

Sounddesign für VR

Elisabeth Drobny
Tonseminar WS 19/20

Inhalt

Was ist VR?

Hardware

Lokalisation

Uncanny Valley

Raumwahrnehmung

Lautstärkeabfall

Source Material und Aufarbeitung

Echo & Hall

Mischung von 3D Szenen

Audio Middleware

Was ist VR?

3D Umgebung bei der durch Nutzung von Headset mit Brille und Kopfhörern, die Distanz zum Display vermindert wird und der Nutzer direkt in die Umgebung gesetzt wird. Durch Headtracking sind Bewegungen und Perspektivwechsel des Kopfes umgesetzt.

Vorteile für Sounddesign sind unter anderem, dass es sich um eine uniforme Abhöre handelt, solange Kopfhörer am Headset angebracht sind. Das bedeutet, es muss sich nicht auf Smartphones oder Laptop gut anhören, sondern kann für das spezifische Endgerät gemischt werden.

Wichtige Aspekte, die zu beachten sind, wenn man Ton für VR machen will:

- Visuelles und Auditives muss synchron sein, zeitlich und räumlich. Das bedeutet Drehungen müssen flüssig laufen, Schallquellen sollten sich nicht mitbewegen um Lokalisation zu gewährleisten
- Ansonsten kann es schnell zu Übelkeit kommen
- Isoliertes Hören hilft bei der Immersion, Eindrücke von außen bleiben abgeschirmt

Hardware

Wie bei jeder Art von Content-Produktion sollte man sich der Abhörsituation bewusst sein. Ist es ein öffentliches Event mit vielen Leuten, die sich in der Nähe aufhalten, empfehlen sich möglicherweise geschlossene Kopfhörer. Wird die Anwendung vorwiegend privat genutzt, sind möglicherweise nur günstigere Kopfhörer verfügbar. Die Nutzung von diskreten Lautsprechern in Kombination mit VR-Anwendungen ist schwieriger, da selbst mit 7.1 aufwärts die Lokalisationsschärfe nicht ausreichend bedient werden kann. Auch mit gut funktionierendem HeadTracking könnte man das Springen der Soundquellen zwischen den Lautsprechern hören. Die Vorteile von Kopfhörern sind unter anderem, dass sie generell uniformer sind als Lautsprecher. Auch weil sie oft mit einem VR-Headset kommen. Selbst wenn die bauliche Qualität schlecht ist, ist mit wenig Aufwand (Pegel, Laufzeit, Frequenz) 3D abbildbar. Wichtig ist aber immer, dass auch bei denselben Gegebenheiten hört jeder durch Training, Wohnort/Gewohnheit, Schädigung/Alter und Präferenzen anders.

Im Vergleich zu linearen Medien wie Filme und Serien, weisen Spiele oder andere 3D/VR Anwendungen andere Schwierigkeiten auf. In linearen Medien kann man sich darauf verlassen, dass der narrative Erzählstrang immer in Vollständigkeit wiedergegeben werden kann und die Emotionen und Spannungen aufgebaut und ausbalanciert werden können. Bei Anwendungen, bei denen der Nutzer verschiedene Möglichkeiten hat sich zu bewegen, ist es immer eine Frage der Aufmerksamkeit, ob er oder sie alles mitbekommt oder ob Lücken in der Geschichte auftreten.

Hier muss der Realismus und die komplett freie Entscheidungsmöglichkeit mit der Narrative ausbalanciert werden.

Lokalisation

Die Ortung von Schallquellen im Dreidimensionalen Raum funktioniert durch die Fähigkeit, die interaurale Differenzen auszuwerten. Dabei werden Laufzeit-, Pegel- und Frequenzunterscheide zwischen linkem und rechten Ohr verglichen. Beim Designen von Content kann man die Schwierigkeiten des menschlichen Hörens auch zum Vorteil nutzen: das Ausnutzen von „Front/Back-Confusion“ in Kombination mit dem Visuellen. Auf der Medianebene ist eine Triangulation zwischen den Ohren nicht möglich, dadurch kann nicht unterschieden werden ob ein Signal vorne oder hinten ist. Umso weiter über/unter dem Kopf sich eine Schallquelle (real oder virtuell) befindet, desto schlechter die Hörenauflösung. Da wir aber direkt davon ausgehen, dass etwas von hinten kommt, wenn wir es nicht sehen, kann mit Hinzufügen von Filtern ein sehr überzeugender Eindruck einer Quelle hinter dem/der NutzerIn hervorgerufen werden.

Uncanny Valley

Oft, wenn etwas so realistisch abgebildet ist, dass man nicht mehr ausmachen kann welche Teile real sind, aber man trotzdem wahrnimmt, dass etwas nicht stimmt, kann das ein unangenehmes Gefühl erzeugen. Eine mögliche Lösung kann es sein, die kritischeren Konzepte „Bigger then life“ anzusetzen.

Raumwahrnehmung

Ein "leerer Raum " hat (u.a) folgende Eigenschaften:

- Materialeigenschaften der Wände/Boden
- Deckenhöhe
- Entfernung zu den Wänden
- Objekte
- Kleidung/Haare
- Bewegung

Das Nachbilden eines Raumes mit allen Parametern, Dämpfung und Reflexion ist mit der Verwendung von Ray Tracing sehr rechenaufwändig. Deswegen empfiehlt es sich generell Methoden wie Raycasting nur für bestimmte Effekte zu verwenden (wenn z.B. ein wichtiges Element im Weg steht und nicht nur Sicht sondern auch Geräusche verdecken soll.). Auch die sollte dann definiert werden, ab welchem Abstand die Rechnung durchgeführt werden soll, dass nicht vom Start des Levels die Rechenleistung ausgelastet ist.

Man kann aber durch psychoakustische Effekte viel umgehen und simplifizieren. Da wir das Hören gelernt haben, haben wir viele Aspekte schon abgespeichert und rufen diese ab, wenn eine bestimmte Empfindung angenähert wird. Hören wir jemanden schreien, das Signal ist allerdings recht leise, wissen wir aus Erfahrung, dass sich die Schallquelle weit weg befinden muss; auch wenn wir keinen realistischen Raumhall und interaurale Differenzen wahrnehmen können.

Das Verhältnis von Direktschall zu Nachhall ist auch ein identifizierender Faktor für die wahrgenommene Entfernung.

Der Mensch lokalisiert die Position von Schallquellen im ihn umgebenden Raum durch Analyse der Unterschiede zwischen den Signalen die von den beiden Ohren empfangen werden. Hier spielen auch anatomische Parameter (Ohrabstand, Schulter, Haare...) eine Rolle, die den Sound verändern, aber uns bekannt sind. Dadurch, kann kein perfektes, allgemein anwendbares 3D-Audio für Kopfhörer produziert werden. Es kann sich aber angenähert werden. Mit 3D Platzierungen über rechnerische Annäherung befinden wir uns immer ein Stück weg vom Kopf allerdings können virtuelle Schallquellen dieser Art im Raum fest positioniert werden, sodass sie sich nicht mit der Kopfdrehung mitbewegen. Im Gegensatz dazu gibt es auch die Möglichkeit, mit Kunstkopfaufnahmen binaural Geräusche wiederzugeben, die sich zum Beispiel direkt im Nacken oder am Ohr befinden. Der wahrgenommene Unterschied in Abständen ist deutlich unterscheidbar. Allerdings kann diese Methode nur in Kombination mit platzierten Geräuschen und nur für kürzere Clips verwendet werden. Diese sind nämlich am Kopf positioniert und drehen sich mit.

Begriffe, die häufiger auftreten, wenn man sich mit Ton im dreidimensionalen Raum beschäftigt:

» HRTF - Head Related Transfer Function

beschreibt die spektrale Veränderung eines Signals auf dem Weg von der Quelle bis zum Gehörgang

» ITD - Interaural Time Difference

ist die unterschiedliche Ankunftszeit desselben Signals am linken und rechten Ohr

» IID - Interaural Intensity Difference

ist die Pegeldifferenz der beiden Signale an den Ohren

» IPD - Interaural Phase Difference

die Veränderung in Phase zwischen den beiden Signalen

» Channel Based Audio

Audio-Signale die aufgeteilt sind auf Spuren, die spezifischen diskreten Lautsprechern zugeordnet sind

Lautstärkeabfall

Wird auch Attenuation genannt. Definieren des Lautstärkeabfalls bildet Soundquellen realistischer im Raum ab, hilft aber auch bei der narrativen Einbindung. Wenn man einen Gegenstand von sich wirft und er dementsprechend leiser wird, oder ein Radio, das bestimmte Informationen vermitteln soll, aber erst ab einer bestimmten Position hörbar sein soll.

Die Attenuation der Lautstärke lässt sich auch kombinieren mit dem entfernungsabhängigen Verändern von Frequenzen durch Filter oder anderen Parametern. Eine Geräuschquelle die weiter entfernt ist hat beispielsweise weniger hohe Frequenzen, da diese schneller abfallen als tiefe.

Source Material und Aufarbeitung

Um Audio sinnvoll im Raum platzieren und nachbearbeiten zu können, sollte man mit Mono-Dateien starten. Bei der Aufnahme sollte man natürlich auf möglichst viel Rauschabstand und gute Qualität achten, aber zusätzlich möglichst wenig Raum mit aufnehmen. Je nach Gegenstand kann man aber nicht grundsätzlich ganz nah mit dem Mikrofon herangehen, verschiedene Gegenstände können sich sehr unterschiedlich anhören 5cm oder 1m entfernt (Bsp. Ein Wecker, dessen Uhrwerk man hört oder nur das Ticken). Auch hier spielt das Vorwissen wie o.g. mit ein, nur dass es sich negativ auswirken kann.

Beim Bearbeiten in der DAW sollte man alle Clips säubern, dass außer dem Geräusch das verwendet werden soll wenig Leere ist. Längere Clips sollten gut loopbar sein, kürzere Clips (von denen, für die Variationsmöglichkeiten oft mehrere verwendet werden) möglichst homogen zu sortieren, sodass keines heraussticht.

Falls es nicht möglich ist, das Geräusch zu mikrofonieren, ist ein Reference Recording zu empfehlen.

Echo & Hall

Es gibt Möglichkeiten, Echo und Hall auszurechnen, das benötigt allerdings viel Prozessorkapazität, was teuer und nicht für alle Projekte realistisch ist.

Wenn die Geräuschquellen auf Objekte im Raum positioniert werden kann aber auch ein räumlicher Höreindruck erstellt werden. Außerdem ist es möglich über Plugins das ausgehende Signal der Kopfhörer zu binauralisieren indem man HRTFs aufrechnet.

Zu den positionierten Geräuschquellen kann ein zugeordneter Reverb an anderer Stelle im Raum positioniert werden. Das funktioniert nicht für jede Umgebung, aber stehe ich an einem Canyon und habe einen Gegenstand in der Hand, der ein Geräusch abgibt, kann zeitverzögert das ähnliche Signal aus der Entfernung ein sehr überzeugendes Echo imitieren.

Geräusche können auch aufgeteilt werden, zum Beispiel der Auslöser und Schuss einer Pistole auf der Position der Waffe, der Knall/Impact auf dem Ziel und der Nachhall aus dem Raum.

Der Mensch nimmt selten alle Umgebungsgeräusche wahr und filtert nur die für ihn relevanten heraus. Das gilt auch für den Nachhall. Das bedeutet bei guter Gestaltung fehlt dem Höreindruck nichts, das man im Vergleich mit der echten Welt wahrnehmen würde, obwohl deutlich weniger Information am Ohr ankommt.

Mischung von 3D Szenen

Das Mischen von Anwendungen in 3D und auf „First Person“ Basis erfolgt zum Teil erst bei Wiedergabe am Nutzer. Bei jedem Durchgang werden die Trigger zu unterschiedlichen Zeitpunkten und verschiedenen Reihenfolge und Menge erreicht. Würde man das aufnehmen, hätte man jedes mal eine andere Mischung. Bis zu einem gewissen Grad lässt sich allerdings beeinflussen was NICHT gehört werden soll. Beschränkungen, wie viele Sounds der gleichen Art abgespielt werden können sind zum Beispiel sehr hilfreich. Wozu eine Mischung auch auf jeden Fall relevant ist, ist die Lautstärken anzugleichen aufeinander. Da es keinen linearen Ablauf gibt, ist es am schnellsten, wenn man zu zweit simultan an der Mischung arbeitet. Das erfordert

Audio Middleware

Eine Software, die die Differenz zwischen Digital Audio Work Stations und Game Engines überbrückt. Sie ermöglicht es Sounddesignern, bei verschiedenen Projekten mit unterschiedlichen Game Engines mitzuarbeiten, ohne die speziellen Eigenheiten der verschiedenen Softwares zu kennen oder sogar Elemente aus verschiedenen Projekten zusammenführen zu können.

Die üblichsten Middleware Softwares sind:

- Wwise (Zielgruppe Game Sound Designer, Toolbox mit Bausteinprinzip)
- Fmod (weniger programmierähnlicher, eher wie DAW Workflow)
- Fabric (von Unity, browserkompatibel, Einbindung mit Reaper möglich)

In Audio Middleware können Tonhöhen und Lautstärken zufällig variabel gesetzt werden, generieren von zufälliger Reihenfolge, Fading und Loops und das erstellen von Sound Pools, die zusammengehörige Clips unter verschiedenen Bedingungen beinhalten können.

Als Sounddesigner arbeitet man mit bekannten Tools und muss sich nicht unbedingt mit der Programmierseite der verschiedenen Engines auseinandersetzen.

Bei Fragen gerne melden unter elisabeth.drobny@gmail.com