes, immer anderes Geräusch entstehen.

Die Geräusch-Hierarchie wird im *Actor-Mixer* organisiert. Hier werden alle Geräusch- und Bewegungseigenschaften festgelegt.

Hat man Musik in der Anwendung, so wird diese separat in der *Interactive Music* Hierarchie angelegt. Das ermöglicht es Nutzer*innen musikspezifische Parameter gezielt zu ändern. Einstellungen sind möglich in Hinsicht auf Tempo und möglichen Grids, sowie "Stinger" zu definieren. Stinger sind kurze Musik-Einspieler, die nicht Teil einer festen Komposition sind und der Verstärkung bestimmter Effekte und damit in Verbindung stehender Emotionen dienen. Die Stinger sollen häufig als Teil der Musik wahrgenommen werden, oder sie werden genutzt, um Wechsel zwischen Musikstücken zu überblenden. Ein Stinger kann ein Gong, ein dramatischer Akkord, oder ein Trommelwirbel sein (Somberg, 2017).

Übergreifend regelt der *Master-Mixer* das Routing und die Outputs der beiden oben beschriebenen Mixer. In diesem Mixer kann beispielsweise festgelegt werden, dass alle Raumgeräusche, die im *Actor-Mixer* angesiedelt sind, durch ein Reverb geschickt werden, aber die Musik aus dem Hauptmenü, die im *Interactive Music* Mixer liegt, trocken bleiben soll. "Trocken" bezeichnet hier das Fehlen aller Hall-Effekte und Reflexionen, es handelt sich also nur um den Direktschall.
Jedes Projekt hat mindestens einen Bus, den *Master-Mixer-Bus*, durch den alle Sounds laufen, die abgespielt werden. Hier lässt sich auch das Sound-Level in dB anzeigen. Die Lautstärke kann man zwar auch in den einzelnen Clips anzeigen, da aber die Audio-Files den *Actor-Mixer* nie durchlaufen, gibt es dort kein Level-Meter.

4.3.1 Aufbau einer Session

Im *Wwise Project Explorer* lassen sich die vorbereiteten Files in verschiedene Strukturen einordnen.
In jedem neuen Projekt gibt es bereits eine Default Work Unit. Eine *Work Unit* ist die kleinste Einheit, die die Basis für jedes Element, das angelegt wird, bildet. Die *Default Work Unit* kann auch leer bleiben, wenn neue *Work Units* angelegt werden, um spezielle Hierarchien zu integrieren oder die Aufgabenteilung zwischen mehreren Sounddesigner*innen zu definieren. Allen *Work Units* können eigene Farben zugeordnet werden.

Der *Project-Explorer* ist in sieben Tabs aufgeteilt. Anhand dieser Tabs werden im Folgenden die grundlegenden Funktionselemente erklärt.
An dieser Stelle wird nicht auf spezifische Einstellungen eingegangen, sondern nur auf die Möglichkeiten der Strukturierung.
Wie diese Einstellungen verwendet werden, um bestimmte akustische Effekte zu erzielen, wird im folgenden Kapitel "Spezielle Funktionen" näher beschrieben.
Die Tabs, die in unserem Projekt häufiger zur Anwendung kamen, werden im Verlauf der Arbeit detaillierter beschrieben.
Durch Doppel-Klick auf das gewünschte Element öffnet man den dazugehörigen Editor, falls ein solcher jeweils existiert, um die Eigenschaften der Elemente zu ändern.

- *Audio*

  Im Audio Tab sind die oben genannten *Mixer* angesiedelt. Hier werden Audio Dateien importiert und Busse oder Outputs geroutet. Außerdem können verschiedene Typen von Containern angelegt werden: Random, Sequence, Switch und Blend Container. Erst wenn im *Audio Tab* die Existenz und Hierarchie dieser Elemente definiert ist, können sie in den anderen Tabs weiterverwendet werden.

- *Events*

  Im Events Tab sind alle erstellten Events gelistet. Durch Drag und Drop lassen sich diese in andere Elemente integrieren oder mit ihnen verknüpfen. Das ist speziell in Verbindung mit den Soundbanks eine schnelle Möglichkeit, dafür zu sorgen, dass alle benötigten Events in den richtigen Soundbanks liegen.

- *Soundbanks*

  Hier können neue Soundbanks erstellt werden. Die meisten Tätigkeiten der Soundbanks fanden bei Nanna allerdings im *Soundbank Manager* statt, hierzu im folgenden Kapitel mehr.

- *Game Syncs*

  In diesem Tab können bei Bedarf Switches, States und Trigger belegt und bearbeitet werden, nachdem sie im Actor Mixer angelegt wurden. Das ist dann nötig, wenn bestimmte Areale oder Zustände verschiedene Geräusche erfordern, und Audio-Clips aus unterschiedlichen Ordnern gezogen werden sollen.

- *ShareSets*

  Die Effekt-Plugins, die bei der Installation ausgewählt wurden, sind in dem *ShareSets* Tab aufgeführt. Alle Effekte können aber auch im zum Clip gehörigen *Property Editor* unter *Effects* gerufen werden. Wie sich bei der Arbeit an Nanna gezeigt hat, waren für die grundlegende Verwendung von Attenuation und Effekten, die von der Software vorgeschlagenen Einstellungsmöglichkeiten ausreichend.

- *Sessions*

  Für größere Projekte kann es durchaus notwendig sein, während der Produktion den aktuellen Stand vorab testen zu können. Hierzu kann Wwise Simulationen des Projekts in einer *Soundcaster Session* erstellen. Unterschiedliche Konzepte können probeweise gemischt und verglichen werden. Ebenso können *Events* getestet werden. Der *Soundcaster* kann sich mit *Game Syncs* sowie Objekten und *Events* verbinden.

- *Queries*

  Der *Queries* Tab gibt eine Übersicht über alle bisher erstellten *Queries*, die nach selbst definierten Kriterien bestimmte Objekte finden können. Diese Suchfunktion

zu nutzen ist im Besonderen dann hilfreich, wenn mehrere Leute an einem großen Projekt arbeiten, man nicht alle Objekte selbst angelegt hat, aber sich einen Überblick verschaffen muss.

Für Nanna wurden speziell in den *Audio*- und *Events*-Tabs Objekte angelegt und strukturiert. Da wir mit nur einer Soundbank gearbeitet haben, waren im dazugehörigen Tab keine Hierarchien zu erstellen. Die Funktionen der restlichen Tabs ließen sich objekt-spezifisch rufen, da wenig globale Einstellungen und Übersichtsstrukturen notwendigen waren.
An dieser Stelle war es sehr hilfreich, dass man "Shortcuts" nutzen kann und für kleinere Projekte nicht Hierarchien erstellen muss, die ungenutzt bleiben.

Container
bieten die Möglichkeit ausgewählte Audio-Dateien in Beziehung miteinander zu setzen. Dies geschieht unabhängig von der Gruppierung nach Typ. Eine Einteilung nach Typ erleichtert den Umgang mit Containern.
Bei einer Ordnung wie beispielsweise:
Umgebung > Höhle > Wasser > Tropfgeräusche
können innerhalb dieser Einteilung für beliebig viele *Work Units* in denen Audio Dateien abgelegt sind, folgende Container angelegt werden:

- *Random*
  Random Container ermöglichen es, alle in ihnen enthaltenen Dateien in zufälliger Reihenfolge wiederzugeben. Außerdem kann definiert werden, wie oft ein Geräusch nicht mehr ausgewählt werden darf, nachdem es abgespielt wurde.
  Das kann Wiederholungen und Mustererkennung der Nutzer*innen vorbeugen und je nach Ähnlichkeit und Menge der im Container enthaltenen Audio Clips entschieden werden. Sind sich die Clips ähnlicher, kann die "Wiederholungssperre" kleiner gesetzt werden. Gibt es eine große Anzahl von Clips, kann die "Wiederholungssperre" größer gesetzt werden, da dann nicht die Gefahr besteht, dass die Clips in mehrmals in der gleichen Reihenfolge abgespielt werden.

- *Sequence*
  Alle in diesem Container enthaltenen Objekte können in einer bestimmten Abfolge angeordnet und wiedergegeben werden. So lässt sich eine Endlosschleife erstellen, die im gleichen Muster wiederholbar abgespielt wird. Am Ende einer so erstellten Playlist muss definiert werden, ob zum ersten Objekt des Containers gesprungen wird oder die Liste nach oben, in umgekehrter Reihenfolge wiedergegeben wird.
  Der Vorteil gegenüber einem zusammengeschnittenen Clip, für den zuvor die Audioaufnahmen in einer DAW arrangiert und in einer Datei herausgerechnet wurden, liegt in der Flexibilität.
  Wird beispielsweise ein Mechanismus gestaltet, dem zusätzliche sichtbare Elemente hinzugefügt werden, können die dazugehörigen Geräusche auch dem Sequence Container hinzugefügt werden, ohne dass das zuvor angelegte Material verändert werden

muss. Auch können so Verzögerungen oder Reihenfolge angepasst werden, um im Nachhinein das Gehörte dem Visuellen anzupassen und zu synchronisieren.

Die verschiedenen Arten der Container lassen sich miteinander kombinieren. So ist es möglich, einen Sequence Container mit mehreren Random Containern zu füllen. Diese enthalten wiederum Audio Clips verschiedener Typen.
Dadurch lässt sich beispielsweise immer ein Knall, gefolgt von einem Einschlag, gefolgt von herunterfallenden Steinen abspielen. Die einzelnen Elemente klingen aber bei jedem Durchlauf unterschiedlich, und kommen in unterschiedlichen Kombinationen vor.

- *Switch* Container werden immer in Kombination mit einer *Switch Group* verwendet. In der Gruppe sind die verschiedenen Switches, also die möglichen Stadien die der Switch einnehmen kann definiert. Diese Stadien können im Switch Container angewählt und befüllt werden.
  So kann beispielsweise jede Lichtstimmung mit einem Switch verbunden sein und jeder Farbe ein anderes Grundrauschen zugeordnet werden. Die Switches werden von *game parameter* gesteuert. Dabei werden Informationen aus der Anwendung zurückgeleitet und deren Einfluss kann durch die sogenannten RTPC (Real-time Parameter Controls) Einstellungen definiert werden.

- *Blend* Objekte, die in Blend Containern abgelegt sind, werden alle gleichzeitig aktiviert oder abgespielt. Das können einzelne Audio Clips sein, oder weitere Container. Durch bereits erwähnte Integrierung von RTPCs lassen sich Eigenschaften definieren, die es erlauben, das sich nicht alle Geräusche überlagern, sondern diejenigen stärker hervortreten, die relevant zur Handlung sind.
  Die RTPC Definition erfolgt immer mithilfe einer Kurve. Diese stellt das Verhältnis des ausgewählten *Game Parameter*s zur jeweiligen Einstellung, wie zum Beispiel dem Lautstärkepegel, dar. Wird in der Anwendung eine Tageszeit festgelegt, kann diese als Game Parameter beeinflussen, welche Geräusche aus dem Blend Container zu hören sind, um Objekten morgens und abends unterschiedliche Eigenschaften zuzuordnen.

Es ist hervorzuheben, dass die oben genannten Funktionen dazu beitragen, die Tongestaltung realistischer zu machen. Durch Erstellung von Ordnung und Unordnung kann die Wahrnehmung der Geräusche wesentlich beeinflusst werden.
Auch wenn das Ziel oft darin besteht, neue wissenschaftliche Methoden anzuwenden, um das Verhalten von Schall nachzubilden, ist es notwendig eine unvorhersehbare Zufälligkeit beizubehalten. Diese Zufälligkeit ermöglicht, dass das Audio als authentisch wahrgenommen wird, als würde es, wie in der echte Welt von vielen einzelnen Faktoren beeinflusst werden (Schütze & Irwin-Schütze, 2018).

Wie bereits zu Beginn erwähnt, beruht die Funktion von Wwise zu einem großen Teil auf dem Strukturieren und Definieren der Work Units.

Neben den Work Units, die die Bausteine bilden, sind die Container die Verbindungsglieder dazwischen. Sie ermöglichen es, in kleinen Einheiten Eigenschaften zu definieren und Abläufe zu programmieren. Die Funktion der Container ist an jedem Punkt überprüfbar. Durch den Play-Button aktiviert ein Container alle Abläufe und es lässt sich wiederholt überprüfen und simulieren, wie der Ablauf in einer Anwendung sein könnte.
Durch präzise Konzeption und Klärung der Abläufe, sowie Aktionen und Reaktionen mit Objekten, lässt sich aus jedem Schritt eine *Parent-Children* Abhängigkeit mit den oben genannten Elementen bilden. Diese Vorgehensweise unterscheidet sich deutlich von der linearen und zeitdefinierten Arbeitsweise in einer klassischen DAW. Entscheidungen müssen oft vordefiniert sein, um einen Ablauf überhaupt zu implementieren. Hier ist das Arbeiten dem Programmieren deutlich ähnlicher, jedoch immernoch geprägt vom Ausprobieren und Probehören.

4.3.2 Leistung und CPU-Auslastung

Um zu gewährleisten, dass man das erstellte Sounddesign auch problemlos nutzen kann, gibt es die Möglichkeit, die Menge der Assets und die aktive Verwendung aller Assets im Blick zu haben. Durch die Einstellungen zum Memory Pool und den Soundbanks kann die Auslastung überwacht und gesteuert werden.

Die Memory Pool Größe und die Anzahl der Voices kann zur Entlastung der CPU angepasst werden. Die Default Größe für Memory Pools beträgt 16 MB, das ist für die meisten Spiele ausreichend, laut Audiokinetics Dokumentation der Software.
Erst, wenn man alle Sounds implementiert hat, kann man herausfinden, wieviel Speicher aktiv benötigt wird.
Werden viele Geräusche gleichzeitig aktiviert, wird deutlich mehr Speicher benötigt, als in einer Anwendung, in der zwar einen großen Speicher von Audio Assets belegt wird, aber nur Wenige davon gleichzeitig abgespielt werden. Müssen außerdem noch Hall, Echo oder andere Umgebungseffekte errechnet werden, erhöht sich der Unterschied noch.
Im *Total Used Memory Value* hat man die Möglichkeit, sich anzeigen zu lassen, wieviel Speicherplatz im Memory Pool maximal (Peak Used) benötigt wird. Man kann diesen entsprechend herabsetzen, um den genutzten Speicherplatz optimal zu nutzen.
Nicht zu verwechseln sind diejenigen Voices, welche für alle Geräusche, mit gleichem Abspielzeitpunkt stehen, mit den Voice Objekten, die erstellt werden können, um Dialog in verschiedenen Sprachen anwählen zu können.

Um den benötigten Speicherplatz möglichst klein zu halten, kann man entweder für alle Sounds einzeln, oder für Gruppen von ähnlichen Audio Files die Anzahl der Voices limitieren.
Die minimal nötige Nähe der Geräuschquelle zum *Listener*, bei der die Geräuschquelle hörbar ist, kann durch Priorisierungen und Einstellungen definiert werden. Dies zu definieren gewährleistet, dass in Szenarien, in denen viele Soundquellen die verfügbare Menge an

Voices belegen, nur diejenigen aussortiert und nicht abgespielt werden, die vernachlässigbar sind.
Beispielsweise könnte statt des inneren Monologs des Protagonisten, eher das dritte oder vierte Schussgeräusch, das überlappend aktiviert wird, vernachlässigt und nicht abgespielt werden, oder die Stimme eines sehr weit entfernten Passanten, der nicht mehr im Blickfeld ist. Auf Letzteren ist mehr Wert zu legen, da der Monolog wichtige Informationen für Nutzer*innen bieten kann, die auch bei technischer Begrenzung nicht fehlen dürfen.

4.3.3 Soundbanks

Im Soundbank Manager können neue Soundbanks generiert, benannt und via "drag and drop" befüllt werden.
Im Edit Tab wird detailliert aufgelistet, welche Inhalte in der Soundbank vorhanden sind. Die eingefügten Inhalte können im Nachhinein wieder abgewählt werden, bleiben aber trotzdem verlinkt. Das kann sich als hilfreich erweisen, wenn verschiedene Varianten eines Events zur Auswahl stehen aber noch nicht klar ist, welche am Ende verwendet werden soll.
Soundbanks müssen immer generiert werden und wieder neu generiert werden, wenn sie verändert werden. Alle in den Soundbanks enthaltenen Elemente werden zu einem Paket geschnürt, das von der Engine verwendet werden kann.
Die Engine lädt immer nur die benötigte Soundbank. Es ist allerdings möglich, bei kleineren Projekten auch nur mit einer Soundbank zu arbeiten.
Wenn man weiß, dass das Projekt von überschaubarer Größe ist, oder wie im Fall von Nanna nicht auf anderen Computern laufen muss, kann man diese Arbeitsschritte einsparen. Anderenfalls muss die benötigte Leistung so beschränkt werden, oder der Speicherplatz optimiert sein, dass auch Computer, die eine schwächere Leistung haben die Anwendung nutzen können. Das sollte von der vorgesehenen Nutzung abhängig gemacht werden.
Ein Vorteil des Arbeitens mit einer einzelnen Soundbank ist, dass man sicher gehen kann, dass mit der einen Soundbank immer alle benötigten Events und Audio Files geladen werden. Weiterhin muss in der Game Engine nicht festgelegt werden, wann "in-game" bestimmte SoundBanks geladen sein oder neu geladen werden müssen.
Um mehrere Soundbanks zu verwalten und zum richtigen Zeitpunkt zu laden ist es erforderlich, Tracking zu implementieren. Das Tracking muss auswerten, welche Sounds verfügbar und welche benötigt sind. Dies erfordert zusätzliche Programmierung in Unity.
Als Zwischenweg, in dem Fall, es werden verschiedene Level erstellt, kann für jedes Level eine Soundbank angelegt werden. Auf diese Weise lässt sich das System beliebig erweitern, und immer wenn ein neues Level geladen wird, wird auch die dazugehörige Soundbank von Unity gerufen. Die Spieler*innen könnten keine Events oder Sounds aus anderen Leveln durch ihre Handlung aktivieren und die Trennung zwischen den Soundbanks wäre deutlich.

4.3.4 Events

Events sind Auslöser bestimmter Aktionen. Sie fungieren als Bindeglied zwischen der Aktion in der Anwendung, die durch Eingreifen der Nutzer*innen ausgelöst wird und der Aktion die in Wwise gerufen wird, durch die Audio abgespielt wird.
Der Name eines Events muss auf beiden Plattformen identisch sein, damit zuordenbar ist, welches Event in der Game Engine gerufen, oder in Wwise verändert wurde.
Nachdem ein Event erstellt wurde, müssen Aktionen angelegt werden und Eigenschaften im Event Editor vorgenommen werden. Die simpelste Aktion ist das Abspielen eines Audio Clips. Es können aber auch Container aktiviert, sowie pausiert, gefiltert oder in ihrer Lautstärke verändert werden; darüber hinaus können auch Parameter aus der Anwendung, sowie Switches oder Trigger mit Events verknüpft werden.
Events definieren alle Aktionen, die von der Software ausgelöst werden. Daher müssen auch alle Events, die verwendet werden sollen, in die *Action List* aufgenommen werden. So kann zwischen verschiedenen Möglichkeiten einer Problemlösung gewechselt werden und es bleibt ersichtlich, welche Aktionen ausgelöst werden und welche passiv sind. Die Events aus der *Action List* werden dann mit den Soundbanks in die Game Engine transportiert, wo sie mit den gleichnamig definierten Game-Events interagieren können. So können die Programmierer*innen die Game Events schon vorher definieren und platzieren und den Auslöser festlegen. Daran ändert sich auch nichts, wenn in Wwise das Event noch einmal angepasst wird.

4.4 Spezielle Gestaltungsmöglichkeiten

4.4.1 Orientierung durch Geräusche

Bei der Orientierung durch Geräusche stellt sich die Frage, ob die Ortung im Vordergrund stehen soll, oder die Hinweise, die Nutzer*innen aufmerksam auf eine bestimmte Richtung machen sollen, um sich an eine vorbestimmte Position begeben zu können. Da steht sich die Neutralität der Repräsentation gegebenüber. Die Neutralität wird von einer perfekt realistischen Umgebung hervorgerufen, bei der keinerlei Illusion oder Manipulation im Spiel ist. Die Repräsentation ist eine Annäherung an die Realität, die aber auf vorbestimmten Schwerpunkten basiert (Rumsey, 2001).
In einem Versuch zur Aufmerksamkeit in einer VR-Umgebung wird festgestellt, dass Objekte, die mit Geräuschen in Verbindung stehen, deutlich attraktiver sind und ihnen mehr Beachtung geschenkt wird, als anderen Objekten, die nicht auditiv gestalten wurden (Rothe et al., 2018).
Wenn ein Objekt also sichtbar ist, aber keine merklichen Geräusche produziert, wird diesem deutlich weniger Aufmerksamkeit geschenkt.
Ist allerdings etwas nicht sichtbar, aber gut zu hören, leidet darunter nicht die Aufmerksamkeit. Hört man beispielsweise einen Helikopter in der Stadt, existiert dieser zweifellos. Eine Diskrepanz entsteht, wenn der Helikopter zu sehen ist und über einem fliegt, ohne dass er zu hören wäre. Dies verdeutlicht die Relevanz der Tongestaltung und edr Auswirkung auf Glaubhaftigkeit und Orientierungsmöglichkeiten einer Anwendung.
Wie Axel Berndt in einem Kapitel zur diegetischen Musik beschreibt (Grimshaw, 2011), wird auf den Bereich interaktiver Medien Diegese vor allem in ihrem engeren Sinne verwendet: sind Geräuschquellen im Raum des Geschehens platziert und alle Charaktere sowie die Nutzer*innen können die Geräusche hören, im Kontext der Umgebung, handelt es sich um ein diegetisches Geräusch.
Ist die Geräuschquelle kein im Raum platziertes Objekt, sondern beispielsweise ein Hinweiston, Earcon, begleitende Musik oder innerer Monolog, handelt es sich um ein nichtdiegetisches Geräusch.
Diese Unterscheidung ist bei interaktiven, also nicht-linearen Medien teilweise nicht möglich, so Berndt, da die Trennung zwischen innerem, diegetischen Ton und den Zuschauer beeinflussenden, nicht-diegetischen Ton nicht so klar zu machen ist. So kann ein nichtdiegetisches Geräusch, wie zum Beispiel eine Musik-Cue Spieler*innen vor nahender Gefahr warnen.
Da die Spieler*innen den Charakter steuern, reagiert dieser auf Informationen, die innerhalb der Realität des Spiels nicht existieren.
Darum möchte ich mich auch nur auf die engere Definition der Diegese beziehen, die sich auf die Positionierung der Geräuschquelle bezieht und nicht auf Information und Rezipient, also den Einflussbereich des Geräuschs.

Um Wwise zu ermöglichen, Sound Objekte in einem dreidimensionalen Kontext abzuspielen,

muss das sogenannte *3D positioning* im Actor Mixer angewählt werden. Dadurch wird die Distanz und die Richtung der Kopfbewegung im Raum miteinbezogen.

4.4.2 Atmo

Atmos, oder in der englischen Literatur Ambiences, sind diegetische und nicht diegetische Hintergrundgeräusche.
Atmos können also eine Mischung realer Geräuschquellen sein, seien es sichtbare, wie das Rascheln der Äste und Blätter in einem Wald, oder unsichtbare, wie das ferne Wasserplätschern in einer Tropfsteinhöhle, und ungegenständliche Geräusche, seien es Drones und Stings oder Collagen aus wirrem Flüstern.

Entscheidung der Ambience-Geräusche bei Nanna
Obwohl die Mehrheit der Geräusche eigens für das Projekt aufgenommene Foleys waren, mussten bei einzelnen Geräuschen, wie dem Feuer auf Libraries zurückgegriffen werden. Das Knistern der Feuerschalen wurde in einem Studio aufgenommen, mit Plastikfolie die zerknüllt und entfaltet wurde. Ein echtes, loderndes Feuer wurde allerdings nicht eigens entfacht. Zusätzlich waren auch wenige, von Instrumenten erzeugte Geräusche Teil des Ambiente.
Für den Teil des Raumes, der rau, warm und organisch klingen sollte, wurden diese besagten Adjektive in den Fokus gerückt und eigene Aufnahmen eines Kontrabasses hinzugefügt. Diese sollten die Aufgabe übernehmen, die Atmo aufzufüllen und gerade die tiefen Frequenzbereiche stark zu charakterisieren. Die Entscheidung fiel auf ein echtes Instrument, da sich die Möglichkeit bot und die Klangtextur eines hölzernen Instruments entspricht genau dem Konzept für diesen Teil des Raumes.
Es wurden lange Flächen gehaltener Töne, sowie einige nicht tonale Aufnahmen verwendet. Bei letzteren wurde der Musiker gebeten, soweit es ihm möglich war, einige Striche mit der Bogenrückseite, sowie langsame Glissandi, das stufenlose Gleiten zwischen Tonhöhen, einzuspielen. Dies erzeugte die erwünschte Klangtextur.
Die Verwendung der Kontrabassaufnahmen erfolgte ausschließlich in dem warmen, rauen Teil des Raumes, um eine stärkere Abgrenzung zur dünneren, höheren Wirkung der Atmo des anderen Raumteiles zu schaffen. Die tiefen Flächen bezeichnet man gemeinhin als *Drones*:

> The drone (bourdon) effect refers to the presence of a constant layer of stable pitch in a sound ensemble with no noticeable variation in intensity. (Augoyard, 2005)

Diese Definition nach Augoyard bezeichnet Drones als Effekt, der eine durchgehende, variationslose Fläche mit gleichbleibender Tonhöhe erzeugt. Weiterhin beschreibt er die Aufgabe der Drones als Hilfe eine Atmo zu kreieren, die “more comfortable”, also angenehmer ist. Dies sollte im Fall des Projektes dafür Sorgen, dass der Raum als wärmer wahrgenommen wird.

4.4.3 Raumbezüge herstellen

Reflektionen für jedes Objekt zu errechnen ist zwar eine Möglichkeit, gestaltet sich aber in Echtzeit eher schwierig. Daher eignen sich Real Time Reflekionen zum jetzigen Stand der Technik noch nicht für die Verwendung in VR.

Um eine aktuelle Einschätzung zum technischen Stand der Game Engines für VR von 2019 einzubringen:

> Eine physikalisch exakte echtzeitfähige Berechnung der Schallabsorption, -reflexion und -beugung durch beliebige Hindernisse stellt – ähnlich wie die Reflexion und Brechung von Lichtstrahlen – sehr hohe Performanzanforderungen und wird daher von Game Engines und Szenengraphsystemen nicht unterstützt. (Dörner et al., 2019)

Allerdings gibt es andere Möglichkeiten, Raumbezüge herzustellen, zum Beispiel das Einfügen eines passenden Halls, der eine annähernde Darstellung des Diffusschalls wiedergeben kann.
Hier kann die Entscheidung getroffen werden, ob man sich an dem Maßstab des virtuellen Raumes direkt orientieren möchte oder eine Verstärkung hervorrufen möchte. Soll der Raum eine bestimmte Wirkung bekommen, kann man als eine mögliche kreative Entscheidung mit einem Hall arbeiten, der "bigger than life" also beeindruckender als die Realität ist.
Ein Echo, also ein Nachhall, der durch seine starke Zeitverzögerung als eigenes Hörereignis wahrgenommen wird, kann vom Hall separiert genutzt werden.
Für bestimmte Ereignisse kann dieser Effekt verstärkend wirken. Beim Echo ist keine Lokalisierung nötig, da es keine Auskunft über die Position des Schallereignisses gibt, sondern über die enorme Weite des Raumes.
Um ein Beispiel zu nennen: Schlägt man in einer VR-Anwendung einen Stein an einen Brunnen und es wird durch dieses Ereignis ein Echo ausgelöst, werden die Dimensionen des Brunnens wahrnehmbar.
Wie bei der Wahrnehmung des Echos, das nicht direkte Rauminformationen sendet, ist es oft eine Frage des Kontextes und des gelernten und erwartenden Hörens.

Attenuation

Die Schalldämpfung tritt in Wwise unter dem Begriff *Attenuation* auf. Unter diesen Einstellungen findet man die Möglichkeit, das Signal in seiner Lautstärke zu dämpfen. Die Dämpfung steht im Bezug zur Distanz von Geräuschquelle und Ohr.
Um Objekten eine Position im Raum und einen Bezug zum Hörenden zu geben, muss sich die Lautstärke im Bezug zur Entfernung ändern. Dieser Effekt kann durch *Attenuation* gesteuert werden.
Bei kugelförmiger Schallausbreitung nimmt die Schallintensität mit dem Quadrat des Abstandes ab. Jede Verdopplung des Abstandes von der Schallquelle führt also zu einer Abnahme des Schallpegels um 6 dB (Raffaseder, 2010). Die Abnahme der Lautstärke

sollte beim Mischen überprüft werden, jedoch ist dies eine gute Annäherung für einen Ausgangswert, der in den meisten Fällen einer realistischen Wahrnehmung entspricht.

Frequenzfilter
Die Abstrahlung kann abhängig vom Winkel unterschiedlich gefiltert werden. Durch Übertragung von Schall durch die Luft können diese Filtereffekte entstehen. So kann nachempfunden werden, dass ein Objekt von vorne, von der Seite und von hinten eine unterschiedliche Klangcharakteristik besitzt. Für jeden Winkelabschnitt kann diese beliebig detailliert festgelegt werden.

4.4.4 Interaktionsgeräusche

Berührung und Kollision von Objekten und dem Nutzenden können mit Events belegt werden. Man kann entscheiden, welche Kollisionen ein Geräusch hervorrufen sollen: Die Controller, die die Platzierung der Hände einnehmen, oder der Listener, der auf der Position des Headsets, also des Kopfes liegt. Möglich sind auch beide Positionen.

Beispiel Kopf Würfel
Im Foyer befanden sich zwei Würfel, die beim Nähertreten leise Geräusche aussenden. Steckt man seinen Kopf in das scheinbar feste Objekt, hört man die Geräusche und sieht das Innere des Würfels.
Damit sollte die Möglichkeit verdeutlicht werden, dass je nach Positionierung des Kopfes andere Perspektiven wahrgenommen werden können.
Anders als beim gewohnten Konsumieren von Medien vor einem flachen Bildschirm, bei dem die Kopfbewegung selbst, keine Interaktion mit dem Inhalt ermöglicht.

Game Objekte
Objekte wie das oben genannte Beispiel können in beliebiger Größe in jeder Position verteilt werden. Die Events, die an Objekte geheftet sind, sind nicht raumabhängig. Das bedeutet, bewegt sich das interaktive Objekt durch den Raum, bleibt die Interaktionsmöglichkeit bestehen.

Unsichtbare Events
Game Objekte können auch unsichtbar sein und ohne Interaktionsmöglichkeit. Diese Objekte können als Raumtrennung dienen, wobei die definierten Raumobjekte, auf die bestimmte Atmos angewendet werden können, nicht mit den sichtbaren Raumgrenzen übereinstimmen müssen. Die mit ihnen verbundenen Events lösen bei der Kollision, also der Überschreitung der unsichtbaren Grenze eine Aktion aus. Diese Aktionen definieren dann die jeweiligen Umgebungsgeräusche.

4.5 Mischung

Ein Teil des klassischen Mixens besteht aus der Balancierung der einzelnen Spuren. Es wird festgelegt, zu welchem Zeitpunkt eine Spur welche Lautstärke besitzt. Das ist eine sehr vereinfachte Darstellung, da bei dem Prozess zu jedem Zeitpunkt Filterungen, Panning und andere Parameter eingestellt werden. Allerdings verdeutlicht diese Grundsätzlichkeit den Unterschied zum Mischen des Sounds in Virtual Reality.
Weder der Zeitpunkt, zu dem ein Audio-Objekt abgespielt wird ist reproduzierbar, noch die Lautstärke in welcher das Geräusch zu hören ist. Das liegt daran, dass sich beide Parameter auf Position im Raum, im Verhältnis zu Event-Auslösern und den geräusch-emittierenden Objekte beziehen. Diese Position ist frei wählbar und nicht exakt vorhersehbar. Es ist die Aufgabe der Mischenden, vorher festzustellen, welche Positionen besonders relevant sein könnten:
Schlüsselpositionen, die für das Weiterkommen in der Anwendung relevant sind, sowie Positionen die Auslöser für Events sind, als auch Positionen dazwischen, an denen sich mehrere Umgebungen, Atmos oder andere Geräusche überlappen können.
Einen noch stärkeren Kontrast zum klassischen Mischen von Musikstücken bildet das Prozedurale Erstellen von Audio.
Das Mischen der Geräusche in Nanna lässt sich dazwischen ansiedeln. In einigen Aspekten ähnelte es eher der linearen Mischung, in anderen der prozeduralen. In diese Unterteilung sollen verschiedene Elemente im Folgenden eingeordnet:
Das zuvor aufgenommene Material existiert sowohl bei einer linearen Mischung, als auch bei Nanna.
Abgesehen von dem Raum, in dem prozedurales Audio gezielt eingesetzt wurde, basierte alles auf Foleys und Clips aus Aufnahmen und Libraries. Für das Mischen des prozeduralen Raumes war ein anderes Teammitglied zuständig, deswegen wird dieser Part aus dem Vergleich ausgeklammert.
Die Zeitpunkte sind im prozeduralen Erstellen von Audio ebenso unklar wie bei Nanna. Bei beiden Varianten gibt es Parameter die es zu erfüllen gilt, um ein Geräusch zu platzieren oder zu erzeugen, die enger oder weiter definiert werden können.
Allerdings weiß man bis zum Moment der Wiedergabe nicht, wann das tatsächlich sein wird, da das Element der Nutzer*innen miteinbezogen wird. Es gab bei Nanna nur ein Geräusch, das getimt war, und zwar die Länge der Aufzug-Musik zu Anfang.
Da der Aufzug immer die gleiche Strecke zurücklegt, war es am einfachsten, diese über einen Zeitrahmen zu definieren und dann auszublenden.
Das Zusammenspiel der einzelnen Tracks oder Stems kann beim linearen Mischen sehr fein abgestimmt werden.
Überlagerungen verschiedener Bereiche mit ähnlichen Frequenzen, oder sogar Verdeckungen einzelner Elemente können so vermieden werden. Außerdem können einzelne Elemente der (Film-)Musik hervorgehoben werden, durch die Platzierung im Stereobild, durch Veränderung der Lautstärke oder indem in den Frequenzbändern der gleichzeitig abgespielten

Elemente “Platz geschaffen” wird.
Diese feine Balancierung ist beim Mischen im virtuellen Raum deutlich schwerer zu erzielen und kann nur mit flächendeckender Überprüfung der fertigen Anwendung angenähert werden.
Man bekommt einen anderen Zugang zu Balance im Vergleich zum Vorgehen beim linearen Mischen, die Nutzer*in selbst Prioritäten setzen kann. Beispielsweise können Nutzer*innen bei Verdeckung des Geräuschs eines gut sichtbaren Objekts näher zu diesem hinbewegen. Alternativ haben Nutzer*innen bei Interesse die Möglichkeit, durch Drehen des Kopfes verschiedene Perspektiven miteinander zu vergleichen.
Diese Aspekte sind in der Mischung selbst schwieriger umzusetzen, können aber nicht prinzipiell als Einschränkung definiert werden, da sie, wie in den Beispielen dargestellt, neue Herausforderungen an den Vorgang des interaktiven Mischens bieten.
Ein weiterer Punkt, der mit dem letztgenannten zusammenhängt, ist das Überprüfen der fertigen Mischung.
Beim linearen Mischen, kann man sicher gehen, dass das Produkt, das man ganz zum Schluss hört, abgesehen von einem Mastering, das ist, was jeder Nutzer und jede Nutzerin, die das Musikstück oder den Film zu hören bekommen. Man gibt nach dem linearen Mischen ein Produkt aus der Hand, das genau so reproduziert werden wird. Diese Kontrolle muss man bei der Arbeit mit einem interaktiven Medium zu einem gewissen Grad abgeben. Es ist umso wichtiger, bei den vorherigen Schritten, der Konzeption und den Aufnahmen, auf ein schlüssiges und einheitliches Bild abzuzielen. Dies ist relevant, um sich nicht erst in der Postproduktion beim Zusammenmischen auf einen Stil oder eine Wirkung zu orientieren.

Eine Unterscheidung, die sich zwischen dem Sound in der VR-Anwendung und prozedural erzeugtem Audio machen lässt, ist, wie synthetisch die Geräusche scheinen.
Bei der Arbeit mit Wwise werden zwar, genau wie beim prozeduralen Vorgehen, viele Parameter eingeführt, die für Variabilität sorgen, allerdings immer auf Basis von Mikrofonaufnahmen. Das bedeutet, dass die Geräusche nie komplett trocken sind und immer einen eigenen Klangcharakter besitzen.
Bei der Bearbeitung kann dies mehr Sorgfalt erfordern, um ungewünschte Elemente nicht zu verstärken, wie beispielsweise ein Grundrauschen. Gleichtzeitig ist der Aufwand insoweit geringer, nicht bei Null zu starten, sondern ein bereits existierenden Clip an die Anforderungen des Objekts, das die Schallquelle sein soll, anzupassen.

Beim Mischen in Nanna wurde schnell klar, das einige Teile die nach klassischeren Work-Flows dem Mischen zuzuordnen sind, schon früher geschehen, bevor die Geräusche aufeinander treffen.

Um dem entgegenzuwirken, wurde die im vorherigen Teil dieser Arbeit beschriebene Soundskizze ins Projekt eingeführt. Zu diesen Elementen gehören:

- die Auswahl der Clips
- das Festlegen der Reihenfolge oder Gruppierung unter ähnlichen Geräuschen
- die zeitliche Abhängigkeit von Events oder anderen Sounds
- Filtern der Frequenzen

Durch Definition der Charakteristiken und Einteilungen in Gruppen wird bereits im Vorhinein nötigerweise einiges festgelegt.
Man hat die Möglichkeit, die Sounds in der Middleware bereits abzuspielen, das geschieht auf Knopfdruck und simuliert das Abspielen in der Anwendung nur notdürftig.
Das tatsächliche Mischen, bei dem gehört, abgeändert und wieder gehört werden kann, passiert erst beim Testen der Anwendung. Daher wurde der Prozess des Mischens bei Nanna stark zweigeteilt: Vor der Positionierung und Integrierung in die Game Engine und nach dieser.
Beim Testen in der Game Engine ließen sich dann einzelne Geräusche im Kontext ihrer Umgebung hören, also so wie sie später von Nutzer*innen auch gehört werden würden.
Von einem Objekt bewegt man sich zum Nächsten, um auch den Übergang testen zu können.
Alle platzierten Geräusche lassen sich so nie gemeinsam hören, aber sie werden auch im Endprodukt nie zum selben Zeitpunkt gehört. Vergleichbar mit Musikstücken, in denen ähnlich wenig Relevanz besteht, alle Noten gleichzeitig hören zu können. Der Dialog wird klassischerweise, der Verständlichkeit halber, in den Center gelegt.

Vorteil der Plattformgebundenheit

Durch die Plattformgebundenheit der VR-Anwendung lässt sich gewährleisten, dass alle, die das Spiel oder die Anwendung nutzen, mit ähnlicher Hardware abhören. Zurückzuführen ist das auf die Schwierigkeit, ein mehrkanaliges System so aufzubauen, dass es für die räumliche Wiedergabe einer 3D-Anwendung mit Headtracking ausreicht. Die deutlich einfachere und weitverbreitete Methode ist in diesem Fall das Nutzen von Kopfhörern.
Es gibt keine Down-Mixes, wie es zum Beispiel bei Video on Demand oft der Fall ist. So ist es oft möglich, dass Filme, die fürs Kino gemacht sind, auf der Heimanlage und Flachbildschirm, auf dem Laptop oder auch am Handy angeschaut werden können.
Für alle diese Geräte zu produzieren, ist eine Herausforderung, die man beim Sounddesign für VR vernachlässigen kann. Man muss nicht davon ausgehen, dass sich jemand das Ergebnis auf einem Mono-Handy-Lautsprecher anhört.
Bei anderen “neuen Realitäten” kann dies der Fall sein, doch sind AR-Anwendungen oft spezifisch im App Format, für mobile Endgeräte ausgelegt. Beim 360 Grad Video, das auch von Youtube oder Facebook unterstützt wird, muss man miteinbeziehen, dass die Ton-Wiedergabe abhängig vom Endgerät sehr unterschiedlich ausfallen kann.

Live Mix zu Zweit

Zum Ende der Produktion ist, wie im linearen Workflow, das Mischen ein essentieller Bestandteil. Da es keine Version gibt, die am Stück abgespielt werden kann, muss die Anwendung gestartet werden. Über den Mixing Desk kann eine Wwise Session in die Game Engine geladen werden. Über diese Verbindung lässt sich die Anwendung starten und gleichzeitig live in Wwise verändern. Dadurch ist es möglich, Inhalte zu mischen und den Effekt der getätigten Einstellungen direkt abzuhören.
Dieser Vorgang nennt sich Live Mixing, im Gegensatz zu Offline Mix (Somberg, 2019).

Das gemeinsame Mischen mit zwei Personen war bei Nanna unerlässlich. Nachdem alle Geräuschquellen platziert und vordefiniert waren, konnte man diese zwar einzeln abhören und Attenuation Kurven annähern, sowie Lautstärkepegel voreinstellen, man konnte diese jedoch nicht in der Anwendung selbst überprüfen.
Das Problem hierbei ist, dass es schlecht möglich ist, sich gleichzeitig in VR zu befinden und dann bei Bedarf schnell genug die Ausrüstung abzulegen, um eine Änderung vorzunehmen. Um etwaige Änderungen zu überprüfen und anzugleichen, müssen das verkabelte Headset und die Controller, die es an den Handgelenken zu sichern gilt angelegt werden. Anschließend muss die vorherige Position im Raum, die etwa im gleichen Bezug zum Objekt, das es zu überprüfen gilt, wieder eingenommen werden. Allein arbeitend muss man Sekunden später wieder die gesamte Technik ablegen und sich wieder an die Software zu setzen.
Auch spielt die in der Literatur häufig erwähnte Immersion eine relevante Rolle:
selbst wenn man den technischen Aufbau einer Anwendung kennt, ist es möglich sich auf diese einzulassen. Aber sich auf die Umgebung und das Gehörte einzulassen, benötigt, wie bei allen anderen Nutzer*innen, Zeit. Diese bekommt man, wenn man sich in einem Zweierteam an die Mischung wagt. Bei dieser Teamlösung müssen Kommunikationshürden überwunden werden. Es muss alles abgestimmt werden, so dass die Person die am Regler sitzt, versteht welche Änderungen nötig sind. Genauso müssen der Person, die sich in VR befindet, die gezielte und produktive Fragen gestellt werden.
So lässt sich deutliche effektiver arbeiten. Indem zwei verschiedene Meinungen abgeglichen werden und bei Ermüdung der Ohren oder des Gleichgewichtssinnes die Plätze getauscht werden können.

Die Person, die das VR Headset auf dem Kopf trägt, kann nach und nach den Raum abtasten, um Problemstellen zu finden, an denen zu viel passiert oder auch Löcher, an denen zu wenig passiert. Mit dem Wissen, welche Objekte einen Sound emittieren, ist es auch deutlich klarer möglich, eine erwartete Dämpfung der Geräusche über die Entfernung hinweg festzulegen. Diese Einschätzung ist auf dem kleinem Bildschirm in der Game Engine nur grob abschätzbar.
Es zeigte sich, dass die räumlichen Dimensionen eine große Rolle dabei spielen, auch wenn sie nur simuliert sind. Wie wir Ton wahrnehmen und welche Lautstärken sich als glaubhaft annehmen lassen, kann je nach Immersionslevel sehr unterschiedlich ausfallen.
In den meisten Fällen waren die auffälligsten Geräusch-Ungereimtheiten zu harte Cut-Offs.

Beim ersten Zusammmenmischen scheint beispielsweise ein Geräusch bereits zu leise zu sein, um es noch zu hören. Doch im virtuellen Raum bemerkt man es noch deutlich, wenn eine weiter entfernte, aber dennoch sichtbare Geräuschquelle plötzlich nicht mehr hörbar ist.

Es kann sich lohnen, vorher eine Strategie auszumachen und im Team abzusprechen, welche Elemente der Mischung von entscheidender Wichtigkeit sind. Die sich daraus ergebende Reihenfolge kann festgehalten werden und später als Vorlage diesen. Das Notieren hilft, um einen Überblick zu behalten und bei der Überprüfung festzustellen, was schon geändert wurde.
Auch bei Tests, bei denen Menschen außerhalb des Teams als Testpersonen herangezogen wurden, lohnt es sich sehr festzuhalten, welche Sachen beanstandet wurden oder welche ungewöhnlichen Postionen eingenommen wurden. Kennt man die Anwendung bereits, bewegt man sich ganz anders fort. Man ist häufig zielstrebiger als ungeübte Testpersonen. Dies kann aber ein großer Nachteil für das Finden von eventuellen Ungereimtheiten bieten. Es besteht die Gefahr, die immer gleichen Wege zu gehen und relevante Schwachstellen zu übersehen.
Während des Mischens kann am Rechner das Sichtfeld des Headsets überprüft werden. Mithilfe der kleinen Ansicht können dem "Misch-Partner" spezifische Fragen gestellt werden oder verständlich Richtungen und Objekte benannt werden. Dabei bietet sich eine Objektorientierung an, da es aus der First-Person Perspektive schwieriger sein kann, exakte Entfernungen abzuschätzen, als in der Overhead Perspektive, die man beim Gestalten des Raumes einnimmt.
Wichtig zu erwähnen ist, dass es im Gegensatz zum zuvor erwähnten linearen Mischen nur wenig "globale" Änderungen gibt .
In Wwise selbst, kann man einige Einstellungen global vornehmen, und damit auf alle Files eines bestimmten Typs anwenden.
Beim Mischen gestaltet sich das allerdings schwieriger. Natürlich lässt sich der Pegel der Ausgabelautstärke im Masterbus einstellen. Jedoch sind Änderungen, wie allgemeine Frequenzbänder zu filtern oder Lautstärken desselben Objekttyps schwer beim Mischen zu überprüfen und sollten gut begründet sein, da es meist zu umfangreich wäre, alle Auswirkungen der Einstellungsänderung auch abzuhören. Das würde zu einer weiteren flächendeckenden Überprüfung führen, die viel Zeit in Anspruch nehmen würde.
Neben der objektorientierten Vorgehensweise kann man auch die einzelne Räume für sich überprüfen. Raumtöne können verglichen werden und Übergänge getestet werden. Im "Basic-Emotions" Raum war ein starker Unterschied gewünscht, zwischen jeweils dem rechten und linken mit dem mittleren Verbindungsraum. Der Anspruch hierbei war, es vorsätzlich klingen zu lassen, dass eine Veränderung des Raumes deutlich merkbar ist. Gleichzeitig sollte es nicht nach einem technischen Fehler, wie dem Aussetzen der Kopfhörer, klingen.
Diese Balance zu treffen, verlangte mehrere Versuche und einige Bewegungstests durch den

virtuellen, wie auch den realen Raum. Die Art und Weise der Fortbewegung im virtuellen Raum musste beachtet werden. Sich mit dem Teleporter ein ganzes Stück versetzen zu lassen, klingt ganz anders, als die drei Schritte, einen nach dem anderen, zu gehen. Ein harter Cut ist hier unerwünscht. Um ein bewusstes Fortbewegen mit eigenem Überschreiten der Raumgrenze zu veranlassen, wurde der Teleport so programmiert, dass man speziell die Felder an der Grenze nicht erreichen kann. So wurde ein Problem der Mischung durch das bewusste Programmieren einer Mechanik gelöst.

5 Ausblick

Im Folgenden wird betrachtet, an welchen Punkten bei den Nutzungs-Tests Probleme aufgetreten sind. Durch Vorschläge, wie dies durch Sound verbessert werden könnte, soll ein Ausblick gegeben werden. Da der zeitliche Projektrahmen beschränkt war, werden die Probleme, die auf fehlende Zeit zurückzuführen sind außenvor gelassen. Mit mehr Zeit wären mehr Tests und Überprüfungen möglich gewesen. Im ersten Teil werden die Problematiken beschrieben, die nicht wie geplant durchgeführt wurden.
Im zweiten Teil werden mögliche Erweiterungen beschrieben, die im Laufe des Projekts oder bei der Recherche im Nachhinein auftauchten. Diese sollen nicht als fehlende Elemente dargestellt werden, da ihre Implementierung in Nanna nie geplant war. Bei einer Weiterentwicklung oder eines größeren Projekts sind diese Erweiterungen eine vielversprechende Möglichkeit, die Raumwahrnehmung zu erweitern oder zu optimieren.
Diese werden kurz umrissen und beschrieben, da sie einen relevanten Unterschied in der Entwicklung eines Sounddesigns in Wwise machen können.

5.1 Was hat nicht geklappt

- Orientierung zwischen den Räumen
  Wie oben beschrieben sollten sich die Leute zwischen den Räumen frei bewegen können. Das Fehlen einer Narration oder deutlicher, visueller Hinweise wurde zum Nachteil. Oft wussten Nutzer*innen nicht, wo sie als nächstes hingehen sollten, oder wie sie zwischen den Räumen wechseln können.
  Entweder sollte eine bessere Balance zwischen freier Erkundung und geführter Erzählung gefunden werden, oder visuell und mit Geräuschen eindeutiger auf die Besonderheit der unterschiedlichen Plattformen hingewiesen werden.
  Da das Teleportieren über die Plattformen ein essentieller Teil der Fortbewegung in Nanna war, wäre der Rest des virtuellen Raums, ohne Eingreifen des Teams, nicht erkundbar gewesen. Durch diesen, von außen gegebenen Hinweis, wird kurzzeitig das Gefühl der Immersion und Präsenz zerstört.

- Fehlende Sound Cues für den Teleport
  Durch hörbare Hinweise, genannt Sound Cues, hätte einige Verwirrung über die Fortbewegung in Nanna vermieden werden können. Es gab ein Geräusch für den "Suchstrahl", das durch die Teleport-Funktion ausgelöst wurde. Allerdings wurde der Moment des Springens nicht gekennzeichnet. Das bedurfte einiger Gewöhnung der Tester*innen. Auch die Plattformen, welche die Verbindung zwischen den Räumen bildeten, waren nicht ausreichend gekennzeichnet. Sie waren nicht als "Durchgang" oder Bewegungsmöglichkeit gekennzeichnet, außer durch die farbliche Abhebung, und eine erfolgreiche Teleportation sorgte häufig für Verwirrung. Geräusche, möglicherweise angelehnt an Science-Fiction Medien, hätten hier deutliche Assoziationen hervorrufen können, die bei der Orientierung helfen.

5.2 Mögliche Erweiterungen in zukünftigen Anwendungen

Bei den möglichen Erweiterungen sind zwei Richtungen erkennbar. Eine Anwendung kann stetig realistischer und naturgetreuer simuliert werden, durch realistischere Grafiken sowie realistischere Geräuschabbildung. Zu einem gewissen Grad ist das mit kreativen Entscheidungen zu verbinden, allerdings ließe sich dann nie das ganze Potenzial ausnutzen. Das Potenzial einer virtuellen Realität ist es, Umgebungen und Situationen zu schaffen, die im echten Leben nicht erreichbar sind. Die "true identity" einer Sache, also die wahre Identität, kann durch ein ausreichend akkurates Phantombild gezeigt werden. Bei dieser Abbildung muss man sich zwischen der Repräsentation aller Geräuschquellen und Reflektionen und einer Auswahl auf die wichtigsten Eigenschaften beschränken (Rumsey, 2001). Der Anspruch an Perfektion muss abgewogen werden. Um dieser Perfektion oder der Herausarbeitung einzelner wichtiger Elemente nahe zu kommen, gibt es zahlreiche Tools, von denen im Folgenden einige vorgestellt werden:

Nutzung des Mikrofons des Headsets
In dem HTC vive Headset ist ein Mikrofon eingebaut. Dieses könnte Sprachsteuerung ermöglichen oder einen überraschenden Echo-Effekt entstehen lassen. Die Spracherkennungssysteme für mobile Engeräte ständig weiterentwickelt werden, ist es gut vorstellbar, das auch simple Interaktionen über Dialog möglich sind. Dieser zusätzliche Input könnte das Immersionsgefühl deutlich verstärken, da Nutzer*innen aus dem Alltag die sprachliche Kommunikation gewohnt sind, diese allerdings selten in Unterhaltungsmedien genutzt wird.

Vertonen von Schritten
Eine mögliche Erweiterung wäre das Vertonen der Schritte. Dazu benötigt man einen Sensor, der jeden Schritt registrieren kann. Da diese Technik in den meisten Handys eingebaut ist, wäre es hier nur eine Frage der Verbindung der verschiedenen Geräte. So kann der Bezug zu einem weichen oder harten Untergrund noch verstärkt werden. Ohne die genaue Synchronization mit den echten Fußbewegungen lässt sich kein zufriedenstellendes Ergebnis erzielen, da von einer Bewegung des Kopfes nicht auf Schritte zu schließen ist.

Plugins
Beim Installieren der Software Wwise ist eine Vielzahl von Plugins von Software-Partnern und von Wwise selbst zu finden. Darunter finden sich einige bekannte Namen größerer Hersteller.
Zu einigen Plugins wird im Folgenden eine knappe Einordnung gemacht:
Wwise Convolution ermöglicht das Erstellen von Hallräumen aus Aufnahmen echter Räume. Die Raumimpulse können aus Libraries verwendet, oder es können eigene, importierte RIRs verwendet werden.
Wwise Motion ist eine Erweiterung des Authoring Tools, die auch Zugriff auf das haptische Feedback von verwendeten Controllern bietet.
So können die Sinne der Ohren und der Hände gleichzeitig angesprochen werden. Durch abgestimmtes Rütteln der Controller an den Händen, bei lauten Geräuschen, könnte sich

so ein Effekt verstärken, ohne die Lautstärke unangenehm anheben zu müssen.
Wwise Reflect implementiert eine Simulation der Erstreflexionen, bei der die Zeitverzögerung der Reflektionen errechnet wird. Dies kann man sich wie eine simple Spiegelung vorstellen, die leicht verzögert wird.
Mit dem *Soundseed* (Impact, Air, Grain) ist eine Granularsynthese möglich, durch deren Implementierung die Erzeugung "zufälliger" Geräusche aus existierenden Soundclips möglich ist.
Es können also Hallräume, haptisches Feedback, Reflexionen, und Variabilität erzeugt werden. Diese Funktionen unterstützen bereits erwähnte relevante Effekte, die die räumliche Wahrnehmung unterstützen oder erweitern. Es gibt so die Möglichkeit sich einer realistischen Darstellung noch weiter anzunähern, ohne eigenen Code zu programmieren.
Weitere Plugins geben die Möglichkeit Raumimpulsantworten zu importieren oder andere Audio-Software wie Izotope oder Nuendo zu integrieren.

Occlusion Rays

Occlusion Rays sind ein Verfahren, das genutzt werden kann, um Verdeckung von Schallquellen zu ermitteln. Nach dem gleichen Prinzip, wie das Raycasting, durch das Lichstrahlen verfolgt werden können und Lichtverhältnisse oder Sichtbarkeiten verschiedener Objekte errechnet werden können.
Statt der Lichtstrahlen wird die simulierten Schallausbreitung verfolgt. Mit höherem Rechenaufwand können mehr Reflektionen und Abdämpfungen simuliert werden. Diese Simulation erfolgt in Echtzeit und bezieht die exakten Dimensionen des Raumes und die darin enthaltenen Objekte ein.

HRTFs anwenden

Head Related Transfer Functions, die Kopfbezogene Übertragungsfunktion beinhalten die Personenbezogene Interaurale Zeitdifferenz (IATD), sowie die Lautstärkedifferenz zwischen den Ohren (Interaural Level Difference) und deren individueller Frequenzfilterung (Suzuki et al., 2011). Die Nutzung von HRTFs kann sich im Nahfeld als schwierig gestalten, da bei einer geringen Entfernung zwischen Geräuschquelle und Kopf die Genauigkeit der Abbildung nicht ausreicht (Brungart et al., 1996). Das Verwenden von HRTFs ermöglicht die Nachbildung des individuellen Hörerlebnisses. Wwise selbst bietet keine Möglichkeit, HRTFs direkt zu integrieren. Das ist nur in Kombination mit einem "Spatializer" Tool möglich.

Prozessor Optimierung

Die im Kapitel "Leistung und CPU-Auslastung" erwähnte *Memory Pool*, kann zu Beginn beliebig festgelegt werden. Um zu vermeiden, dass es trotz ausreichender Kapazitäten zu Leistungsschwierigkeiten kommt, empfiehlt es sich den tatsächlich benötigten Speicherplatz zu ermitteln. Dieser kann dann freigehalten werden ohne zu viel zu beanspruchen. Auch sollte bei Bedarf die maximale Anzahl von Voices festgelegt werden. Das war beim Projekt Nanna allerdings nicht notwendig, da die Anzahl aller gleichzeitig abgespielten Objekte

recht konstant war. Nur einzelne Geräusche wurden zusätzlich durch Interaktionen aktiviert.

Tuning Room-Tones

Eine Aufnahme eines *Roomtones* zu nehmen, das Grundrauschen eines Raumes, einen Loop zu erstellen und mit einem EQ bearbeiten. Das ist die Grundidee die in einem Vortrag von Vincent Gagnon im Rahmen des Montreal International Game Summit vorgestellt wird (Audiokinetic, 2018).

"Tuning a Room Tone" nennt der Vortragende diesen Vorgang. Hierbei wird eine Frequenz gewählt, die den Anteilen eines musikalischen Akkords entsprechen. Diese Methode zeigt, wie sich kreative Entscheidungen in die Tongestaltung einbetten lassen, ohne dass das Gesamtbild an Realismus verliert. Virtual Reality ermöglicht oder erfordert nicht nur die Nachbildung der Realität, sondern lässt die Freiheit zu gestalten. Anhand eines Beispiels wird im oben genannten Vortrag erklärt, wie die Wirkung eines Moll-Akkords eine Atmo beeinflussen kann. Die Grenze zwischen Musik und Atmo verschwimmen dahingehend, dass die Frequenzen des Akkordes sowohl im Raumton verstärkt werden, als auch in Elementen der Komposition auftreten.

Diese Methode wäre bei Nanna ein Schritt gewesen, der Musik zu einem gewissen Grad miteinbindet, ohne sich auf eine klassische Komposition zu stützen. Im Ansatz findet man diese Idee in den Cello-Flächen, die unter die Umgebungsgeräusche gemischt wurden.

Schlusswort

Es gibt noch unzählige weitere Möglichkeiten, eine Tongestaltung kreativer oder auch realistischer zu gestalten. Zusammenfassend lässt sich sagen, dass die Arbeit mit einer Middleware viele Arbeitsschritte vereinzelt und präzise bearbeitbar macht. Allerdings erweitern sich die Möglichkeiten auch maßgeblich.

Diese Arbeit soll klarer machen, welche Entscheidungen unter welchen Bedingungen gemacht werden können.

Die vorgestellten Techniken und Verfahren befähigen Gestalter*innen auf verschiedenen Wegen ihrer Arbeit Ausdruck zu verschaffen.

Dass es nicht den einen Weg gibt, der zu einem "richtigen" Ergebnis führt, erschwert die Arbeit nicht, sondern verdeutlicht, dass es sich trotz der technischen Arbeit um eine kreative Gestaltung handelt, die nach Einbeziehen aller Einschränkungen und Projektgegebenheiten dem Geschmack der Schaffenden entsprechen muss. Viele subjektive Entscheidungen können durch den Grad der Realisierbarkeit und Nutzer-Tests erleichtert werden, aber bleiben trotzdem gestalterische Entscheidungen.

Literaturverzeichnis

Alberti, P. W. (2006). *The Anatomy And Physiology Of The Ear And Hearing*. 11.

Audiokinetic. (2018). *Musical Theory For Sound Design*.

Augoyard, J.-F. (2005). Sonic Experience. *McGill-Queen's University Press*, 237.

Bezmalinovic, T. (2020). *Diese VR-Brillen haben das größte Sichtfeld*.

Bopp, K. (2009). *Serious Games Ein Literaturbericht*. KIT Karlsruher Institut für Technologie.

Brungart, D. S., Rabinowitz, W., & Durlach, N. I. (1996). Auditory Localization of a Nearby Point Source. *Journal of the Acoustical Society of America*, *100*, 2593.

Bryson, S. (2013). Virtual Reality: A Definition History - A Personal Essay. *arXiv:1312.4322 [cs]*. http://arxiv.org/abs/1312.4322

Bucher, H.-J. (2001). Wie Interaktiv Sind Die Neuen Medien? In H.-J. Bucher & U. Püschel (Hrsg.), *Die Zeitung Zwischen Print Und Digitalisierung* (S. 139–171). VS Verlag für Sozialwissenschaften.

Cancellaro, J. (2005). *Exploring Sound Design for Interactive Media (Design Exploration Series)*. Thomson Delmar Learning.

Dörner, R., Broll, W., & Grimm, P. (2019). *Virtual und Augmented Reality (VR/AR)* (B. Jung, Hrsg.). Springer Vieweg.

Ellermeier, W., Eigenstetter, M., & Zimmer, K. (2001). Psychoacoustic Correlates of Individual Noise Sensitivity. *The Journal of the Acoustical Society of America*, *109*(4), 1464–1473.

Görne, T. (2019). The Emotional Impact of Sound: A Short Theory of Film Sound Design. *KLG 2017. Klingt Gut! 2017 International Symposium on Sound*, 17–30. https://doi.org/10.29007/jk8h

Grimshaw, M. (Hrsg.). (2011). *Game Sound Technology and Player Interaction: Concepts and Development*. Information Science Reference.

Pfisterer, U. (2019). Virtualität Und Interaktivität. In U. Pfisterer (Hrsg.), *Metzler Lexikon Kunstwissenschaft: Ideen, Methoden, Begriffe Sonderausgabe* (S. 459–461). J.B. Metzler.

Raffaseder, H. (2010). *Audiodesign* (U. Schmidt, Hrsg.; Second). Carl Hanser Verlag München.

Rothe, S., Hußmann, H., & Allary, M. (2018). Diegetic Cues for Guiding the Viewer in Cinematic Virtual Reality. In L. T. De Paolis & P. Bourdot (Hrsg.), *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Bd. 10850, S. 101–117). Springer International Publishing. https://doi.org/10.1007/978-3-319-95270-3_7

Rumsey, F. (2001). *Spatial Audio*. Focal Press.

Schütze, S., & Irwin-Schütze, A. (2018). *New Realities in Audio: A Practical Guide for VR, AR, MR and 360 Video*. CRC Press. https://doi.org/10.1201/b22146

Somberg, G. (Hrsg.). (2017). *Game Audio Programming: Principles and Practices*. CRC Press, Taylor & Francis Group.

Somberg, G. (Hrsg.). (2019). *Game Audio Programming 2 - Principles and Practice*. CRC Press, Taylor & Francis Group.

Streppel, M., & Robert-Koch-Institut (Hrsg.). (2006). *Hörstörungen und Tinnitus*. Robert Koch-Inst.

Suzuki, Y., Brungart, D., Iwaya, Y., Iida, K., Cabrera, D., & Kato, H. (Hrsg.). (2011). *Principles And Application Of Spatial Hearing*. World Scientific Publishing Co. Pte. Ltd.

VIVE Pro | The Professional-Grade VR Headset. (2021). https://www.vive.com/eu/product/vive-pro/.

Wwise Dokumentation. (2021). https://www.audiokinetic.com/library/edge/?source=Help&id=wwise_help.

Abbildungsverzeichnis

1 Foyer Raum in Nanna 8
2 Rhythmus-Raum in Nanna 9
3 Basic-Emotions-Raum in Nanna 10