

Audio in AR – Akzeptanz von ergänzenden Audiosignalen in realer Umgebung

Rodrigo Silvetti Murillo¹, Prof. Dr. Eva Wilk²

¹ Student im Master Digital Reality HAW, 22081 Hamburg, Email: rodrigo.silvetti@gmail.com

² Department Medientechnik HAW, 22081 Hamburg, Email: Eva.Wilk@haw-hamburg.de

Einleitung

Im Rahmen dieses Forschungsprojekts wird untersucht, wie glaubwürdig die unterschiedlichen Audiosignale in Augmented-Reality-Anwendungen sind und wie sie die Immersion beeinflussen. Unerwartete Geräusche wie beispielsweise ein Motorsound oder das Geräusch eines Fahrzeugs oder LKWs können die Wahrnehmung verzerren und einen Schreck- oder Orientierungsreflex auslösen. In dieser Studie wird diese Idee untersucht. Aus diesem Grund wurde ein experimentelles Augmented-Reality-System entwickelt. Die Probanden nehmen an einem interaktiven Spiel teil, während verschiedene Geräusche ihre Aufmerksamkeit ablenken. Das Augmented-Reality-Audio-System wurde auf einem Microsoft HoloLens 2-Gerät entwickelt. Die Studienergebnisse sollen zeigen, ob es Unterschiede in der Wahrnehmung von Klängen in diesem Szenario gibt. Diese Ergebnisse sind bedeutsam, da sie zur Verbesserung der Gestaltung von Audio-Augmented-Reality-Anwendungen beitragen können, um ein möglichst natürliches Erlebnis zu schaffen. Zudem können sie Aufschluss darüber geben, ob weitere Untersuchungen notwendig sind und ob Unterschiede in der Wahrnehmung zwischen passenden und unpassenden Klängen bestehen. Solche Projekte können auch in Museen [7] eingesetzt werden, um den Realismus von Führungen zu erhöhen, wie das iArt-Projekt 'Augmented Reality in Your Headphones' zeigt (verfügbar unter: <http://www.iart.ch/en/next/augmented-reality-audio/>).

Stand der Wissenschaft

Wir erfassen die Reaktionen der Testpersonen in unserem Projekt durch Beobachtung ihrer Kopfbewegungen. Das Konzept wurde im Artikel "3D Audio Augmented Reality: Implementation and Experiments" von Sundareswaran V., Wang K. und Chen S. [1] vorgestellt. In diesem Artikel wird über ein Projekt mit einem 3D-Audio-Wearable-System berichtet. Das System ermöglicht dem Benutzer, Positionen zu teilen und Alarmer auszulösen. Das in diesem Paper beschriebene System verwendet den CharmIT-Computer und umfasst 3D-Audio, drahtlose Kommunikation, Voice-over-IP, Spracherkennung, GPS-Positionierung und Head-Tracking. In Tests waren Benutzer erfolgreich in der Lage, mithilfe von 3D-Audio virtuelle Positionen zu navigieren. Eine noch präzisere Navigation könnte durch eine verfeinerte akustische Darstellung erreicht werden. Die Daten für diese Studie wurden durch die Bewegungen und Drehungen der Köpfe der Testpersonen erfasst. Aus diesem Grund kam uns die Idee, die Drehung und Bewegung der Kamera der HoloLens als Hilfsmittel zu nutzen.

Eine weitere Quelle für unser Projekt ist die Studie "Au-

ditary display design for exploration in mobile audio-augmented reality" von I. Oakley [4], die sich mit der Analyse der aufgezeichneten Daten der Teilnehmer beschäftigt und beschreibt, dass der Grad der Immersion mit dem verstärkten Anhalten und Abtasten der Umgebung zusammenhängt, was anhand der Gehgeschwindigkeit und der Kopfbewegung quantifiziert werden kann. Die Versuchspersonen sollten glauben, dass die Objekte tatsächlich in der Realität existieren.

Diese Projekte haben gezeigt, dass bestimmte technische Voraussetzungen, wie spezielle Brillen und Kameras zur Verfolgung der Person im Raum, ein mobiles Gerät mit GPS, Sensoren wie Gyroskop und Accelerometer, die Simulation des Raumes, von großer Bedeutung sind. Basierend auf der bereits erwähnten Literatur, die als Grundlage für diese Untersuchung dient, kann davon ausgegangen werden, dass die Verwendung der HoloLens 2 in Spielen die Immersion durch die Verwendung von 3D-Sound mittels HRTF erhöht. Das Headset der HoloLens erleichtert die Datenaufnahme. Zudem wird keine WLAN-Verbindung benötigt, da die Anwendung offline betrieben werden kann.

Studien-Design

Die Hypothesen der Studie lauten wie folgt:

Hypothesen

Für die **quantifizierte Analyse** haben wir die folgenden Hypothesen:

Ha 0: Die Probanden reagieren unterschiedlich, wenn sie mit passenden Klängen und unpassenden Klängen konfrontiert werden.

Ha 1: Die Reaktion der Probanden zeigt keinen Unterschied zwischen passenden und unpassenden Klängen.

Wenn Ha0 gilt:

Hb 0: Die Probanden bewegen sich mit passenden Klängen mehr, als mit unpassenden Klängen.

Hb 1: Die Probanden bewegen sich mit unpassenden Klängen mehr, als mit passenden Klängen.

Für die **qualitative Analyse** haben wir die folgende Hypothesen:

Hc 0: Die Probanden glaubten, dass die virtuellen Geräusche zur realen Welt gehörten.

Hc 1: Die Probanden glaubten, dass die virtuellen Geräusche nicht zur realen Welt gehörten.

Benutzer - Erfahrung

Audio Augmented Reality bietet eine 3D-Benutzererfahrung, die eine fortschrittliche Interaktion

mit digitalen Informationen ermöglicht und für eine noch bessere Immersion in der virtuellen Welt sorgt. Ein Spiel wird eingesetzt, da es laut der Studie "SSoundPacman: Audio Augmented Reality in Location-Based Games" [2] die Immersion erhöht. Die Reaktionen und Bewegungen der Probanden werden für weitere Analysezwecke erfasst, um Unterschiede zu finden, die unsere Hypothese unterstützen.

Methoden und Materialien

Für diese Studie werden die erfassten Daten der Hololens 2 von Microsoft und ein Fragebogen verwendet.

Hololens 2

Die Hololens 2 wurde für dieses Projekt ausgewählt, da sie Spatial Audio durch HRTF unterstützt und das Tracking der Person und der Umgebung vereinfacht. Die Microsoft Hololens 2 ist eine hochauflösende Mixed-Reality-Brille, die Raumklang in nahezu realistischer Qualität ermöglicht. Die Brille kann mit Hilfe von vier Tiefenkameras die Umgebung erfassen und die Position von Schallquellen bestimmen. Zusätzlich sind IMU-Sensoren integriert, die die Kopfdrehung und -bewegung des Probanden erfassen.

Sound Entwicklung

Es wurde ein Spiel mit passenden und unpassenden Umgebungsgeräuschen erstellt. Die Probanden/innen befinden sich in einer Hochschule, in der es verschiedene Klänge gibt. Die folgende Liste zeigt einige der in diesem Projekt verwendeten Sounds:

Passende Klänge	Unpassende Klänge
Vogelstimmen	Moto Harley
Baby jammert	Fahrradklingel
Hundegebell	Auto Hupt
Beifall klatschen	Löwe
um Hilfe bitten	Grillen
Hallo sagen	Drums
Gesprächen	Kojote
Regen	Froschgeräusche
Katze miaut	Wolfsgheul

Die Aufgabe des Probanden besteht darin, im Spiel drei Hinweise zu finden. Es gibt auch Anleitungen, die durch das Spiel führen. Die Audios wurden als .wav-Dateien aufgenommen und heruntergeladen, um eine gute Qualität sicherzustellen. Um Klangwiederholungen oder ähnliches zu vermeiden, kann eine dynamische Änderung der Tonhöhe helfen, eine realistische Wahrnehmung zu erreichen.

Fragebogen

Diese Fragen zielen darauf ab, die Immersion des Spiels für die Probanden zu erfassen. Die Antworten werden für eine qualitative Analyse ausgewertet. Der Fragebogen Slather-Usch-Steed [3] wurde für diese Studie angepasst und den Proband/innen zur Verfügung gestellt, damit die Präsenz in virtuellen Erlebnissen gemessen werden kann.

In Anwendungen von virtueller oder erweiterter Realität wird der Begriff Präsenz definiert als: *Präsenz steht in engem Zusammenhang mit Immersion und beschreibt das Gefühl "dort zu sein". Es beschreibt die Erfahrung des Nutzers im Spiel, als ob er im Spiel wäre. Und er empfindet diese Erfahrung als real.*[6]

Methoden der Studie

Dieser Abschnitt beschreibt, wie die Daten für die Studie bearbeitet und gesammelt wurden:

Durchführung

Die Teilnehmer/innen sollen mit dem Hololens 2 Gerät durch das Foyer der Neubauten der Hochschule für Angewandte Wissenschaft (HAW - Fakultät DMI) gehen. Dabei begleitet sie eine virtuelle Assistenz durch das Spiel. Die Bewegungen der Testpersonen lösen passende oder unpassende Geräusche aus, die im Hintergrund des Spiels zu hören sind. Daher werden die Position und Rotation der Probanden/innen erfasst, damit sie auf die Klangquelle reagieren. Deshalb ist es möglich, ihre Position und Rotation zu bestimmen und die Daten für die quantitative Analyse der Studie zu nutzen. Abschließend beantworten die Teilnehmer/innen noch einen Fragebogen.

Datensammlung

12 Probanden/innen haben mit oder ohne Erfahrung in Mixed Reality Devices an der Studie teilgenommen. Alle Versuche wurden an der Hochschule HAW-DMI in Hamburg gemacht. Jeder Versuch hat ungefähr 15 Minuten gedauert. Die Probanden waren zwischen 20 und 60 Jahre alt und 33.3 % von ihnen haben regelmäßig ein Augmented Reality und/oder Virtual Reality Headset genutzt, 41.7 % haben selten ein Headset benutzt und 25 % haben nie ein Headset getragen.

Statische Auswertung

Nach jedem Versuch wird eine .csv-Datei in Hololens 2 erstellt. Diese Datei beinhaltet die folgenden Items: die Position des Headsets, die Rotation des Headsets, ob es sich um einen passenden Klang bzw. einen unpassenden Klang handelt. Diese Daten werden jede halbe Sekunde gespeichert. Die Position der Teilnehmer im Raum wird durch einen dreidimensionalen Vektor (X, Y, Z) beschrieben. Die Drehung hingegen wird durch einen vierdimensionalen Vektor (X, Y, Z, W) beschrieben, der als Quaternion bezeichnet wird. Das Euler-Winkel-Verfahren wird verwendet, um ein Quaternion in einen Vektor mit drei Positionen umzuwandeln. Anschließend werden die Daten klassifiziert, um die Kategorien "passende Klänge", "unpassende Klänge" und "ohne Klänge" zu unterscheiden.

Nachdem die Daten erhoben wurden, führen wir eine quantitative Analyse durch. Dafür nehmen wir die Daten der Position und der Rotation. Jeder neue Wert (die aktuelle Position oder die aktuelle Rotation) x wird von dem vorherigen Wert (die vorherige Position oder die vorherige Rotation) subtrahiert, um eine Differenz (genannt auch Reaktionsstärke) zu bestimmen. Je größer

die Differenz, desto größer war den Schreckreflex bzw. Orientierungsreflex der Probanden/innen. Wir benötigen die Mittelwerte der Differenzen für jede Gruppe auf jeder Achse (x,y,z). Das wird für jeden Probanden/in wiederholt. Anschließend führen wir eine Anova RM-Test-Analyse für alle Probanden/innen zwischen den Mittelwerten der Bewegung ohne Klänge, der Bewegung mit passenden Klängen und der Bewegung mit unpassenden Klängen durch. Die beantworteten Fragen des Fragebogens werden aufgelistet und die Auswertung der Daten wird mittels SLATER-USOH-STEED-Technik [3] durchgeführt.

Ergebnisse

Der ANOVA-RM Test hat folgendes ergeben:

Position		
	f-Wert	p-Wert
Position in x-Achse	0.032	0.967
Position in y-Achse	7.681	0.002
Position in z-Achse	0.187	0.829

Rotation		
	f-Wert	p-Wert
Rotation in x-Achse	0.999	0.379
Rotation in y-Achse	5.939	0.006
Rotation in z-Achse	0.522	0.598

Aus diesen Ergebnissen können wir sagen, dass nur die Position und Rotation in Y signifikant sind. Denn für beide gilt der p-Wert < 0.05 . Deshalb ist die Hypothese von ANOVA H_0 und die Hypothese H_{a0} wahr: Die Probanden reagieren unterschiedlich, wenn sie mit passenden Klängen und unpassenden Klängen konfrontiert werden. Die Hypothese H_{b0} wird mittels Post-hoc-Test geprüft.

Das Testergebnis wird ausgegeben:

```

Anova Test für Rotation in Y
      df  sum_sq  mean_sq      F  PR(>F)
C(treatments)  2.0  0.006538  0.003269  5.939481  0.006712
Residual      30.0  0.016511  0.000550      NaN      NaN
-----
Tukey's HSD Test für Rotation in Y
  group1 group2  Diff  Lower  Upper  q-value  p-value
0   Real  Normal  0.031688  0.007032  0.056344  4.479874  0.009561
1   Real  Unreal  0.027609  0.002953  0.052265  3.903214  0.025646
2   Normal Unreal  0.004079 -0.028577  0.028735  0.576660  0.900000
    
```

Abbildung 1: Post-Hoc Tukey Test mit p- und q-Werten für jede Gruppe. Real bezeichnet Bewegungen mit passenden Klängen, Unreal bezeichnet unpassende Klänge und Normal steht für Bewegungen ohne Klänge

In Übereinstimmung mit den Forschungsmethoden [5] kann aus der post-hoc-Interpretation geschlossen werden, dass der p-Wert in unserer Studie für passende Klänge und ohne Klänge (Real - Normal) sowie für passende Klänge - unpassende Klänge (Real - Unreal) kleiner als 0,05 ist. Die folgende Hypothese kann daher als wahr betrachtet werden:

H_{b0} : Die Probanden bewegen sich mit passenden Klängen mehr, als mit unpassenden Klängen.

In den folgenden Abbildungen können wir die Variation innerhalb jeder Gruppe auf der Y-Achse besser erkennen.

Die roten Punkte sind die Mittelwerte der Differenzen für jeden Teilnehmer.

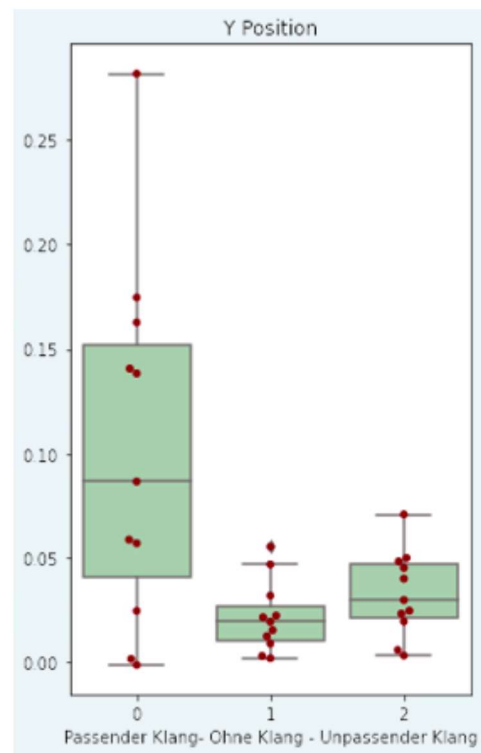


Abbildung 2: Boxplot der Mittelwertdifferenzen für die Position auf der Y-Achse

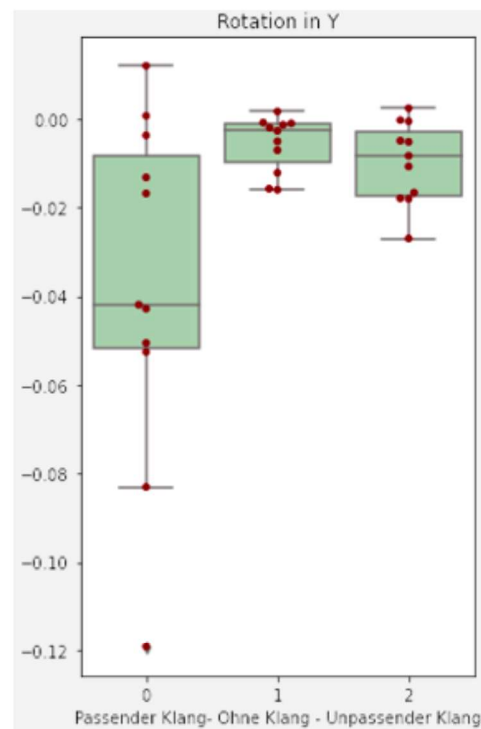


Abbildung 3: Boxplot der Mittelwertdifferenzen für die Rotation auf der Y-Achse

Da ist zu beobachten, dass die Variation des Durchschnitts für passende Geräusche in diesem Fall größer ist

als in anderen Fällen.

Der SLATER-USOH-STEED-Test besteht aus Fragen auf einer Skala von 1 bis 7, und nur die Antworten 6 oder 7 sollen gezählt werden, um eine eindeutige Bewertung zu erhalten. Die Fragen in unserem Test erhielten für jede Frage einen Wert zwischen 25% und 33%. Dieser Test misst die Präsenz der Teilnehmer in der Augmented-Reality-Anwendung. Die Ergebnisse unserer qualitativen Analyse haben nicht gezeigt, dass die Probanden/innen die Geräusche als besonders immersiv empfanden. Deshalb kann die Hypothese Hc1 als wahr betrachtet werden.

Diskussion

Bei diesem Experiment wurden folgende Erkenntnisse gewonnen:

Die Entwicklung eines Audio-Augmented-Reality-Projekts war erfolgreich und lief nach Plan, überraschenderweise war die Orientierung in der Szene auch ohne Internet möglich und präzise.

Obwohl die qualitative Analyse kein eindeutiges Ergebnis bezüglich der Wahrnehmung der Geräusche durch die Versuchspersonen ergab, zeigten die quantitativen Ergebnisse, dass die Reaktion der Versuchspersonen bei den passenden Geräuschen stärker war.

Die Suche nach Hinweisen in einem Raum war zu komplex, um gleichzeitig auf Geräusche und Anweisungen zu achten. Ein einfaches Spiel könnte hier besser geeignet sein.

Da das Projekt mit einer Mixed-Reality-Brille die Hololens 2 umgesetzt wurde, erwarteten die Probanden eine visuelle Rückmeldung für jedes Objekt.

Idealerweise sollten die Spielumgebungen indoor sein, da die Kameras der Hololens 2 noch Schwierigkeiten bei der Anpassung an die Helligkeit der Sonne oder andere Wetterbedingungen im Freien haben.

Nicht alle unpassende Geräusche lenken die Aufmerksamkeit auf sich – manche können auch einfach ignoriert werden.

Die Reaktion auf unpassende Töne ist von Person zu Person sehr verschieden – manche Menschen reagieren überhaupt nicht anders auf unpassende Klänge als gewohnte Klänge.

Zusammenfassung

In dieser Studie wurde eine Anwendung für Audio-Augmented-Reality entwickelt, um die Unterschiede in Reaktionsstärken bei passenden und unpassenden Geräuschen in Bezug auf die Umgebung zu untersuchen. Es wurde ein Spiel in der Hochschule mit verschiedenen passenden und unpassenden Klängen erstellt, um Bewegungsdaten zu sammeln. Das Headset Microsoft Hololens 2 mit seinen integrierten Sensoren ermöglichte sowohl die Datenerfassung als auch die Durchführung des Spiels. Die Ergebnisse der quantitativen Analyse zeigen, dass die Verwendung von zur Umgebung passenden Geräuschen zu einer höheren Reaktionsintensität führt.

Literatur

- [1] Sundareswaran V., Wang K., Chen S., Behringer R., McGee J., Tam C., Zahorik P., "3D audio augmented reality: implementation and experiments", The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, (2003). Proceedings., Tokyo, Japan, 2003, pp. 296-297, doi: 10.1109/ISMAR.2003.1240728.
- [2] Chatzidimitris Thomas, Gavalas Damianos, Michael Despina, SoundPacman: Audio augmented reality in location-based games, (2016) 18th Mediterranean Electrotechnical Conference (MELECON), pp.1-6, doi:10.1109/MELCON.2016.7495414
- [3] Usuh Martin, Catena Ernest, Arman Sima, Slater Mel, Using Presence Questionnaires in Reality, (2000), Presence, v.9, pp.497-503, doi: 10.1162/105474600566989
- [4] Vazquez-Alvarez Yolanda, Oakley Ian, Brewster Stephen, Auditory display design for exploration in mobile audio-augmented reality Personal and Ubiquitous Computing, (2011), 16. 1-13, doi:10.1007/s00779-011-0459-0.
- [5] Nicola Döring, Jürgen Bortz, Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, Springer Berlin, (2016), Heidelberg, doi:10.1007/978-3-642-41089-5
- [6] Freina L., Ott M., A literature review on immersive virtual reality in education: state of the art and perspectives. In The international scientific conference elearning and software for education, (2015, April), (Vol. 1, No. 133, pp. 10-1007).
- [7] Fatima Zahra Kaghat, Ahmed Azough, Mohammed Fakhour, Mohammed Meknassi, A new audio augmented reality interaction and adaptation model for museum visits, Computers & Electrical Engineering, Volume 84, (2020), 106606, ISSN 0045-7906, doi: 10.1016/j.compeleceng.2020.106606.