**Visualisierung von Audio-Daten in VR**

Manuel-Philippe Hergenröder (12085)

Vertiefende Studienarbeit, WS 2019

Prof. Dr. Damon T. Lee

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ehrenwörtliche 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. | | | | |
|  | | | | |
| Karlsruhe, der xx.02.2010 |  | Manuel-Philippe Hergenröder |  |  |
| Ort, Datum |  | Vorname Nachname |  | Unterschrift |

[1 Einleitung 4](#_Toc31714892)

[2 Konzeption 4](#_Toc31714893)

[3 Entwicklung der Visualisierung von Audio-Daten in VR 4](#_Toc31714894)

[3.1 Hardwarevoraussetzungen 4](#_Toc31714895)

[3.2 Softwarearchitektur 5](#_Toc31714896)

[3.3 Implementation und Fallstricke 6](#_Toc31714897)

[4 Ausblick und Möglichkeiten 6](#_Toc31714898)

[5 Fazit 6](#_Toc31714899)

[6 Literatur 7](#_Toc31714900)

# 1 Einleitung

Schon lange ist die Visualisierung von Informationen eine wichtige Methode der Vermittlung von Informationen. Dabei besitzt der Sehapparat die größte Bandbreite zum Gehirn im Vergleich zu den anderen Sinnesorganen. Im Gegensatz zum Beispiel zur akustischen Verarbeitung können eine große Menge an Informationen bzw. Daten „auf einem Blick“ verarbeitet werden. Insbesondere auch in Bezug auf den zeitlichen Verlauf, wohingegen beim Hören immer nur der gegenwärtige Moment bewusst erfasst werden kann. Das heißt die Anreicherung von gehörtem Ton mit visuellen Informationen, kann unser Verständnis über die Struktur des Klangmaterials sinnvoll ergänzen.

# 2 Konzeption

Die angestrebte Softwarelösung soll ausgehend von Offline-Audio-Daten, die über eine Importfunktion aus einer Audiodatei heraus in einen Buffer geladen werden können, eine in VR „begehbare“ Visualisierung ermöglichen. Dafür soll eine frequenzbasierte Darstellung verwendet werden, um Eigenschaften der Audio-Daten in visueller Form zu offenbaren, die evtl. beim bloßen Hören verborgen bleiben. Durch die Nutzung von Kopfhörern und eines Head-Mounted-Display (HMD) soll eine Immersion hergestellt werden, welche aufgrund der Kombination von Hörsinn und Sehsinn das Erfassen und Erleben von Klanginformationen intensiviert. Die physische Bewegung in Form einer „VR-Room-Scale-Experience“ dient einer intuitiven Navigation durch den virtuellen Raum und wirkt außerdem dem Problem der „motion sickness“ entgegen.

# 3 Entwicklung der Visualisierung von Audio-Daten in VR

## 3.1 Hardwarevoraussetzungen

Für die Visualisierung in VR ist zum einen das HMD HTC VIVE erforderlich und zum anderen ein leistungsfähiger PC mit dedizierter 3D-Grafikkarte. Die von HTC VIVE angegebenen Mindestanforderungen[[1]](#footnote-1) setzen als CPU „Intel Core i5-4590/AMD FX 8350 oder besser“, als GPU eine „NVIDIA GeForce GTX 1060, AMD Radeon RX 480 oder besser“, „4 GB RAM oder mehr“ Arbeitsspeicher, „HDMI 1.4, DisplayPort 1.2 oder neuer“, „1x USB 2,0 oder höher“ und als Betriebssystem „Windows 7 SP1, Windows 8.1 oder später, Windows 10“ voraus. Dies ist insbesondere wichtig, da zu schwache Hardware für Bildaussetzer sorgt, was im Zusammenhang mit dem Head-Tracking zu Übelkeit und Kopfschmerzen führen kann. Im Idealfall sollten immer mind. 90 Bilder pro Sekunde ohne sogenannte „Frame-Drops“ gewährleistet sein.

Ohne HMD lässt sich die Visualisierung auch auf einem 2D-Monitor mit Tastaturnavigation mit verringerter Immersion verwenden. Die Hardwareanforderungen fallen dabei geringer aus.

Für die schnelle Berechnung der Fourier-Transformation ist generell eine schnelle CPU mit guter Multi-Threading-Performance ratsam, um Wartezeiten zu verkürzen.

Bei dem System, welches zur Entwicklung verwendet wurde, handelt es sich um einen PC mit Intel i9-9900K mit 32GB RAM und einer NVIDIA Geforce RTX 2070, welcher sich als adäquat herausgestellt hat.

## 3.2 Softwarearchitektur

Als Grundlage für die Visualisierung wurde die Laufzeit- und Entwicklungsumgebung Unity gewählt. Diese Spiele-Engine ist für nicht-kommerzielle Entwicklungen kostenlos nutzbar und bietet ein weitreichendes Ökosystem aus Assets, Erweiterungen und Anbindungen an externe Bibliotheken. Außerdem lässt sich das Verhalten durch selbstgeschriebene Skripte (u.a. in der .NET-Sprache C#) beliebig anpassen. Über das SteamVR Plugin[[2]](#footnote-2) aus dem Unity Asset Store lässt sich das Head Mounted Display HTC VIVE, für das dieses Projekt konzipiert ist, ansteuern.

Für den Import von Audio-Daten wird die Open-Source-Bibliothek NAudio[[3]](#footnote-3) für .NET (entwickelt von Mark Heath) verwendet.

Für das Zerlegen der Audio-Daten in seine Frequenzanteile wurde die C-Bibliothek FFTW[[4]](#footnote-4) gewählt. Ausschlaggebend dafür war die im Vergleich zu anderen Bibliotheken schnelle Performance, das freie Lizenzmodell, welches mit einer umfangreichen Dokumentation hergeht, sowie die Tatsache, dass FFTW viele verschiedene Algorithmen implementiert (u.a. verschiedene Varianten der diskreten und schnellen Fourier-Transformation – auch in umgekehrter Richtung).

Für die verteilte Versionsverwaltung wird Git[[5]](#footnote-5) auf einer selbst-betriebenen Gitlab-Installation, sowie Github verwendet.

## 3.3 Implementation und Fallstricke

Zunächst wurde SuperCollider[[6]](#footnote-6) als Audio-Engine verwendet, was sich insbesondere für den Datenaustausch mit Unity als problematisch herausgestellt hat. Über die OpenSoundControl[[7]](#footnote-7)-Spezifikation ließ sich der Datenaustausch implementieren. Da das Protokoll allerdings eher für das Bereitstellen von Steuerdaten und nicht für den Austausch großer Datenmengen konzipiert ist, musste ein eigener Handler zum Aufteilen der Audio-Daten in mehrere OSC-Messages implementiert werden. Problematisch erwies sich dabei vor allem der Verlust von einzelnen OSC-Messages, da OSC UDP-Pakete verwendet. Unter anderem musste eine Pause zwischen den OSC-Messages eingehalten werden, weil sonst – vermutlich durch Begrenzungen von Puffergrößen der OSC-Bibliothek und/oder des TCP/IP-Stacks – ein Großteil der Pakete nicht beim Empfänger (Unity) ankam. Außerdem wurden die Pakete durchnummeriert – ähnlich zu „sequence numbers“ beim Transmission Control Protocol –, um einerseits die korrekte Reihenfolge der Daten zu gewährleisten und andererseits verlorengegangene Pakete identifizieren zu können. Insgesamt betrachtet ist die Methode OSC zu nutzen als langsam, fehleranfällig und aufwendig bzgl. der Implementation zu betrachten.

Ein anderes Problem – bezogen auf die Verwendung von SuperCollider – ist, dass es sich als komplex herausgestellt hat, auf die FFT-Rohdaten innerhalb von sclang zugreifen zu können, um sie dann via OSC zu Unity weiterzuleiten. Die eigentlichen FFT-Daten liegen im Speicherbereich von scserver, der wiederum mit sclang via OSC kommuniziert. Ein Austausch über das Dateisystem oder einen Shared-Memory-Bereich wäre prinzipiell denkbar gewesen. Aufgrund der spärlichen Dokumentation und des hohen Aufwands wurde schlussendlich davon abgesehen SuperCollider als Audio-Engine zu verwenden.

Nach einem kompletten Rewrite setzt das Projekt nun auf die NAudio-Bibliothek für den Import der Audiodaten und als Playback-Engine. Dabei werden Mono-Daten aller gängigen Sampleraten, sowie 8, 16, 24, 32 oder 64bit Bittiefe unterstützt. Die Audio-Daten werden zunächst aus einer WAV PCM Audio-Datei eingelesen (durch die Nutzung von NAudio werden auch andere Containerformate und Codecs wie AIFF, MP3, etc. unterstützt).

Danach wird die Fouriertransformation sequenziell auf einzelne Blöcke der Audio-Daten angewendet. Die Ergebnisse bilden die Grundlage zur Generierung eines Meshes. Es wurde mit verschiedenen Darstellungsweisen experimentiert, u.a. einer zylindrischen, die die Werte der einzelnen FFT-Bins radial in Zackenform anordnet, was sich in der Praxis als unergonomisch erwiesen hat, da dies zu vielen Kopfdrehungen und Bücken geführt hat. Letztendlich hat sich eine „Terraindarstellung“, also das Darstellen der Daten wie bei einer begehbaren Gebirgsformation, bewährt. Zusätzlich ist eine Teleportation möglich, um einerseits einen großen virtuellen Raum, der die Größe des physischen Raumes übersteigt, zu ermöglichen – und andererseits lange Laufwege zu vermeiden. Die Teleportation ist dabei auch einer der gängigen „best practices“, um „motion sickness“ zu vermeiden.

Das Mesh wird dabei iterativ in Unity zur Laufzeit generiert. Problematisch war zunächst das Rendern aller Daten innerhalb eines einzelnen Meshes. Unity nutzt standardmäßig nur 16bit für die Mesh-Indizes, so dass nur 65,536 vertices indexiert werden können. Dies lässt sich zwar mit dem Attribut .indexFormat auf 32bit umkonfigurieren, um alle Daten im Mesh unterzubringen – allerdings führte dies zu einer extrem schlechten Performanz bis hin zum Absturz von Unity. Erst durch das Erzeugen individueller Meshes pro FFT-Block konnten stabile Frameraten und genügend Stabilität erreicht werden. Vermutlich hängt dies mit einer stark verbesserten Parallelisierbarkeit der Render-Pipeline zusammen.

# 4 Ausblick und Möglichkeiten

Die Immersion und Interaktion, die durch den Aspekt der „Begehbarkeit“ der Audio-Daten entsteht, stellt eine

# 5 Fazit

# 6 Literatur

**Jerald, Jason,** NextGen Interactions. **–** *The VR Book – Human-Centered Design for Virtual Reality*, A publication in the ACM Books series, #8, First Edition, Association for Computing Machinery and Morgan & Claypool, 2016, San Rafael, ISBN 978-1-97000-112-9

1. VIVE - “Wie sind die Systemanforderungen?” <https://www.vive.com/de/support/vive/category_howto/what-are-the-system-requirements.html> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-1)
2. Unity Asset Store – SteamVR Plugin - <https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-2)
3. NAudio – Audio and MIDI library for .NET - <https://github.com/naudio/NAudio> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-3)
4. FFTW Home Page - <http://fftw.org/> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-4)
5. Git –fast-version control - <https://git-scm.com/> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-5)
6. SuperCollider - <https://supercollider.github.io/> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-6)
7. OpenSoundControl - <http://opensoundcontrol.org/> (letzter Abruf: 04.02.2020) [↑](#footnote-ref-7)