

# Entwurf und Evaluation eines Kommunikationskanals für Gesten in kollaborativen virtuellen 3D-Welten

Stefan Wittek<sup>1</sup>, Niels Pinkwart<sup>2</sup>

Institut für Informatik, Technische Universität Clausthal<sup>1</sup>  
Institut für Informatik, Humboldt-Universität zu Berlin<sup>2</sup>

## **Zusammenfassung**

Diese Arbeit beschäftigt sich mit der Frage, wie sich ein direkter und kontinuierlicher Kommunikationskanal für Gesten in kollaborativen virtuellen 3D Welten auf die Benutzer dieser Welten auswirkt. Als Basis für die Untersuchung dieser Frage wird ein Prototyp vorgestellt, der Gesten mit Hilfe der Microsoft Kinect erfassen und übertragen kann. Weitere Interaktionen mit der 3D-Welt sind durch die Wii Remote möglich. Eine durchgeführte Studie zeigt, dass der Gestenkommunikationskanal für bestimmte Aufgaben intensiv genutzt wird und deutet zudem auf eine Steigerung der durch die Benutzer wahrgenommenen Präsenz in der virtuellen Welt hin.

## 1 Einleitung

Kollaborative virtuelle Welten sind synchrone, persistente Netzwerke von Menschen, die diese Menschen durch Avatare repräsentieren, ihnen gemeinsame, kollaborative Interaktionen erlauben und von Computernetzwerken bereitgestellt werden (in Anlehnung an (Bell, 2008, p. 2)). Beispiele für Einsatzgebiete dieser Welten im privaten Umfeld sind MMORPGs wie etwa World of Warcraft, dreidimensionale soziale Netzwerke wie beispielsweise Second Life und auf professionelle virtuelle Meetings ausgelegte Systeme wie etwa OpenWonderland. Kooperation basiert in allen drei Beispielen auf der Möglichkeit, innerhalb dieser Welten miteinander kommunizieren zu können. Für die unvermittelte zwischenmenschliche Kommunikation spielen nonverbale Elemente eine essenzielle Rolle. Da Avatare in virtuellen Welten die Nutzer nicht nur grafisch darstellen, sondern ihnen auch erlauben, durch sie mit der Welt und mit anderen Nutzern zu interagieren, ist es naheliegend, auch nonverbale Kommunikation auf Avatare abzubilden.

Der einfachste Ansatz hierzu besteht darin, Bewegungsabläufe der Avatare bereits vor einer konkreten Kommunikationssituation zu definieren und es dem Nutzer zu ermöglichen, diese

vordefinierten Abläufe in einer konkreten Situation durch einen Trigger auf dem Avatar abzuspielen. Beispiele für solche Trigger sind Textkommandos (Salem and Earle, 2000), Elemente der GUI (Guye-Vuillème et al., 1999) oder dedizierte Eingabegeräte wie etwa der Beschleunigungssensor der Wii Remote (Sreedharan et al., 2007). Um die Bewegungsabläufe der Avatare besser an konkrete Situationen anpassen zu können, werden in einigen Ansätzen Trigger gewählt, aus denen sich bestimmte Parameter wie Geschwindigkeit und Dauer extrahieren lassen. Beispiele für derartige Trigger sind ein Grafiktablett (Barrientos and Canny, 2002) oder die Microsoft Kinect (Thai, 2011). Alle diese Ansätze haben gemeinsame Schwächen. So muss sich der Nutzer bewusst für das Senden eines nonverbalen Signals entscheiden, weshalb unbewusste nonverbale Signale nicht abgebildet werden können. Zudem muss er die Trigger kennen und während der Kommunikationssituation Aufmerksamkeit darauf verwenden, ein passendes Signal auszuwählen. Diese Probleme können umgangen werden, indem man dem Nutzer die direkte Kontrolle über die Bewegungsabläufe des Avatars gibt. Dies ist über verschiedene Eingabegeräte möglich. Allerdings sind diese Ansätze meist auf einzelne Körperteile beschränkt (Barrientos, 2000) oder verwenden am Nutzer angebrachte Sensoren (Chua et al., 2003; Duchowski et al., 2004; Steptoe et al., 2009), was die Gefahr birgt, den Nutzer abzulenken oder zu behindern. Zudem führt eine detailliertere Abbildung der Nutzerbewegungen auf seinen Avatar zu der Frage, wie es möglich ist, Akte der Kommunikation, wie etwa Gesten, von Bewegungen zu Steuerung in der Welt, beispielsweise über Maus und Tastatur, zu trennen.

Unser Ansatz besteht darin, die Bewegungsabläufe des Nutzers direkt, kontinuierlich und vollständig auf seinen Avatar abzubilden. Wir untersuchen in diesem Artikel die Fragestellung, ob die Möglichkeit zur direkten Gestenübertragung durch die Nutzer verwendet wird und ob sich für die Nutzer ein Mehrwert nachweisen lässt. Unser konkreter Untersuchungsgegenstand in diesem Artikel sind die Konzepte der Präsenz und der Kopräsenz. Mit dem Begriff der Präsenz wird dabei nach Witmer und Singer (1998) ein mentaler Zustand beschrieben, in dem sich ein Individuum geistig an einen anderen Ort versetzt fühlt. Analog hierzu bezeichnet Kopräsenz die subjektive Wahrnehmung, mit jemand anderem zusammen zu sein, selbst wenn man physikalisch voneinander entfernt ist (Mennecke et al., 2010; Slater et al., 2000). Es konnte bereits gezeigt werden, dass ein rudimentärer, triggerbasierter Ansatz die Kopräsenz der Nutzer einer virtuellen Welt steigern kann (Casanueva et al., 2001) und auch für die wahrgenommene Präsenz ist dieser Zusammenhang naheliegend. Deshalb formulieren wir die Hypothese, dass die Einführung der Gestenübertragung die von den Nutzern wahrgenommene Präsenz und Kopräsenz erhöht. Für die Überprüfung dieser Hypothese wurde ein auf der Microsoft Kinect basierender Prototyp entwickelt. Dieser verwendet zusätzlich die Wii Remote, um dem Nutzer die Navigation durch und Interaktion mit der virtuellen Welt zu ermöglichen, ohne den Kommunikationskanal für Gesten zu beeinflussen.

Abschnitt 2 dieses Artikels beschreibt den Aufbau des von uns erstellten Prototyps. In Abschnitt 3 wird die zur Untersuchung der aufgeworfenen Fragestellungen durchgeführte Studie dargestellt. Abschnitt 4 diskutiert die gewonnenen Erkenntnisse.

## 2 Systembeschreibung

Als Grundlage für unseren Ansatz dient eine Server-Client Architektur, da diese auch bei vielen kollaborativen virtuellen Welten genutzt wird. Die Aufgabe des Servers liegt in der Simulation der 3D-Welt, während der Client die Welt darstellt und die Eingaben des Benutzers entgegennimmt. In unserem Ansatz werden zwei verschiedene Arten von Benutzereingaben unterschieden. Eine Eingabeart ist dabei die Körperhaltung des Nutzers, die mit der Kinect aufgenommen werden kann. Aus dieser wird ein sogenanntes Skelett extrahiert. Der Client überträgt diese Eingabe in ein Format, das auf Avatare angewendet werden kann und sendet dies über den Server an alle angeschlossenen Clients. Hier wird sie auf die entfernte Instanz des Avatars angewendet. Somit werden alle mit der Kinect erfassten Bewegungen eines Benutzers für die anderen Benutzer (als Bewegungen des Avatars) sichtbar. Da die Beine des Avatars weiterhin genutzt werden sollen, um eine „laufen“ Animation darzustellen, wenn der Nutzer durch die Welt navigiert, konzentriert sich unser Ansatz auf die Übertragung von Bewegungen des Oberkörpers. Hierdurch wird der größte Teil möglicher Gesten abgedeckt. Die zweite Eingabeform, die der Client verarbeitet, sind Signale zur Navigation und Interaktion. Die naheliegende Lösung, auch hierfür die Kinect Eingabe zu verwenden, gestaltet sich aufgrund der möglichen wechselseitigen Überlagerung von Gesten zur Kommunikation und Bewegungssteuerung für Navigation und Interaktion schwierig. Auch die sonst übliche Verwendung von Maus und Tastatur für diese Aufgabe, ist nicht möglich, da der Nutzer für die Erfassung durch die Kinect frei im Raum stehen muss. Deshalb übernimmt in unserem Ansatz eine Wii Remote diese Aufgabe.

### 2.1 Technische Umsetzung

Der Prototyp verwendet den Second Life Viewer als Client und OpenSimulator als Server. Beide Systeme wurden für den Prototypen erweitert. Abbildung 1 zeigt die Architektur des Systems und die wesentlichen Informationsflüsse.

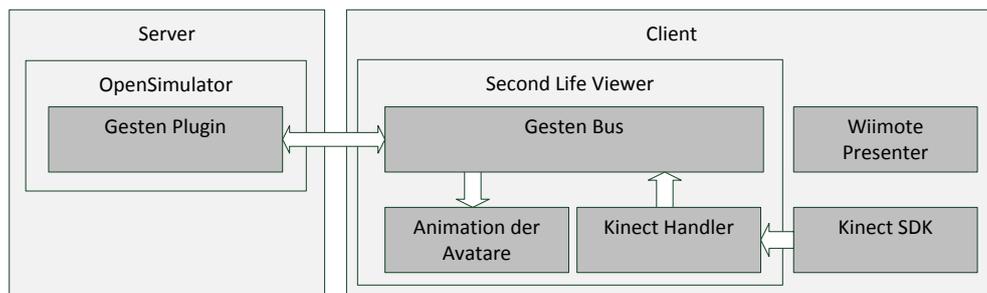


Abbildung 1: Architektur und Informationsfluss des Systems

Die Bereitstellung des Kommunikationskanals für Gesten zerfällt in zwei Aufgaben. Zunächst müssen die Daten der Kinect ausgelesen und umgewandelt werden. Danach müssen diese Daten auf die entfernten Kopien des betreffenden Avatars, welche die anderen Nutzer mit ihren Clients sehen, angewendet werden. Das Lesen und Umwandeln der Daten erfolgt

im Client durch den Kinect Handler. Er ist in die Hauptschleife des Viewers integriert und liest durch das Kinect SDK generierte Kinect-Skelette (KS) aus.

Das KS unterscheidet sich in einigen Punkten von dem für den Viewer verwendbaren Viewer-Skelett (VS). Zunächst sind die Koordinatensysteme, die beiden Skeletten zugrunde liegen, nicht identisch. Dies kann durch eine einfache lineare Abbildung korrigiert werden.

Die Funktion  $f: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -z \\ -x \\ y \end{pmatrix}$  beschreibt diese Abbildung.

Zudem werden im KS lediglich die Positionen der Gelenke beschrieben. Um einen Avatar in die Pose eines Skelettes bringen zu können, ist allerdings eine Beschreibung des Skeletts über die Drehungen der Gelenke in Form von Quaternionen notwendig. Um diese Beschreibung zu erhalten, werden pro Gelenk die beiden zugehörigen Knochen eingeschlossene Winkel und die zugehörige Drehachse bestimmt. Seien  $\overrightarrow{AB}$  und  $\overrightarrow{BC}$  die beiden Knochen die das Gelenk B bilden. Dann sind  $\alpha = \arccos\left(\frac{|\overrightarrow{AB} \cdot \overrightarrow{BC}|}{|\overrightarrow{AB}| \cdot |\overrightarrow{BC}|}\right)$  und  $\vec{V} = \overrightarrow{AB} \times \overrightarrow{BC}$  der Winkel und die Drehachse, aus denen sich ein entsprechendes Quaternion konstruieren lässt. Abbildung 2 illustriert diese Umwandlung an einem Beispiel.

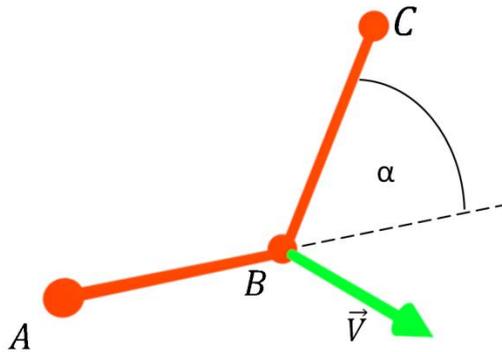


Abbildung 2: Beispiel für die Bestimmung eines Rotationsquaternions.

Wenn keinerlei Animation abgespielt wird, befindet sich ein Avatar in einer Grundstellung. Diese Grundstellung muss bei der Bestimmung der Drehungen berücksichtigt werden. Hierzu lassen sich aus einem Skelett in dieser Grundstellung Korrekturquaternionen bestimmen, die als Konstanten gespeichert werden können. Diese Konstanten lassen sich dann auf die ermittelten Gelenkdrehungen anwenden. Schließlich ist zu beachten, dass die Struktur von KS und VS nicht identisch sind. Aus diesem Grund wurde ein Mapping definiert, das bestimmt, welche Gelenke des KS in welche Gelenke des VS umgewandelt werden.

Alle Avatare verfügen über eine eindeutige ID. Die ID des zum aktuellen Benutzer gehörenden Avatars wird mit dem ermittelten VS zusammen an den Gesten-Bus übergeben. Der Bus sendet diese Daten mit Hilfe eines entsprechend definierten Nachrichtentyps an den Server. Im Server läuft ein Plug-In, welches die Skelett-Nachrichten an alle Clients

weiterleitet. Der Gesten-Bus der anderen Clients nimmt die Daten entgegen und speichert sie unter der ID des Avatars. Für jeden Avatar wurde bei dessen Erstellung eine Animation angelegt, welche über den Gesten-Bus angefragt werden kann. Das Animationsframework des Viewers wird schließlich verwendet, um Avatare in die beschriebene Pose zu bringen.

Um die Wii Remote für Navigation und Interaktion nutzen zu können, wird der sogenannte Wiimotepresenter (Smith and Kitzmann, 2008) verwendet. Die Software verfolgt das in (Yang and Li, 2011) beschriebene Konzept, um den Controller zur Emulation von Maus und Tastatur zu verwenden. Für den Viewer ist nicht zu erkennen, dass die Wiimote als Quelle der Eingaben dient. Die Hardware des Controllers ermöglicht es, die ungefähre Position auf der Leinwand zu ermitteln, auf die das Gerät zeigt. Diese Information nutzt der Präsenter, um den Mauszeiger an diese Stelle zu bewegen. Das erzeugte Verhalten ähnelt dem eines Laserpointers. Mit der „B“-Taste des Controllers kann der Nutzer einen Linksklick auslösen. Klickt er dabei auf ein 3D-Objekt im Viewer, verarbeitet dieser dies als eine „Berührung“ des Objektes. Dies reicht als grundlegende Interaktionsmöglichkeit aus. Zur Navigation in der 3D-Welt werden die Tasten des Steuerkreuzes der Wii Remote auf die Pfeiltasten der Tastatur abgebildet. Dies ermöglicht es dem Nutzer, den Avatar nach rechts und links zu drehen sowie ihn vorwärts und rückwärts zu bewegen.

### 3 Evaluation

Die hier vorgestellte Studie soll zwei Fragen beantworten. Zum einen ist dies die Frage, ob der neuartige Kommunikationskanal für Gesten durch die Nutzer verwendet wird. Zum anderen soll überprüft werden, ob die Benutzer durch das System eine größere Präsenz und Kopräsenz wahrnehmen. Die Studie wurde mit 57 Teilnehmern durchgeführt. Unter ihnen waren 18 Frauen und 39 Männer.

#### 3.2 Studiendesign

Für die Studie wurden die Teilnehmer in 19 Gruppen zu jeweils drei Personen zusammengefasst. Hierbei wurde auf eine gleichmäßige Verteilung der Geschlechter geachtet. Die Gruppen durchliefen nacheinander zwei Szenarien. Szenario 1 durchliefen 10 Gruppen mit aktivierter Gestenübertragung, bei 9 Gruppen war die Gestenübertragung nicht aktiv. In diesem Szenario wurden die wahrgenommene Präsenz und Kopräsenz der Teilnehmer vergleichend zwischen diesen beiden Bedingungen untersucht. Um die Steuerung mit der Wiimote als Störgröße auszuschließen, wurde diese in beiden Bedingungen deaktiviert (und war für die Aufgabe der Gruppen auch nicht nötig). In Szenario 2 waren die Gestenübertragung und die Wii Remote-Steuerung für alle Teilnehmer aktiv. Der Fokus dieses Szenarios lag auf der Eignung der Steuerung mit Hilfe der Wii Remote in Zusammenspiel mit dem Gestenkommunikationskanal. Es wurden neben der Usability des Gesamtsystems auch gezielte Fragen zu den beiden Eingabegeräten untersucht. Tabelle 3 gibt einen Überblick über das gewählte Studiendesign.

|            | 19 Gruppen                                 |   |
|------------|--|---|
|            | 10 Gruppen                                 | 9 Gruppen                                   |
| Szenario 1 | mit Gestenübertragung,<br>keine Wii Remote | ohne Gestenübertragung,<br>keine Wii Remote |
| Szenario 2 | mit Gestenübertragung, mit Wii Remote      |   |

Tabelle 3: Überblick über das Studiendesign

### 3.3 Aufbau

Die Teilnehmer einer Gruppe wurden für beide Szenarien in je einem von drei separaten Räumen untergebracht. Es wurde darauf geachtet, dass sie sich vor der Studie nicht begegneten. In den Räumen befand sich jeweils ein Laptop. An diesen war ein Beamer zur Ausgabe des Bildschirminhaltes auf einer Leinwand angeschlossen. Zudem waren eine Kinect, eine Wii Remote und ein Headset mit dem Gerät verbunden. Vor Beginn des Szenarios 1 wurden die Teilnehmer gebeten, sich an einen Punkt etwa 2,5 Meter vor der Leinwand aufzustellen und dort während der Szenarien zu verweilen. Die Kinect wurde unter der Leinwand positioniert und auf den Nutzer ausgerichtet. Zudem wurden die Teilnehmer gebeten, das Headset aufzusetzen. Mit dem Headset waren sie durch die Voice-over-IP Software Teamspeak miteinander verbunden. Zum Start der Szenarien wurde auf dem Laptop der Viewer gestartet und maximiert, so dass dieser die gesamte Leinwand einnahm. Die Teilnehmer erhielten eine Einweisung in die Gestenübertragung (falls diese aktiviert war) und zudem ein Handout, welches die jeweilige Aufgabe erläuterte.

Der Studienleiter befand sich während der Szenarien in einem vierten Raum und war von dort ebenfalls in die virtuelle Welt eingeloggt, hielt sich aber für den Verlauf der Szenarien im Hintergrund und beschränkte sich darauf, den Verlauf zu beobachten. Von seinem Client aus wurde der Studienverlauf in der Welt inklusive der Audiokommunikation aufgezeichnet.

### 3.4 Szenario 1

Für dieses Szenario befanden sich die drei Avatare in einem gemeinsamen virtuellen Raum. Sie standen in einem Kreis und die Kamera der Viewer war so ausgerichtet, dass jeder Teilnehmer die Avatare der beiden anderen Teilnehmer komplett sehen konnte. Die Navigation und Interaktion mit der 3D-Welt waren für alle Teilnehmer deaktiviert und die Wii Remote war versteckt.

Die Aufgabe der Teilnehmer bestand darin, eine Projektauswahlentscheidung innerhalb eines kommunalpolitischen Szenarios zu treffen. Jeder der Teilnehmer verfügte über Informationen zu einer von drei möglichen Projektideen, die er den anderen vorstellen und die er innerhalb der folgenden Diskussion vertreten sollte. Obwohl jeder Teilnehmer die Anweisung hatte, seine Projektidee möglichst durchzusetzen, wurde allen Teilnehmern auch mitgeteilt, dass keine Entscheidung zu treffen der denkbar schlechteste Ausgang des Szenarios sei. Sie konnten also auch eine der anderen Projektideen wählen, falls sie für ihre eigene Idee keine Mehrheit finden konnten. Die Bearbeitungszeit für das Vorstellen und die Diskussion der Projektideen betrug 30 Minuten. Nach 25 Minuten wurden die Teilnehmer darauf

hingewiesen, dass die Bearbeitungszeit bald endet. War nach 35 Minuten noch keine Entscheidung getroffen, wurde das Szenario beendet.

Nach dem Szenario füllten die Teilnehmer einen Fragebogen aus. Dieser enthielt 22 Aussagen, welche die Teilnehmer auf eine Likert-Skala mit sieben Abstufungen bewerten sollten. Hierin waren 10 Items des IGroup Presence Questionnaire aus (Schubert et al., 2001) enthalten. Diese wurden anschließend im Präsenzscore  $V_{IPQ}$  zusammengefasst. Dieser bewegt sich im Bereich von 0 und 60, wobei ein hoher Wert einer hohen Präsenz entspricht. Zudem enthielt der Fragebogen die 6 Aussagen des Fragebogens aus (Casanueva et al., 2001). Diese Fragen wurden ins Deutsche übersetzt und zu einem Kopräsenzscore  $V_{KP}$  zusammengefasst. Dieser bewegte sich im Bereich von 0 bis 36, wobei ein höherer Wert einer höheren Kopräsenz entspricht. Zudem enthielt der Fragebogen weitere Elemente bezüglich der Diskussion und der Sympathie, welche die Teilnehmer sich entgegenbrachten.

### 3.5 Szenario 2

In Szenario 2 waren Navigation und Interaktion über die Wii Remote aktiviert. Der eigentlichen Aufgabe wurde eine Eingewöhnung vorangestellt, in welcher die Teilnehmer in separaten Räumen der 3D-Welt mit der Steuerung experimentieren konnten. Nachdem alle Teilnehmer bestätigten, dass sie sich an die Steuerung gewöhnt hatten, wurden sie aufgefordert sich in einen vierten virtuellen Raum zu begeben.

In diesem Raum befand sich die eigentliche Aufgabe des Szenarios. Die Teilnehmer sollten ein kooperatives, dreidimensionales Memory-Spiel lösen. Das Spiel bestand aus 48 Karten auf die sich 24 Motive - jeweils in Paaren - verteilten. Die Karten waren zufällig auf die drei Seiten einer Pyramide aufgeteilt. Diese war so groß, dass jeder Teilnehmer nur eine Seite gleichzeitig erkennen konnte. Es wurde sichergestellt, dass sich Paare nie auf derselben Seite der Pyramide befanden. Die Karten waren zu Beginn alle verdeckt. Wenn ein Teilnehmer eine der Karten mit Hilfe der Wii Remote „berührte“, wurde diese aufgedeckt. Danach konnte ein anderer Spieler eine weitere Karte aufdecken. Waren die Motive gleich, wurden die Karten entfernt. Das Ziel war es, alle Karten zu entfernen. Da ein einzelner Spieler immer nur eine Karte aufdecken und damit kein Paare entfernen konnte, war hierfür Zusammenarbeit notwendig. Die maximale Bearbeitungszeit betrug 20 Minuten, danach wurde das Spiel abgebrochen. Abbildung 3 zeigt einen Screenshot des Szenarios.

Nach dem Ende des Szenarios beantworteten alle Teilnehmer einen zweiten Fragebogen. Dieser enthielt 17 Aussagen, welche anhand eine Likert-Skala mit fünf Abstufungen bewertet werden sollten. Die Ankerpunkte der Skala waren „trifft gar nicht zu“ und „trifft völlig zu“. Der Fragebogen enthielt die 10 Elemente des System Usability Scale (SUS) Fragebogens (Brooke, 1996). Hieraus wurde der SUS Score  $V_{SUS}$  entsprechend der Vorschriften von Brooke berechnet. Zudem enthielt der Fragebogen Elemente zur Qualität der Gestenübertragung, zu Steuerung mit der Wii Remote und zur allgemeinen Bewertung der Aufgabe.



Abbildung 3: Szenario 2 (aus Vogelperspektive – die Teilnehmer hatten je nur Sicht auf eine Pyramidenseite)

### 3.6 Auswertung

Im Fragebogen zu Szenario 1 lag der Durchschnitt von  $V_{IPQ}$  für Teilnehmer mit Gestenübertragung mit 21,73 (SD=9,87) über dem der Teilnehmer ohne Gestenübertragung mit 16,26 (SD=11,41). Diese Ergebnisse weisen tendenziell in die erwartete Richtung, sind allerdings statistisch (knapp) nicht signifikant ( $p=0,0596$ ). Der Durchschnitt der Variable  $V_{KP}$  liegt für Teilnehmer mit Gestenübertragung bei 26,87 (SD=4,97) und damit nur minimal über dem der Teilnehmer ohne Gestenübertragung mit 26,04 (SD=5,20). Dieser Unterschied ist nicht signifikant ( $p=0,5418$ ). Im Fragebogen zu Szenario 2 lag der durchschnittliche durch die Teilnehmer vergebene SUS-Score bei 81,58. Dieser Wert ordnet sich in der Skala von Bangor (Bangor et al., 2009) zwischen „good“ (71,4) und „excellent“ (85,5) ein. Die Aussage „Mit der Steuerung durch den Controller bin ich gut zurechtgekommen.“ wurde im Durchschnitt mit 4,47 (SD=0,73) von 5 bewertet und damit deutlich bestätigt. Dies spricht dafür, dass die Steuerung mit der Wiimote durchaus für Navigation und Interaktion in virtuellen Welten geeignet ist. Die Aussage „Die Genauigkeit, mit der der Bewegungssensor meine Bewegungen übertragen hat, fand ich ausreichend“ wurde im Durchschnitt mit 3,51 (SD=1,10) von möglichen 5 Punkten bewertet. Um die Nutzung von Gesten innerhalb beider Szenarien bewerten zu können, wurden die Aufzeichnungen des Studienverlaufes codiert und das Auftreten verschiedene Arten von Gesten untersucht. Es wurde dabei zwischen rhythmische Gesten zur Strukturierung und Betonung des Gesprochenen, Gesten zur Erläuterung des Gesprochenen (wie z.B. das Beschreiben einer Wendeltreppe durch eine spiralförmige Bewegung der Hand) und Zeigegesten unterschieden. Da diese Differenzierung keine weiterreichenden Erkenntnisse liefern konnte, soll hier nur auf die Gesamtzahl der Gesten eingegangen werden. Für Szenario 1 wurden nur Gruppen mit aktiver Gestenübertragung betrachtet, für Szenario 2 wurden alle Gruppen untersucht. Es wurde jeweils die Anzahl an Gesten

aller drei Teilnehmer einer Gruppe summiert. Um die unterschiedlichen Bearbeitungsdauern zu berücksichtigen, wurden die Werte mit dieser Dauer in Bezug gesetzt. Die Angaben sind dementsprechend in Gesten pro Stunde. Der Durchschnitt dieses Wertes ist mit 123,9 (SD=66,7) für Szenario 1 signifikant ( $p=0,0003$ ) größer als bei Szenario 2 mit 8,2 (SD=9,5).

## 4 Diskussion und Fazit

In diesem Artikel wurde ein neuartiger, auf der Kinect basierender Kommunikationskanal für Gesten in virtuellen 3D-Welten vorgestellt. Navigation und Interaktion mit der 3D-Welt erfolgen in diesem Ansatz getrennt hiervon mit Hilfe der Wii Remote. Die durchgeführte Studie hat gezeigt, dass Nutzer die Usability des Ansatzes sehr positiv bewerten. Auch einzeln werden Steuerung und Gestenübertragung positiv bewertet. Insbesondere erscheint die stark modifizierte Steuerung mit Hilfe der Wii Remote für 3D-Welten geeignet. Die ebenfalls mit dieser Studie untersuchte Hypothese, dass der Gestenkommunikationskanal die von den Benutzern wahrgenommene Präsenz und Kopräsenz steigert, konnte nicht bestätigt werden. Für die Präsenz liegt dies vermutlich an der doch recht geringen Stichprobengröße, da die vorliegenden Daten eine starke Tendenz in die vermutete Richtung aufweisen. Das Fehlen einer solchen Tendenz für die Kopräsenz der Teilnehmer steht im Widerspruch zu den Ergebnissen von Casanueva (Casanueva et al., 2001). Möglicherweise ist der eingesetzte Fragebogen nicht sensibel genug, um eine eventuell dennoch vorhanden Unterschied zu detektieren. Zudem ist es möglich, dass die insgesamt gesteigerte Qualität der Darstellung der Avatare seit den Untersuchungen von Casanueva ein bereits sehr hohes Maß an Kopräsenz erzeugt und so den Unterschied, den die Gestenübertragung erzeugt, verdeckt.

Während Gesten in Szenario 1 durchaus rege genutzt wurden, gab es fast keinen Gesten-Gebrauch in Szenario 2. Eine Beeinflussung durch die nur in Szenario 2 aktive Wiimote erscheint aufgrund der hohen Usability und der guten direkten Bewertung der Steuerung unwahrscheinlich. Auch die Bewertung des Items „Der Controller in meiner Hand hat mich nicht daran gehindert, Gesten auszuführen.“ mit im Durchschnitt 3,18 (SD=1,31) von 5 spricht eher gegen eine solche Beeinflussung. Ähnlich wie bei unmittelbarer Kommunikation scheint es, dass die Nutzung des Gestenkanals stark von der jeweiligen Aufgabe abhängig ist. Dies wird auch durch die Bewertungen der Aussage „Ich glaube, die Gestenübertragung hat uns bei unserer Aufgabe geholfen“ aus dem Fragebogen zum zweiten Szenario mit im Durchschnitt 1,73 (SD=0,82) von 5 gestützt. Zukünftige Untersuchungen im Bereich Gestensteuerung in 3D-Welten sollten sich daher stärker auf das Zusammenspiel zwischen Aufgabe und Gestennutzung fokussieren.

### Literaturverzeichnis

- Bier, E.A. & Stone, M.C. Snap-Dragging. In *13th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH 1986* (18.-22. Aug., Dallas, TX). ACM Press, NY, 1986. pp. 233-240.
- Bangor, A., Kortum, P., Miller, J., 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Stud.* 4, 114–123.
- Barrientos, F., 2000. Continuous control of avatar gesture, in: *Proceedings of the 2000 ACM Workshops on Multimedia, MULTIMEDIA '00*. ACM, New York and NY and USA, pp. 5–8.

- Barrientos, F.A., Canny, J.F., 2002. Cursive: controlling expressive avatar gesture using pen gesture, in: Proceedings of the 4th International Conference on Collaborative Virtual Environments, CVE '02. ACM, New York and NY and USA, pp. 113–119.
- Bell, M., 2008. “Virtual Worlds Research: Past, Present & Future” July 2008: Toward a Definition of “Virtual Worlds”. In: Journal of Virtual Worlds Research.
- Brooke, J., 1996. SUS-A quick and dirty usability scale. Usability Eval. Ind. 189, 194.
- Casanueva, J., Blake, E., others, 2001. The effects of avatars on co-presence in a collaborative virtual environment, in: Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT2001). Pretoria, South Africa.
- Chua, P.T., Crivella, R., Daly, B., Hu, N., Schaaf, R., Ventura, D., Camill, T., Hodgins, J., Pausch, R., 2003. Training for Physical Tasks in Virtual Environments: Tai Chi, in: Proceedings of the IEEE Virtual Reality 2003, VR '03. IEEE Computer Society, Washington and DC and USA, p. 87.
- Duchowski, A.T., Cournia, N., Cumming, B., McCallum, D., Gramopadhye, A., Greenstein, J., Sadasivan, S., Tyrrell, R.A., 2004. Visual deictic reference in a collaborative virtual environment, in: Proceedings of the 2004 Symposium on Eye Tracking Research & Applications, ETRA '04. ACM, New York and NY and USA, pp. 35–40.
- Guye-Vuillème, A., Capin, T., Pandzic, S., Thalmann, N.M., Thalmann, D., 1999. Nonverbal communication interface for collaborative virtual environments. Virtual Real. J. 4, 49–59.
- Mennecke, B.E., Triplett, J.L., Hassall, L.M., Conde, Z.J., 2010. Embodied Social Presence Theory, in: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, HICSS '10. IEEE Computer Society, Washington and DC and USA, pp. 1–10.
- Salem, B., Earle, N., 2000. Designing a non-verbal language for expressive avatars, in: Proceedings of the Third International Conference on Collaborative Virtual Environments, CVE '00. ACM, New York and NY and USA, pp. 93–101.
- Schubert, T., Friedmann, F., Regenbrecht, H., 2001. The Experience of Presence: Factor Analytic Insights. Presence Teleoper Virtual Env. 10, 266–281.
- Slater, M., Sadagic, A., Usoh, M., Schroeder, R., 2000. Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. Presence Teleoperators Virtual Environ. 9, 37–51.
- Smith, J., Kitzmann, I., 2008. Wiimotepresenter. Online verfügbar unter: <https://sites.google.com/site/jasonlpsmith/wiimotepresenter>
- Sreedharan, S., Zurita, E.S., Plimmer, B., 2007. 3D input for 3D worlds, in: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI '07. ACM, New York and NY and USA, pp. 227–230.
- Stephoe, W., Oyekoya, O., Murgia, A., Wolff, R., Rae, J., Guimaraes, E., Roberts, D., Steed, A., 2009. Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty Interaction in Immersive Collaborative Virtual Environments, in: Proceedings of the 2009 IEEE Virtual Reality Conference, VR '09. IEEE Computer Society, Washington and DC and USA, pp. 83–90.
- Thai, P., 2011. Using Kinect and OpenNI to Embody an Avatar in Second Life. Online verfügbar unter: <http://www.youtube.com/watch?v=ehTvtkybubM>
- Witmer, B.G., Singer, M.J., 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence Teleoperators Virtual Environ. 7, 225–240.
- Yang, Y., Li, L., 2011. Turn a Nintendo Wiimote into a Handheld Computer Mouse. IEEE Potentials 30, 12–16.