

# **Akustischen Szenen in VR - Lärmbelastung und Natur in der Stadt**

Immersive Ambisonics Klanglandschaften-Komposition  
mit 360° Videoaufnahmen in Graz

**Tobias Bettke**

Universität für Musik und darstellende Kunst Graz  
Institut für Elektronische Musik und Akustik

FH Joanneum  
Institut für Design & Communication

Sound Design

Betreuer: Univ.Prof. Dipl.-Ing. Dr.techn. Alois Sontacchi

Diese Arbeit wird zur Erlangung des Grades eines Master of Arts eingereicht.

# Ehrenwörtliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Masterarbeit selbstständig angefertigt und die mit ihr verbundenen Tätigkeiten selbst erbracht habe und keine anderen als die angegebenen Hilfsmittel benutzt habe. Ich erkläre zudem, dass ich mich bei der Erstellung der Arbeit an die Richtlinie der FH JOANNEUM zur Sicherung Guter Wissenschaftlicher Praxis und zur Vermeidung von Fehlverhalten gehalten habe. Insbesondere erkläre ich, dass ich alle aus gedruckten, ungedruckten oder dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Regeln für gutes wissenschaftliches Arbeiten zitiert und durch Fußnoten bzw. durch andere genaue Quellenangaben gekennzeichnet habe. Die vorliegende Originalarbeit ist in dieser Form zur Erreichung eines akademischen Grades noch keiner anderen Hochschule vorgelegt worden. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben kann.

Graz, am 07. Juni 2022

Tobias Bettke

# Inhaltsverzeichnis

**4 Abstract**

**4 Kurzfassung**

**5 Einleitung**

**6 Akustische Szenen**

**7 Wahrnehmung von Schallereignissen**

**8 Ökologie der Klangwelt**

**9 Schallfeldaufnahmen**

**11 Klanglandschaften-Komposition**

**12 Immersion**

**14 360° Aufnahmen und Reproduktion  
von Klanglandschaften**

**15 Produktion von Ambisonics Audio**

**17 360° Videoaufnahmen**

**20 Werkstück:**

**Graz - Stadt vs. Natur**

**21 Einordnung und Vorgehen**

**22 Auswahl der Orte und Zeitpunkt der  
Aufnahme**

**26 Technische Umsetzung**

**28 Aufbau**

**30 Audioaufnahmen**

**29 Audioverarbeitung**

**32 Videoaufnahmen**

**32 Bildverarbeitung**

**34 Interaktion der Software**

**36 Analyse und Einordnung der  
Schallfeldaufnahmen**

**40 Kreativer Prozess**

**42 Abspielmedium/ Ergebnis**

**44 Diskussion**

**46 Fazit**

**48 Dateianhang**

**49 Literatur**

## Abstract

This master thesis and the corresponding workpiece deal with the recording and production of audiovisual 360° media. The topics Ambisonics, immersion in virtual reality and the acoustic ecology in soundscape compositions are presented. Recorded acoustic three-dimensional scenes in image and sound are immersively experienced with regard to noise pollution in a city. In addition, the material is artistically processed and thus enriched by an extended level of reflection. The work consists of six 360° scenes from three locations within the city of Graz and deals with the relationship between city and nature. A special feature is the precise spatial localisation of the real sound sources through recordings with a Zylia ZM-1 microphone, which supports 3rd order Ambisonics. The documentation of the workpiece is intended to provide insight into the technical and creative production process.

## Kurzfassung

Die vorliegende Masterarbeit und das dazugehörige Werkstück beschäftigen sich mit der Aufnahme und Produktion von audiovisuellen 360° Medien. Es werden die Themengebiete Ambisonics, Immersion in virtueller Realität und der akustischen Ökologie bei Klanglandschaft-Kompositionen vorgestellt. Aufgenommene akustische dreidimensionale Szenen in Bild und Ton werden hinsichtlich der Lärmbelastung in einer Stadt immersiv erfahrbar gemacht. Zusätzlich wird das Material künstlerisch bearbeitet und somit um eine erweiterte Reflexionsebene bereichert. Das Werkstück besteht aus sechs 360° Szenen von drei Orten innerhalb der Stadt Graz und beschäftigt sich mit der Beziehung zwischen Stadt und Natur. Besonders dabei ist die genaue räumliche Verortung der realen Klangquellen durch Aufnahmen mit einem Zylia ZM-1 Mikrophon, das Ambisonics 3. Ordnung unterstützt. Die Dokumentation des Werkstücks soll einen Einblick in den technischen und kreativen Produktionsprozess bieten.

# Einleitung

Heutzutage werden technische und kreative Realisationen von neuen Ideen in der Medienwelt immer greifbarer. Es ist bereits mehr technisch möglich als bei leicht zugänglichen Medien bislang verfügbar ist, wie z. B. bei YouTube oder Facebook, vor allem bei 360° Virtual-Reality (VR) Inhalten. Dafür werden einige Themengebiete näher betrachtet, um anschließend ein Werkstück zu erstellen und in das Thema einzuordnen. Dazu gehört die Wahrnehmung und Aufnahme von 3D Audio und Video in 360°, sowie kreative Anstöße, wie mit Klanglandschaften immersiv gearbeitet werden kann.

Das Werkstück behandelt die akustische Beziehung zwischen der Stadt und der Natur. Durch die Lärmbelastung des Straßenverkehrs im Zentrum der Stadt besteht ein höherer Grundgeräuschpegel, der mit dem natürlichen Klang der Natur, wie Rufe von Tieren, konkurriert.

Die Motivation der vorliegenden Arbeit liegt darin, immersive audiovisuelle Umgebungen unter Verwendung von Ambisonics dritter Ordnung zu erstellen, welche kreativ auf die Lärmbelastung in der Stadt Graz aufmerksam machen und über die akustische Umwelt reflektieren lassen.

Folgend werden zunächst die Wahrnehmung und der Umgang von Schall-

ereignissen und Videoaufnahmen vorgestellt, sowie die Bearbeitung und Reproduktion dieser. Außerdem wird auch die kreative Herangehensweise von Kompositionen mit Schallfeldaufnahmen betrachtet. Zuletzt wird das Werkstück vorgestellt, in das Thema eingeordnet und der Produktionsprozess eingehend vorgestellt und dokumentiert.

# Akustische Szenen

Akustischen Ereignissen wird nach individuellen und kulturell geprägten Eigenschaften einer Person unterschiedlich viel Aufmerksamkeit geschenkt, wobei es jedoch in allen Kulturen im Wesentlichen die drei Empfindungen Tonhöhe, Lautstärke und Klangfarbe gibt.<sup>1</sup>

Bei akustischen Ereignissen werden aber auch die Eigenschaften unserer physikalischen Umwelt durch Energieabgabe von Körpern, Substanzen oder Strukturen in Form von Schall wahrgenommen. Es können viele Informationen, z. B. über die Art der Schallerzeugung, den Raum oder das Material aus dem Gehörten entnommen werden. Im alltäglichen Leben sind wir ständig mit Situationen konfrontiert in denen Informationen aus der akustischen Umwelt entnommen werden müssen, wie beim Motorengeräusch im Straßenverkehr oder dem Weinen eines Kindes vor Zorn oder Schmerz.

Eine Schallquelle kann viele Aspekte an Informationsgehalt vermitteln. Dazu gehören laut Raffaseder Ort, Form, Größe, Material oder die Bewegung. Bei einem Raum kann die Art oder die Größe erkannt werden. Bei der Anregung kann sogar zwischen Art, Stärke, Rhythmus, Geschwindigkeit und Beschaffenheit des Auslösers differenziert werden.<sup>2</sup> Die Hörwahrnehmung kann dabei ganz gezielt auf bestimmte Schallsignale gelenkt werden, obwohl eine Vielzahl weiterer Parameter gleichzeitig wahrgenommen werden. Es ist allerdings auch möglich viele verschiedene akustische Ereignisse gleichzeitig zu hören und die Veränderung der akustischen Szene wahrzunehmen. Isoliert man einen Klang aus dieser Szene und stellt den Klang in einen neuen Kontext, kann sich laut Raffaseder auch seine Bedeutung verändern. Diese Bedeutung von Klängen in einem größeren Gesamtkontext kann als Möglichkeit der Täuschung für die Nachbearbeitung in Filmen genutzt werden.<sup>3</sup>

# Wahrnehmung von Schallereignissen

Bei der Wahrnehmung von breitbandigen Schallereignissen können die Ohren Parameter wie Lautstärke-, Phasen- oder Laufzeitunterschiede getrennt voneinander erkennen, wobei das Unterscheiden von Verhältnissen eine große Rolle spielt.

Bei der Einschätzung der Entfernung einer Schallquelle hat die Lautstärke zunächst eine große Bedeutung. Da bei einem unbekanntem Klang jedoch keine Referenzlautstärke aus unmittelbarer Nähe zum Vergleich herangezogen werden kann (manche Klänge sind näherungsweise im Gedächtnis gespeichert), müssen weitere Parameter analysiert werden. Wichtig dafür sind das Pegelverhältnis und die Verzögerungszeit von Direktschall, Erstreflexionen und Nachhall.

Für die Wahrnehmung der Richtung nutzt der Mensch die Pegelunterschiede<sup>4</sup> (interaurale Pegeldifferenz) und die Laufzeit- bzw. Phasenunterschiede<sup>5</sup> (interaurale Laufzeitdifferenz) zwischen beiden Ohren. Diese entstehen durch die Entfernung der Ohren voneinander am Kopf. Für die Unterscheidung von oben und unten, sowie vorne und hinten (es ergeben sich eigentlich keine Laufzeitunterschiede für diese Richtungen) ist die Form der Ohrmuschel zuständig, die das Schallsignal

leicht filtert, bevor sie das Trommelfell erreicht.<sup>6</sup>

Besonders für die Produktion von audiovisuellen Medien ist ein Verständnis der Wahrnehmungspsychologie des Menschen sehr wichtig, um die Möglichkeiten optimal zu nutzen, indem sie an die menschliche Wahrnehmung angepasst werden.

<sup>1</sup> Raffaseder, *Audiodesign*. S. 20f.

<sup>2</sup> Raffaseder. S. 24f.

<sup>3</sup> Raffaseder. S. 26f.

<sup>4</sup> „Pegeldifferenz“.

<sup>5</sup> „Laufzeitdifferenz“.

<sup>6</sup> Raffaseder, *Audiodesign*. S. 110f.

# Ökologie der Klangwelt

Unter der Klangwelt (später im Text auch Klanglandschaft bezeichnet) versteht man die gesamte akustische Umgebung, die aus biotischen und abiotischen Geräuschen besteht, wobei sich biotische Geräusche auf Geräusche beziehen, die von lebenden Tieren, einschließlich Menschen, erzeugt werden, während abiotische Geräusche alle natürlichen physikalischen Geräusche (z. B. Wasser, Wind, Regen) und vom Menschen erzeugte, nicht natürliche, mechanische Geräusche umfassen.<sup>7</sup>

Man könnte auch zwischen biophonen, geophonen und antropophonen Klängen unterscheiden. Unterschiede der akustischen Klangwelt werden bei verschiedenen Orten, Jahres- und Tageszeiten deutlich, wobei viele Faktoren mit einbezogen werden müssen. Auch bei dem Versuch die Klangwelt digital oder analog abzubilden und sie aufzunehmen, bestehen viele verschiedene Möglichkeiten, wobei sich meist auf bestimmte Klangparameter konzentriert werden muss. Auf die Möglichkeiten für eine möglichst realistische Aufnahme eines Schallfeldes wird in einem folgenden Kapitel eingegangen.

Die Geräusche von rufenden Tieren (biophon) sowie die nichtbiologischen Geräusche von fließendem Wasser und rauschendem Wind (geophon) stammen aus natürlichen Landschaften. In städtischen Landschaften hingegen dominieren vom Menschen erzeugte Geräusche, die von einer Vielzahl von Quellen ausgehen, z. B. von Maschinen, Sirenen, die Reibung von Reifen auf dem Asphalt oder Flugzeuge.<sup>8</sup> Ein Ziel in der angewandten Ökologie und Naturschutzbiologie ist es, zu verstehen, wie vom Menschen verursachter Lärm unterschiedliche Organismen sowohl in marinen als auch in terrestrischen Umgebungen beeinflusst.

Lärm ist ein räumlich ausgedehnter Schadstoff, und es gibt immer mehr Hinweise darauf, dass er schädliche Auswirkungen auf natürliche Lebensgemeinschaften haben kann. Dennoch fehlt bisher laut Barber ein gemeinsamer Rahmen, um die ökologischen Folgen von Lärm auf Menschen und Tiere komplett zu verstehen.<sup>9,10</sup>



# Schallfeldaufnahmen

Schallfeldaufnahmen (field recordings) nehmen die unzähligen Geräusche der Welt auf. Dabei kann jede Aufnahme als akustisches Abbild von Ort, Gegebenheiten und Zeit sein. Sie werden jedoch so unterschiedlich aufgenommen, dass zwischen den verschiedenen Möglichkeiten aufzunehmen unterschieden werden muss.

Schallfeldaufnahmen werden aufgrund der vielfältigen Einsatzmöglichkeiten am besten als eine Reihe von Techniken und nicht als Kunstform an sich verstanden. Außerdem umfasst das „field recording“ auch Praktiken des Zuhörens, der Überprüfung und Bearbeitung von Aufnahmen, des Komponierens und Abmischens, des Abspielens und Anhörens. Diese Prozesse sind wohl bedeutsamer als der Prozess der Aufnahme selbst, da eine einzige Aufnahme, wenn sie auf unterschiedliche Weise präsentiert wird, verschiedene Räume erzeugen kann.<sup>11</sup>

Gallagher hat die verschiedenen Stile der Aufnahme von Schallfeldaufnahmen in den Natur-Stil, den Klanglandschaft-Stil, den Akusmatik-Stil und den Klangkunst-Stil gegliedert. Nachfolgend werden diese kurz angesprochen.

Der Natur-Stil wird verwendet, um die Schwingungen von Tieren, Pflanzen, Lebensräumen und Ökosystemen „ein-

zufangen“. Bei diesem Stil wird die hörbare Präsenz des Menschen in der Regel so weit wie möglich ausgelöscht, indem menschliche Stimmen, der Lärm von Städten, Verkehrssystemen usw. vermieden werden. Einige Naturdokumentaristen bezeichnen ihre Arbeit als ein Mittel, um eine vom Menschen „bedrohte“ Natur hervorzuheben, zu schützen und zu erhalten. Resultierend daraus entstehen Aufnahmen, in denen die „Natur“ ästhetisiert und manchmal als unberührtes, exotisches „Anderes“ romantisiert wird. Diese Art von Feldaufnahmen stellt somit eine besondere Art von politischer Ökologie und politischer Ökonomie dar.<sup>12</sup>

Bei dem Klanglandschaft-Stil, wie er in der akustischen Ökologie und der Klanglandschaften-Komposition entwickelt wurde, geht es darum, die Geräusche einer bestimmten Umgebung zu dokumentieren und darzustellen. Arbeiten mit Klanglandschaften stellen einen klaren Bezug zwischen den Aufnahmen und den Kontexten her, in denen sie entstanden sind.<sup>13 14</sup>

<sup>7</sup> Blumstein u. a., „Acoustic Monitoring in Terrestrial Environments Using Microphone Arrays“.

<sup>8</sup> Francis und Barber, „A framework for understanding noise impacts on wildlife“.

<sup>9</sup> Francis und Barber.

<sup>10</sup> Pijanowski u. a., „Soundscape Ecology“.

<sup>11</sup> Gallagher, „Field Recording and the Sounding of Spaces“.

<sup>12</sup> Gallagher.

<sup>13</sup> Gallagher.

<sup>14</sup> Levack Drever, „Soundscape Composition“.

Akusmatische Feldaufnahmen machen sich Dekontextualisierung zunutze, um zu einem „reduzierten Hören“ einzuladen, bei dem die Ästhetik der aufgenommenen Klänge im Vordergrund steht und nicht der Versuch, ihre Quellen zu erkennen.<sup>15</sup> Der Akusmatische-Stil schafft abstrakte Räume, die sehr suggestiv sein können. Bei der künstlerische Art Schallfeldaufnahmen anzufertigen, werden alle verschiedenen Techniken verwendet. Es gibt einige besondere Arten der Verwendung von Feldaufnahmen, die in der Klangkunst üblich sind. Dazu gehört das kreative Experimentieren mit überhörten, versteckten oder normalerweise unhörbaren Klängen, oft durch Eingriffe in klangliche Umgebungen in Form von ortsspezifischen Installationen, Performances oder „Audiowalks“.

# Klanglandschaften-Komposition

Die Klanglandschaften-Komposition (übersetzt *Soundscape Composition*) ist eine Form der elektroakustischen Musik, die sich durch das Vorhandensein erkennbarer Umgebungsgeräusche und -kontexte auszeichnet und die Assoziationen, Erinnerungen und die Vorstellungskraft des Hörers in Bezug auf die Klanglandschaft anregen soll. Es entstand aus der pädagogischen Absicht das Bewusstsein die Klanglandschaft zu fördern.

Bei der Komposition von Klanglandschaften gibt es die Möglichkeit die Klänge aus ihrem „Rahmen“ zu entfernen oder hinzuzufügen, um einen kaum merklichen Übergang zu schaffen. Andere Werke verwenden Umformungen von Umgebungsgeräuschen, und hier kommt die gesamte Palette analoger und digitaler Studientechniken zum Einsatz, wobei der Abstraktionsgrad zwangsläufig zunimmt. Die Absicht ist jedoch immer, eine tiefere Ebene der dem Klang innewohnenden Bedeutung zu enthüllen und die semantischen Assoziationen des Hörers zu wecken, ohne die Wiedererkennbarkeit des Klangs zu verwischen.<sup>16</sup>

Bei der Erstellung von Klanglandschaft-Kompositionen sollte laut Truax darauf geachtet werden wenige Prinzipien einzuhalten. Es sollte die Wiedererkennbarkeit eines Stückes beibehal-

ten und das Wissen des Hörers über die Umwelt und den psychologischen Kontext berücksichtigt werden. Das Werk sollte bestmöglich unser Verständnis der Welt verändern, sodass sein Einfluss sich auf die alltäglichen Wahrnehmungsgewohnheiten überträgt. Wichtig ist dabei auch das die Form der Komposition auf jeder Ebene durch das Wissen und den ökologischen und psychologischen Kontext des Komponisten beeinflusst wird.<sup>17</sup>

Die Komposition von Klanglandschaften ist in einem breiten Spektrum von Möglichkeiten angesiedelt, von denen jede ihre eigene Praxis der Abbildung oder Darstellung der Welt hat. Die aktuellen technischen Möglichkeiten und ethische Fragen im Zusammenhang mit dem Produktionsprozess spielen bei der Abbildung eine große Rolle. Da Klanglandschaft-Kompositionen sowohl narrative als auch abstrakte Elemente besitzen, besteht eine weitere Möglichkeit darin, völlig imaginäre, jedoch hyperrealistische Räume zu schaffen. Rein synthetisches Sounddesign für Umweltgeräusche bleibt hinter den in den letzten Jahrzehnten entwickelten Synthesetechniken für Instrumental- und Vokalklänge zurück, insbesondere was die mögliche Komplexität und den Realismus des Klangmaterials betrifft.

<sup>15</sup> Chion, Gorbman, und Murch, *Audio-Vision*.

<sup>16</sup> B. Truax, „*Simon Fraser University - Soundscape Composition*“.

<sup>17</sup> B. Truax, „*Genres and Techniques of Soundscape Composition as Developed at Simon Fraser University*“.

## Immersion

Das Konzept der Immersion ist in den letzten Jahren weit verbreitet und diskutiert worden, da „head mounted displays“ (HMD - kopfgetragene Bildschirme)) und Computertechnologien nun in der Lage sind, qualitative virtuelle visuelle Szenen darzustellen.

Das Wort Immersion stammt aus dem Englischen und bedeutet nach dem Oxford Dictionary übersetzt so viel wie „eintauchen“ oder „einbetten“, nach dem Cambridge Dictionary „sich vollständig auf etwas einlassen“.<sup>18 19</sup>

Leider reicht diese Definition kaum aus, um die gesamte sensorische Erfahrung zu beschreiben, die insbesondere im Zusammenhang mit VR-Inhalten (Virtual Reality) ein Gefühl der Immersion vermittelt. Der Begriff „Immersion“ ist zu weit gefasst, um ihn ohne zusätzliche Informationen zu rechtfertigen.

Aus diesem Grund könnte es sinnvoll sein, die Wahrnehmung von Immersion in Bezug auf bestimmte Sinne zu quantifizieren, z. B. auditive Immersion, visuelle Immersion oder Immersion in einer multisensorischen Kapazität, um zu versuchen, die grundlegenden Wahrnehmungsmerkmale, die diese Erfahrungen verursachen, klarer zu umreißen.<sup>20</sup>

Seit den 70er Jahren ist der Begriff Immersion geläufig und die technischen Voraussetzungen begrenzen dabei den Grad der Immersion. Laut Eaton ist das Thema des immersiven Audios in den letzten Jahren immer aktueller geworden, mit Entwicklungen in der binauralen und Ambisonics-Wiedergabe über Kopfhörer sowie Mehrkanal-Lautsprecherwiedergabesystemen, die behaupten, das bisher immersivste Audioerlebnis zu bieten.

Während der Bereich der multisensorischen Immersion im Kontext von Videospielen recht gut erforscht ist, ist die Frage, welche Aspekte der auditiven Wahrnehmung am meisten zur Immersion beitragen, noch weitgehend unerforscht.<sup>21</sup>

In der Audiobranche werden auch Überlegungen zum Konzept des „räumlichen Klangs“ und seiner Beziehung zur Immersion angestellt.

Roginska & Geluso zum Beispiel definieren Immersion einfach als Klang, der aus der Umgebung des Hörers kommt. In diesem Beispiel wird davon ausgegangen, dass räumliches Audio immersiv sein muss, aber es wird nicht kommentiert, welche Systeme mehr oder weniger immersiv sein können.<sup>22</sup>

Berg & Rumsey (2001) beschreiben die Einhüllung des Hörers als eine Möglichkeit, das Gefühl der Immersivität in einer Umgebung mit viel Hall zu beschreiben, in der der Klang von überall her zu kommen scheint.<sup>23</sup>

In einem Experiment von Hamasaki, et al. wurde untersucht, welchen Unterschiede der Wahrnehmung zwischen einem 2-Kanal zu einem 5.1 und einem 22.2 System bestehen. Dieses Experiment ergab, dass bei Attributen wie Lokalisierungsgenauigkeit, Natürlichkeit und Grad der Freude und des Interesses nur geringe Unterschiede zwischen 22.2 und 5.1 Systemen wahrgenommen wurden. Deutlichere Unterschiede wurden jedoch bei Attributen wie Präsenz, Umhüllung, Tiefe, Breite und vorne/hinten-Verwirrung festgestellt.<sup>24</sup>

<sup>18</sup> *Oxford Dictionary*. „immersion noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearners-Dictionaries.com“.

<sup>19</sup> „immersion - Cambridge Dictionary“.

<sup>20</sup> Eaton, „Quantifying Factors of Auditory Immersion for Virtual Reality“.

<sup>21</sup> Eaton. S.14f.

<sup>22</sup> Roginska und Geluso, *Immersive Sound*.

<sup>23</sup> Berg und Rumsey, *Verification and Correlation of Attributes Used for Describing the Spatial Quality of Reproduced Sound*

<sup>24</sup> Hamasaki u. a., „A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)“.

## 360° Aufnahmen und Reproduktion von Klanglandschaften

Für ein immersives System ist es wichtig, einen Teil der audiovisuellen Umgebung zu rekonstruieren. In einem klassischen audiovisuellen System werden Audio und Video mit einem einzigen Mikrofon und einer einzigen Videokamera aufgenommen. Dies ist gleichbedeutend mit der Erfassung einer einzelnen räumlichen Ebene aus einem akustischen Wellenfeld bzw. einem Lichtfeld. Daher bedeutet die immersive Medienerfassung die Erfassung vieler räumlicher Proben aus diesen Feldern, die eine Rekonstruktion wesentlicher Teile dieser Umgebung ermöglichen.<sup>25</sup> Mit den Fortschritten der räumlichen Audiotechnologien, sowohl bei der Aufnahme als auch bei der Wiedergabe, gibt es immer mehr Anwendungen, die 360°-Audio einbeziehen, um ein immersives Klangerlebnis zu schaffen.

Die Gestaltung und Bewertung von Klanglandschaften für die Stadtplanung kann nun auf die umfangreichen „spatial audio tools“ zur Klangerfassung und 3D-Klangwiedergabe (wie z.B. IEM Plugin Suite<sup>26</sup>, Sparta, uvm.) über Kopfhörer und Lautsprecherarrays zurückgreifen.

Heutzutage wird das Thema der 360° Tonaufnahmen vor allem durch unterstützende Medien, wie YouTube, Facebook oder VR Spiele wieder interessant. Dafür wird meist ein Mikrofon für Ambisonics erster Ordnung verwendet, da diese einfach zu bedienen und sehr kompakt sind. Mikrofone dieser Art nehmen vier Kanäle auf, was bei Bewegungen auch hörbar einen Unterschied zu herkömmlichen Methoden darstellt. Jedoch sind die Richtungen sehr grob und bieten keinen genauen räumlichen Eindruck.

Die korrekte Richtungsverteilung der Audioaufnahmen bei kompakten Ambisonics-Mikrofonen höherer Ordnungen beruht auf den Lösungen der Helmholtz-Gleichung. Bei der Berechnung werden die Signale in frequenzunabhängige Kugelfächenfunktionen zerlegt, welche eine frequenzabhängige, radiale Schärpenfilterung beinhalten, die mit jeder Kugelfächenfunktion verbunden ist. Das entstehende Produkt sind die Ambisonics-Signale.<sup>27</sup>

# Produktion von Ambisonics Audio

Die 3D-Audioproduktion kann sich je nach Umgebung schwierig gestalten, vor allem bei realen Audioszenen mit vielen sich möglicherweise bewegenden Quellen in verschiedenen Richtungen und Höhen. Für gängige Medien werden oft Surround-5.1<sup>28</sup> Systeme verwendet, um einen Audioeindruck zu erlangen, der den Hörer umgibt. Bei 5.1 Systemen sind jedoch alle Lautsprecher in der Horizontalebene, wodurch der Richtungseindruck und die Reflektionen in der Horizontale deutlich angehoben sind. Von Ober- und Unterhalb der Hörerposition gibt es keine Informationen.

Ambisonics bietet dafür als periphere Audioproduktionstechnologie einen großen Vorteil.<sup>29</sup> Es gibt verschiedene Hauptmikrofon-Arrays, die für die Aufnahme von räumlich wirkenden Audioszenen, räumlicher Musik oder Räumen mit interessantem Nachhall eingesetzt werden können.<sup>30</sup>

Ambisonics ist eine 3D-Aufnahme- und Wiedergabemethode, die auf der Methode der Schallfeldbeschreibung durch Kugelflächenfunktionen beruht. Diese spezielle Schallfeldbeschreibung ermöglicht die Trennung von Aufnahme- und Wiedergabesystem. Die räumliche Klangerzeugung kann auf das jeweils verfügbare Wiedergabesystem

angepasst werden. Die Anpassung an ein gegebenes Wiedergabesystem (Lautsprecher oder bewegungsgesteuerte Kopfhörer) wird durch einen geeigneten Decoder erreicht. Die räumliche Auflösung der Wiedergabe und die damit benötigten Lautsprecherkanäle werden durch die Ambisonics Ordnung N festgelegt.<sup>31</sup> Der Bereich um das Zentrum eines Ambisonics-Systems, in dem ein Schallfeld genau wiedergegeben werden kann (der physikalische „Sweet Spot“), vergrößert sich mit der Ambisonics-Ordnung und nimmt mit der Frequenz ab. Bei höheren Frequenzen ist der Sweet Spot nur wenige Zentimeter groß.

Ambisonics kann auch als Amplituden-Panning-Methode (engl. vector base amplitude panning, VBAP) verstanden werden, welche den Wahrnehmungseffekt der Phantomschallquellen ausnutzt und diese sehr präzise erzeugt.

Es ist bekannt, dass Pegelunterschiede zwischen den Lautsprechern zu Laufzeitunterschieden bei tiefen Frequenzen führen, die in der horizontalen Ebene die wichtigsten Lokalisierungshinweise sind.<sup>32</sup> Für die Produktion von Ambisonics über Kopfhörer (binaurales Audiosignalpaar) wird mit einem Binaural Decoder das Signal mit gewichteten, linear kombinierten Außen-

<sup>25</sup> Assunção und Gotchev, *3D Visual Content Creation, Coding and Delivery*.

<sup>26</sup> Rudrich, „IEM Plug-in Suite“.

<sup>27</sup> Zotter und Frank, *Ambisonics*.

<sup>28</sup> „5.1“.

<sup>29</sup> Gerzon, „Periphone (with Height Sound Reproduction)“.

<sup>30</sup> Frank, Zotter, und Sontacchi, *Producing 3D Audio in Ambisonics*.

<sup>31</sup> Frank, Zotter, und Sontacchi, *Producing 3D Audio in Ambisonics*.

<sup>32</sup> Frank, „How to Make Ambisonics Sound Good“.

ohrübertragungsfunktionen (engl. head related transfer functions, HRTF)) gefaltet, die den Abstand der Ohren und die verschiedenen Laufzeit- und Frequenzunterschiede miteinbeziehen. Für die Ambisonics Produktion stehen in der Anwendung einige Programme zur Verfügung, wie die IEM Ambisonics Plugin Suite, mit Encodern, Multichannel-Effektplugins, Binaural- und dem AllRADecoder (zum Dekodieren von Ambisonics auf nahezu jedes Lautsprecher-Setup).<sup>33</sup>



# 360° Videoaufnahmen

Wenn von immersiven Videos gesprochen wird, muss sich auch auf das 360°-Video bezogen werden, welches sich derzeit in einer umfassenden technologischen Entwicklung befindet. Das 360°-Video ermöglicht es, die Szenerie in alle Richtungen, um eine bestimmte virtuelle Position des Betrachters herum anzuschauen. Fortgeschrittene Versionen des 360°-Videos erlauben es dem Betrachter auch, das Video in jeder Richtung von seinem virtuellen Standort aus zu betrachten und den virtuellen Standort zu wechseln.<sup>34</sup>

Dazu gehört das monoskopische 360°-Video, bei dem in der Regel das Video mehrerer Kameras zu einem Panorama zusammengesetzt wird, das stereoskopische 360°-Video, das es dem Betrachter ermöglicht, in einer beliebigen Position mit verschiedenen räumlichen Empfindungen zu schauen (3D Video) und das 6DoF (six degrees of freedom; die Bewegungsfreiheit eines Körpers -translation, rotation) 360°-Video, das dem Betrachter die Möglichkeit bieten kann, seinen Standort frei zu wechseln. Auf einem zweidimensionalen Bildschirm werden diese Videos meist equirectangular angezeigt.

In der Regel bestehen bei einem Aufnahmeprozess für omnidirektionale Videos zwei Möglichkeiten. Zum einen kann eine spezielle Linse verwendet werden, die in der Lage ist, sofort eine Aufnahme in alle Richtungen aufzunehmen (Fisheye Objektiv oder Parabolspiegel), die zweite Möglichkeit ist, mehrere Kameras zu verwenden und das Bildmaterial anschließend zusammenzufügen (stiching). Der Nachteil dabei ist der erhöhte Arbeitsaufwand, der Vorteil eine deutlich verbesserte Qualität durch höhere Pixeldichte.<sup>35</sup>

Für die Aufnahme und anschließende Verarbeitung von 360°-Videos gibt es bereits verschiedene Tools und Software auf dem Markt. Die Kamerahersteller, wie GoPro oder Insta 360 bieten die Möglichkeit mit der eigenen Software wie z. B. GoPro Fusion Studio<sup>36</sup> das Videomaterial der beiden Kameras zu einem equirectangularen Video zusammenfügen zu lassen und anschließend zu rendern.

Aber auch in fast jedem professionellen Videobearbeitungsprogramm kann mit 360°-Videos gearbeitet werden. Das Video kann dann z. B. in dem GoPro eigenen Videoplayer, oder in jedem anderen Player abgespielt werden, der diese Art von Videos unterstützt.

<sup>33</sup> Rudrich, „IEM Plug-in Suite“.

<sup>34</sup> Assunção und Gotchev, *3D Visual Content Creation, Coding and Delivery*.

<sup>35</sup> Huang u. a., „A 360-degree panoramic video system design“.

<sup>36</sup> GoPro Fusion Studio, „GoPro Fusion | GoPro“.

# Abspielmedium

Als Abspielmedium für Ambisonics bestehen zwei verschiedene Möglichkeiten. Zum einen mit Kopfhörern, oder mit entsprechend vielen Lautsprechern in geeigneter Position um den Hörer. Letzteres ist z. B. mit dem IEM AllRADecoder möglich, der mit einem Algorithmus für unterschiedliche Lautsprecheranordnungen geeignete Dekoder findet. Die Dekodierung von Ambisonics-Signalen für Kopfhörer erfolgt in der Regel ähnlich wie bei Lautsprechern, mit dem Unterschied, dass die Lautsprechersignale durch Faltung mit den Kopf-bezogenen Impulsantworten (Head Related Impulse Response, HRIR) der entsprechenden Wiedergaberichtung für Kopfhörer wiedergegeben werden.<sup>37</sup> Bei Kopfhörern besteht jedoch oft das Problem, dass Richtungsunterschiede der Höhe - bei Verwendung von standardisierten HRIRs - nicht so genau wahrgenommen werden können, da für die minimalen Phasenunterschiede zur Erkennung auch die Ohrmuschel eine Rolle spielt.

Für die Wiedergabe auf Lautsprechersystemen ist laut Frank der AllRADecoder eine praktikable, perceptiv angepasste Lösung, die auch die Möglichkeit bietet, zusätzlich zu den bestehenden Lautsprechern ergänzende imaginäre Lautsprecher einzufügen und damit perceptiv die Wiedergabe optimiert.<sup>38</sup> Für die Dekodierung zu weniger regelmäßigen Lautsprecheranordnungen ermöglicht der AllRADecoder die Bearbeitung von Lautsprecherkoordinaten und die automatische Berechnung eines Decoders innerhalb des Plugins.

Als Abspielmedium für 360° Videos gibt es die Möglichkeit einer VR-Brille (Virtual Reality), in welcher das Video virtuell sphärisch um den Betrachter projiziert wird. Durch die Bewegung des Kopfes bewegt sich auch das Bild in der Brille mit, sodass die visuelle Illusion eines Ortes in 360° geschaffen wird. Zu diesen Arten von VR- Brillen gehören z. B. die HTC Vive, oder die Oculus Quest. Kostengünstigere Versionen sind die Google Cardboard Linse in Kombination mit einem Smartphone. Aber auch ohne Brille, können sich die Videos durch Mausbewegungen an einem Bildschirm in alle Richtungen angeschaut werden.<sup>39</sup>

<sup>37</sup> Zotter und Frank, *Ambisonics*.

<sup>38</sup> Zotter und Frank.

<sup>39</sup> Olivieri, Peters, und Sen, „*A Technology Overview and Application to Next-Generation Audio, VR and 360° Video*“.

**20/21**

**Werkstück:**

**Graz - Stadt vs. Natur**

# Einordnung und Vorgehen

Nach dem kurzen Überblick über den Themenkomplex und die Produktion von Klanglandschaften wird nun näher auf die Umsetzung des Werkstücks eingegangen. Das Ziel der Arbeit liegt darin, immersive, audiovisuelle Umgebungen zu erstellen, basierend auf dem Ambisonics-Ansatz unter Verwendung von Ambisonics dritter Ordnung. Das resultierende natürlich-real wahrgenommene Schallfeld soll kreativ auf die Lärmbelastung in der Stadt Graz aufmerksam machen. Durch Ambisonics dritter Ordnung ist der Klang räumlich genauer aufgelöst und trägt damit zu einer immersiveren Erfahrung des Tons bei als bei herkömmlichen Methoden. Mit der Arbeit wird neben dem kreativen Produktionsprozess auch der technische und analytische Prozess beleuchtet, der für diese Art von Medien notwendig ist.

Für das Werkstück wurden zunächst die 360° Bild- und Tonaufnahmen an verschiedenen Orten in Graz aufgenommen, um anschließend diese zu analysieren und kreativ neu zu interpretieren. Die Dokumentation der aufgetretenen technischen Probleme soll auch dabei helfen zukünftige 360° Produktionen mit Ambisonics-Audio zu erleichtern.

## Auswahl der Orte und Zeitpunkt der Aufnahme

Für die Auswahl der verschiedenen Aufnahmeorte wurden Orte in Graz gesucht, die in sich in Stadtteilen mit unterschiedlichen Entfernungen zum Stadtzentrum befinden. Dadurch wird ein Licht auf den Klang der Natur und Lärmbelastung im Ballungsraum der Stadt und in den Randgebieten geworfen.

Bei den Aufnahmezeitpunkten wurde darauf geachtet, alle Aufnahmen innerhalb von einer Stunde am selben Tag durchzuführen, wobei sich für den 30.3.2022 um 16-17 Uhr entschieden wurde. Die Wetterlage war gleichbleibend für einen Monat mit sonnigem, fast wolkenlosem Himmel bei etwa 17°C. Durch die Beständigkeit des Wetters kurz vor Frühlingsanbruch erschien die Situation gut für den Zeitpunkt der Aufnahme.

Ort A: Der erste Ort befand sich im Grazer Stadtpark nahe der Ententeichkurve. Die Aufnahme wurde wenige Meter entfernt vom Weg unter einem Baum aufgenommen mit einigen Menschen in näherer Umgebung. Es sind Vogelgeräusche im Baum über der Hörposition zu vernehmen und Menschen, die sich auf Fahrrädern und zu Fuß an einem vorbei bewegen. Es handelt sich um einen vergleichsweise ruhigen Tag und es sind durchschnittlich viele Leute unterwegs.

Ort B: Der zweite Ort liegt am Rand des Naturerholungsgebietes Rosenhain, wobei als Standort eine Stelle in der Nähe eines Kindergartens gewählt wurde. Es befinden sich eine Wiese, Wald und ein Weg in unmittelbarer Nähe, es gehen ab und zu ein paar Leute vorbei und spielende Kinder sind zu hören. Der Weg führt Richtung Osten weiter in das Naturerholungsgebiet, und Richtung Westen innerhalb von wenigen hundert Metern in die Stadt, bzw. Wohnstraßen.

Ort C: Als dritter Ort wurde ein Wanderweg durch die Rettenbachklamm gewählt. Dieser befindet sich am Stadtrand von Graz im Stadtteil Mariatrost. Dort wurde eine Stelle gewählt die etwa 300 Meter innerhalb der Klamm liegt. Diese ist am Aufnahmeort am Boden nur wenige Meter breit. Der Aufbau wurde neben dem Wasser an der Seite des Wanderweges platziert. Durch die Lage in der Klamm dringen kaum Geräusche von außerhalb des Waldes durch das Wasserrauschen des Baches und es gibt Reflektionen von den Wänden und Bäumen. In dieser Szene kommen keine anderen Menschen vor.



*Abbildung 1: Ort A (Stadtpark): Ausschnitt des equirectangularen Videos*



*Abbildung 2: Ort B (Rosenhain): Ausschnitt des equirectangularen Videos*



*Abbildung 3: Ort C (Rettenbachklamm): Ausschnitt des equirectangularen Videos*

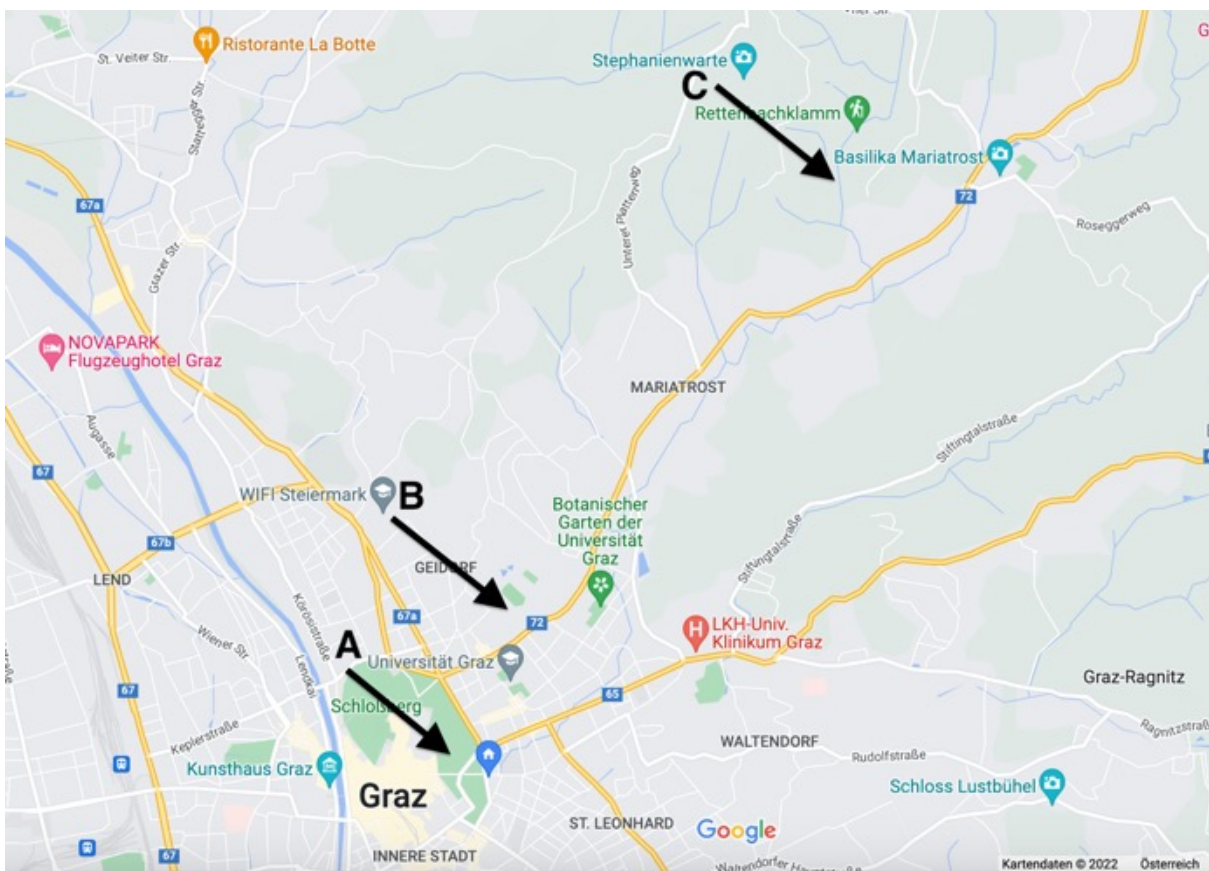


Abbildung 4: Aufnahmeorte Stadtpark (A), Rosenhain (B), Rettenbachklamm (C)





Abbildung 5: Aufnahmeort Stadtpark, Ententeich (A)

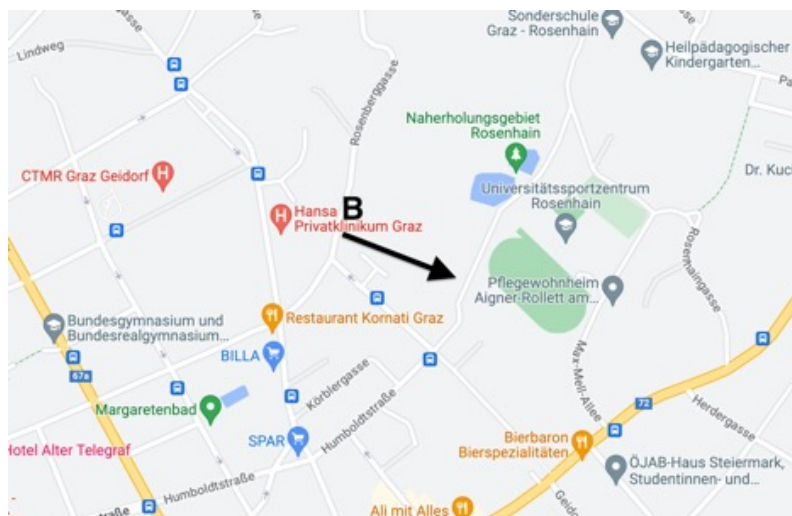


Abbildung 6: Aufnahmeort Rosenhain (B)



Abbildung 7: Aufnahmeort Rettenbachklamm (C)

## Technische Umsetzung

Für die technische Umsetzung des Projektes mussten einige neue Möglichkeiten in Betracht gezogen werden, da für diese Arbeit kein geeigneter Videoplayer zur Verfügung stand, der 360° Videos in Verbindung mit Ambisonics höherer Ordnungen abspielen kann. Die Plattformen, auf denen 360° Videos angeboten werden, wie Youtube oder Facebook, bieten nur 4 kanaliges Audio (somit 1. Ordnung Ambisonics). Es gibt, wie bereits im Kapitel „Produktion von Ambisonics Audio“ genannt, zwar einige Programme (G-Audioworks, Facebook Spatial Audio, Audio Easy 360 Pan Suite, usw.), die entweder Ambisonics erster Ordnung mit 360° Videos verarbeiten können, jedoch wird ein geeigneter binaural Decoder benötigt, um bei Ambisonics Spuren das Signal korrekt für Kopfhörer zu berechnen.

Um diese Lücke zu umgehen, wird in der vorliegenden Umsetzung Bild und Ton aufgeteilt, um beide Komponenten optimal bearbeiten zu können. Anschließend müssen beide Komponenten wieder zusammengeführt, um einerseits zeitlich und räumlich synchron daran arbeiten zu können und andererseits die Szene in VR erfahrbar zu machen.

In Abbildung 8 sind der Bearbeitungs- und Interaktionsprozess für die beiden Komponenten Bild und Ton dargestellt. Zudem zeigt die Abbildung auch jene Wege, die basierend auf einem macOS Betriebssystem, nicht funktioniert haben.

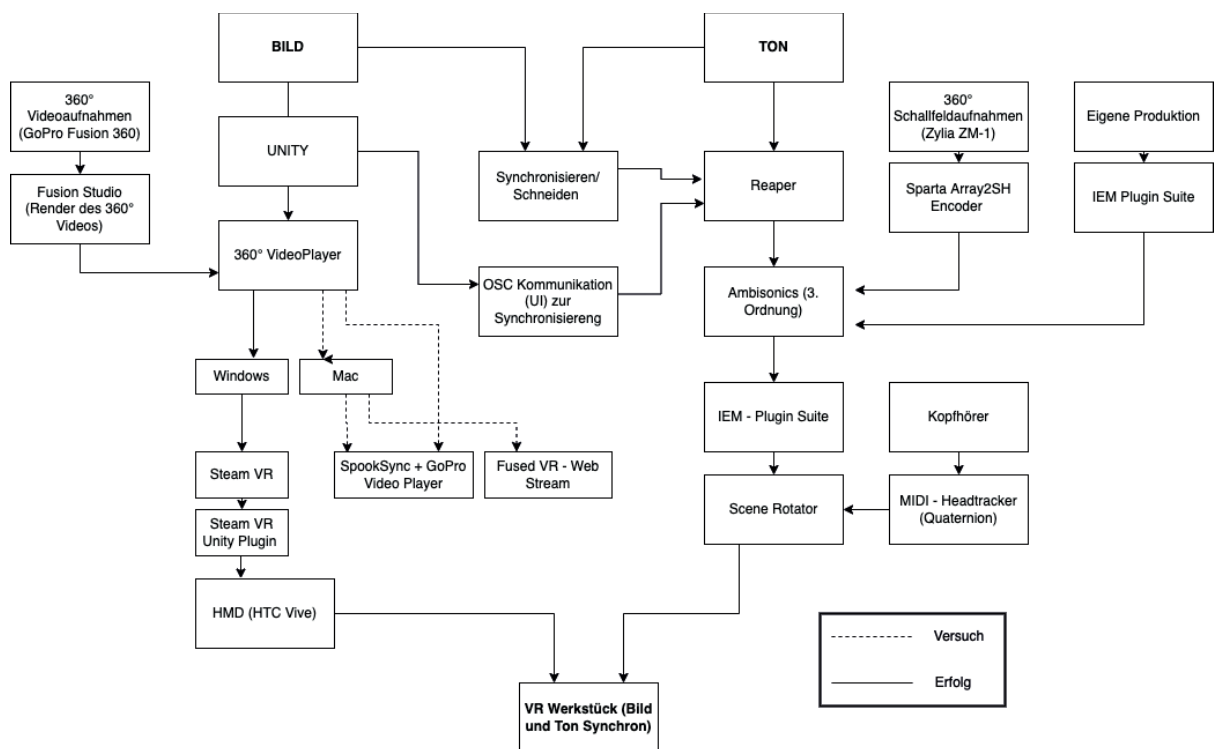


Abbildung 8: technische Umsetzung - Interaktion von Bild und Ton

## Aufbau

Um Ton und Videoaufnahmen der verschiedenen Orte gleichzeitig und vergleichbar aufzunehmen wurde ein Aufbau aus zwei Stativen genutzt. Die Stative hatten unterschiedliche Größen, wobei das kleinere genau unter dem größeren stand. Auf dem unteren Stativ wurde das Mikrofon montiert, auf dem oberen Stativ die 360° Kamera. Die Kamera war etwa zwei Meter über dem Boden positioniert, das Mikrofon etwa 40 cm tiefer als die Kamera. Durch diese Art des Aufbaus können von demselben Ort gleichzeitig hochwertige Video- und Audioaufnahmen gemacht werden, ohne sich gegenseitig zu beeinflussen. Das Mikrofon ist nicht im Bild zu sehen, und umgekehrt verdeckt die Kamera nicht das Mikrofon. Das Mikrofon wurde zusätzlich mit einem Windschutz bedeckt, um Windgeräuschen vorzubeugen. Die Ausrichtung und Synchronisation der Ton- und Bildaufnahmen konnte, als Eigenschaft der 360° Aufnahmen, in der Postproduktion angeglichen werden.

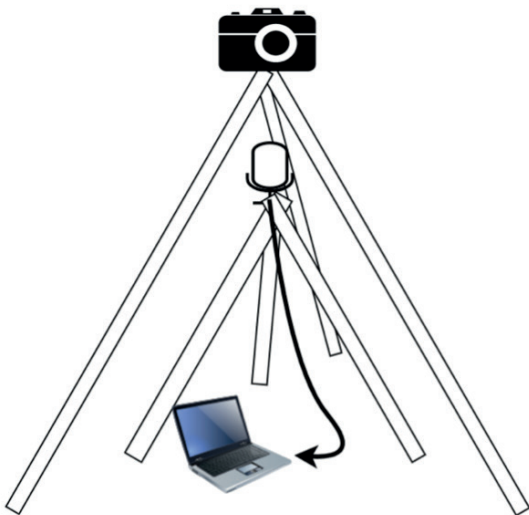


Abbildung 9: schematische Abbildung des Aufnahmeaufbaus

# Audioaufnahmen

Die Audioaufnahmen in diesem Projekt wurden mit einem Zylia ZM-1 <sup>40</sup> Mikrofon aufgenommen. Für die Aufnahmen wurde dieses mit einem Windschutz versehen. Das Mikrofon nimmt das Schallfeld mit 19 Kapseln unter Verwendung der digitalen Abtastrate von 48 kHz-Abtastrate und einer 24-Bit-Auflösung auf. Die 19 Spuren können z. B. mit dem Sparta Array2SH Encoder zu Ambisonics dritter Ordnung verarbeitet werden. Dieser bietet bereits eine Voreinstellung für das Zylia ZM-1 Mikrofon, mit der die Kanäle korrekt ihren sphärischen Koordinaten zugewiesen werden. Bei allen Aufnahmen wurde der aufgenommene Pegel nicht verändert, um die Vergleichbarkeit zwischen den Lautstärken der Szenen bei einer späteren Analyse zu erhalten.

<sup>40</sup> ZYLIA, „ZYLIA ZM-1 Microphone“.

## Audioverarbeitung

Für das Routing in Reaper wurden die Kanäle mit den Audioaufnahmen des Zylia ZM-1 Mikrofons (19 Kapseln) auf 20 Spuren gesetzt, wie in Abbildung 11 unter den Spuren „Ort 1/2/3“ zu sehen ist. Alle anderen Kanäle haben 16 Spuren für Ambisonics dritter Ordnung. Die Ausgänge aller Kanäle werden auf den Haupt-Bus geleitet, welcher das Signal an den Masterkanal sendet. Auf dem Haupt-Bus ist als Effektkette ein IEM Scene Rotator und danach ein IEM Binaural Decoder.

Der Scene Rotator wird mit einem Headtracker MIDI-Device (IEM MrHeadTracker<sup>41</sup>) verbunden, welcher auf den Kopfhörern montiert ist (Abbildung 10). Für die Verbindung des Headtrackers wird dieser als MIDI Device mit Reaper verknüpft, anschließend werden die Bewegungsdaten in dem SceneRotator Plugin auf die Quaternionen übertragen. Der Headtracker steuert die Quaternionen gegenläufig im und kann durch einen Knopfdruck kalibriert werden. Anschließend wird das Signal durch den IEM Binaural Decoder für die Kopfhörer aufbereitet. Es werden nun die Kopfbewegungen auf die Szene übertragen und die Schallquellen drehen sich gegenläufig zu den eigenen Bewegungen mit.



Abbildung 10: Headtracker auf Kopfhörern  
(Beyerdynamics DT 770 (80 Ohm))

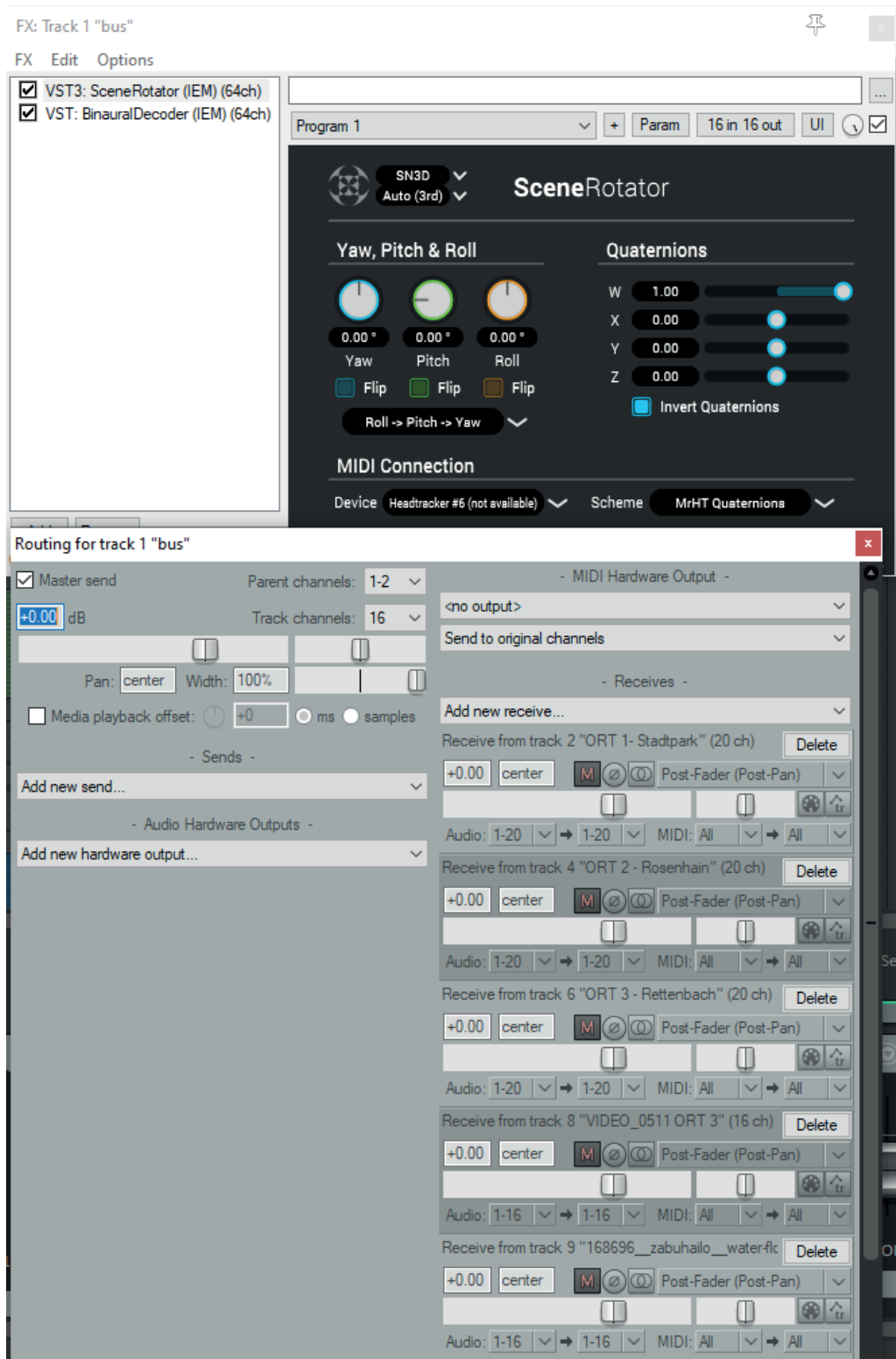


Abbildung 11: Routing des Bus-Kanals in Reaper. Empfängt alle Spuren - Master send. FX: IEM Scene Rotator mit Verbindung zu Headtracker; IEM Binaural Decoder

<sup>41</sup> Romanov u. a., *Implementation and Evaluation of a Low-cost Head-tracker for Binaural Synthesis.*

## Videoaufnahmen

Für die Videoaufnahmen wurde als Kamera die „GoPro Fusion 360“ verwendet. Die Kamera nimmt mit einer Auflösung von maximal 5,2K-Videos auf, welche anschließend in der dazugehörigen Schnittsoftware „GoPro Fusion Studio“ zu 360° Videos gerendert werden können.

Bei diesem Projekt wurde wegen der Größe der Dateien die Qualität auf 4K (3840 x 1920) gestellt und als Video-Codec der H.264 ausgewählt.

Das resultierende equirectangulare Video lässt sich mit einem geeigneten Videoplayer (z. B. GoPro Video Player, Unity) abspielen und in alle Richtungen betrachten.

## Bildverarbeitung

Die Software Unity wurde für dieses Projekt als 360°-Videoplayer verwendet, wegen der Möglichkeit OSC (Open Sound Control) zu versenden. Dazu wurde ein Videoplayer als Game Objekt erstellt, welches das equirectangulare Video entzerrt und auf eine 360° Oberfläche projiziert.<sup>42</sup> Die Oberfläche besteht dabei aus einer Rendering Textur, welche dieselbe Pixeldichte hat wie die aufgenommenen Videos (3840 x 1920).

Unity war zudem die Schnittstelle für die Verbindung mit dem „head mounted display“ (HMD) über Steam VR und dem Steam VR<sup>43</sup> Unity Plugin. Dafür wird der vorgefertigte „Player“ Game Objekt in das Projekt geladen und steuert direkt die Kamera und die Verbindung zum HMD. Über das HMD kann die Game-Ansicht in 360° mit den eigenen Kopfbewegungen betrachtet werden.



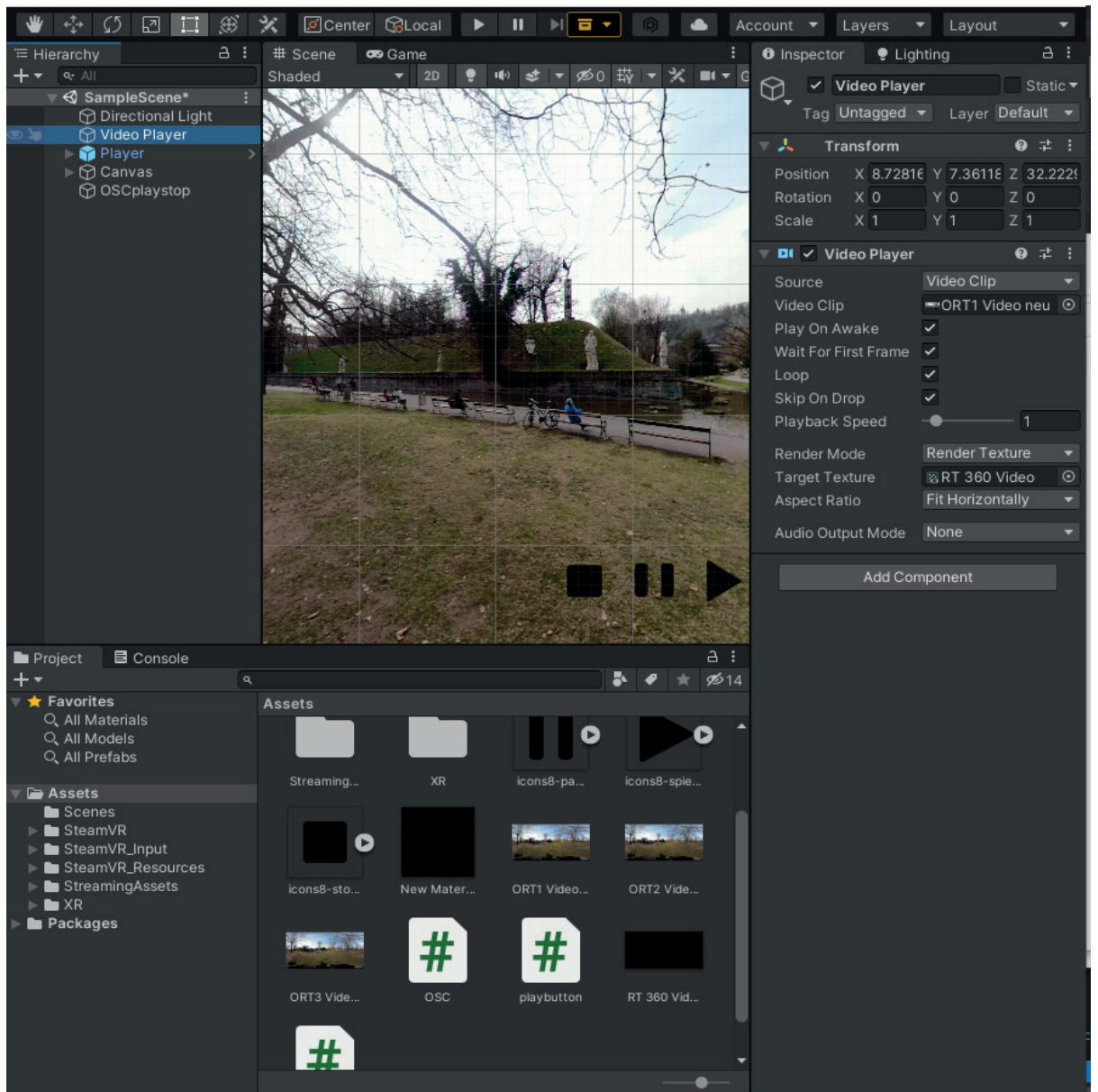


Abbildung 12: Screenshot des Videoplayers in Unity (Sceneview, Szene Ort A)

<sup>42</sup> Valem, *360 VIDEO PLAYER (with VR mode) - Unity Tutorial*.

<sup>43</sup> „SteamVR Plugin | Integration | Unity Asset Store“.

## Interaktion der Software

Für ein User Interface und die Verbindung zu Reaper wurden in Unity drei Knöpfe für Start, Pause und Stopp eingefügt. Diese sind jedoch nicht für Interaktionen mit einem VR Interface optimiert und sind daher nur auf der Oberfläche (Canvas) des 2D Computerbildschirms sichtbar. Diese Knöpfe übertragen ihre Funktionen über das Open Sound Control (OSC) an Reaper. Der dazugehörige Code für die OSC Verbindung stammt von Thomas Fredericks.<sup>44</sup> Dieser wurde angepasst an die Übertragung der Knopffunktionen und durch weitere Scripte ergänzt.

Für eine erfolgreiche Verbindung sind in beiden Programmen die IP-Adresse und Portnummern richtig einzustellen (Abbildung 13). Durch diese Verbindung ist ein gleichzeitiges Starten und Stoppen des Videos in Unity und des Tons in Reaper, und damit Synchronität, möglich. Der Code für die OSC Verbindung der Knöpfe wurde angepasst, um die richtigen OSC Daten, wie „starten“ oder „pausieren“, an Reaper zu senden.

Zunächst wurde auch versucht die Rotationsdaten der Kamera in Unity zu verwenden, um auch den Ton in Reaper über OSC zu steuern, was jedoch nur mit der Maus und einem Kamerascript in Unity, und nicht der HTC Vive funktioniert hat.

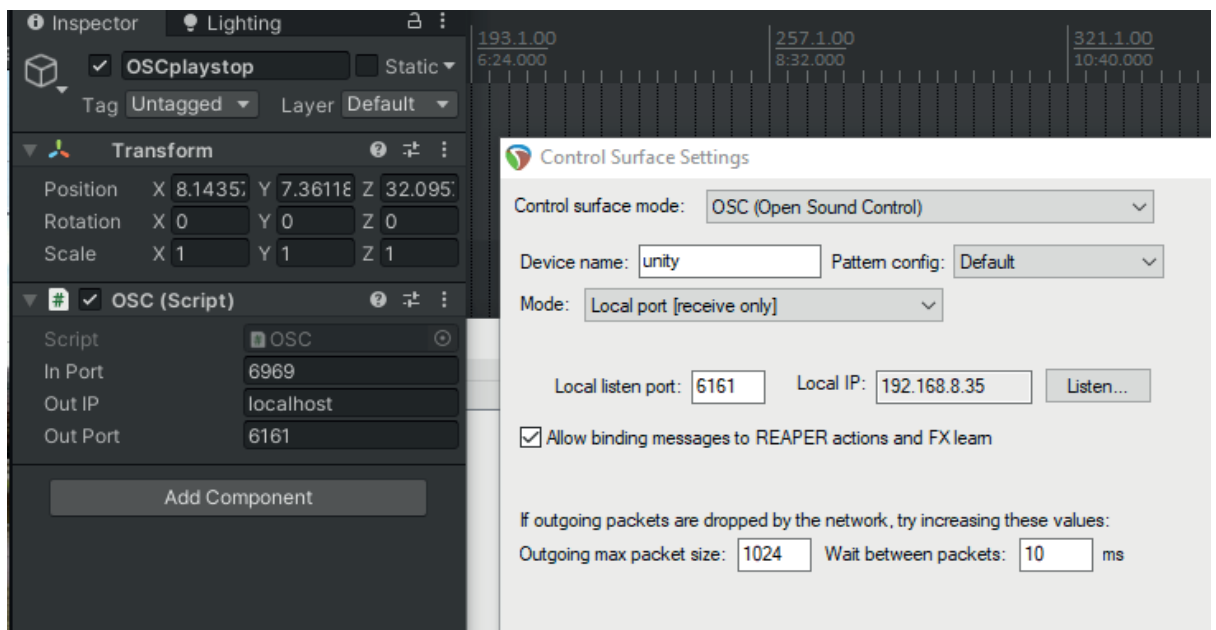


Abbildung 13: OSC Schnittstellen in Unity (links) und Reaper (rechts)

<sup>44</sup> Fredericks, Thomas, „UnityOSC - Open Sound Control (OSC) for Unity 3D“.

## Analyse und Einordnung der Schallfeldaufnahmen

Bei der Analyse und Einordnung von Schallfeldaufnahmen sollte immer der Ort und Zeit mit einbezogen werden. Da der Fokus des Werkstückes auf die Beziehung zwischen dem Klang der Stadt und der Natur gelegt wurde, wurden die Aufnahmeorte in verschiedenen Stadtteilen von Graz, unterschiedlich weit vom Zentrum entfernt, festgelegt. Dabei konnten vor Ort unterschiedliche Verhältnisse von Verkehr und Naturgeräusche wahrgenommen werden.

Besonders deutlich erfahrbar wird die Nähe zum Zentrum bei der Lage der Aufnahmeorte im Vergleich mit der Verkehrslärmbelastung.

Dies wird erkennbar in den Abbildungen 14-17, welche die geografische Verteilung des Verkehrs in der Stadt Graz veranschaulichen.

Im Zentrum nahe dem Stadtpark liegt die Glacisstraße, eine der lautesten Straßen Graz mit etwa 85dB (Tagesschnitt), direkt neben dem Erholungsort Stadtpark (Ort A). In Richtung des Randgebietes der Stadt nimmt die durchschnittliche Lautstärke ab und liegt am Rosenhain (Ort B) und in der Rettenbachklamm (Ort C) nur noch bei etwa 50dB pro Tag. Die Daten der Abbildungen sind aus dem Jahr 2011, und da in den letzten 10 Jahren die Stadt Graz um etwa 13% gewachsen ist (2011: 258 000 Einwohner; 2021: 291 000 ), kann heutzutage mit einem vergleichbaren, wenn nicht sogar höherem Verkehrsaufkommen gerechnet werden.<sup>45</sup>

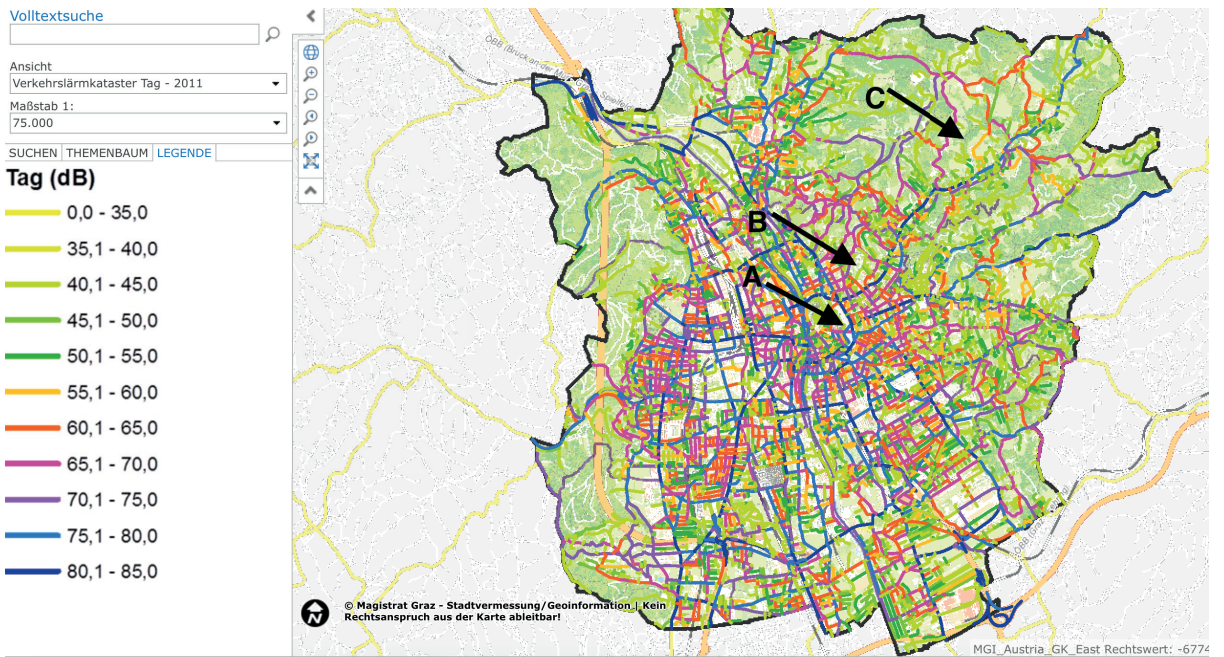


Abbildung 14: Verkehrslärmkataster Stadt Graz (2011), Aufnahmeorte (A), (B), (C)

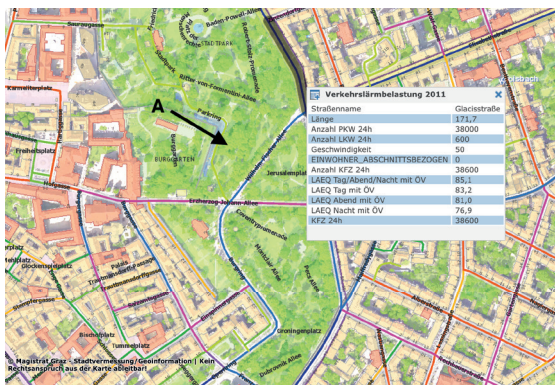


Abbildung 15: Lärmbelastung im Detail: Stadtpark (A) (Daten der Glacisstraße 2011)

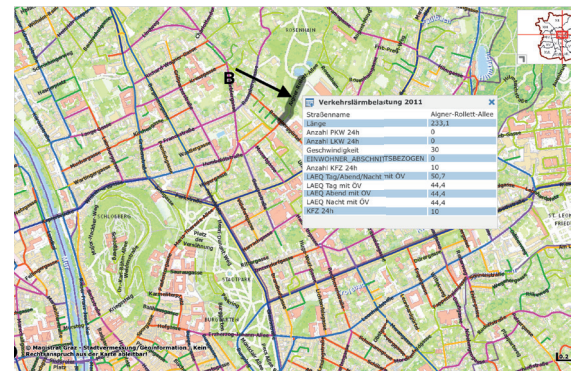


Abbildung 16: Lärmbelastung im Detail: Rosenhain (B) (Daten der Aigner-Rolett-Allee 2011)

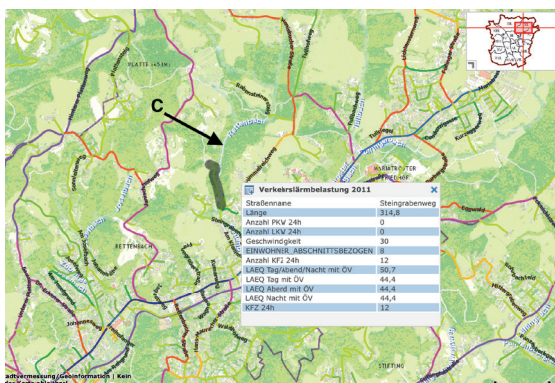


Abbildung 17: Lärmbelastung im Detail: Rettenbachklamm (C) (Daten des Steingrabenweg 2011)

<sup>45</sup> WIBIS-Steiermark, „Einwohner Graz, Steiermark“.

In den Abbildungen 18 (Ort A, 19 (Ort B), 20 (Ort C) sind die Wellenformen der Audioaufnahmen zu sehen. Dafür wurde nur die richtungsunabhängige Ambisonics-Repräsentation, der W-Kanal der Ambisonics-Datei, analysiert. Da alle Aufnahmen mit denselben Mikrofoneinstellungen (Vorverstärkung) aufgenommen wurden, sind die Werte der Lautstärke vergleichbar. Dabei ist auch zu beachten, dass die angeführten Pegel nicht die Lautstärke (dB SPL/ Phon) beschreiben, sondern auf die digitale maximale Vollaussteuerung bezogen sind (full-scale range). Es können dadurch zumindest die relativen Pegelunterschiede als Hinweis auf große Lautstärkeunterschiede in Erfahrung gebracht werden.

Deutlich wird der Lautstärkeabfall mit der Entfernung zum Zentrum. Bei Ort A bewegt sich die Lautstärke zwischen -35 dB und -50 dB, hingegen bei Ort B zwischen -40 dB und -65 dB und an Ort C sogar nur zwischen -47 dB und -62 dB. Diese Lautstärkeunterschiede sind auch subjektiv eindeutig und offensichtlich wahrnehmbar, da in der Stadt der Verkehr und andere Klänge sich zu einem Geräuschteppich vermischen.

Für die Analyse der Video- und Audioaufnahmen lassen sich an jedem Ort markante Klänge erkennen.

In der Szene im Stadtpark (Ort A) ist man von einigen Menschen umgeben, welche auch hörbar sind. Es sind Bruchstücke von Unterhaltungen erkennbar und Geräusche von Fahrrädern sowie der nahe liegenden Straße. Es sind auch Vögel zu sehen und zu hören. Ein Raabe setzt sich während der

Szene auf einen Ast über der Aufnahme-position und gibt Laute von sich.

Am Rosenhain (Ort B) hingegen sind die Stadtgeräusche deutlich weniger zu hören und es sind mehr Bäume und Vögel zu hören. Außerdem sind durch die anliegende Wiese und den Kindergarten sehr viele Kinderstimmen und -schreie zu hören sowie ab und zu vorbeigehende Passanten und Fahrradfahrer.

In der Rettenbachklamm (Ort C) zeigt die Szene akustisch und visuell wenige Veränderungen. Man befindet sich innerhalb der Klamm direkt neben einem Bach und ist umgeben von Wald und großen Steinen. Durch die eng zulaufenden Wände der Klamm entstehen leichte Reflektionen, die von der gegenüberliegenden Seite des Baches zurückgeworfen werden. Neben dem Rauschen des Baches und der Vögel im Wald sind kaum andere Geräusche zu hören, bis auf ein sehr entferntes Straßengeräusch.

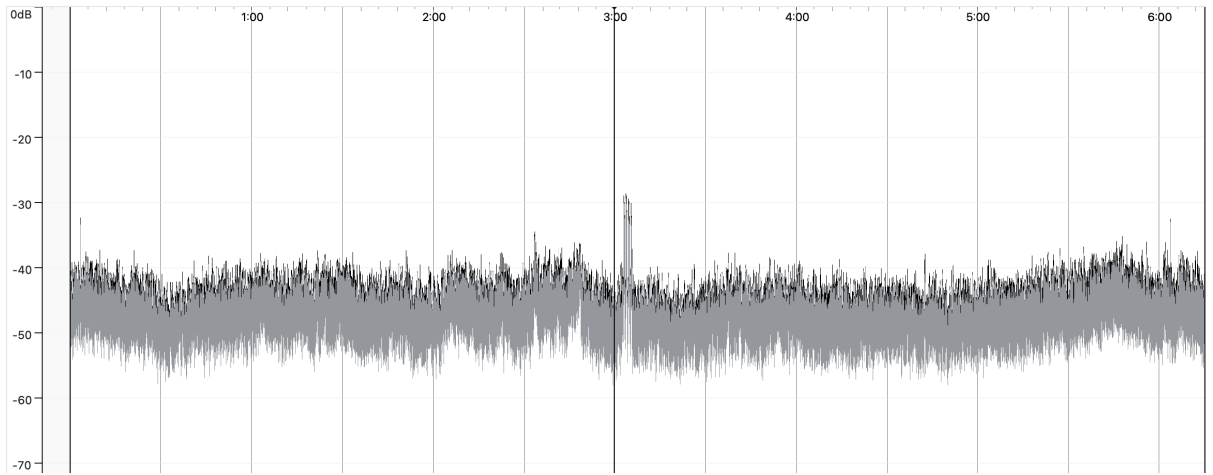


Abbildung 18: Stadtpark (A): Wellenform der Audioaufnahme des W-Kanals

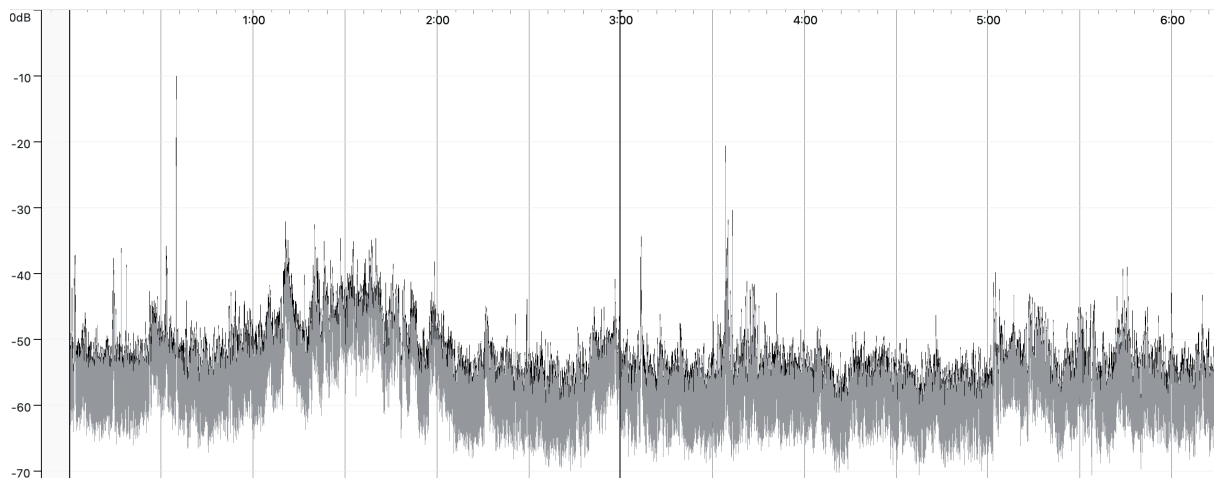


Abbildung 19: Rosenhain (B): Wellenform der Audioaufnahme des W-Kanals

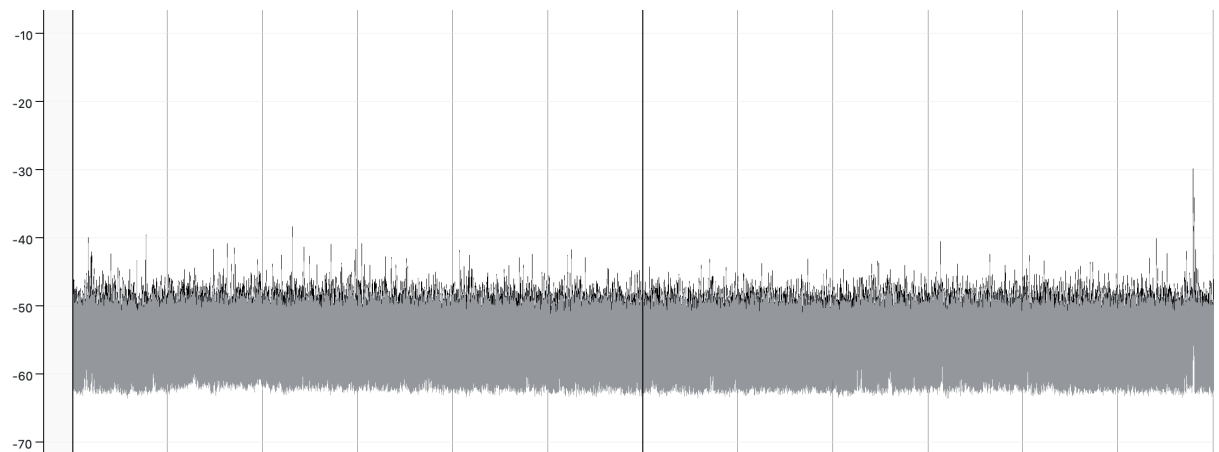


Abbildung 20: Rettenbachklamm (C): Wellenform der Audioaufnahme des W-Kanals

# Kreativer Prozess

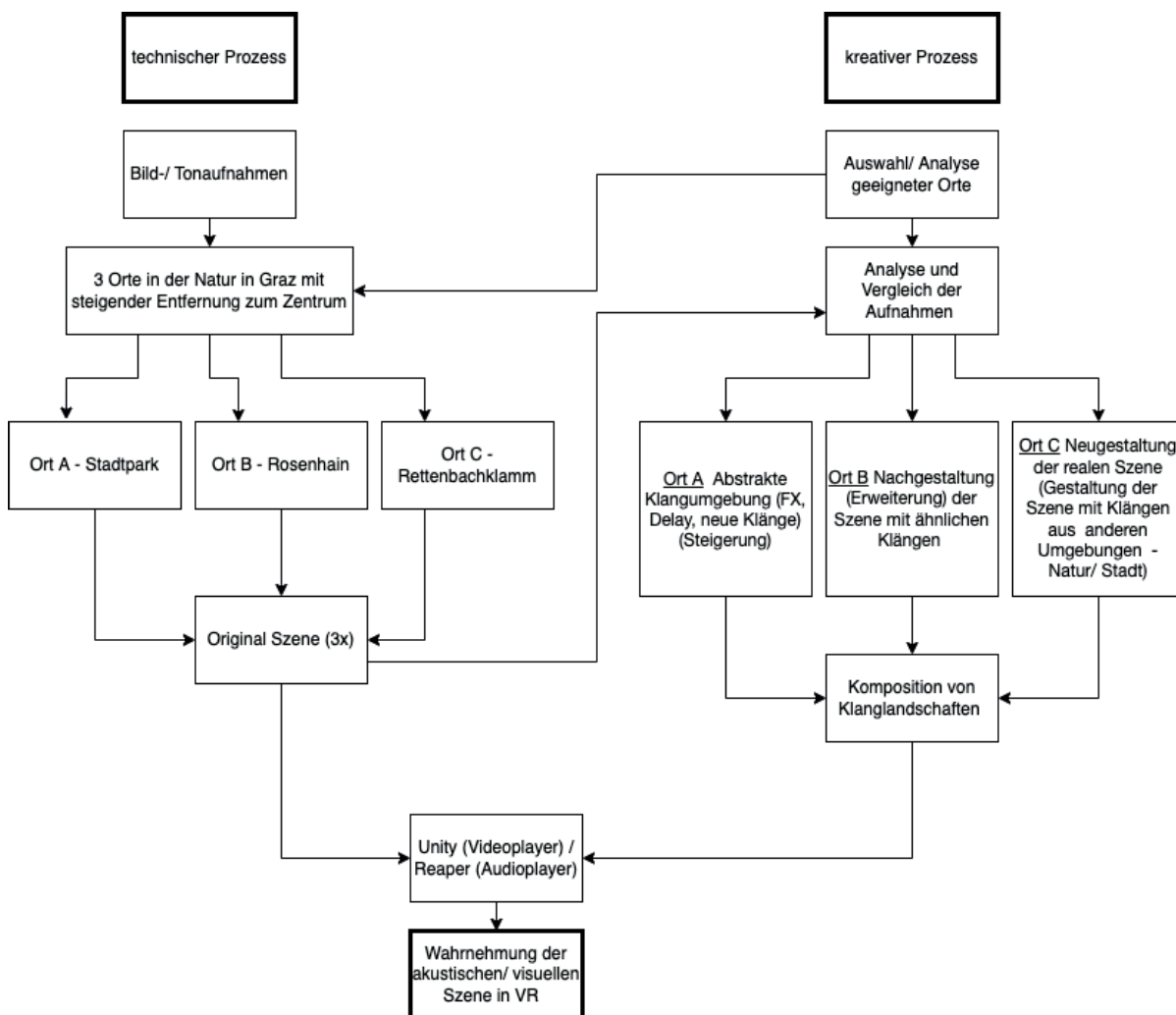


Abbildung 21: Verbindung von kreativem und technischem Prozess



Der kreative Produktionsprozess wurde erst nach den Aufnahmen und der Analyse gestartet. Nach dem erfolgreichen technischen Aufbau konnte mit dem HMD und den Kopfhörern mit Headtracker produziert werden, was den Produktionsprozess in VR erleichtert, da man sich währenddessen umhören/umschauen kann. Nach der Analyse der verschiedenen Klangumgebungen der Orte wurden die charakteristischen und prägnanten Klänge festgestellt und um diese herum eine Klangwelt neu erstellt. Dabei wurde für jede Szene ein eigener Ansatz gewählt. Da in der Szene an Ort A viele einzelne Schallquellen wahrzunehmen sind, wie Stimmen der vorbeigehenden Personen, Rabenlaute oder die naheliegende Straße, wurde eine abstrakte Art der Vertonung gewählt. Diese basiert hauptsächlich auf Echo- und Halleffekten. Durch das Plugin IEM Dual Delay und IEM FDN Reverb wurden gleichmäßig ansteigend sphärische Mehrkanal-Echo- und Halleffekte mit dem Originalaudiomaterial vermischt. Es erscheint durch die sich ständig vermischenden und sich wiederholenden Klänge ein Klangteppich aus dem weniger zu erkennen ist als zuvor. Damit soll verdeutlicht werden, wie die Überlagerung von Schall in der Stadt zu einer Verringerung der akustisch wahrnehmbaren Auflösung beiträgt.

An Ort B wurde bedingt durch die deutlich wahrnehmbaren Kinderlaute und die entfernten Stadt- und Naturgeräusche, wie Verkehr oder Blätterrascheln auf eine Verstärkung des Ausgangsmaterials konzentriert. Dafür wurden teilweise neue Klänge von ähnlichen Umgebungen verwendet, wie z.B. Kin-

dergarten-, Stadt- oder Naturgeräusche. Durch die Verstärkung dieser dominanten Geräusche bleiben zwar die bildsynchronen Klänge, wie Schritte und Fahrräder erhalten, jedoch wird die Hintergrundszenerie verstärkt.

Für Ort C wurde aufgrund der Klangumgebung mit wenigen Veränderungen eine komplett neue akustische Szene erstellt. Für diese wurden alle am Originalort vorkommenden Klänge ersetzt und um mehr Klänge, wie Tiergeräusche, ergänzt. Durch diese neuartige Klanglandschaft mit sphärisch korrelierenden Richtungen der Klangquellen, wie der Bach oder die Vögel, entsteht eine „realistisch wirkende“ Umgebung. Dabei können vom Sound Designer Klangquellen ersetzt oder ausgetauscht werden, ohne, bis zu einem gewissen Grad, den akustischen Realismus zu verlieren. Vorausgesetzt ist dabei jedoch, dass das Audio mit dem visuellen korreliert. Für weitere kreative Möglichkeiten bleibt sehr viel Spielraum, da das Arbeiten im 360° Raum neue Wege eröffnet. Dazu gehören räumliche Effektplugins, wie Hall oder Echo, die sich verändern können, aber auch die Möglichkeit, Klänge um den Hörer herum zu platzieren oder zu entfernen.

Die verwendeten Klänge für die Erstellung der neuen Klangumgebungen wurden von [freesound.org](https://freesound.org) heruntergeladen und an die jeweilige Szene in Reaper angepasst. Da es sich dabei hauptsächlich um Stereoaufnahmen von gänzlich anderen Orten handelt, wurden die Richtungsinformation mit dem IEM Stereo Encoder erzeugt und die Klänge somit in der 360° Szene platziert.

## Abspielmedium/ Ergebnis

Da das Projekt audiovisuell für den 360° Raum erstellt wurde, kann es nur mit einem HMD und Kopfhörern entsprechend wahrgenommen werden. Als HMD wird die HTC Vive verwendet. Um die Bewegungen des Kopfes auf die Bewegungen des Videos in der HTC Vive zu übertragen, werden zwei gegenüberliegende Infrarotsensoren der Brille im Raum platziert, wodurch die Brille mit den angebrachten Sensoren die Rotation ermittelt.

In Unity übernimmt das SteamVR Plugin die Kameraführung und passt das Bild bei entsprechenden Kopfbewegungen an. Für die Übertragung der Rotationsdaten des Kopfes für den Ton wurde ein MIDI Headtracker an den Kopfhörern befestigt, welcher durch eine USB-Verbindung direkt mit dem Scene-Rotator in Reaper verbunden ist. Dort kann zwischen verschiedenen Optionen der Gyroskopdaten gewählt werden. Für

dieses Projekt wurde die Einstellung MrHT Quaternions (Invert) verwendet. Beim Abspielen des Videos durch das User Interface in Unity startet alles synchron und man kann in VR die Szenen wahrnehmen.

Für die Demonstration der Installation, ohne die notwendigen technischen Voraussetzungen, wurden für jede mögliche Szene (dreimal die Originalszene, dreimal neues Sound Design) Aufnahmen angefertigt, die es einem ermöglichen, mit Kopfhörern und Bildschirm die Kopfbewegungen und die Szene nachzuempfinden. Da jedes Video etwa fünf Minuten dauert (30 Minuten für alle Beispiele), sind die aufgenommenen Szenen etwa halb so lang wie die virtuelle Umgebung, um einen Eindruck der Erfahrung zu verschaffen.

Diese aufgenommenen Beispiele sind jedoch nicht vergleichbar mit der Erfahrung durch das HMD und Kopfhörer, da die vom Kopf aufgenommenen Kamerabewegungen viel zu schnell sind für einen üblichen Kameraschwenk. Dadurch fallen einem mit statischem Kopf die Phasenüberlagerungen und minimalen Frequenzunterschiede der Aufnahme bei der Drehung des Kopfes viel mehr auf, als wenn die Drehung von einem selbst gesteuert wird. Bei der Erfahrung in der virtuellen Installation dreht man automatisch den Kopf, um die verschiedenen Schallquellen räumlich mit dem visuellen zu verknüpfen.

Dieses „Erkunden“ der akustischen Landschaft geht ohne VR verloren und ist durch einen zweidimensionalen Bildschirm schwer nachzuvollziehen. Für eine realistische Erfahrung des Tons muss man sich selbst in der Szene bewegen, anstatt mit einer Maus das Bild zu bewegen oder den Kopfbewegungen einer anderen Person zu folgen. Dennoch ist die Darstellung in diesem Fall gewählt worden, um einen Eindruck des Werkstücks zu verschaffen, der ohne den technischen Aufbau erfahrbar ist.

# Diskussion

Auch wenn es derzeit noch wenige Möglichkeiten gibt, mit 360° Audio zu arbeiten, gibt es immer mehr Alternativen. Für dieses Projekt wurden Ambisonische dritter Ordnung und eine GoPro Fusion 360 (4k Auflösung auf 360° Projektionsfläche) verwendet. Mit heutigem technischem Stand könnten bereits sehr viel höher aufgelöste Videos, bspw. mit der Insta 360 Pro, aufgenommen werden, sowie Ambisonics fünfter Ordnung mit dem Eigenmike Mikrofon. Da dieses Equipment jedoch schwer zu bekommen ist, sowie die Rechenleistung, die benötigt wird, um die Videos zu rendern, außerhalb des Rahmens dieser Masterarbeit liegt, wurde mit nicht ganz so hoher Qualität aufgenommen.

In zukünftigen Installationen könnte eine leicht höhere Audio- und Bildqualität erreicht werden.

Dieses Werkstück hebt sich in dem Sinn von anderen Arbeiten ab, das mit der Produktion von akustischen 360° Szenen in analytischem, sowie kreativen Kontext gearbeitet wurde. Der technische Aufbau ist mit ähnlichen technischen Voraussetzungen rekonstruierbar und zukünftige Umsetzungen könnten aus dem Arbeitsprozess dieser Arbeit lernen, welche Lösungsansätze funktionieren und welche nicht funktionieren. Die akustische und visuelle Umgebung der Stadt Graz wurde an unterschiedlichen Orten aufgenommen, um auf eine gewisse Art den Charakter der Stadt einzufangen. Wahrscheinlich könnten selbst in Graz, oder in anderen Städten ganz andere Muster (z.B. Verkehr oder Veranstaltungen) zu anderen Zeiten auftreten und neu im akustischen Bild einer Stadt interpretiert werden.

Auch im Bereich der Klanglandschaften-Komposition gibt es meiner Meinung nach viele Möglichkeiten diese in neuen Medien, wie Spielen in virtueller Realität, oder Kunstinstallationen unterzubringen, da Menschen ein neuartiges immersives Erlebnis geboten wird. Im Vergleich zu anderen künstlerischen Arbeiten mit Bezug zu Schallfeldaufnahmen ist das Ergebnis dieser Arbeit eher unauffällig und subtil, da eine eher realistische bis leicht abstrakte akustische Darstellung verfolgt wurde. Bei der Sound Design Vertonung von Ort A wurde es abstrakter, was jedoch auch extremer umgesetzt werden könnte. Interessante Ergebnisse könnte man vielleicht auch mit animierten Videos erzielen, da dann das Bild auch auf den Ton reagieren könnte und nicht nur umgekehrt.

Bei der Präsentation des Werkstücks unbedingt mit einbezogen werden, dass die Erfahrung mit dem HMD und den Kopfhörern ein gänzlich anderes Erlebnis bieten als die im Nachhinein aufgenommenen Bildschirmaufnahmen zur Veranschaulichung. Da-

bei geht auf der einen Seite die immersive akustische Erfahrung verloren, da die Bewegungen des Tons nicht von den eigenen Kopfbewegungen kommen. Als einzigen Anhaltspunkt bleiben die Bewegungen der dazugehörigen Bildschirmaufnahme des HMD. Die Kameraführungen sind jedoch sehr schnell, da sie mit den Kopfbewegungen einer Person korrelieren, die sich in der Szene umschaute und diese „erforscht“. Durch die Neugier dieses neuartigen Mediums wird von Teilnehmern automatisch versucht den Kopf zu drehen, um die Schallquellen besser lokalisieren zu können. Dennoch fiel die Entscheidung auf diese Variante, da sich das Werkstück mit der Immersiven Erfahrung beschäftigt, und dies nur ein Abbild zu einem gewissen Grad davon darstellen kann.

Aufgrund der Vielzahl technischen Möglichkeiten, wie so ein Projekt umgesetzt werden kann (Kameras, Mikrofone, Software), sowie dem Mangel an Software für diese Art von Medien, gibt es auch andere Wege, die zum Ziel führen können.

Für eine Ausstellungssituation in der Öffentlichkeit müssen einige technischen Details optimiert werden, um auf der einen Seite mehrere Menschen gleichzeitig teilhaben zu lassen, und auf der anderen Seite über ein kompaktes System die Installation kabellos gestalten zu können.

Zudem dienen die Originalaufnahmen der Orte auch als historisches Zeugnis, da Sie wie Bilder eines Ortes auch in Zukunft noch dazu die akustische und visuelle Szene genau dieser Orte festhalten. Vor allem akustisch gibt es be-

reits viele Arten der Schallfeldaufnahmen, jedoch wenige omnidirektionale Videos mit der dazugehörigen räumlichen akustischen Umgebung.

Darüber hinaus war die Arbeit an dem Werkstück eine Übung und Analyse, wie Menschen im Alltag mit Akustik umgehen und welche Klänge einem mehr/ weniger auffallen, wenn man vor Ort ist.

Das Themengebiet beschäftigt sich mit der Beziehung zwischen dem Menschen an urbanen Orten und der Präsenz von Natur. Natur ist fast in jeder Stadt zu finden, dennoch hängt es stark von dem Ort ab, ob bei der Stadtplanung Natur mit einbezogen ist, oder wie wirksam diese ist. Menschen gewöhnen sich an gleichbleibende Klänge schnell und konzentrieren sich eher auch ein Gespräch des Freundes im Park als die Straßengeräusche. Dennoch verschwinden die Klänge nicht und können sich auf das allgemeine Wohlbefinden, wie Stress, auswirken.

## Fazit

Zusammenfassend kann gesagt werden, dass es möglich ist, eine reale Umgebung für interaktive virtuelle Erfahrung in Bild und Ton in hochwertiger natürlicher/ realistischer Qualität technisch zu erfassen. Die Verbindung von Ton und Bild ist durch die noch fehlenden, geeigneten Videoplayer kompliziert. Wie gezeigt, kann mit Reaper und Unity eine Alternative erstellt werden, in der Ton und Bild synchron abgespielt werden, beide aber auch auf die Bewegungsdaten des Kopfes reagieren.

Bei der Analyse der Aufnahmen kann die Theorie bestätigt werden, dass die Stadt Graz in der Innenstadt mehr Lärmbelastung hat als außerhalb der Stadt. Dies geht auch durch Verkehrslärmbelastungsanalysen der Stadt Graz hervor und steht im Kontrast zu den Klängen der Natur. Diese Beziehung kann durch das Werkstück immersiv erfahren werden, sowie eine erweiterte künstlerische Ebene, bei der mit Klängen der Szenen kreativ verschiedene Eindrücke durch Sound Design behandelt werden.

Als Sounddesigner steht man sowohl vor künstlerischen als auch vor technischen Herausforderungen. Die Technik muss beherrscht werden, um künstlerische Entscheidungen frei treffen zu können. Technische Anweisungen können weitergegeben werden, aber die künstlerische Freiheit liegt beim Schöpfer. Geschmack und Ästhetik liegen immer beim Betrachter und die Vorlieben können sehr unterschiedlich sein.

Das begleitende räumliche Werkstück hat sich fast mehr als eine technische Umsetzung denn als künstlerische Leistung herausgestellt, dennoch können durch die künstlerischen Anstöße mit dieser Art von Technik neue Denkanstöße erzeugt werden, da Menschen akustische Szenen aus anderen Blickwinkeln betrachten können. Der kreative Teil wurde bewusst subtil gestaltet, um das Ohr des Zuhörers auf die leichten Veränderungen zu schärfen. Obwohl in einer 3D-Audioumgebung mehr Platz für zusätzliche Klänge vorhanden ist als in einem begrenzten Stereobild, ist auch bei 360° Audio weniger manchmal mehr.

Eine weitere wichtige Erkenntnis ist, dass die Abbildung dieses immersiven Werkstücks kaum möglich ist, auf einem zweidimensionalen Bildschirm mit Kopfhörern. Die Verbindung der Kopfbewegungen mit den Bewegungen von Bild und Ton machen die Erfahrung realistischer und interessant. Ohne diese Verbindung lässt sich die Wirkung von virtueller Realität nur erahnen.

Da die meisten Plattformen, die 360° Medien anbieten, nur Ambisonics erster Ordnung unterstützen ist diese Art von Technik bisher nicht für die breite Masse erfahrbar, jedoch von den technischen Voraussetzungen im Bereich des Möglichen.

Da die Arbeitsphase für diese Masterarbeit zeitlich begrenzt war, bleiben noch einige Aufgaben für zukünftige Forschungen.

Ich denke das diese Art von Medien vor allem für die Spiele- und Filmindustrie interessant sein können, durch ein steigendes Maß an immersiven Umgebungen, sowie Kunstinstallationen. Technisch müssen noch ein paar kleine Schritte passieren, um das Medium für mehr Menschen und Künstler zugänglich zu machen, wobei auch das Interesse danach gefordert ist.

# Dateianhang

Link zum Dateianhang:

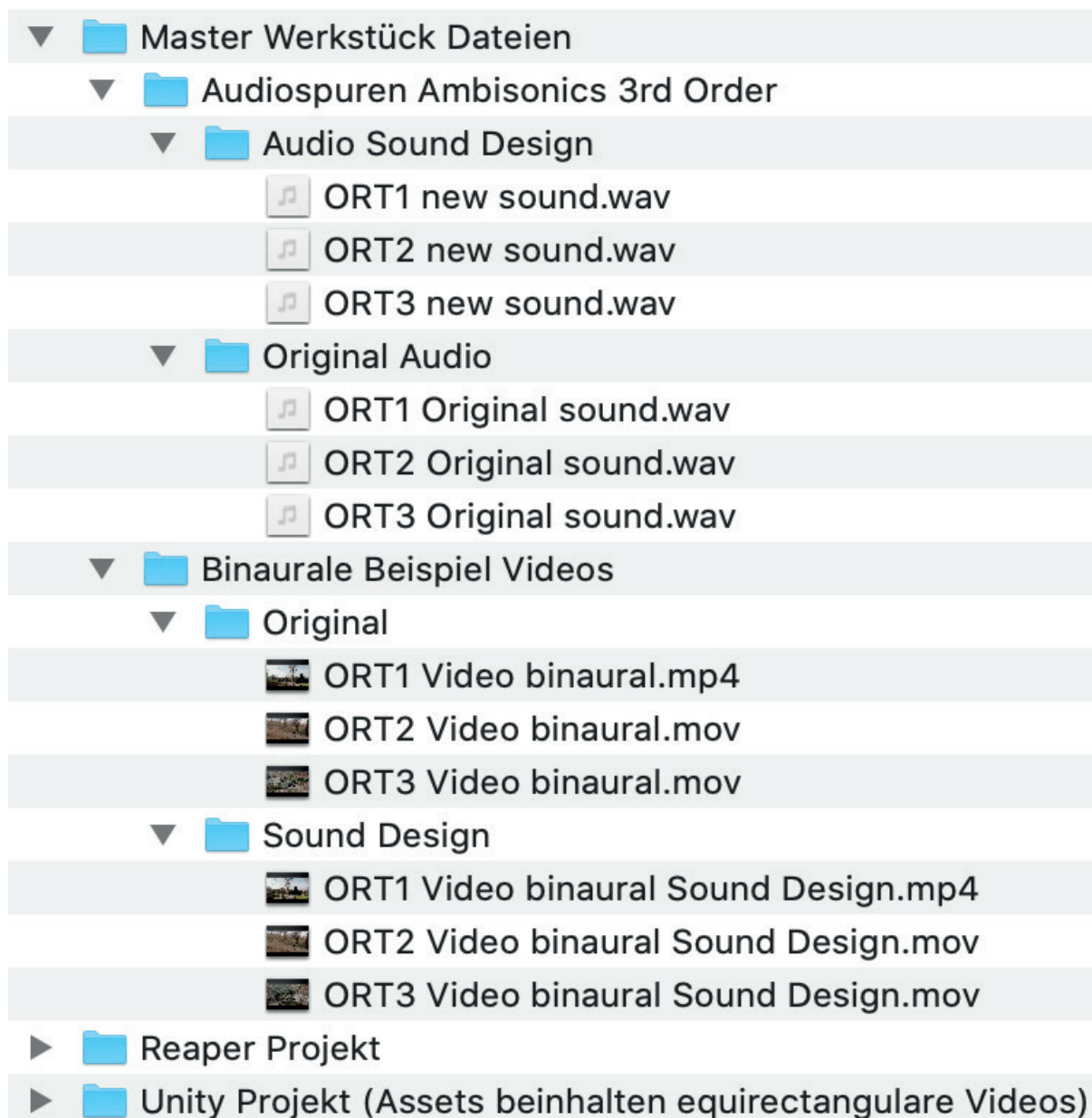


Abbildung 22: Werkstück-Dateianhang; beinhaltet Original- und durch Sound Design veränderte Ambisonics-Audiospuren, binaurale Beispielvideos (2D), equirectangulare Videos, Reaper Projekt, Unity Projekt und C# Skripte,



# Literaturverzeichnis

- „5.1“. In Wikipedia, 9. Juli 2020. <https://de.wikipedia.org/w/index.php?title=5.1&oldid=201727730>.
- Assunção, Pedro Amado, und Atanas Gotchev, Hrsg. 3D Visual Content Creation, Coding and Delivery. 1st ed. 2019. Signals and Communication Technology. Cham: Springer International Publishing : Imprint: Springer, 2019. <https://doi.org/10.1007/978-3-319-77842-6>.
- Blumstein, Daniel T., Daniel J. Mennill, Patrick Clemins, Lewis Girod, Kung Yao, Gail Patricelli, Jill L. Deppe, u. a. „Acoustic Monitoring in Terrestrial Environments Using Microphone Arrays: Applications, Technological Considerations and Prospectus“. *Journal of Applied Ecology* 48, Nr. 3 (2011): 758–67. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>.
- Chion, Michel, Claudia Gorbman, und Walter Murch. *Audio-Vision: Sound on Screen*. New York: Columbia University Press, 1994.
- Eaton, Callum. „Quantifying Factors of Auditory Immersion for Virtual Reality“. Masters, University of Huddersfield, 2020. <http://eprints.hud.ac.uk/id/eprint/35372/>.
- Francis, Clinton, und Jesse Barber. „A framework for understanding noise impacts on wildlife: An urgent conservation priority“. *Frontiers in Ecology and the Environment* 11 (8. August 2013). <https://doi.org/10.1890/120183>.
- Frank, Matthias. „How to Make Ambisonics Sound Good“, 2014, 6.
- Frank, Matthias, Franz Zotter, und Alois Sontacchi. *Producing 3D Audio in Ambisonics*. Proceedings of the AES International Conference. Bd. 2015, 2015.
- Fredericks, Thomas. „UnityOSC - Open Sound Control (OSC) for Unity 3D“. Zugegriffen 12. Mai 2022. <https://thomasfredericks.github.io/UnityOSC/>. freesound.org, abgerufen im April 2022
- Gallagher, Michael. „Field Recording and the Sounding of Spaces“. *Environment and Planning D: Society and Space* 33, Nr. 3 (Juni 2015): 560–76. <https://doi.org/10.1177/0263775815594310>.
- Gerzon, Michael J. „Periphone (with Height Sound Reproduction)“. *Audio Engineering Society*, 1972. <https://www.aes.org/e-lib/browse.cfm?elib=1835>.
- GoPro Fusio Studio. „GoPro Fusion | GoPro“. Zugegriffen 4. April 2022. <https://gopro.com/de/at/update/fusion>.
- Hamasaki, Kimio, Toshiyuki Nishiguchi, Reiko Okumura, Yasushige Nakayama, und Akio Ando. „A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)“. *SMPTE Motion Imaging Journal* 117 (1. April 2008): 40–49. <https://doi.org/10.5594/J15119>.

- Huang, Kai-Chen, Po-Yu Chien, Cheng-An Chien, Hsiu-Cheng Chang, und Jiun-In Guo. „A 360-degree panoramic video system design“. In *Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test*, 1–4, 2014. <https://doi.org/10.1109/VLSI-DAT.2014.6834863>.
- „immersion - Cambridge Dictionary“. Zugegriffen 14. März 2022. <https://dictionary.cambridge.org/de/worterbuch/englisch/immersion>.
- „immersion noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com“. Zugegriffen 14. März 2022. <https://www.oxfordlearnersdictionaries.com/definition/english/immersion>.
- „Laufzeitdifferenz“. In Wikipedia, 2. April 2022. <https://de.wikipedia.org/w/index.php?title=Laufzeitdifferenz&oldid=221714485>.
- Levack Drever, John. „Soundscape Composition: The Convergence of Ethnography and Acousmatic Music“. *Organised Sound* 7, Nr. 1 (April 2002): 21–27. <https://doi.org/10.1017/S1355771802001048>.
- Olivieri, Ferdinando, Nils Peters, und Deep Sen. „A Technology Overview and Application to Next-Generation Audio, VR and 360° Video“, o. J., 28.
- „Pegeldifferenz“. In Wikipedia, 15. April 2018. <https://de.wikipedia.org/w/index.php?title=Pegeldifferenz&oldid=176549228>.
- Pijanowski, Bryan C., Luis J. Villanueva-Rivera, Sarah L. Dumyahn, Almo Farina, Bernie L. Krause, Brian M. Napoletano, Stuart H. Gage, und Nadia Pieretti. „Soundscape Ecology: The Science of Sound in the Landscape“. *BioScience* 61, Nr. 3 (März 2011): 203–16. <https://doi.org/10.1525/bio.2011.61.3.6>.
- Raffaseder, Hannes. *Audiodesign: mit 101 Bildern, 32 Tabellen und einer CD-ROM. Medientechnik*. München: Wien u.a., 2002.
- Roginska, Agnieszka, und Paul Geluso, Hrsg. *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. New York: Routledge, 2017. <https://doi.org/10.4324/9781315707525>.
- Romanov, Michael, Paul Berghold, Daniel Rudrich, Markus Zaunschirm, Matthias Frank, und Franz Zotter. *Implementation and Evaluation of a Low-cost Head-tracker for Binaural Synthesis*, 2017.
- Rudrich, Daniel. „IEM Plug-in Suite“. Zugegriffen 23. März 2022. <https://plugins.iem.at/>.
- „SteamVR Plugin | Integration | Unity Asset Store“. Zugegriffen 12. Mai 2022. <https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647>.
- Truax, Berry. „Genres and Techniques of Soundscape Composition as Developed at Simon Fraser University“. *Organised Sound*, 7(1), 5–14 Auflage. Zugegriffen 6. April 2022. <https://www.sfu.ca/~truax/OS5.html>.
- Truax, Berry. „Simon Fraser University - Soundscape Composition“. Zugegriffen 6. April 2022. <https://www.sfu.ca/~truax/scomp.html>.
- Valem. *360 VIDEO PLAYER (with VR mode) - Unity Tutorial*, 2018. <https://www.youtube.com/watch?v=RxlQnPcOoYc>.

- Verkehrslärmkataster 2011“. Stadtportal der Landeshauptstadt Graz. Zugegriffen 13. Mai 2022. [https://www.graz.at/cms/beitrag/10295955/8115447/Online\\_Karte\\_Verkehrslaermkataster.html](https://www.graz.at/cms/beitrag/10295955/8115447/Online_Karte_Verkehrslaermkataster.html).
- WIBIS-Steiermark. „Einwohner Graz, Steiermark“, 18. August 2017. <https://wibis-steiermark.at/bevoelkerung/struktur/einwohner-gesamt/>, 1:<https://wibis-steiermark.at/bevoelkerung/struktur/einwohner-gesamt/wibis/einwohner/steiermark-13-bezirke/zeitreihe/>.
- Zotter, Franz, und Matthias Frank. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Bd. 19. Springer Topics in Signal Processing. Cham: Springer International Publishing, 2019. <https://doi.org/10.1007/978-3-030-17207-7>.
- ZYLIA. „ZYLIA ZM-1 Microphone“. ZYLIA - 3D AUDIO RECORDING & POST-PROCESSING. Zugegriffen 11. April 2022. <https://www.zylia.co/zylia-zm-1-microphone.html>.