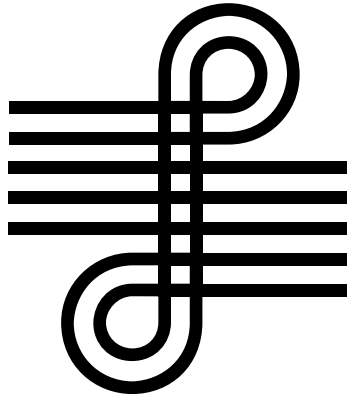


Hochschule für Musik Karlsruhe

IMWI - Institute für Musikinformatik und Musikwissenschaft



Bachelorarbeit zum Thema:

Interaktion mit Audio-Daten in VR mithilfe von Unity und der htc Vive

Zur Erlangung des Grades Bachelor of Arts

Vorgelegt von:

Manuel-Philippe Hergenröder, 12085

mail@manuelhergenroeder.de

Betreuender Dozent:

Prof. Dr. Damon T. Lee

Studiengang:

Musikinformatik (Hauptfach) / Musikwissenschaften (Nebenfach)

Abgabe:

xx.07.2020

1	Inhaltsverzeichnis	
2	Einleitung	3
3	Grundlagen	3
3.1	Virtual Reality	3
3.1.1	Definition, Historie und Abgrenzung	3
3.1.2	Anforderungen an VR-Software und -Hardware	4
3.2	Diskrete Fourier-Transformation	7
4	Vorstellung VrAudioSandbox	7
4.1	Architektur und externe Bibliotheken	8
4.1.1	Unity als Laufzeitumgebung	9
4.1.2	SteamVR	9
4.1.3	NAudio Bibliothek	9
4.1.4	DFT mit FFTW3	9
4.1.5	FFTWSharp	10
4.2	Implementation	10
4.2.1	FFT und Audio Engine	10
4.2.2	Visualisierung der Spektrum-Daten	10
4.2.3	User Interface und Steuerung	10
5	Kritische Betrachtung der Ergebnisse	10
6	Fazit	11
7	Literatur	12
8	Sonstige Quellen	13
9	Abbildungsverzeichnis	14
10	Eidesstattliche Erklärung	15
11	Anhang - Programmcode	16

2 Einleitung

Virtual Reality (VR) bietet das Potential unserem Geist direkten und erlebbaren Zugriff auf digitale Daten – losgelöst von der Umgebung in der wir uns aufhalten – zu geben. Dabei

In dieser Arbeit sollen die Möglichkeiten der Interaktion mit Audio-Daten im virtuellen Raum am Beispiel der Darstellung und Manipulation von FFT-Audio-Daten mithilfe der HTC Vive und den dazugehörigen Motion-Tracking-Controllern untersucht werden. Gegenstand der Arbeit ist die praktische Implementation dieses Konzepts und **<Herausforderungen, Bewertung, Ausblick>**.

3 Grundlagen

3.1 Virtual Reality

3.1.1 Definition, Historie und Abgrenzung

Der Begriff *Virtual Reality* impliziert, dass eine virtuelle Realität mithilfe einer Simulation geschaffen wird, welche für den Nutzer als real wahrgenommen wird. Bereits 1965 beschrieb der Computergrafikpionier Ivan E. Sutherland mit dem „Ultimate Display“ einen Raum, in dem die Materie durch den Computer gesteuert wird. Auf einen Stuhl in diesem Raum könne man sich setzen, eine Pistolenkugel wäre tödlich.¹

Dieses Ziel der kompletten Immersion konnte bisher nicht umgesetzt werden. Ein weiterer Aspekt ist aber auch nicht nur der Versuch eine komplett reale und glaubwürdige Umgebung zu schaffen, sondern

Gerade im Bereich der Hardware spielte die Entwicklung des sogenannten *Head-Mounted-Display* (engl. für „Am-Kopf-befestigter-Bildschirm“) – oder kurz *HMD* – eine wichtige Rolle. Ein *HMD* besteht aus einem oder mehreren Displays und stellt am Kopf befestigt eine Schnittstelle zwischen der visuellen Wahrnehmung des Menschen und der Grafikausgabe eines Computers her. Ivan E. Sutherland hat 1968 ein solches System mit dem Namen „The Sword of Damocles“ mithilfe seiner Studenten realisiert.² Dies legte die Grundlagen zum einen für die stereoskopische Darstellung des im Computer erzeugten Raums in Vektorgrafik, zum anderen für die notwendige schnelle Synchronisierung der angezeigten Bilder mit den Kopfbewegungen

¹ Vgl. Sutherland, Ivan: „The Ultimate Display“ – Konferenzband: Information Processing 1965: proceedings of IFIP Congress / Wayne A. Kalenich [Hrsg.]. International Federation for Information Processing, Amsterdam u.a., Washington u.a. (1965), S. 508 – <http://worrydream.com/refs/Sutherland%20-%20The%20Ultimate%20Display.pdf>, letzter Abruf: 09.06.2020

² Vgl. Sutherland, Ivan: „A head-mounted three dimensional display“ – AFIPS '68 (Fall, part I): Proceedings of the December 9-11, 1968, fall joint computer conference, part I, December 1968 Pages 757–764, S. 757 – <http://www.medien.fh-lmu.de/lehre/ss09/ar/p757-sutherland.pdf>, letzter Abruf: 10.06.2020

des Benutzers. Ein Sensor mit mechanischem Arm hat zunächst die Kopfposition ermittelt. Später wurde dies für bessere Bewegungsfreiheit mithilfe von Ultraschall-Emittern und -Empfängern umgesetzt.³ Dieses sogenannte *Head-Tracking* und die zeitnahe Abstimmung der angezeigten Bilder an Kopfposition und -bewegungen ist auch heute noch ein wichtiger Aspekt bei *HMDs* um Übelkeit (sogenannte *Motion-Sickness*), die in unserem Gehirn entsteht, wenn widersprüchliche Reize – in diesem Fall die visuellen Informationen unseres Sehapparates und die körperliche Wahrnehmung bezüglich der Bewegung und Stellung des Kopfes verarbeitet werden – beim Benutzer vorzubeugen.

Die Entwicklung von Technologien für VR reicht weit zurück. <NASA Datenhandschuh, VR Hype 90er, 2. Welle Oculus Rift/Vive, Nintendo Virtual Boy, Gaming Industry, Professionelle Anwendungen Architektur, CAD, Ergonomie>

Ein mit der *Virtual Reality* verwandtes Feld ist die sogenannte *Augmented Reality* bei der die echte Realität mit zusätzlichen Informationen angereichert wird. Im Gegensatz zur VR werden reales Umfeld und virtuelle Bestandteile vom Computer vermischt – z.B. durch ein transparentes Display oder mithilfe eines eingeblendeten Live-Kamera-Bildes. Bei einer starken Interaktion zwischen Bestandteilen realer und virtueller Welt wird zudem auch der Begriff *Mixed-Reality* verwendet.

3.1.2 Anforderungen an VR-Software und -Hardware

Gängige etablierte Bedienparadigma, die auf der der Computer- und Medientechnik basieren, welche sich über die letzten Jahrzehnte entwickelt und etabliert haben, sind für VR-Anwendungen nicht in Gänze übertragbar. Insbesondere die Gestaltung in Form einer zweidimensionalen Bedienoberfläche (*User Interface*), welche auf einem zweidimensionalen Display dargestellt wird und mit Maus, Tastatur oder via Touch bedient wird, lässt sich nur mit Einschränkungen bzgl. der Ergonomie und Nutzerfreundlichkeit bei Virtual Reality anwenden. Verschachtelte Menüstrukturen, die viel Feinmotorik zur Interaktion benötigen, sind nicht für die Bedienung mit *Motion-Controllern* ausgelegt.

Ein weiteres Beispiel sind in vielen Computerspielen – aber auch in Software fernab des Unterhaltungssektors – vorhandene *HUD-UI-Elemente*. *HUD* steht für *Head-Up-Display* und beschreibt das Platzieren von informativen Elementen oder Elementen zur Steuerung im sogenannten *Screen-Space* – d.h. überlagernd ohne Bezug zur dreidimensional dargestellten Szene. Durch die stereoskopische Darstellung in VR kann dieser fehlende Bezug zur Geometrie

³ Vgl. Ebd. S. 760-761

des virtuellen Raumes irritierend wirken. Auch dass der Benutzer sich im Raum orientieren muss und seine Blickrichtung dabei frei wählbar ist, kann dazu führen, dass UI-Elemente im Sichtbereich stören können. Das VR-Computerspiel „Half-Life Alyx“ der Firma Valve zeigt den Ansatz traditionelle HUD-Elemente durch virtuell-physische Elemente zu ersetzen.



Abbildung 1 – Vergleich HUD-basierte Darstellung (oben) von Gesundheit und Munition mit im virtuellen Raum als physische Objekte manifestierte Darstellung (unten)

Einzelquellen (von oben n. unten): Steam Workshop Half-Life: Alyx [HUD], Screenshot von „Alex“ – <https://steamuserimages-a.akamaihd.net/ugc/1016065725020200218/B47282DABAF0C22E8A6C5770FE46BCA8FCFFBF718/>; IGDB Half-Life: Alyx Press kit – https://images.igdb.com/igdb/image/upload/t_original/sc7dad.png; Screenshot von SuperQGS, reddit – <https://i.redd.it/yojg6f875rg41.jpg> – alle Screenshots © 2019 Valve Corporation, letzter Abruf 17.06.2020

Gesundheitszustand des Spielers und Munitionsanzeige sind Teil der Spielwelt und können durch entsprechenden Blick des Spielers bei Bedarf erfasst werden ohne zu stören (siehe Abbildung 1 unten, oben die traditionell genutzte Methode durch HUD-Elemente).

Dadurch, dass der Benutzer seine Blickrichtung jederzeit frei wählen kann bzw. muss, ergeben sich Anforderungen an die Steuerung der Aufmerksamkeit durch das Design der VR-Applikation. Das Cone-Of-Focus-Modell (siehe Abbildung 2) teilt dabei das Umfeld des Spielers in verschiedene Zonen ein. Alle wichtigen Ereignisse sollten sich im Zentrum der Aufmerksamkeit und Blickrichtung des Benutzers abspielen. Sollte es gewünscht sein, dass der Nutzer seine Aufmerksamkeit auf eine andere Zone lenkt, bieten sich visuelle Anhaltspunkte über die primären und sekundären peripheren Sichtbereich an, um den Nutzer zur Änderung seiner Blickrichtung zu motivieren.⁴

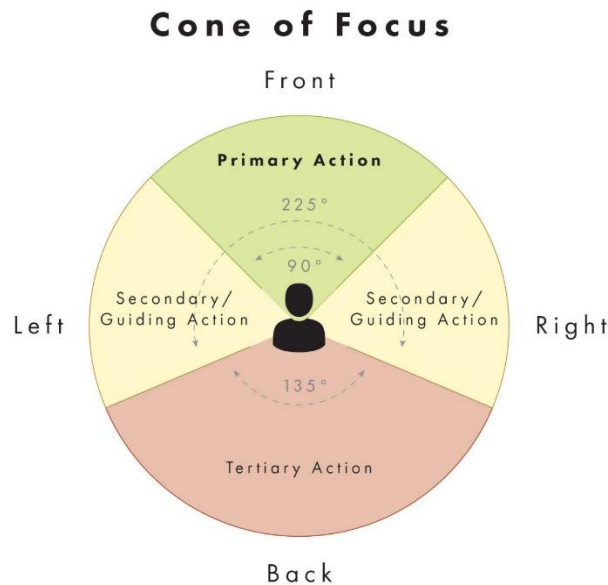


Abbildung 2 – Cone of Focus, UploadVR Copyright 2019 UVR Media LLC –
<https://mk0uploadvrcom4bcwhj.kinstacdn.com/wp-content/uploads/2016/07/cone-of-focus.jpg>,
 letzter Abruf 17.06.2020

Ein weiterer Aspekt – bei dem insbesondere das Zusammenspiel zwischen Software und Hardware wichtig ist – ist die Latenz zwischen Bewegung (beispielsweise des Kopfes) und der entsprechenden Berechnung des angezeigten Bildes, welche die ausgeführte Bewegung berücksichtigt. Wichtig ist dabei Latenz des gesamten Systems bestehend aus Motion-Tracking, Verarbeitung im Computer (Applikation, Rendering) und die Ausgabe auf dem HMD. Hier spielen auch Faktoren wie die Frequenz des Trackings, Bildwiederholrate des HMD und die Verwendung von Framebuffern eine Rolle. Diese Latenz muss möglichst gering ausfallen, da ansonsten das Präsenz-Gefühl des VR-Nutzers leidet und es schnell zu Übelkeit bei der Nutzung

⁴ Vgl. Jerald, Jason (Ph.d.): „The VR Book: Human-Centered Design for Virtual Reality“, A publication in the ACM Book series #8, Association for Computing Machinery and Morgan & Claypool Publishers, San Rafael (California) (2016), S. 254

kommen kann.⁵ Experimente am NASA Ames Research Center⁶ legen nahe, dass einerseits Probanden unterschiedlich sensibel auf verschiedene Latenzen reagieren und teilweise bei einem VR-System mit 7.4 ms Latenz noch Unterschiede im Bereich von 3.2 ms wahrgenommen werden konnten. Gängige HMDs (Stand 2015) liegen laut Messungen von Kjetil Raaen und Ivar Kjellmo bei einer Latenz von etwa 40 ms.⁷ Auch wenn die Fähigkeiten der eingesetzten Hardware vom Softwareentwickler nur bedingt beeinflusst werden können, sollte die Software auf stabile und hohe Frameraten hin optimiert werden, um Irritationen und Übelkeit zu vermeiden.

Immersion -> Warning VR Space <_> real space

- Motion sickness, Latenz
- Refresh rate
- Bedienparadigma (2D UI vs intuitiver Umgang mit Objekten in VR)

3.2 Diskrete Fourier-Transformation

Die diskrete Fourier-Transformation ist ein wichtiger Algorithmus in der digitalen Signalverarbeitung und überführt Signal-Daten aus dem Zeitbereich in den Frequenzbereich. **Der Frequenzbereich bietet Möglichkeiten Eigenschaften des Signals abzuleiten oder das Signal zu bearbeiten, welche im Zeitbereich nur schwer oder gar nicht möglich sind.** Fast-Fourier-Transform (kurz. FFT) beschreibt dabei eine Untermenge von Implementationen der DFT, die besonders effizient und schnell zu berechnen sind.

4 Vorstellung VrAudioSandbox

Im Rahmen dieser Arbeit wurde die Möglichkeit der Darstellung und Manipulation von Audio-Daten innerhalb der Virtual Reality praktisch implementiert. Dazu wurde neben den technischen Voraussetzungen ein Bedienkonzept zugeschnitten auf die Motion Tracking Controller der htc Vive entwickelt. Der Benutzer kann beliebiges vorhandenes Audiomaterial importieren, mithilfe des htc Vive HMD und der dazugehörigen Motion-Controller modifizieren und das Ergebnis

⁵ Vgl. Ebd. S. 183-184

⁶ Ebd. Vgl. S. 184-185 –Adelstein et al. 2003 & 2006, Jerald 2009, Ellis et al. 1999, 2004. Mania et al. 2004

⁷ Vgl. Raaen, Kjetil; Kjellmo, Ivar: „Measuring Latency in Virtual Reality Systems“, 14th International Conference on Entertainment Computing (ICEC), Sep 2015, Trondheim, Norway. pp.457-462, 10.1007/978-3-319-24589-8_40. hal-01758473

Anmerkung: Stand 2015, zwischen 35-45 ms wurden bei der Oculus Rift DK2 gemessen bei aktivierten V-Sync – d.h. Synchronisierung der Bildwiederholrate mit der Grafikkarte, um Tearing-Effekte zu vermeiden. Dazu wurde eine Unity Szene und eine lichtempfindliche Photozelle verwendet. Latenz durch Eingabegeräte ist nicht berücksichtigt.

probezuhören oder als Audio-Datei zu exportieren. Die Bedienung findet dabei vollständig in VR statt.

4.1 Architektur und externe Bibliotheken

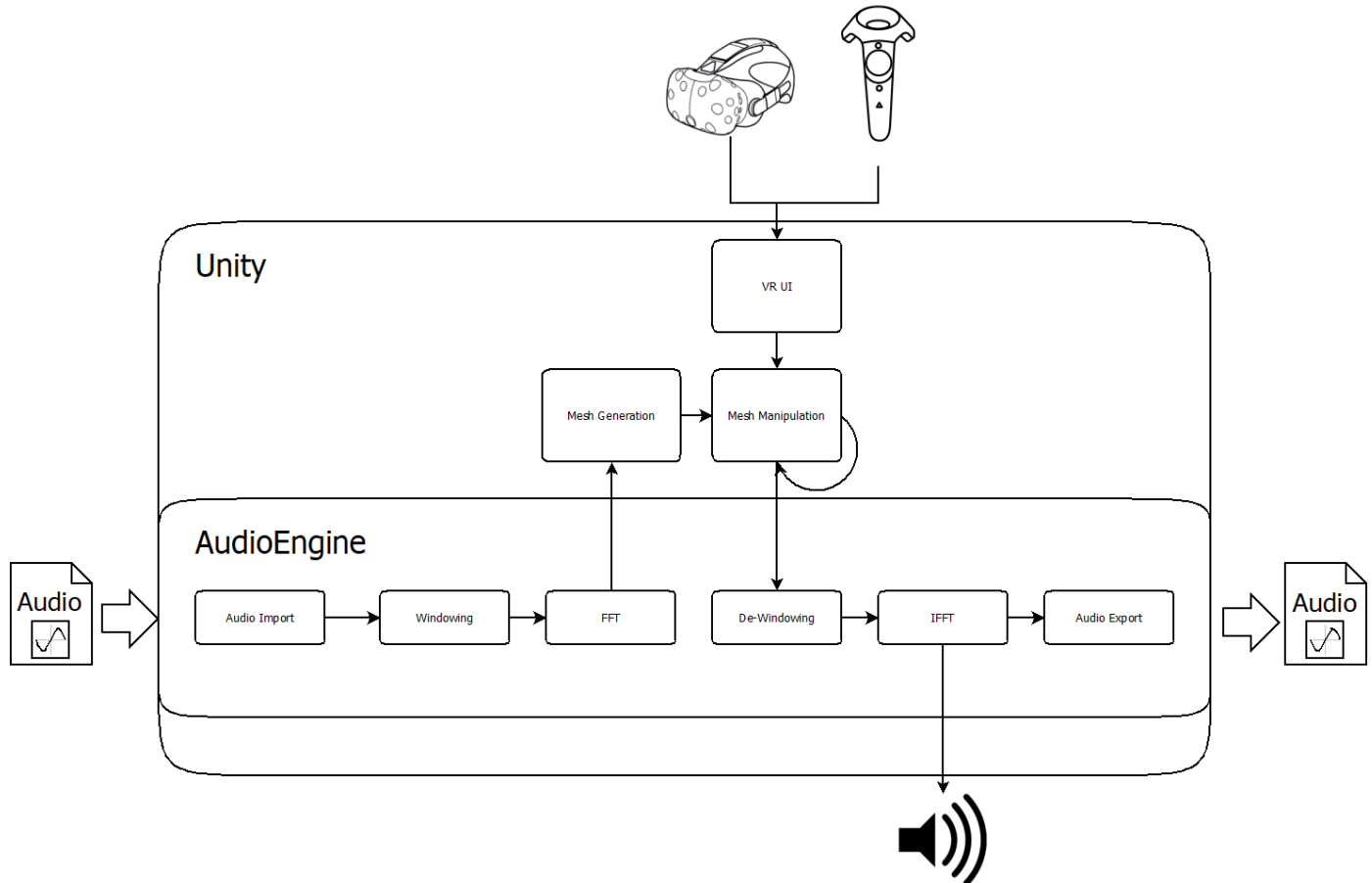


Abbildung 3 – Eine abstrakte Darstellung der Architektur - Eigene Darstellung

Das Herzstück für das Rendering und die Einbindung von SteamVR bildet dabei die Laufzeitumgebung Unity.

Im Folgenden werden die genutzten externen Bibliotheken vorgestellt und näher erläutert.

4.1.1 Unity als Laufzeitumgebung

Die Laufzeit- und Entwicklungsumgebung Unity ist eine weitverbreitete Spiele-Engine mit etwa 50% Marktanteil für den Bereich PC/Konsole/Mobile Games.⁸ Das dazugehörige Unternehmen „Unity Technologies“ (ehemals „Over the edge“) wurde 2004 von David Helgason, Nicholas Francis und Joachim Ante gegründet. Auch außerhalb der kommerziellen Spiele-Entwicklung ist Unity sehr beliebt – u.a. auch da es eine für nicht-kommerzielle Zwecke kostenlose „Personal“-Lizenz gibt, welche den Hauptteil des Funktionsumfangs von Unity bereitstellt.⁹

Unity bildet das Herzstück für das Rendering, <Erklärung Zusammenhang C#, Unity, etc>

4.1.2 SteamVR

4.1.3 NAudio Bibliothek

NAudio ist eine von Mark Heath entwickelte Open-Source-Audio-API für .NET und ist in C# geschrieben.

4.1.4 DFT mit FFTW3

FFTW (Akronym für „The Fastest Fourier Transform in the West“) ist eine freie FFT-Bibliothek zur Berechnung der diskreten Fourier-Transformation und wurde von Matteo Frigo und Steven G. Johnson am Massachusetts Institute of Technology entwickelt.¹⁰ Die in C und OCaml geschriebene Bibliothek ist im Quelltext verfügbar, sehr portabel und unterstützt daher viele Plattformen. Außerdem bietet sie laut den Autoren im Vergleich zu anderen Implementationen eine sehr gute Performance.¹¹

FFTW bietet eine Vielzahl an Algorithmen und Funktionen an. Für die DFT und inverse DFT in VrAudioSandbox wird auf die „1d Discrete Fourier Transform (DFT)“ aus der FFTW3-Bibliothek mit Double-Precision zurückgegriffen, welche durch folgende mathematische Summenformel beschrieben wird:¹²

⁸ Vgl. Unity Technologies: „Sie fragen sich was Unity ist? Entdecken Sie, wer wir sind, wo wir angefangen haben und wohin wir uns entwickeln | Unity“ – <https://unity.com/de/our-company>, letzter Abruf: 11.06.2020

⁹ Vgl. Unity Technologies: „Powerful 2D, 3D, VR, & AR software for cross-platform development of games and mobile apps.“ – <https://store.unity.com/#plans-individual>, letzter Abruf: 11.06.2020

¹⁰ Vgl. Frigo, Matteo; Johnson, Steven G.: „The Design and Implementation of FFTW3“ – Proceedings of the IEEE, Volume 93, Number 2, 2005, S. 216-231, S. 231 – <http://www.fftw.org/fftw-paper-ieee.pdf>, letzter Abruf: 13.06.2020

¹¹ Vgl. FFTW: „FFT Benchmark Results“ – <http://www.fftw.org/speed/>, letzter Abruf: 13.06.2020

¹² Vgl. FFTW: 3.3.8 Manual, Kapitel 4.8.1 „The 1d Discrete Fourier Transform (DFT)“ – http://www.fftw.org/fftw3_doc/The-1d-Discrete-Fourier-Transform-_0028DFT_0029.html, letzter Abruf: 13.06.2020

$$Y_k = \sum_{j=0}^{n-1} X_j e^{-2\pi j k \sqrt{-1/n}}$$

Die Eingabe X ist dabei ein ein-dimensionales Array komplexer Zahlen der Größe n und Y das Ausgabe-Array, wobei das k -te Element der Frequenz k/n entspricht.

Bei der inversen Funktion handelt es sich um die gleiche Funktion ohne negatives Vorzeichen im Exponenten der DFT:

$$Y_k = \sum_{j=0}^{n-1} X_j e^{2\pi j k \sqrt{-1/n}}$$

FFTW berechnet eine unnormalisierte Transformation ohne Koeffizienten vor der Summe der DFT, d.h. eine vorwärtsgerichtete DFT gefolgt von der rückwärtsgerichteten DFT ergibt als Ergebnis die Eingabe multipliziert mit n .

4.1.5 FFTWSharp

4.2 Implementation

4.2.1 FFT und Audio Engine

- Implementation, Windowing, Phase Amplitude, etc.
- Codebeispiele

4.2.2 Visualisierung der Spektrum-Daten

- Vertices, Triangles, Mesh
- Codebeispiele

4.2.3 User Interface und Steuerung

- Motion-Controller, Actions
- Konzept: Interaktion ohne verschachtelte 2D-Menüs, Werkzeugfarben, Ablesen im virtuellen Raum

5 Kritische Betrachtung der Ergebnisse

- Verbesserungsvorschläge
- Ideen für weitere Werkzeuge / Funktionen

6 Fazit

7 Literatur

Frigo, Matteo; Johnson, Steven G.: “The Design and Implementation of FFTW3” – Proceedings of the IEEE, Volume 93, Number 2, 2005, S. 216-231 – <http://www.fftw.org/fftw-paper-ieee.pdf>, letzter Abruf: 13.06.2020

Jerald, Jason (Ph.d.): „The VR Book: Human-Centered Design for Virtual Reality”, A publication in the ACM Book series #8, Association for Computing Machinery and Morgan & Claypool Publishers, San Rafael (California) (2016), ISBN 978-1-97000-112-9

Sutherland, Ivan: „A head-mounted three dimensional display” – AFIPS '68 (Fall, part I): Proceedings of the December 9-11, 1968, fall joint computer conference, part I, December 1968 Pages 757–764, S. 757 – <http://www.medien.ifl.lmu.de/lehre/ss09/ar/p757-sutherland.pdf>, letzter Abruf: 10.06.2020

Sutherland, Ivan: „The Ultimate Display” – Konferenzband: Information Processing 1965: proceedings of IFIP Congress / Wayne A. Kalenich [Hrsg.]. International Federation for Information Processing, Amsterdam u.a., Washington u.a. (1965)

8 Sonstige Quellen

FFTW: „FFT Benchmark Results“ – <http://www.fftw.org/speed/>, letzter Abruf: 13.06.2020

FFTW: 3.3.8 Manual, Kapitel 4.8.1 „The 1d Discrete Fourier Transform (DFT)“ – http://www.fftw.org/fftw3_doc/The-1d-Discrete-Fourier-Transform-_0028DFT_0029.html, letzter Abruf: 13.06.2020

Unity Technologies: „Powerful 2D, 3D, VR, & AR software for cross-platform development of games and mobile apps.“ – <https://store.unity.com/#plans-individual>, letzter Abruf: 11.06.2020

Unity Technologies: „Sie fragen sich was Unity ist? Entdecken Sie, wer wir sind, wo wir angefangen haben und wohin wir uns entwickeln | Unity“ – <https://unity.com/de/our-company>, letzter Abruf: 11.06.2020

9 Abbildungsverzeichnis

Abbildung 1 – Vergleich HUD basierte Darstellung (oben) von Gesundheit und Munition mit im virtuellen Raum als physisches Objekt manifestierter Darstellung (unten).....	5
Abbildung 2 – Cone-of-Focus-Modell, UploadVR Copyright 2019 UVR Media LLC – https://mk0uploadvrcom4bcwhj.kinstacdn.com/wp-content/uploads/2016/07/cone-of-focus.jpg , letzter Abruf 17.06.2020.....	6
Abbildung 3 – Eine abstrakte Darstellung der Architektur - Eigene Darstellung	8

10 Eidesstattliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist, insbesondere dass ich alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, durch Zitate als solche gekennzeichnet habe.

Ich versichere auch, dass die von mir eingereichte schriftliche Version mit der digitalen Version übereinstimmt. Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde/Prüfungsstelle vorgelegen hat. Ich erkläre mich damit nicht einverstanden, dass die Arbeit der Öffentlichkeit zugänglich gemacht wird.

Ich erkläre mich damit einverstanden, dass die Digitalversion dieser Arbeit zwecks Plagiatsprüfung auf die Server externer Anbieter hochgeladen werden darf. Die Plagiatsprüfung stellt keine Zurverfügungstellung für die Öffentlichkeit dar.

Oberboihingen,
der xx.xx.2020

Manuel-Philippe
Hergenröder



Ort, Datum

Vorname Nachname

Unterschrift

11 Anhang - Programmcode