# Lokalisierung von Schallquellen in Unity

Jan Lorenz, Vincent Schnoor, Tim Zschage

24. April 2020

Projektlink: Audio Localization Project

## 1 Motivation

Die Außenohr-Übertragungsfunktion oder Head-Related Transfer Function, kurz HRTF, ist dafür da, Sound möglichst realistisch darzustellen, indem es die Form von Außenohr, Kopf und Rumpf mit einbezieht.
Da die HRTF sehr individuell ist, braucht man eine gute Durchschnitts-HRTF, die für eine breite Masse funktioniert. Deshalb benötigt man in Studien zur Untersuchung von HRTFs möglichst viele Probanden um ein breites Ergebnis zu bekommen. Es ist jedoch schwierig genügend Studienteilnehmer zu finden, weshalb man eine Alternative finden muss, um nicht zwangsläufig von der Zahl der Probanden abhängig zu sein.
In unserem Projekt wollen wir im Rahmen einer Simulation in Unity eine Anwendung entwickeln, welche die Position einer Geräuschquelle mit höchstmöglicher Genauigkeit bestimmt, um damit die Notwendigkeit von physischen Probanden zu verringern.

## 2 Herangehensweise

Um eine Schallquelle anhand verschiedener Audiospuren möglichst genau lokalisieren zu können, haben wir uns mit zwei Verfahren vertraut gemacht, welche in der Realität zu diesem Zweck verwendet werden:

### 2.1 Pegeldifferenzbestimmung

Bei der Pegeldifferenzbestimmung wird der Pegel des linken Kanals und rechten Kanals miteinander verglichen. Der linke und rechte Kanal stehen hierbei für das linke und rechte Ohr. Sind die Pegel identisch, lässt dies darauf schließen, dass sich die Audioquelle direkt vor oder direkt hinter dem Audio Listener befindet. Dreht sich der Audio Listener komplett um die eigene Achse, wird es zwei Punkte geben, an denen die Pegeldifferenzen ein Maximum haben. Dies ist das Resultat davon, dass sich die Audioquelle entweder links oder rechts vom Audio Listener befindet.
Die Pegeldifferenzbestimmung kann man auch dazu nutzen, die Elevation der Audioquelle zu bestimmen. Dazu wird nach Bestimmung des Azimuts, dem Horizontalwinkel, der Audio Listener um die horizontale Achse relativ zur Audioquelle gedreht. Dies erreicht dieselbe räumliche Anordnung von Audioquelle und Audio Listener wie bei der Bestimmung des Azimuts und das obig beschriebene Verfahren kann zur Bestimmung der Elevation wiederholt werden.

### 2.1.1 Erwartung

Mit diesem Lösungsansatz erwarten wir eine simple und verlässliche Umsetzung, die jedoch eine geringere Lokalisationsgeschwindigkeit mit sich bringt. Unsere Umsetzung soll für eine gegebene Rotationsposition den linken und rechten Kanal vergleichen und sich anschließend in Richtung des lauteren Kanals bewegen. Anschließend wird der Vorgang so lange fortgeführt, bis der vorher leisere Kanal zum lauteren Kanal wird, denn dies bedeutet, dass die Audioquelle direkt vor dem Audio Listener liegt.

## 2.2 Interaurale Kreuzkorrelation

Bei der Interauralen Kreuzkorrelation können die Interaurale Leveldifferenz (ILD) und die Interaurale Zeitdifferenz (ITD) für die Bestimmung des Azimuts der Audioquelle genutzt werden [1].
Die ILD ist die wahrgenommene frequenzabhängige Pegeldifferenz in dB zwischen beiden Ohren.
Die ITD ist die frequenzabhängige zeitliche Phasenverschiebung $\omega$ zwischen beiden Ohren.
ITD liefert genauere Ergebnisse für niedrige Frequenzen, da Phasenverschiebungen genauer zu identifizieren sind, da die Schallwellen, aufgrund ihrer Länge, um den Kopf herum gebogen werden. ILD liefert genauere Ergebnisse bei hohen Frequenzen, deren Wellenlänge wesentlich kleiner als der Abstand zwischen den Ohren ist, da der Kopf hochfrequente Schallwellen absorbiert, bzw. reflektiert und somit ein Lautstärkeunterschied zwischen beiden Ohren entsteht.
Da beide Verfahren der Bestimmung des Azimuts dienen, können sie gleichzeitig verwendet werden, um die Azimutbestimmung zu verbessern [3].

### 2.2.1 Erwartung

Da die Interaurale Kreuzkorrelation einen Zusammenhang zwischen beiden Ohren herstellt, erwarten wir eine sofortige und genaue Bestimmung des Azimuts. Diese sollte nicht über eine sukzessive Annäherung erfolgen, sondern eine direkte Angabe des Winkels bei jeder gegebenen Position liefern.

## 2.3 Head-related Transfer Function (HRTF)

Die Head-related transfer function function (HRTF) ist eine Transferfunktion, welche beschreibt, wie ein Hörereignis von einem individuellen Kopf und dessen Ohrmuscheln und Oberkörperform verändert wird, bevor es in den Gehörgang tritt.
Der Einsatz einer HRTF wirkt sich positiv auf die Lokalisationsfähigkeit von Geräuschen einer Person in virtuellen Umgebungen aus.

# 3 Umsetzung

## 3.1 Unity-Szene

Um unsere Anwendung testen zu können, musste eine Unity-Szene erstellt werden, in welcher folgende Components benötigt werden: Audio Source und Audio Listener. Auf der Audio Source liegt ein Spawn-Skript, welches bestimmt, an welcher zufälligen Position in einem Kreis um den Audio Listener herum die Audio Source beim Start der Simulation erscheint (siehe Abbildung 2). Dabei muss der Audio Source-Component wie in Abbildung 1 zu sehen konfiguriert sein.

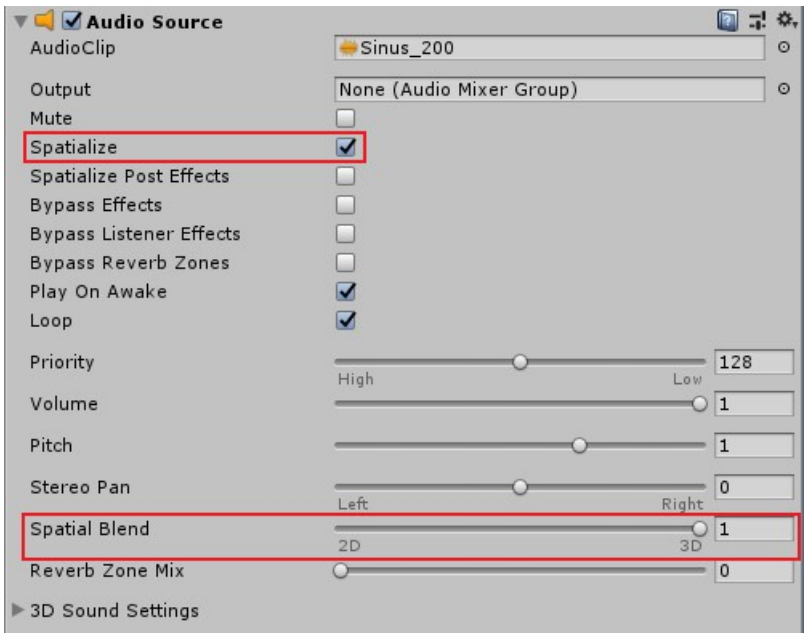


Abbildung 1: Audio Source-Konfiguration

Auf dem Audio Listener liegt das Skript, welches zur Lokalisation der Audioquelle dient. Weiterhin muss in den Projekteinstellungen, im Reiter Audio, ein Spatializer-Plugin (HRTF) angegeben werden.
In dem Projekt wurden verschiedene Spatializer verwendet. Die Unity Engine stellt mit dem „Windows Mixed Reality Package und dem „Oculus Desktop Package", zwei installierbare Pakete mit Spatializern nativ zur Verfügung. Zusätzlich wurde der „Steam Audio Spatializer" als Drittanbieter Software eingebunden. Steam Audio ist ein Audio Plugin, das seit 2017 von der Firma Valve angeboten wird und darauf abzielt, den Sound in Videospielen so realistisch wie möglich klingen zu lassen.

Wir ließen die Anwendung mit den drei Spatializern, sowie ein weiteres Mal ohne Spatializer, durchlaufen und die daraus resultierenden Daten geschätzter Winkel, tatsächlicher Winkel und Winkelabweichung in einer CSV-Datei speichern.

## 3.2 Skript

### 3.2.1 Spawn-Skript

Das Skript um die Audio Source an einer zufälligen Stelle in einem Kreis um den Audio Listener herum zu platzieren nimmt als einstellbares Argument den Radius als Fließkommazahl an. Daraufhin wird eine zufällige Position auf dem Kreis mit dem eingestellten Radius mit dem Audio Listener als Mittelpunkt gewählt und die Audio Source an diese Position gesetzt (siehe Abbildung 2).

### 3.2.2 Lokalisationsskript

Das Skript für die Lokalisation der Audioquelle geht folgende Schritte durch:

1. Auslesen der Spannungswerte der einzelnen Kanäle und Umwandlung in Pegel (dB)
2. Vergleich der Pegelwerte beider Kanäle
3. Rotation in Richtung des lauteren Kanals
4. Nach x Sekunden wird der aktuelle Winkel dokumentiert und das Programm beendet oder zurückgesetzt.

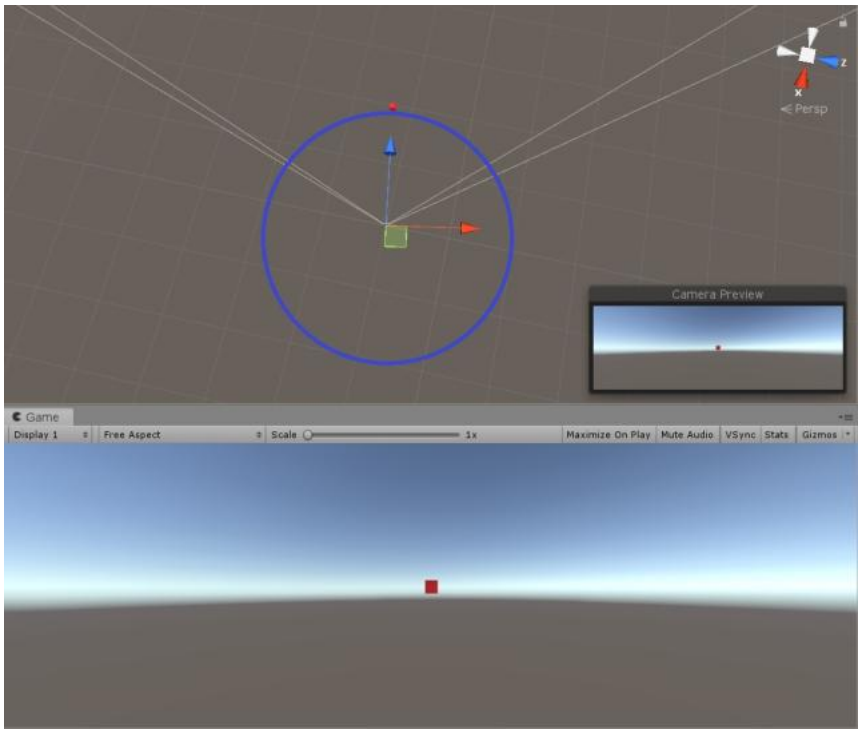


Abbildung 2: Audioquelle (rot) dreht befindet sich auf einer randomisierten Position auf dem Kreis (blau) um den Audio Listener innerhalb des Kreises

## 3.3 Wahl der Audioclips

Um die Funktionalität für verschiedene Audioclips zu verifizieren, wurden sowohl Mono- als auch Stereo-Clips verwendet.
Die Frequenzabhängigkeit der Interauralen Kreuzkorrelation wird geprüft, indem Sinussignale mit mehreren Frequenzen verwendet werden.
Ebenso werden Audioclips mit rosa und weißem Rauschen, Musik und Sprache geprüft, um die häufigsten Soundarten abzudecken.

Verwendet wurden:

- Rosa Rauschen
- Weißes Rauschen
- Sinus 200 Hz
- Sinus 1000 Hz
- Sinus 2000 Hz
- Sinus 3200 Hz (Mono)
- Sinus 3200 Hz (Stereo)
- Sprache (Stereo)
- Orchesteraufnahme (Stereo)

Die Auswahl der Sinusfrequenzen leitet sich aus der Wellenlänge ab. Basierend auf einem durchschnittlichen Ohrenabstand von 21 cm leitet sich eine Frequenz von ca. 1600 Hz ab. Demnach wählen wir Frequenzen über und unter 1600 Hz aus, um zu prüfen, ob unterschiedliche Ergebnisse bei der Lokalisation mit der Interauralen Kreuzkorrelation entstehen.

Die Unterscheidung von Mono- und Stereospuren wurde vorgenommen, um zu prüfen, ob die Unity-Engine diese anders verarbeitet und andere Ergebnisse daraus resultieren.

# 4 Durchführung des Versuchs

Wird das Programm gestartet, fängt der Audio Listener an sich zu drehen. Dies tut er in 1° Inkrementen in die Richtung des lauteren Kanals. Nach 10 Sekunden stoppt der Listener und der momentane Winkel wird gespeichert und der nächste Durchgang startet. Insgesamt wird jeder Audioclip 10 mal wiederholt und die Audio Source erscheint jedes Mal an einer anderen Stelle. Wir ließen die Anwendung mit den drei Spatializern, sowie ein weiteres Mal ohne Spatializer, durchlaufen und die daraus resultierenden Daten geschätzter Winkel, tatsächlicher Winkel und Abweichung wurden in einer CSV-Datei gespeichert.

# 5 Resultate

## 5.1 Kreuzkorrelationsfunktion

Bei dem Versuch, den Azimut mit der interauralen Kreuzkorrelation zu bestimmen, stellte sich heraus, dass die Unity-Engine Audiosignale zeitunabhängig berechnet. Eine Lokalisation mit der interauralen Kreuzkorrelation war deshalb nicht möglich, da diese ausschließlich zeitabhängig arbeitet.

## 5.2 Pegeldifferenz

Das Lokalisieren der Audioquellen hat mit den verschiedenen Spatializern durchgehend gut funktioniert. Alle Spatializer schafften es, die Winkeldifferenz zwischen tatsächlicher Position und geschätzter Position in einem kleinen Bereich, mit einer unerheblichen Zahl an Ausreißern, zu halten.

Mit der MS HRTF lag die Winkeldifferenz beispielsweise meistens zwischen -19° und 9° und bei Steam Audio zwischen -7° und 12°. Besonders gut funktionierten „Speech“ und „Symphony Sounds“. Nur bei einem Sinuston von 200Hz konnte fast kein Spatializer die Quelle genau orten. Mit einem Fehlerbereich von -4° bis 8° war der Oculus Spatializer der einzige, der die Quelle bei 200Hz ziemlich genau orten konnte. Die gemessenen Werte lassen darauf schließen, dass eine

Soundquelle mit 200Hz mit dem Einsatz von Spatializern schwer zu lokalisieren ist.
Die Abweichungen bei 200Hz waren zwar unter den Spatializer recht weit auseinander aber die Werte lagen dennoch ziemlich konstant beieinander. Der MS HRTF Spatializer wies beispielsweise eine Abweichung im Bereich -51° bis -54° und der Native Oculus Spatializer im Bereich -4° bis 8° auf. Lediglich der Steam Audio Spatializer war ziemlich ungenau mit Abweichungen im Bereich von -172° bis 167° .

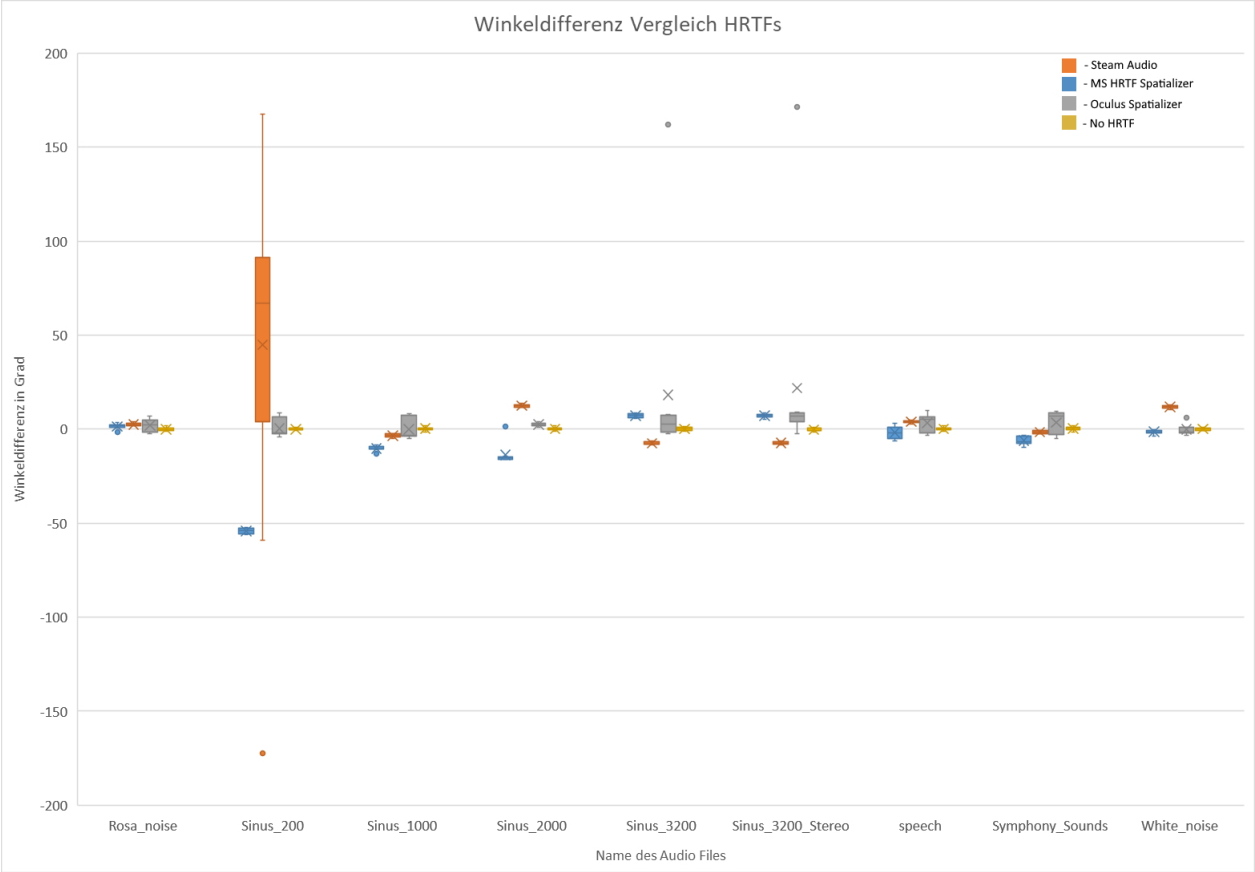


Abbildung 3: Vergleich der verschiedenen HRTFs

Man erkennt in der Abbildung, dass das Lokalisieren von Audioquellen ohne Spatializer anhand der Pegeldifferenzbestimmung sehr genau ist. Die größte Abweichung wurde hier mit 1,9° gemessen.

# Literatur


[1] C. Faller and J. Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *Acoustic Society of America*, Vol. 116(5), 2004.

[2] D. Grelaud, N. Bonneel, M. Wimmer, M. Asselot, and G. Drettakis. *Efficient and Practical Audio-Visual Rendering for Games using Crossmodal Perception*. Association for Computing Machinery Inc, 2009.

[3] M. Raspaud, H. Viste, and G. Evangelista. Binaural source localization by joint estimation of ILD and ITD. *IEEE Transactions on Audio, Speech, and language processing*, 18(1), 2010.

[4] S. Weinzierl. *Handbuch Der Audiotechnik*. Springer, 2008.

[5] U. Zölzer. *DAFX: Digital Audio Effects*. John Wiley and Sons Ltd, 2011.