

Evaluierung der sozialen Akzeptanz verschiedener Interaktionsarten für Augmented-Reality-Datenbrillen

Nils Adrian Mack
Fachgebiet Mensch-Maschine-Systemtechnik
Universität Kassel
Kassel Germany
N.Mack@uni-kassel.de

Ludger Schmidt
Fachgebiet Mensch-Maschine-Systemtechnik
Universität Kassel
Kassel Germany
L.Schmidt@uni-kassel.de

ZUSAMMENFASSUNG

Die soziale Akzeptanz ist neben der praktischen Akzeptanz ein wichtiger Bestandteil der Akzeptanz eines Systems durch die Nutzer. Es ist möglich, dass trotz hoher praktischer Akzeptanz ein System nicht genutzt wird, da dieses sozial nicht akzeptabel ist. Für Augmented-Reality-Datenbrillen (AR-Datenbrillen) wurden schon verschiedene Faktoren determiniert, welche die soziale Akzeptanz beeinflussen können, um zu vermeiden, dass diese von den Nutzern abgelehnt werden. Ein wichtiger Faktor, welcher die soziale Akzeptanz einer Datenbrille beeinflussen kann, ist die Interaktionsart. Es ist anzunehmen, dass die Verwendung unterschiedlicher Interaktionsarten in dem gleichen sozialen Kontext mit dem gleichen Gesamtsystem nicht in der gleichen sozialen Akzeptanz resultiert. Nachfolgend wird eine Studie mit 10 Probanden und 6 aktuell auf AR-Datenbrillen verwendeten Interaktionsarten durchgeführt, um deren soziale Akzeptanz in verschiedenen Orts- und Zuschauerkontexten vergleichend zu evaluieren. Die Studie zeigt, dass die Interaktionsarten, insbesondere diverse Arten der Sprachinteraktion, mit steigender Distanz der sozialen Beziehung zu dem Zuschauer in der Bewertung degradieren und, dass der Ort des voraussichtlichen Einsatzes der Interaktionsart berücksichtigt werden muss.

SCHLÜSSELWÖRTER

Augmented Reality, Interaktionsarten, Datenbrillen, Soziale Akzeptanz

1 Hintergrund und Motivation

Prognosen von Experten zufolge werden AR-Datenbrillen in naher Zukunft die gleiche Stellung wie Smartphones im alltäglichen Leben einnehmen [1]. Als alltäglicher Begleiter werden die AR-Datenbrillen in den verschiedensten Szenarien Anwendung finden, z. B. beim Sightseeing oder bei der

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MuC'19 Workshops, Hamburg, Deutschland

© Proceedings of the Mensch und Computer 2019 Workshop on Virtual and Augmented Reality in Everyday Context (VARECo).

Copyright held by the owner/author(s).

<http://doi.org/10.18420/muc2019-ws-575>

Fußgängernavigation [2]. Unter Annahme der breiten Verfügbarkeit ist neben den technischen Gesichtspunkten auch die soziale Akzeptanz der AR-Datenbrillen zu hinterfragen. Der Fall der Google Glass [3] und auch Studien zur sozialen Akzeptanz von AR-Datenbrillen deckten insbesondere das Problem der verdeckten Videoaufnahme auf [4].

Neben der praktischen Akzeptanz, in die u. A. die Gebrauchstauglichkeit einfließt, bestimmt nach Nielsen [5] die soziale Akzeptanz die Gesamtsystemakzeptanz. Diese ist vor allem dadurch gekennzeichnet, dass alle Bedürfnisse des Nutzers und weiterer Personen, welche von dem System betroffen sind, erfüllt werden.

Aufbauend auf Brewster et al. [6] definieren Montero et al. [7] zwei Teilaspekte der sozialen Akzeptanz: Der erste Teilaspekt ist die soziale Akzeptanz des Nutzers, welche den Eindruck oder das Gefühl beschreibt, welches die Aktion bei dem Nutzer hinterlassen hat. Unter Anderem wird dies dadurch beeinflusst, wie natürlich, peinlich oder komisch die Aktion war. Der zweite Teilaspekt der sozialen Akzeptanz ist die soziale Akzeptanz des Zuschauers. Von Bedeutung ist, welchen Eindruck die Aktion des Nutzers bei einem Zuschauer hinterlassen hat. Beeinflusst wird dies u. A. durch das Verständnis der Aktionen des Nutzers bzw., ob dessen Verhalten aus Beobachterperspektive normal oder komisch war.

Neben verdeckten Videoaufnahmen beeinflusst die verwendete Interaktionsart die soziale Akzeptanz von Datenbrillen, insbesondere wenn diese im öffentlichen Raum eingesetzt wird [8; 4].

Um Faktoren zu identifizieren, welche die soziale Akzeptanz einer Interaktionsart beeinflussen, kann die Interaktion von Personen untereinander als Anhaltspunkt benutzt werden. Nach Goffman [9] ist jede öffentlich stattfindende Aktion, die eine Person ausführt, eine Aufführung zur Selbstdarstellung. Dabei gibt es unterschiedliche Anstandsregeln für verschiedene Orte, an welche sich die Person bei ihrer Aufführung halten muss. So gelten z. B. in einer Kirche andere Regeln als am Arbeitsplatz.

Eine weitere Feststellung von Goffman [9] ist, dass die Person sich je nach Zuschauer anders verhält. So segregieren Personen die Zuschauer und entscheiden, vor welchen der Zuschauer sie bestimmte Aktionen ausführen. Es ist annehmbar, dass sowohl Zuschauer als auch der Ort der Interaktion wichtige Faktoren für die soziale Akzeptanz von Interaktionsarten sein werden.

Häufig wird versucht, die Akzeptanz eines Systems mit dem Technologieakzeptanzmodell (TAM) nach Davis [10] zu erfassen. Dieses berücksichtigt aber nur die empfundene Einfachheit der Benutzung sowie die empfundene Nützlichkeit. Soziale Einflüsse fehlen im TAM und Versuche diese zu berücksichtigen [11], sind nicht in der Lage, Faktoren zu identifizieren, welche die soziale Akzeptanz einer Technologie beeinflussen können [12].

Ein weiteres Modell, welches häufig verwendet wird, um die Akzeptanz eines Systems zu erfassen ist das Modell der Nutzerakzeptanz von Informationstechnologien (UTAUT), welches von Venkatsh et al. [13] entwickelt wurde, um die Schwächen anderer Akzeptanzmodelle zu adressieren. Unter anderem wird der Faktor des sozialen Einflusses betrachtet. Dieser entspricht aufgrund der Fragestellungen eher dem sozialen Druck zur Benutzung einer Technologie, insbesondere im Rahmen des Arbeitsplatzes. Aus diesem Grund wurde hier auf die Verwendung dieses Modells verzichtet.

Im Rahmen der nachfolgenden Studie werden verschiedene Interaktionsarten verglichen, welche aktuell auf AR-Datenbrillen eingesetzt werden. Unter anderem soll eine Vergleichbarkeit derer sozialer Akzeptanz in verschiedenen Orts- und Zuschauerkontexten ermöglicht werden. Dies scheint insbesondere für aktuelle Studien sinnvoll, welche außerhalb von Laborbedingungen stattfinden. In solchen Situationen sollte nicht nur die praktische Akzeptanz der Interaktionsart betrachtet werden, da sonst die Gesamtakzeptanz des Systems geringer ausfallen kann.

2 Stand der Forschung

Akzeptanz von Interaktionsarten auf AR-Datenbrillen:

Eine der wenigen Studien, welche sich mit der Akzeptanz von Interaktionsarten für AR-Datenbrillen im öffentlichen Raum beschäftigt, wurde von Tung et al. [8] für den Ortskontext eines Cafés durchgeführt. Die Nutzer sollten dabei die Interaktion aus den Kategorien Handheld (z. B. Smartphones), Touch (Berührung auf Körperteilen oder Wearables) und Non-Touch (non-Handheld, z. B. Sprache und Freihandgesten) demonstrieren, welche sie mit der gezeigten Manipulation auf der AR-Datenbrille verbinden. Es konnte gezeigt werden, dass die soziale Akzeptanz von Interaktionsarten in öffentlichen Räumen eine wichtige Rolle spielt. Auch wurde festgestellt, dass Gesten auf Gesichtshöhe generell gemischt und nicht sozial akzeptabel sind.

Akzeptanz von Gesten und Gestensteuerung: Rico & Brewster [12] untersuchten die soziale Akzeptanz mehrerer geräte- und körpergebundener Gesten auf deren soziale Akzeptanz in verschiedenen Orts- und Zuschauerkontexten in einer Labor- sowie ergänzenden Feldstudie. Es konnte gezeigt werden, dass sowohl der Ort als auch die anwesenden Zuschauer bei der Bereitschaft, eine bestimmte Geste zu benutzen, eine Rolle spielen. Je privater der Kontext war, desto wahrscheinlicher ist die Bereitschaft durch Gesten zu interagieren. Diese Abhängigkeit der sozialen Akzeptanz von dem Orts- und Zuschauerkontext konnte auch für die Gesteninteraktion im Umfeld von handgehaltenen mobilen Geräten [14] und in der realen Welt [15] gezeigt werden.

Einen interessanten Einfluss auf die soziale Akzeptanz des Zuschauers, welche Taniberg et al. [16] im Rahmen der Gestensteuerung eines Hi-Fi-Systems aufdecken konnten, ist, dass der Zuschauer die Gestensteuerung schlechter bewertet, wenn deren Ausführung in seine Richtung erfolgt.

Montero et al. [7] konnten zeigen, dass die soziale Akzeptanz des Zuschauers einer Geste davon abhängt, ob dieser deren Effekt bzw. Wirkung versteht. Dies bedeutet, dass spannende Gesten, also solche, welche die Aktion, aber nicht den Effekt offenbaren, sozial nicht akzeptiert werden. Dem entgegengestellt sind die magischen Gesten, bei welchen dem Zuschauer die Aktion verborgen bleibt, aber der Effekt wahrgenommen wird. Diese Art von Gesten hatte eine gute soziale Akzeptanz. Montero et al. [7] vermuteten, dass solche Gesten akzeptabler sind, da der Effekt keinem Verursacher zugeordnet werden kann. Auch geheime Gesten, bei welchen weder Aktion noch Effekt für den Zuschauer sichtbar sind, wurden gut bewertet. Dies stimmt mit den Ergebnissen von Tung et al. [8] überein, welche nachweisen konnten, dass Nutzer subtilere Interaktionsarten bevorzugen. Sowohl die magischen als auch die geheimen Gesten haben eine unscheinbare Aktion durch den Nutzer als Ausgangspunkt. Hierzu ergänzend konnte Ahlström et al. [14] feststellen, dass die Nutzer schnelle, kleine Gesten, gegenüber langen, großen bevorzugen.

Akzeptanz von Spracheingabe: Aus dem Bereich der digitalen Assistenten auf Smartphones ist bekannt, dass Nutzer Hemmungen haben, Spracheingaben vor fremden Personen zu tätigen, auch wenn die Spracheingabe nützlicher bzw. einfacher ist. Diese Hemmungen sind geringer, wenn der Zuschauerkontext Freunde oder Familie sind oder fallen fast komplett weg, wenn der Nutzer alleine ist. Des Weiteren wurde festgestellt, dass das Alter einen tendenziellen Einfluss auf die Bereitschaft zur Nutzung der Spracheingabe hat [17].

Dieses Vermeiden von Spracheingaben in der Öffentlichkeit konnte auch bei einer Anforderungsanalyse für eine AR-Reiseassistentenanwendung für den öffentlichen Verkehr von Klose et al. [18] festgestellt werden. Immerhin 68 % der Befragten wollten keinen oder nur einen geringen Anteil an Spracheingaben in der Anwendung.

Für Smartwatches konnte gezeigt werden, dass die soziale Akzeptanz der Spracheingabe vom Orts- und Zuschauerkontext abhängig ist. So wird eine Texteingabe als Fahrgast oder am Arbeitsplatz bzw. vor Kollegen und Fremden bevorzugt. Auffällig ist auch die bessere Bewertung der Spracheingabe vor dem Partner, der Familie und Freunden. [19]

3 Versuchsaufbau

Der Versuchsaufbau ist an ähnliche Studien zur Evaluierung der sozialen Akzeptanz angelehnt [12; 20]. Jedem Probanden wurde in einer Laborumgebung in zufälliger Reihenfolge sechs Videos einer Interaktion mit unterschiedlichen Interaktionsarten von ca. 1 Minute mit einer Meta-2-Datenbrille gezeigt. In den Videos ist immer der gleiche Darsteller vor dem gleichen neutralen Hintergrund aus der gleichen Perspektive zu sehen. Damit die Interaktionen eine Bedeutung haben, folgt der

Darsteller bei jeder Interaktionsart einem Skript, welches neben Bedeutung der Interaktion auch den ungefähren Bereich der zu nutzenden Schaltfläche im Datenbrillensichtfeld vorgibt (vgl. Abbildung 1).

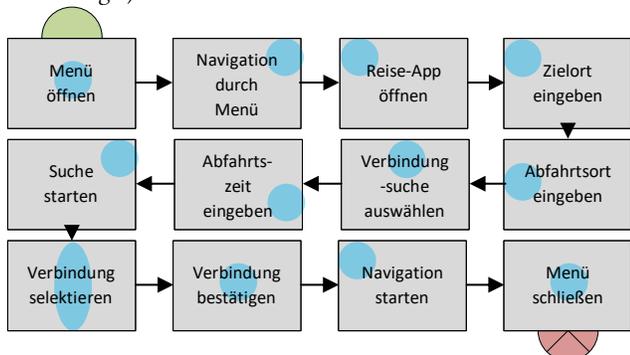


Abbildung 1: Modellierter Ablauf der Interaktion mit der Reiseassistenzanwendung in der Videoaufzeichnung mit den Bereichen, an welchen sich die Interaktionselemente befinden (in Blau)

Die Interaktionsarten wurden aufgrund deren Verwendung auf aktuellen (HoloLens, Meta 2) bzw. bekannter zukünftiger AR-Datenbrillen (HoloLens 2) ausgewählt.

Gaze und Sprache: Adaptiert wurde diese Gaze-and-Commit-Interaktionsart von der HoloLens [21]. Durch Bewegung des Kopfes bewegt der Nutzer einen Cursor, um ein Element zu selektieren [22]. Zum Anklicken des Elementes gibt der Proband ein einzelnes Sprachkommando, wie z. B.: „Auswählen“ oder „Start“. Für die komplexeren Eingaben des Zielortes und der Abfahrtszeit wurden die Kommandos durch die freien Spracheingaben, „Hannover“ und „Fünfzehn Uhr dreißig“ ergänzt.

Sprache: Bei dieser Interaktionsart handelt es sich nicht um eine freie Spracheingabe, sondern um eine mit Kommandos, ähnlich der HoloLens [23]. Zum Öffnen des Menüs wird z. B. das Kommando „Zum Startmenü“ verwendet. Für komplexere Eingaben wurde wie bei Gaze und Sprache vorgegangen.

Gaze und Klicker: Die Interaktionsart wurde von der HoloLens adaptiert [24]. Die Selektion erfolgt abermals durch Gaze. Um das Element anzuklicken, drückt der Nutzer einen kleinen Ein-Knopf-Klicker. Zur Eingabe des Ortes und der Abfahrtszeit, wurde eine Listenauswahl simuliert.

Direkte Manipulation: Die direkte Manipulation ist eine besondere Art der Gestensteuerung [25], welche von der Meta-2-Datenbrille adaptiert wurde, und eine direkte Interaktion mit den virtuellen Objekten ermöglicht, als wären diese real [26]. In der Anwendung navigiert der Darsteller, indem er die Elemente mit seiner Fingerspitze berührt. Dabei wird darauf geachtet, dass der Darsteller immer in dem Interaktionsbereich ein Element auswählt. Zum Öffnen des Startmenüs zieht der Darsteller dieses mit beiden Händen auf.

Gestensteuerung: Da die HoloLens die reine Gestensteuerung nur in einem begrenzten Umfang mit der

Bloom-Geste einsetzt [21], wurde diese Art der Interaktion erweitert. Neben Airtap und Bloom wurden Wischbewegungen verwendet. Die Wischbewegung kann dabei von oben, unten, links oder rechts erfolgen. Der Darsteller simuliert dadurch das Scrollen durch Listen. Bestätigt wird die aktuelle Selektion in der Liste dann durch Airtap. Die Funktion der Bloom-Geste war das Schließen von Menüs.

Gaze und Airtap: Anwendung findet diese multimodale Interaktionsart aktuell auf der HoloLens [21]. Zur Selektion von Elementen wird eine „Blicksteuerung“ durch Kopfbewegungen verwendet. Ausgewählt bzw. angeklickt werden die Elemente anschließend durch die Freihandgeste Airtap auf Gesichtshöhe. Für die Auswahl des Ortes und der Zeit wurde eine Listenauswahl simuliert.

Im ersten Teil des Fragebogens, welcher sich in drei Teile unterteilt, gaben die Probanden ihre demographischen Daten an.

Der zweite Teil des Fragebogens wurde nach jeder Videovorführung wiederholt. Mithilfe einer 7-stufigen Likert-Skala konnten die Probanden angeben, wie sehr sie sich in unterschiedlichen Zuschauer- und Ortskontexten wohlfühlen würden, wie der Darsteller in dem eben gezeigten Video, mit einer Datenbrille zu interagieren. Für den Zuschauerkontext wurden die Personengruppen Familie, Freunde, Kollegen, Bekannte und Fremde mit unterschiedlichen sozialen Nähen zu dem Nutzer selektiert, welche sich an dem Modell von Shneiderman [27] orientieren. Ergänzt wurde „Alleine“, um eine generelle Ablehnung der Interaktionsart zu ermitteln. Diese Auswahl an Zuschauerkontexten stimmt teilweise mit Rico & Brewster [28] überein, welche Lebensgefährte statt Bekannten verwendeten. Für den Ortskontext wurden Orte selektiert, an welchen Experten und andere Forschungsprojekte häufig einen Einsatzort von AR sehen. Um einen Vergleich zu schaffen, wurde als erster Ort zuhause untersucht. Anschließend wurden die Orte Museum [29], Einkaufsladen [2], Arbeitsstätte [30; 31], ÖPNV [18] und Café [8] in dieser Reihenfolge evaluiert.

Nachdem die Probanden alle Videos gesehen hatten, wurde ihnen der letzte Teil des Fragebogens übergeben, in welchem die Probanden bewerten, welche der Interaktionsarten sie selbst, wenn sie einem Nutzer dieser Interaktionsart in der Öffentlichkeit begegnen würden, am wenigsten stören, nerven, verunsichern, belustigen bzw. irritieren würde. Hierdurch wurde versucht die soziale Akzeptanz aus Zuschauersicht zu evaluieren. Zusätzlich sollten sie ordnen, welche Interaktionsarten sie selbst ohne Zuschauer verwenden würden.

Im Anschluss wurden die Probanden noch kurz interviewt, um die Beweggründe und Motivationen für ihre Angaben zu erfassen. Über das Ziel der Evaluation wurden die Probanden erst nach Abschluss der Studie informiert. Für die Teilnahme an der Studie wurde keiner der Probanden monetär entlohnt.

4 Ergebnisse

An der Studie nahmen insgesamt 10 Probanden teil. Der Altersdurchschnitt lag bei 26,7 Jahren (SD = 3,3 Jahre). 40 % der Probanden waren weiblich und 60 % der Probanden männlich. Von den Probanden hatten vier schon einmal eine AR-Datenbrille benutzt.

Mit SPSS 23 wurde mit dem Shapiro-Wilk-Test die Normalverteilung geprüft und entsprechend eine Friedman's ANOVA mit einem Dunn-Bonferroni-Test als Post-hoc-Test bzw. eine einfaktorielle ANOVA mit Messwiederholung durchgeführt. Um Haupteffekte und Interaktionseffekt zu bestimmen, wurden die Daten in R aufgrund fehlender Normalverteilung entsprechend Wobbrock et al. [32] transformiert (aligned ranked transform; ART) und anschließend mit einer mehrfaktoriellen ANOVA mit Messwiederholung und einem paarweise Vergleiche per Wilcoxon-Test mit einer Bonferroni-Korrektur analysiert.

4.1 Zuschauerkontext

Die Probanden wurden gefragt, vor welchen Zuschauerkontext sie sich wohlfühlen würden, die jeweilige Interaktionsart zu verwenden. Die Mittelwerte sowie die Standardabweichungen sind Tabelle 2 zu entnehmen.

Unterschiede der Bewertungen zwischen den Zuschauerkontexten bei einer Interaktionsart: Für Gaze und Sprache ($\chi^2(5) = 45,049$; $p < 0,001$), Sprache ($\chi^2(5) = 42,256$; $p < 0,001$), Gaze und Klicker ($\chi^2(5) = 37,118$; $p < 0,001$), direkte Manipulation ($\chi^2(5) = 36,635$; $p < 0,001$) sowie Gaze und Airtap ($\chi^2(5) = 32,822$; $p < 0,001$) gibt es signifikante Unterschiede zwischen den verschiedenen Zuschauerkontexten. Die signifikanten Unterschiede der paarweisen Vergleiche sind den Spalten in Tabelle 2 zu entnehmen. Im Allgemeinen verschlechtert sich die Bewertung von über die Zuschauer von Alleine nach Fremde.

Unterschiede der Bewertung der Interaktionsarten bei einem Zuschauerkontext: Bei dem Zuschauerkontext Kollegen ist Sphärizität annehmbar ($\chi^2(14) = 11,707$; $p = 0,652$). Eine einfaktorielle ANOVA mit Messwiederholung zeigt, dass es signifikante Unterschiede für die Bewertung bei den Interaktionsarten für den Zuschauerkontext Kollegen gibt ($F(5, 45) = 4,590$; $p = 0,02$). Jedoch war keiner der paarweisen Vergleiche signifikant.

Beim Zuschauerkontext Bekannte gab es signifikante Unterschiede bei der Bewertung der Interaktionsarten ($\chi^2(5) = 17,262$; $p = 0,004$). Paarweise Vergleiche führten zu keinen signifikanten Ergebnissen.

Beim Zuschauerkontext Fremde gab es signifikante Unterschiede zwischen den Interaktionsarten ($\chi^2(5) = 20,649$; $p = 0,001$). Die Probanden würden sich signifikant schlechter fühlen die Interaktionsart Gaze und Sprache vor Fremden zu verwenden als Gaze und Airtap ($p = 0,035$; $r = 0,57$) bzw. als Gaze und Klicker ($p = 0,042$; $r = 0,56$).

Interaktionseffekt zwischen Interaktionsart und Zuschauerkontext: Eine zweifaktorielle ANOVA mit Messwiederholung mit ART-Daten für die Faktoren Interaktionsart und Zuschauerkontext deckt signifikante Haupteffekte für den Faktor Interaktionsart ($F(5, 315) = 23,820$; $p < 0,001$) und den Faktor Zuschauerkontext ($F(5, 315) = 74,203$; $p < 0,001$) sowie einen nicht signifikanten Interaktionseffekt ($p = 0,186$) auf.

Bei den paarweisen Vergleichen sind zwischen den meisten Interaktionsarten signifikante Unterschiede vorhanden, neben den Effektstärken Tabelle 1 zu entnehmen sind.

Der paarweise Vergleich beim Faktor Zuschauerkontext deckt signifikante Unterschiede zwischen allen Zuschauerkontexten auf ($p = [0,04; 0,001]$; $r = [0,27; 0,61]$).

	Sprache	Gaze & Klicker	Dir. Manipulation	Gestensteuerung	Gaze & Airtap
Gaze & Sprache	$p = 1$ $r = -0,04$	$p < 0,001$ $r = -0,46$	$p = 0,018$ $r = -0,29$	$p < 0,001$ $r = -0,44$	$p < 0,001$ $r = -0,52$
Sprache		$p < 0,001$ $r = -0,42$	$p = 0,243$ $r = -0,22$	$p < 0,001$ $r = -0,38$	$p < 0,001$ $r = -0,47$
Gaze & Klicker			$p = 0,001$ $r = 0,35$	$p = 1$ $r = 0,14$	$p = 1$ $r = 0,04$
Dir. Manipulation				$p = 0,004$ $r = -0,33$	$p < 0,001$ $r = -0,42$
Gestensteuerung					$p = 1$ $r = -0,17$

Tabelle 1: Paarweise Vergleiche der Randmittel der Interaktionsarten über die Zuschauerkontexte per Wilcoxon-Test für den Faktor Interaktionsart

4.2 Ortskontext

Anschließend wurden die Probanden gefragt, an welchem Ort sie sich wohlfühlen würden, die Interaktionsart zu verwenden. Die Mittelwerte sowie die Standardabweichungen der Bewertungen sind Tabelle 3 zu entnehmen.

Unterschiede bei der Bewertung der Ortskontexte bei den Interaktionsarten: Zwischen der Bewertung der Ortskontexte gibt es signifikante Unterschiede für die Interaktionsarten Gaze und Sprache ($\chi^2(5) = 30,643$; $p < 0,001$), Sprache ($\chi^2(5) = 28,919$; $p < 0,001$), Gaze und Klicker ($\chi^2(5) = 29,115$; $p < 0,001$), Gestensteuerung ($\chi^2(5) = 24,568$; $p < 0,001$) sowie Gaze und Airtap ($\chi^2(5) = 25,960$; $p < 0,001$). Die signifikanten Unterschiede der paarweisen Vergleiche sind Tabelle 3 zu entnehmen.

Unterschiede der Bewertung der Interaktionsarten für jeden Ort: Bei jedem Ort außer zuhause ($p = 0,446$) gibt es signifikante Unterschiede zwischen den Bewertungen der Interaktionsarten.

Für den Einsatz am Arbeitsplatz ($\chi^2(5) = 18,237$; $p = 0,003$), in einem Laden ($\chi^2(5) = 19,947$; $p = 0,001$) und in einem Café ($\chi^2(5) = 19,699$; $p = 0,001$) sind mit einem paarweisen Vergleich keine signifikanten Unterschiede zu finden.

Jedoch zeigt der Post-hoc-Test signifikante Unterschiede für den Ort Museum ($\chi^2(5) = 29,082$; $p < 0,001$), an welchem sich die Probanden signifikant schlechter fühlen würden, die Interaktionsart Gaze und Sprache zu verwenden als die Interaktionsarten Gaze und Airtap ($p < 0,015$; $r = -0,61$) oder Gaze und Klicker ($p < 0,028$; $r = -0,58$). Ein ebenfalls signifikanter Unterschied ist zwischen der Interaktion durch

	Gaze und Sprache			Sprache			Gaze und Klicker			Dir. Manipulation			Gestensteuerung			Gaze und Airtap		
	M	SD	p. Vgl.															
Alleine	6,40 ^a	0,84		6,00 ^a	1,56		6,90 ^a	0,32		6,50 ^a	0,71		6,70 ^a	0,67		6,80 ^a	0,42	
Familie	5,10	1,52	**	5,10	1,91	**	6,40	0,97	*	5,60	1,65	*	6,20	1,32	*	6,40	0,84	*
Freunde	5,00	1,41	***	4,80	1,93	*	6,00	0,94	**	5,50	1,58	**	5,90	1,29	***	6,30	1,06	**
Kollegen	3,20	1,32	***	3,40	1,43	***	5,40	1,65	**	4,20	1,75	*	4,80	1,99	*	5,10	1,73	*
Bekannte	2,80	1,14	***	3,00	1,49	*	4,90	1,66	*	3,20	1,55	*	4,40	1,71	*	5,10	1,52	*
Fremde	1,50 ^a	0,71		1,70 ^a	0,67		4,30 ^a	1,77		2,50 ^a	2,01		3,70 ^a	1,83		4,20 ^a	2,04	
RM	4,00			4,00			5,65			4,58			5,28			5,65		

^a Normalverteilung nach Shapiro-Wilk annehmbar; paarweise Vergleiche (p. Vgl.): *** p < 0,001; ** p < 0,01; * p < 0,05

Tabelle 2: Mittelwerte (M), Standardabweichungen (SD) und paarweise Vergleiche (p. Vgl.) der Bewertungen (1 am schlechtesten, 7 am besten) der Interaktionsarten je Zuschauerkontext sowie Randmittel (RM) der Interaktionsarten und Zuschauerkontexte

	Gaze und Sprache			Sprache			Gaze und Klicker			Dir. Manipulation			Gestensteuerung			Gaze und Airtap		
	M	SD	p. Vgl.	M	SD	p. Vgl.	M	SD	p. Vgl.	M	SD	p. Vgl.	M	SD	p. Vgl.	M	SD	p. Vgl.
Zuhause	6,60	0,70		6,30	1,34		6,80	0,42		6,40	0,97		6,50	0,85		6,80	0,42	
Arbeit	2,90	1,73	**	3,20 ^a	1,93	*	5,00 ^a	1,49	*	3,20 ^a	2,10	*	4,80	1,69	*	4,80	1,69	*
Laden	2,30	1,89	**	2,70	2,11	*	5,10	1,66	*	3,30 ^a	2,41	**	4,00 ^a	2,00	**	4,80 ^a	1,87	**
Museum	2,40	2,01	**	2,60	2,01	*	5,60	1,43	*	3,60 ^a	2,37	*	5,00	1,89	*	5,60	1,51	*
ÖPNV	2,00	1,25		2,50	1,84		4,70 ^a	1,34		2,40	1,78		4,20 ^a	2,10		4,20 ^a	1,81	
Cafe	1,80	1,03		1,90	1,52		3,90	1,52		2,60 ^a	1,84		3,70 ^a	1,83		3,80 ^a	1,87	
RM	3,00			3,20			5,18			3,58			4,70			5,00		

^a Normalverteilung nach Shapiro-Wilk annehmbar; paarweise Vergleiche (p. Vgl.): *** p < 0,001; ** p < 0,01; * p < 0,05

Tabelle 3: Mittelwerte (M), Standardabweichungen (SD) und paarweise Vergleiche (p. Vgl.) der Bewertungen (1 am schlechtesten, 7 am besten) der Interaktionsarten je Orte sowie Randmittel (RM) der Interaktionsarten und Orte

Sprache und Gaze und Airtap ($p < 0,015$; $r = -0,61$) sowie Gaze und Klicker ($p < 0,028$; $r = -0,58$) zu finden.

Beim Einsatz im öffentlichen Verkehr ($\chi^2(5) = 23,179$; $p < 0,001$) würden die Probanden sich schlechter fühlen, wenn sie Gaze und Sprache anstatt Gaze und Klicker ($p < 0,028$; $r = -0,58$) verwenden würden.

Interaktionseffekt zwischen Interaktionsart und Ortskontext: Eine zweifaktorielle ANOVA mit Messwiederholung bei ART-Daten mit den Faktoren Interaktionsart und Ortskontext deckt signifikante Haupteffekte für die Interaktionsart ($F(5, 315) = 31,749$; $p < 0,001$) und den Ortskontext ($F(5, 315) = 47,368$; $p < 0,001$) sowie einen Interaktionseffekt zwischen der Interaktionsart und dem Ortskontext auf ($F(25, 315) = 1,935$; $p = 0,005$).

Bei den paarweisen Vergleichen für den Faktor Interaktionsart sind die Bewertungen mehrerer Interaktionsarten voneinander signifikant unterschiedlich, wie Tabelle 4 zu entnehmen.

Für den Faktor Ortskontext gibt es keine signifikanten Unterschiede zwischen den Orten Arbeit und Museum bzw. Laden und ÖPNV und Laden bzw. Café. ($p > 0,05$) Ein paarweiser Vergleich der Interaktionen der Faktoren Interaktionsart und Ortskontext konnte keine signifikanten Unterschiede aufdecken.

	Sprache	Gaze & Klicker	Dir. Manipulation	Gestensteuerung	Gaze & Airtap
Gaze & Sprache	$p = 1$ $r = -0,04$	$p < 0,001$ $r = -0,51$	$p = 1$	$p < 0,001$ $r = -0,52$	$p < 0,001$ $r = -0,52$
Sprache		$p < 0,001$ $r = -0,49$	$p = 1$ $r = -0,12$	$p < 0,001$ $r = -0,45$	$p < 0,001$ $r = -0,47$
Gaze & Klicker			$p < 0,001$ $r = 0,47$	$p = 0,84$ $r = 0,17$	$p = 1$ $r = 0,07$
Dir. Manipulation				$p < 0,001$ $r = -0,39$	$p < 0,001$ $r = -0,45$
Gestensteuerung					$p = 0,756$ $r = -0,18$

Tabelle 4: Paarweise Vergleiche der Randmittel der Interaktionsarten über die Orte per Wilcoxon-Test für den Faktor Interaktionsart

4.3 Präferenz

In Anschluss an alle Videos wurden die Probanden gebeten die verschiedenen Interaktionsarten von 1, der besten, bis 6, der schlechtesten, zu ordnen. Es wurde hierbei gefragt, welche der Interaktionsarten sie am wenigsten stören, nerven, verunsichern, belastigen bzw. irritieren würde, wenn ein Passant die Interaktionsart verwenden würde (siehe Abbildung 2). Zwischen den Interaktionsarten gibt es signifikante Unterschiede bezüglich der Ränge ($\chi^2(5) = 32,343$; $p < 0,001$).

Paarweise Vergleiche zeigen auf, dass es signifikante Unterschiede zwischen dem Rang von Gaze und Sprache ($M = 5,2$; $SD = 0,63$) bzw. Sprache ($M = 5,5$; $SD = 0,71$) zu Gaze und Klicker ($M = 1,8$; $SD = 1,32$; $p < 0,001$; $r = -0,83$ bzw. $p < 0,001$; $r = -0,76$), zu Gaze und Airtap ($M = 2,6$; $SD = 0,84$;

$p < 0,008$; $r = -0,64$ bzw. $p < 0,028$; $r = -0,58$) und zur Gestensteuerung ($M = 2,7$; $SD = 0,95$; $p < 0,012$; $r = -0,63$ bzw. $p < 0,042$; $r = 0,56$) gibt. Es konnten keine signifikanten Unterschiede zwischen den Rängen der Interaktionsarten gefunden werden, welche die Probanden selbst privat verwenden würden ($\chi^2(5) = 4,057$; $p < 0,541$).

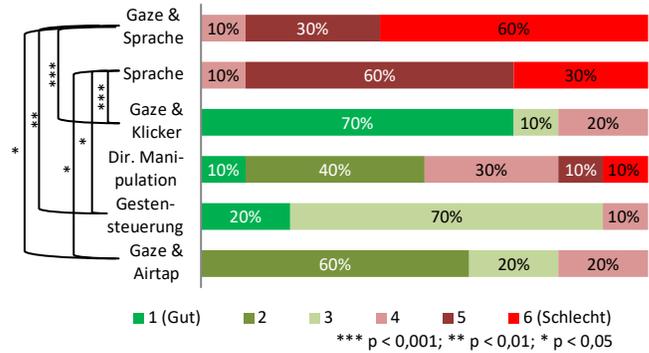


Abbildung 2: Relative Häufigkeiten der Ränge je Interaktionsart

5 Diskussion

Die Ergebnisse zeigen, dass es zwischen dem Ortskontext und der Interaktionsart, jedoch nicht zwischen dem Zuschauerkontext und der Interaktionsart, einen signifikanten Interaktionseffekt gibt. Dies bedeutet, dass die soziale Akzeptanz einer Interaktionsart vom Verwendungsort, aber nicht von der sozialen Nähe zwischen Nutzer und Zuschauer abhängt. Deshalb scheint es empfehlenswert, die Interaktionsart entsprechend des Einsatzortes zu selektieren oder mehrere frei wählbare Interaktionsarten zur Verfügung zu stellen, um die größtmögliche soziale Akzeptanz zu erreichen. So ist z. B. für den Einsatz zuhause die Interaktionsart frei wählbar. Jedoch sollte in einem Museum oder ÖPNV auf eine Eingabe mit Gaze und Sprache verzichtet werden.

Wie erläutert muss auf die soziale Nähe zu den Zuschauern keine Rücksicht bei der Selektion der Interaktionsart genommen werden. Jedoch ist zu beachten, dass mit einer stärkeren sozialen Beziehung die soziale Akzeptanz unabhängig von der Interaktionsart tendenziell steigt und die Bewertung tendenziell homogener wird. Es ist annehmbar, dass verschiedene Personen verschiedene Vorstellungen haben, welche Interaktion vor wem akzeptabel ist bzw. durch die engere soziale Beziehung besser einschätzen können, was die Zuschauer akzeptieren werden.

Auffällig ist, dass die beiden untersuchten Interaktionsarten mit Sprache vor Kollegen, Bekannten und Fremden signifikant schlechter bewertet werden, als wenn diese vor Familienmitgliedern und Freunden benutzt werden. Ein ähnliches Ergebnis hatten Efthymiou & Halvey [19] bei dem Vergleich der Sprach- und Texteingabe. Dabei ist aber zu beachten, dass die Spracheingaben unabhängig des Zuschauer- und Ortskontextes fast immer signifikant weniger sozial akzeptabel sind als die anderen Interaktionsarten.

Dies ist auch aus der Zuschauersicht erkennbar. Die Probanden präferierten als Zuschauer die Interaktion durch Gaze und Klicker bzw. Airtap oder die Gestensteuerung vor den Spracheingaben. Deshalb ist es empfehlenswert, Spracheingaben nach Möglichkeit zu vermeiden und wenn, dann nur vor Personen mit enger sozialer Beziehung, zuhause und/oder bei der Arbeit zu verwenden.

Auffällig ist auch, dass es bei den Randmitteln über alle Zuschauer- sowie Ortskontexte signifikante Unterschiede zwischen den zwei gestenbasierten Interaktionsarten gibt. Dabei wird die direkte Manipulation immer schlechter bewertet. Die Probanden gaben im Interview an, dass der ausgestreckte Finger, mit welchem das Interface benutzt wird, auf Passanten zeigen kann, was diese ansprechen könnte (5 Nennungen) bzw. ein Eingriff in deren Privatsphäre ist (2 Nennungen).

Zusammenfassend ist festzustellen, dass für die Selektion einer Interaktionsart für eine AR-Datenbrille der geplante Einsatzort betrachtet werden muss, um die höchstmögliche soziale Akzeptanz der Interaktionsart zu erreichen. Auf eine Spracheingabe sollte nach Möglichkeit verzichtet werden, insbesondere, da diese von Zuschauern abgelehnt wird. Sollte es Unsicherheit beim Einsatzort der Interaktionsart geben, sind subtile Interaktionsarten zu bevorzugen. Auf aktuellen AR-Datenbrillen wie der HoloLens sind diese Gaze und Klicker bzw. AirTap inkl. der Bloom-Geste.

Im Kontext der Ergebnisse ist zu beachten, dass die soziale Akzeptanz dynamisch ist und sich über längere Zeit verändern wird [7; 33]. Ebenfalls wird die durchgeführte Studie dadurch limitiert, dass nur eine kleine studentische und deutschsprachige Stichprobe verwendet wurde. Insbesondere können das Alter [7] oder die Kultur [7; 34] als Faktoren der sozialen Akzeptanz einen Einfluss haben. Eine weitere Limitierung ist die fehlende scharfe Trennung zwischen der Nutzer- und Zuschauersicht. Idealerweise werden diese Aspekte in nachfolgenden Studien getrennt erhoben. Für die Evaluierung der sozialen Akzeptanz aus der Nutzersicht wäre es vorteilhaft, diese in einem realen Umfeld durchzuführen, da hier keine Transferleistung wie in der durchgeführten Laborstudie nötig ist. Für die Evaluierung aus Zuschauersicht scheint der Laborversuch angemessen.

6 Fazit

Es wurde eine Studie durchgeführt, welche die soziale Akzeptanz verschiedener Interaktionsarten auf aktuellen Datenbrillen (HoloLens, Meta 2) erhebt. Unter Anderem konnte gezeigt werden, dass bei der Auswahl einer aktuellen Interaktionsart für AR-Datenbrillen der Orts-, jedoch nicht der Zuschauerkontext berücksichtigt werden muss. Dabei wurde insbesondere das Problem der direkten Manipulation mit dem ausgestreckten Finger aufgedeckt, welches in naher Zukunft adressiert werden muss, da die HoloLens 2 diese Interaktionsart voraussichtlich nativ einsetzt [35].

DANKSAGUNGEN

Die Autoren danken Frau Melisa Muller-Bjelic für die Versuchsdurchführung.

REFERENZEN

- [1] W. Broll. 2013. Augmentierte Realität. In *Virtual und Augmented Reality (VR/AR). Grundlagen und Methoden der Virtuellen und Augmentierten Realität*, R. Dörner, W. Broll, P. F. Grimm and B. Jung (Eds.). eXamen.press. Springer Vieweg, Berlin, Heidelberg, 241–294.
- [2] T. Lutter, A. Pentsi, M. Poguntke, K. Böhm, and R. Esser. 2015. *Zukunft der Consumer Electronics 2015. Marktentwicklung, Schlüsselrends, Mediennutzung, Konsumverhalten, Neue Technologien* (2015). Retrieved April 2, 2019 from <https://www.bitkom.org/sites/default/files/pdf/noindex/Publikationen/2015/Studien/CE-Studie-2015/150901-CE-Studie-2015-online.pdf>.
- [3] C. Arthur. 2013. *Google Glass: is it a threat to our privacy?* (2013). Retrieved June 11, 2019 from <https://www.theguardian.com/technology/2013/mar/06/google-glass-threat-to-our-privacy>.
- [4] M. Koelle, M. Kranz, and A. Möller. 2015. Don't look at me that way! In *MobileHCI'15. Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services : Copenhagen, Denmark, August 24-27, 2015*. Association for Computing Machinery, New York, New York, 362–372.
- [5] J. Nielsen. 1993. *Usability Engineering*. Academic Press, San Diego.
- [6] S. Brewster, R. Murray-Smith, A. Crossan, Y. Vazquez-Alvarez, and J. Williamson. 2009. *The GAIME project: Gestural and Auditory Interactions for Mobile Environments*. British Computer Society.
- [7] C. S. Montero, J. Alexander, M. T. Marshall, and S. Subramanian. 2010. Would you do that? In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, New York, NY, 275.
- [8] Y.-C. Tung, C.-Y. Hsu, H.-Y. Wang, S. Chyou, J.-W. Lin, P.-J. Wu, A. Valstar, and M. Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, 3327–3336.
- [9] E. Goffman. 2009. *Wir alle spielen Theater. Die Selbstdarstellung im Alltag* (Ungedruckte Taschenbuchausgabe, 17. Auflage). Piper, 3891. Piper, München.
- [10] F. D. Davis. *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Massachusetts Institute of Technology.
- [11] Y. Malhotra and D. F. Galletta. 1999. Extending the technology acceptance model to account for social influence: theoretical bases and empirical validation. In *System Sciences, 1999: Annual Hawaii International Conference on System Sciences (32nd: 1999: Maui, Hawaii)*. IEEE Computer Society Press, [Place of publication not identified], 14.
- [12] J. Rico and S. Brewster. 2010. Usable gestures for mobile interfaces. In *CHI Conference. We are HCI : conference proceedings, Atlanta, Ga, USA, April 10-15, 2010*. Association for Computing Machinery, New York, N.Y., 887.
- [13] Venkatesh, Morris, and Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27, 3, 425.
- [14] D. Ahlström, K. Hasan, and P. Irani. 2014. Are you comfortable doing that? In *MobileHCI'14. Proceedings of the 16th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services : September 23-26, 2014, Toronto, ON, Canada*. Association for Computing Machinery, New York, New York, 193–202.
- [15] J. R. Williamson, A. Crossan, and S. Brewster. 2011. Multimodal mobile interactions. In *Proceedings of the 13th international conference on multimodal interfaces*. ACM, New York, NY, 361.
- [16] A. Taniberg, L. Botin, and K. Stec. 2018. Context of use affects the social acceptability of gesture interaction. In *NordiCHI 2018 Oslo. Revisiting the Life Cycle : proceedings of the 10th Nordic Conference on Human-Computer Interaction : Oslo, Norway, 1-3 October, 2018*. ICPS: ACM international conference proceeding series. ACM, New York, NY, USA, 731–735.
- [17] M. Radziwill, R. Kniewel, and L. Schmidt. 2017. Multimodale Smartphone-Interaktion für Jung und Alt? *atp* 59, 06, 46.
- [18] E. M. Klose, A. Eis, J. Hegeberg, and L. Schmidt. 2018. Nutzerorientierte Anforderungsanalyse für ein adaptiv lernendes Reiseassistenzsystem mit Datenbrillen. *Z. Arb. Wiss.* 72, 1, 3–12.
- [19] C. Efthymiou and M. Halvey. 2016. Evaluating the Social Acceptability of Voice Based Smartwatch Search. In *Information Retrieval Technology. 12th Asia Information Retrieval Societies Conference, AIRS 2016, Beijing, China, November 30 - December 2, 2016, Proceedings*, S. Ma, J.-R. Wen, Y. Liu, Z. Dou, M. Zhang, Y. Chang and X. Zhao (Eds.). Lecture Notes in Computer Science, 9994. Springer International Publishing; Imprint: Springer, Cham, 267–278.
- [20] S. Ronkainen, J. Häkkinen, S. Kaleva, A. Colley, and J. Linjama. 2007. Tap input as an embedded interaction method for mobile devices. In *TEI'07. First*

- International Conference on Tangible and Embedded Interaction : February 15-17 2007 in Baton Rouge, Louisiana, USA, Theme of the CCT 14th Annual Mardi Gras Conference : conference proceedings.* ACM Press, New York, NY, 263.
- [21] rwinj, grbury, J. McCulloch, M. Zeller, and B. Bray. 2019. *Gestures* (2019). Retrieved April 2, 2019 from <https://docs.microsoft.com/en-us/windows/mixed-reality/gestures>.
- [22] A. Turner, J. McCulloch, M. Zeller, rwinj, and B. Bray. 2019. *Gaze* (2019). Retrieved April 2, 2019 from <https://docs.microsoft.com/en-us/windows/mixed-reality/gaze>.
- [23] rwinj, R. K. Gupta, grbury, M. Zeller, and B. Bray. 2019. *Voice design* (2019). Retrieved April 2, 2019 from <https://docs.microsoft.com/en-us/windows/mixed-reality/voice-design>.
- [24] Microsoft Inc. 2017. *Use the HoloLens clicker* (2017). Retrieved April 2, 2019 from <https://support.microsoft.com/en-us/help/12646/hololens-use-the-hololens-clicker>.
- [25] R. Aigner, M. Haller, D. Lindbauer, A. Ion, S. Zhao, and Koh, Jeffrey Tzu Kwan Valino. 2012. *Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI*.
- [26] Meta Company. 2017 (2017). Retrieved April 2, 2019 from <https://www.metavision.com/146-2/>.
- [27] B. Shneiderman. 2002. *Leonardo's laptop. Human needs and the new computing technologies*. MIT Press, Cambridge, Mass.
- [28] J. Rico and S. Brewster. 2009. Gestures all around us. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '09*. ACM Press, New York, New York, USA, 1.
- [29] K. Mase, R. Kadobayashi, and R. Nakatsu. 1996. Meta-Museum: A Supportive Augmented-Reality Environment for Knowledge Sharing. In *ATR Workshop on Social Agents: Humans and Machines*, 107–110.
- [30] L. Schmidt, A. Beu, M. Edelmann, A. Epstein, O. Oehme, P. Quact-Faslem, B. Rottenkolber, G. Triebfürst, S. Wiedenmaier, and W. Wohlgemuth. 2004. Benutzerzentrierte Systemgestaltung. In *ARVIKA. Augmented Reality in Entwicklung, Produktion und Service*, W. Friedrich (Ed.). Publicis, Erlangen, 28–51.
- [31] V. Paelke. 2014. Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment. In *IEEE [International Conference on] Emerging Technologies and Factory Automation (ETFA), 2014. 16-19 Sept. 2014, Barcelona, Spain*. IEEE, Piscataway, NJ, 1–4.
- [32] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *CHI 2011. Conference proceedings and extended abstracts, Vancouver, BC, USA, May 7 - 12, 2011 ; the 29th Annual CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 143.
- [33] J. Rico, A. Crossan, and S. Brewster. 2011. Gesture-Based Interfaces: Practical Applications of Gestures in Real World Mobile Settings. In *Whole Body Interaction*, D. England (Ed.). Human-Computer Interaction Series. Springer London, London, 173–186.
- [34] G. Tan, S. Brave, C. Nass, and M. Takechi. 2003. Effects of voice vs. remote on U.S. and Japanese user satisfaction with interactive HDTV systems. In *Chi 2003*. ACM Press, New York, 714.
- [35] grbury, J. McCulloch, T. Ferrell, D. Songür, and R. Tremblay. 2019. *Getting your existing app ready for HoloLens 2* (2019). Retrieved June 11, 2019 from <https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-porting-guide>.