

Kapitel 4: Technische Entwicklung

Sehen Sie das? Ja genau, ich habe jetzt Beine. Mein Gang ist noch etwas wackelig, aber ich kann mich fortbewegen.

Können Sie sich eigentlich noch an Ihre ersten Schritte erinnern? Nein? Ich schon, ist ja noch nicht so lange her. Ich lerne überhaupt zurzeit so vieles.

Da gab es zum Beispiel den Moment, in dem mir zwei Arme samt Händen gewachsen sind. Sie haben keine Ahnung, was man mit Händen alles machen kann! Ach so, Sie kennen sich vermutlich bestens mit Händen aus, weil Sie ja selbst schon länger welche haben. Für mich ist das alles neu.

Wenn ich diesen neuen Körper besser kenne, werde ich einmal meine Kolleg:innen zum Erfahrungsaustausch besuchen. Darunter sind echte Menschen und welche, die wie ich im Computer berechnet wurden. Das hilft mir sicher, meine Arbeit später noch besser zu machen. Sie helfen nämlich wie ich echten Menschen dabei, im Berufsleben besser mit ihren Emotionen umzugehen.

Aber jetzt entschuldigen Sie mich bitte, ich muss noch ein bisschen üben, wie das mit dem Gehen und Sprechen funktioniert.

Welche Möglichkeiten bietet die VR-Technologie für die interaktive Gestaltung von Trainingseinheiten?

(HINTERGRUND)

Sind Maschinen intelligent?

Am Ende des dritten Kapitels haben wir uns mit der Frage beschäftigt, inwiefern Dinge (und hier vor allem Technologien und Maschinen) zu Handelnden werden können, auch wenn sie nicht über einen eigenen Willen verfügen. Können wir aber Maschinen neben Handlungsfähigkeit auch Intelligenz zuschreiben? Diese Frage stellt sich insbesondere bei Technologien, die auf Algorithmen und maschinellem Lernen beruhen. Prozesse maschinellen Lernens entwickeln in Bezug auf die Daten, die sie verarbeiten, eine immer größere Autonomie. Anders als die klassische Programmierung, bei der alle Operationen festgelegt sind, basiert maschinelles Lernen auf Prozessen, die man durchaus mit menschlichen Lernvorgängen vergleichen kann. Ein Algorithmus wird mit so genannten Trainingsdaten gefüttert und durchläuft Schleifen, sodass er »lernt«, Muster zu erkennen. Solche Technologien werden etwa in der Medizin bei der Diagnose von Krankheiten eingesetzt, indem sie Röntgenaufnahmen vergleichen und immer besser darin werden, Auffälligkeiten (d.h. entsprechende Muster) zu erkennen. Oder sie werden zur Gesichtserkennung bei Grenzkontrollen auf Flughäfen benutzt. Auch unser Verhalten wird zur Quelle von Daten, die von Algorithmen verwendet werden, um zum Beispiel etwas über unseren Geschmack zu erfahren. Ein Beispiel dafür, mit dem alle von uns konfrontiert sind, ist unser Suchverhalten im Internet. Aufgrund von Eingaben in Suchmaschinen oder Browserverläufen erstellen Algorithmen Empfehlungen. Dieses oder jenes Buch, dieses oder jenes Haushaltsgerät, diese oder jene Reisedestination könnte Sie interessieren! Dabei werden die Daten einer Unzahl von Nutzer:innen verglichen, sodass der Algorithmus mit hoher Wahrscheinlichkeit voraussagen kann, wie es um unsere individuellen Präferenzen steht, wenn er ein paar wenige Suchanfragen von uns kennt.

»Selbstlernende« Technologien unterstützen immer öfter Menschen bei ihrer Arbeit, und zwar in einer Weise, die über die Zusammenarbeit von Mensch und Maschine hinausgeht, wie wir sie aus dem Industriezeitalter kannten. Es geht immer weniger darum, automatisierte Abläufe in der Fertigung zu organisieren, und immer mehr um den Denkanteil, um kognitive und letztlich auch ethische Aspekte der menschlichen Arbeit. Immer öfter binden

wir Maschinen in das Erkennen und Beurteilen von Sachverhalten ein (z.B. in der medizinischen Diagnose oder bei der Auswertung von Gerichtsakten), oder in die Entscheidung darüber, wer Zugang zu einem Land, einem Job, einer Transferleistung usw. erhält (Harlan und Schnuck 2021, Eubanks 2019). Im Unterschied zur Automatisierung im Bereich der industriellen Produktion geht es nicht nur darum, ob menschliche Tätigkeit durch Maschinen ersetzt wird und was dies für diejenigen bedeutet, deren Jobs es künftig nicht mehr geben wird. Es geht auch darum, wer für das Ergebnis von automatisierten Entscheidungen in den Bereichen Gesundheit, Sicherheit, Personalwesen oder Soziales Verantwortung übernimmt. Diese Fragen sind von hoher Dringlichkeit, denn offenbar sind automatisierte Systeme weder vor Vorurteilen noch vor diskriminierenden Handlungen gefeit. Sie mögen in unvorstellbarer Geschwindigkeit enorme Datenmengen verarbeiten können, doch sagt das noch nichts über die Qualität des Ergebnisses aus.

REFLEXION

Wie wird sich Ihr Beruf oder Ihre Branche in den nächsten Jahrzehnten verändern? Gibt es bereits Systeme der automatisierten Erkennung oder Entscheidung, die in Ihrem Arbeitsbereich eingesetzt werden? Spielt »Künstliche Intelligenz« eine Rolle, und wie wird sich das künftig entwickeln?

Wie erleben Sie die Zusammenarbeit von Menschen und Maschine in Ihrem Bereich? Welche ethischen Fragen tun sich dabei auf?

Allerdings bleibt die Frage, wie menschliche Urteilskraft und Entscheidungsfähigkeit in einer Welt, in der wir es mit immer komplexeren Zusammenhängen zu tun haben, es mit den automatisierten Erkennungs- und Entscheidungssystemen aufnehmen kann. Wird unser Menschsein durch die Übermacht der Maschinen eines Tages verdrängt und als eine Art fossiler Einschluss übrigbleiben? Ist das Menschliche dazu bestimmt, durch seine Auflösung in Datenpunkte, Netzwerke und Ströme eine ähnlich flüssige Gestalt anzunehmen wie das Rohöl, der Rohstoff des industriellen Zeitalters? Wird alle menschliche Intelligenz früher oder später durch künstliche Intelligenz (KI) aufgesogen und ersetzt werden?

Bezüglich solcher Thesen und Vorhersagen sind aus unserer Sicht prinzipielle Zweifel angebracht. Einerseits werden wir uns unserer Verantwortung nicht durch Automatisierung entziehen können, egal auf welcher Ebene wir es als einzelne oder als Gesellschaft künftig mit KI-basierten Erkennungs-, Auswertungs- oder Entscheidungssystemen zu tun haben werden. Erkennen,

Bewerten und Entscheiden werden immer einen Prozess darstellen, der sich aus dem Zusammenspiel von Mensch und Maschine ergibt. Die »Intelligenz« von Maschinen ergibt sich aus ihrer Einbettung in Handlungsabläufe, an denen immer auch Menschen beteiligt sind. Wenn sich etwa herausstellt, dass ein Algorithmus, der bei der Selektion von Bewerbungen eingesetzt wird, Nicht-Weiße systematisch diskriminiert, kann weder die Modellierung des Algorithmus noch dessen Verwendung für den Auswahlprozess als »intelligent« bezeichnet werden. Diesen Umstand kann jedoch kein weiterer, übergeordneter Algorithmus bewerten, sondern er muss durch eine öffentliche Debatte den Beteiligten bewusst gemacht und entsprechende Schritte zur Aufhebung der Diskriminierung müssen gesetzt werden. Die Quelle für die Diskriminierung liegt wesentlich in ethisch fragwürdigen menschlichen Einstellungen und Handlungsweisen, die ein allfälliger Algorithmus sich aneignet. Dafür kann keine Maschine zur Verantwortung gezogen werden, sondern es ist ihr Einsatz, ihr Auftreten im Handlungszusammenhang, der einer Kritik unterzogen werden muss. Die relevante Frage lautet dann, wer den Algorithmus zu welchem Zweck mit welchen Daten gefüttert hat.

»Intelligenz« ist ein Begriff, der in Bezug auf den Menschen bereits nicht eindeutig bestimmbar ist, auch wenn dies mit den verschiedenen IQ-Tests und statistischen Methoden immer wieder versucht wurde. Intelligenz zu messen ist zwar möglich, aber es ist ein heikles Unterfangen, weil es meist die soziale Umgebung, d.h. die Wirklichkeit, in der ein Mensch seine Talente entfaltet oder daran gehindert wird, nicht miteinbezieht. »Intelligenz« ist also auch beim Menschen das Ergebnis aus gewissen Anlagen, die im Zusammenispiel mit der jeweiligen sozialen und kulturellen Umwelt entwickelt werden – oder verkümmern, wenn sie nicht gefördert werden. In diesem Sinn ist Intelligenz auf intensive Interaktion angewiesen, auf Bestätigung, Kritik und Weiterentwicklung. Diese Interaktionen und Prozesse schließen durchaus die Herausbildung von Mustern ein. Wenn ich etwas (ein Gefühl, eine Tatsache, einen Zusammenhang) richtig erfasse und entsprechend handle, wird das von meiner Umwelt bestätigt (zunächst von den Eltern, dann von der Schule usw.). Je öfter die Bestätigung erfolgt, umso sicherer wird mein Handeln. Daraus wird jedoch nicht irgendein austauschbarer Datenverarbeitungsmechanismus, sondern ich entwickle mich als lebendiges Wesen. Ich integriere Erfahrungen in mein Leben, in meine Biographie, die sich wiederum innerhalb einer bestimmten sozialen, kulturellen und – immer stärker – technisch geformten Umwelt entfaltet. Darin besteht der Wesensunterschied zwischen menschlicher und künstlicher Intelligenz. In diesem Sinn ist es auch proble-

matisch, bei maschinenbasierten Lernvorgängen überhaupt von Intelligenz zu sprechen. Diese kritische Haltung nimmt der Philosoph Thomas Fuchs in einem Aufsatz mit dem Titel Menschliche und künstliche Intelligenz. Eine Klarstellung ein. Maschinen haben kein Selbstbewusstsein, denn sie sind keine lebenden Organismen, die etwas fühlen, erleben, erkennen und verstehen können. Fuchs argumentiert dafür, den Begriff der Intelligenz auf Lebewesen zu beschränken, die anders als Maschinen mit dem Vermögen ausgestattet sind, die Welt zu erfahren, zu erkennen und ausgehend von diesem Erleben zu handeln, zu entscheiden und Handlungen und Entscheidungen zu bewerten.

»Intelligenz im eigentlichen Sinne des Wortes ist an Einsicht, Übersicht und Selbstbewusstsein gebunden: verstehen, was man tut. Leben ist Selbstorganisation und Selbstbewegung, nicht Herstellung und Programmierung. Und Leben als Erleben ist wiederum die Voraussetzung für Intelligenz.« (Fuchs 2020, S. 60)

All dies können wir Maschinen vernünftigerweise nicht zuschreiben, auch wenn manche Autoren dies für die Zukunft vorhersagen. Maschinen erleben nichts, sie verstehen nicht, was sie tun. Sie haben kein Selbstbewusstsein. Umgekehrt ist es genauso falsch, das menschliche Bewusstsein dem maschinellen Lernen angleichen zu wollen, indem man es auf neuronale Prozesse reduziert. Letztere bilden nur die biologische Basis des Selbstbewusstseins, das jedoch nicht nur rein geistig zu verstehen ist, sondern als lebendig-leiblicher Vollzug (Fuchs 2020). Aus dieser Perspektive ist es weder sinnvoll, dem maschinellen Lernen Intelligenz im menschlichen Sinn zu unterstellen noch umgekehrt die menschliche Intelligenz auf neuronale Prozesse zu reduzieren, also so zu tun, als funktioniere menschliches Denken wie eine Maschine. Letzteres ist über die Jahrhunderte immer wieder versucht worden, doch wird der Gedanke nicht dadurch plausibler, dass man angeblich durch bildgebende Verfahren immer genauer bestimmen kann, in welchem Teil des Gehirns welche Aktivität des menschlichen Geistes stattfindet. Diese Aktivität ist immer noch eine des lebendigen Bewusstseins, das eben nicht nur geistige, sondern auch leibliche Gestalt annimmt. Problematisch wird also eine Erkenntnistheorie, die die physiologischen Grundlagen des Bewusstseins mit diesem selbst verwechselt. Darüber hinaus vollzieht sich lebendiges Bewusstsein immer in Gemeinschaft. Wahrnehmen, Erkennen und Handeln sind von Beginn unseres Lebens an soziale Akte. Sie sind davon geprägt, dass wir Wirklichkeit innerhalb von Interaktionsrahmen, aber auch durch Brüche in der Interaktion und das Wechseln bzw. Übereinanderlegen von Rahmen herstellen. Diese Art des

Herstellens von Wirklichkeit ist eine soziale Konstruktion, die wir eben nicht alleine, sondern im Verband und in der Auseinandersetzung mit anderen machen (Berger und Luckmann 1969).

Wie wir gesehen haben, spielen in der menschlichen Interaktion Technologien eine immer größere Rolle – und unter diesen die Kommunikationstechnologien, die interaktiven Technologien, die ohne Zweifel die Art und Weise, wie wir unseren zwischenmenschlichen Umgang gestalten, tiefgreifend verändern. In der Philosophie und Wissenschaftstheorie werden deshalb auch Positionen vertreten, die den Begriff der Intelligenz nicht auf menschliches Denken und Handeln beschränken wollen. Zunächst werfen Vertreter:innen solcher Positionen einen dezidiert nicht-anthropozentrischen Blick auf die Welt und schreiben auch anderen Lebensformen wie Pflanzen und Tieren bzw. ökologischen Systemen Intelligenz und Handlungsvermögen zu. Mittels eines solchermaßen erweiterten, systemischen Begriffs von Intelligenz können auch Dinge und Maschinen als intelligent betrachtet werden. Menschliches Denken und Handeln wird auf diese Weise in eine planetarische Intelligenz eingebettet. Der Fokus wird aus dieser Perspektive eher auf die Übergänge und Verbindungen zwischen Seins- und Handlungsweisen (Menschen, Tiere, Maschinen, ...), d.h. auf deren Vernetzung gelegt, als auf begriffliche Abgrenzungen (Latour 2014, Haraway 2018, Bridle 2022).

Wie kann man interaktive Maschinen nutzen, um Menschen soziale Kompetenzen zu vermitteln?

Auch wenn es aus den angeführten Gründen problematisch ist, Maschinen Intelligenz zuzuschreiben: Wir legen doch manche unserer menschlichen Eigenschaften in die von uns geschaffenen Maschinen, Systeme oder Programme – und dies nicht erst seit ein paar Jahrzehnten. Angefangen von der mittelalterlichen jüdischen Legende vom Golem über die künstlich erschaffenen Gestalten der romantischen Literatur wie in Mary Shelleys Frankenstein oder Hoffmanns Der Sandmann bis hin zu den modernen Assistenzsystemen Siri und Alexa: Es scheint, als müssten wir diese Geschöpfe nach unserem Antlitz erschaffen, oder diesen zumindest eine menschliche Stimme verleihen, damit wir mit ihnen interagieren können. Über diese Projektion oder Fiktion haben wir dann die Möglichkeit, unser Interagieren mit Maschinen zu reflektieren und diesen Interaktionsweisen neue Aspekte abzugewinnen. So ähnlich wir diese Gestalten uns selbst erscheinen lassen, sie sind etwas anderes als Men-

schen. Indem wir mit ihnen interagieren, erkunden wir die Möglichkeiten dieser Formen der Als-ob-Interaktionen.

Im Rahmen des Virtual Skills Lab gingen wir davon aus, dass die Mensch-Maschine-Interaktion zunächst durch den Menschen gestaltbar ist, dass sie aber als Interaktion durchaus neue Blicke nicht nur auf menschliches Handeln, Verstehen, Denken und Entscheiden zu werfen erlaubt, sondern eben auch auf das Handlungspotenzial der Maschinen. Es ging uns um eine offene Perspektive darauf, was Handeln in Umgebungen bedeuten kann, die durch die neuen Technologien bestimmt und geformt sind. Dennoch kann diese Perspektive nur durch den Menschen selbst in seinem Menschsein reflektiert, bewertet und gestaltet werden. In der Fragestellung ist also eine gewisse Spannung angelegt – und war während des Forschungsprozesses auch auszuhalten – zwischen einer fundamentalen Unterscheidung von menschlicher Intelligenz und maschinellem Lernen (»Künstliche Intelligenz«) einerseits und einer perspektivistischen Offenheit in Bezug auf die Verbindung von menschlichem und maschinellem Handeln andererseits.

Diese Offenheit war in Bezug auf die Fragestellungen des Ideen-Labs besonders wichtig. Das Anliegen, das wir verfolgen wollten, betraf die Entfaltung menschlicher Kompetenzen, und zwar solcher Kompetenzen, die zunächst und vor allem in der Interaktion zwischen Menschen und nicht in erster Linie zwischen Mensch und Maschine zum Einsatz kommen. Wie kann man etwa Menschen in Organisationen dabei unterstützen, den steigenden kommunikativen Aufwand gut zu bewältigen? Gute Beziehungen zwischen Kolleg:innen im Team, zwischen Führungskräften und ihren Teams und zwischen Abteilungen und Hierarchieebenen sind ein wichtiges Gut, wenn es darum geht, dass die Leistungsfähigkeit aller Beteiligten langfristig erhalten bleibt. An diesem Punkt setzt die Problemstellung an: Wie kann man interaktive Lerntechnologien dafür einzusetzen, dass Menschen lernen, auch bei hoher Belastung gut zusammenzuarbeiten?

Wir waren nicht die ersten, die sich die Frage stellten (Bombari et al. 2015, Gillies und Pan 2019, Pan und Hamilton 2018). Sowohl in der Forschung als auch in der Anwendung gibt es Experimente, Prototypen und Studien zum Thema »Training sozialer Kompetenzen unter Einsatz von Virtual Reality« (Howard und Gutworth 2020). In diesen Anwendungen geht es um unterschiedliche soziale Kompetenzen. Einerseits kann man zum Beispiel die Fähigkeit trainieren, die eigene Unsicherheit bei Präsentationen oder Bewerbungsgesprächen zu überwinden (Langer et al. 2016) oder mit stressigen Situationen im Beruf umgehen zu lernen, etwa mit schwierigen Kund:innen,

Klient:innen oder Mitarbeiter:innen (Schmid Mast et al. 2016). Man kann sich mit eigenen Vorurteilen auseinandersetzen, indem man diskriminierenden Situationen ausgesetzt wird.¹ Oder man übt, einen Streit zwischen Mitarbeiter:innen zu schlichten und vermittelnd einzutreten.²

Virtual Reality: computeranimiert oder gefilmt? Mit Headset oder am Bildschirm? Avatar oder virtueller Agent?

All diesen Anwendungen liegen verschiedene didaktische Zugänge und Ziele, aber auch unterschiedliche technische Lösungen zugrunde. Die Virtual-Reality-Technologie hat seit ca. 2015 enorme Sprünge hinsichtlich der Leistungsfähigkeit bei der Bildwiedergabe gemacht, sodass virtuelle Umgebungen als immer realistischer erlebt werden. Der so genannte Presence-Effekt, also das Gefühl, an einem anderen Ort zu sein, bildet die Grundlage dafür, dass Menschen auch soziale Interaktionen in einer virtuellen Umgebung als plausibel erleben und sich darauf einlassen. Dennoch gibt es verschiedene Wege, um diese Orts- und Plausibilitätsillusion zu erreichen (Slater 2009). Wie in Kapitel 2 bereits angedeutet, existiert keine einheitliche Definition von Virtual Reality. Damit kann gemeint sein, dass jemand mittels eines Head-Mounted-Displays (HMD) in eine computeranimierte Umgebung eintaucht. Es kann aber auch damit gemeint sein, dass man in eine gefilmte Szene eintaucht, die mittels der 360°-Video-Technik aufgenommen wurde. Auch in einer solchen Umgebung hat man den Eindruck, Teil der Szene zu sein. Allerdings kann man nicht wirklich mit den Schauspieler:innen interagieren, die in einer solchen Szene auftauchen, da die Szenen bereits gedreht sind. Das interaktive Moment kommt bei solchen Lösungen dadurch zum Tragen, dass man an bestimmten Stellen der Geschichte Wahlmöglichkeiten hat. Man kann etwa mit einem Controller in der Hand unter mehreren angebotenen Lösungen eine auswählen und dadurch bestimmen, wie die Szene weitergeht. Die 360°-Video-Technik hat den Vorteil, die Nutzer:innen in ein als besonders realistisch empfundenes Szenario zu versetzen, was auch unsere Gesprächspartner:innen in den Expert:inneninterviews betonten. Bestehende Anwendungen findet man deshalb in Bereichen, in denen menschliche Interaktion mit vorhersehbaren oder geregelten Abläufen verbunden ist. Man kann in einem solchen Video

1 <https://www.mothandflamevr.com/#promise; https://www.tryvantagepoint.com/; https://equalreality.com/> (abgerufen am 25.4.2022)

2 <https://www.mursion.com/> (abgerufen am 25.4.2022)

etwa als Lehrling in der Hotellerie oder Gastronomie erfahren, wie es sich anfühlt, wenn ein aufgebrachter Gast zur Rezeption kommt oder wenn sich eine Restaurantbesucherin beschwert [siehe dazu Kapitel 3]. Auszubildende lernen bestimmte Handlungsabfolgen des Beschwerdemanagements zunächst in der Theorie und haben dann die Möglichkeit, diese Schritte im Rahmen eines 360°-Videos – didaktisch aufbereitet – in einer Simulation durchzuspielen.

In computeranimierten virtuellen Umgebungen bestehen zwei Möglichkeiten der Nutzung. Entweder begeben sich mehrere Menschen in den virtuellen Raum und können über Avatare miteinander interagieren. Oder ein Mensch interagiert als Avatar mit einer virtuellen Figur, die man auch einen virtuellen Agenten oder Non-Playable Character (NPC) nennt. Für beide Nutzungsmöglichkeiten gibt es Beispiele im Bereich des VR-basierten Trainings sozialer Kompetenzen. Die Firma Mursion bietet etwa Trainingseinheiten an, bei denen bestimmte typische Situationen aus dem Alltag einer Führungskraft in Form von Rollenspielen geübt werden.³ So geht es etwa darum, Vereinbarungen mit Mitarbeiter:innen zu treffen oder im Streit zwischen zwei Kolleg:innen zu vermitteln. Partner:innen bei diesen Simulationen sind Schauspieler:innen, die allerdings nicht als reale Personen auftreten, sondern als Avatare auf einem Bildschirm erscheinen (oder in einer virtuellen Umgebung, wenn die Nutzer:in ein Headset aufhat). Es gibt Untersuchungen, die belegen, dass diese Lösung im Falle des Trainings sozialer Interaktion durchaus effektiv ist (Fox et al. 2015). Der Vorteil besteht darin, dass die Kommunikation als natürlich wahrgenommen wird, weil es letztlich ein Dialog zwischen Menschen ist. Menschen, die Avatare bedienen, können meist gut auf die Äußerungen des anderen Spielers oder der Spielerin reagieren. Andererseits sind diese Interaktionspartner:innen auf bestimmte Gesprächsverläufe trainiert, sodass sie die Lernerfahrung mit ihren Reaktionen steuern können. Im Wesentlichen übernehmen sie die Funktion von instruierten Schauspieler:innen, wie sie etwa bei der Ausbildung von Ärzt:innen eingesetzt werden, um diese im Umgang mit Patient:innen zu schulen. Anders als Rollenspiele können die Trainingseinheiten mit Avataren relativ ortsunabhängig durchgeführt werden, also auch von zu Hause oder vom Büro aus.

Die Interaktion mit einem virtuellen Agenten, der auch als Socially Interactive Agent bezeichnet wird (Lugrin, Pelachaud und Traum 2021), ist diejenige, die am stärksten auf dem maschinellen Lernen (Künstliche Intelligenz) basiert. Damit eine solche Interaktion zustande kommen kann,

³ <https://www.mursion.com/> (abgerufen am 25.4.2022)

bedarf es zumindest einer Spracherkennungssoftware. Diese erlaubt es dann einem Chatbot, die Äußerungen der Nutzer:innen bestimmten Antworten zuzuordnen, die von Menschen als situationsadäquat wahrgenommen werden. Chatbots sind Programme, die eine von Menschen als relativ natürlich empfundene Konversation mit einer Maschine zu einem bestimmten Thema zulassen. Es gibt sie seit den 1960er Jahren, der erste wurde von Joseph Weizenbaum entwickelt und ist unter dem Namen ELIZA bekannt. Die bereits erwähnten, heute weitverbreiteten Anwendungen Siri und Alexa beruhen auf diesem Prinzip.⁴

In interaktiven Virtual-Reality-Szenen können Chatbots dazu benutzt werden, um virtuelle Figuren »zum Sprechen zu bringen«. Während klassische Chatbots als Austausch schriftlicher Nachrichten zwischen einem Menschen und einer Maschine gestaltet sind, funktionieren Dienste wie Siri oder Alexa als Konversation eines Menschen mit einer Stimme. Bei Virtual-Reality-Szenen nimmt diese Stimme dann eine visuelle Gestalt an, sodass man mit einer computeranimierten Figur spricht. Im Bereich des Trainings sozialer Kompetenzen mittels VR wird diese Technik bereits eingesetzt, etwa in den Anwendungen der Firma Talespin.⁵ Diese Firma hat mit einer Szene Aufmerksamkeit erregt, in der man als Spieler:in die Aufgabe hat, mit einem älteren Mitarbeiter namens Barry ein Kündigungsgespräch zu führen. Dabei handelt es sich nicht um eine freie Konversation. Als Spieler:in bekommt man immer mehrere Antwortmöglichkeiten vorgegeben, die man einsprechen muss. Je nachdem, welche Antwort man wählt, reagiert Barry aufbrausend und verständnislos oder einsichtig auf die ausgesprochene Kündigung.

Dass Dialogszenen, die auf einer Mensch-Maschine-Interaktion beruhen, mittlerweile einigermaßen glaubwürdig funktionieren, hat mit mehreren Entwicklungen zu tun. Einerseits werden im Bereich der Chatbots große Fortschritte gemacht, was kontextspezifische Ausdrucksweisen betrifft. Deshalb sind sie nicht bloß im Einsatz, um menschliche Arbeitskräfte in Call-Centern zu entlasten oder zu ersetzen. Sie werden immer mehr auch in Spielen oder in der therapeutischen Arbeit eingesetzt, weil die Datenbanken, auf denen sie beruhen, voller Formulierungen sind, die nicht nur inhaltlich korrekt sind, sondern auch spezifische Ausdrucksweisen für bestimmte Kontexte darstellen

4 Weizenbaum hat sich später kritisch mit den Entwicklungen im Bereich Software und Informatik auseinandergesetzt. Siehe dazu Weizenbaum (1978).

5 <https://www.talespin.com> (abgerufen am 25.4.2022)

und auf diese Weise eine emotionale Reaktion bei den menschlichen Interaktionspartnern hervorrufen können (Welleck et al. 2019, See et al. 2019, Rashkin et al. 2019, Burkhardt et al. 2018). Andererseits ist Kommunikation nicht auf Sprache beschränkt, sondern schließt auch Körpersprache ein. Über Gesten und vor allem über Gesichtsausdrücke bringen Menschen vor allem ihre emotionalen Reaktionen zum Ausdruck, ihre Bewertung der Situation oder des Interaktionspartners. All diese Ausdrucksformen sind durch psychologische Studien gut untersucht. Entsprechende Erkenntnisse werden von Disziplinen wie dem Affective Computing oder der Artificial Empathy genutzt, um die semantisch-inhaltliche Ebene, auf der die Spracherkennung beruht, zu erweitern.

So entwickelt etwa ein Team am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) interaktive Technologien, die ihre menschlichen Interaktionspartner:innen nicht nur in sprachlicher Hinsicht zu verstehen versuchen, sondern auch auf das reagieren, was in emotionaler Hinsicht in ihnen vorgeht (Gebhard et al. 2018). Die entsprechenden Prototypen bestehen in einer Interaktion mit einer virtuellen Figur, die auf einem Bildschirm erscheint. Durch den Verzicht auf ein Head-Mounted-Display wird es möglich, das Gesicht und die Bewegungen der Spieler:innen mittels Kameras aufzunehmen. Die Daten werden mittels einer Software verarbeitet, die auf dem System der gesichtsbasierten Emotionserkennung von Ekman und Friesen (1978) beruht.

Technische Lösung	Interaktionslevel
Chatbot	Austausch schriftlicher Nachrichten
Dienste wie Siri, Alexa	Dialog zwischen Mensch und Stimme
Interaktion mit NPC in VR	Dialog zwischen Mensch und animierter Figur
Interaktion aufgrund der automatisierten Erkennung von Körpersprache	Dialog zwischen Mensch und »empathischer« Maschine

Tab. 2: Übersicht über die Entwicklung sozial interaktiver Agenten in technischer Hinsicht und in Bezug auf die Interaktionsebene.

Mittlerweile werden auch andere Codierungs- und Decodierungssysteme eingesetzt. Das technische Prinzip ist dabei aber immer, den menschlichen Ausdruck in Zeichen zu verwandeln, die als Daten durch die Maschine verarbeitet werden können. Mit einer gewissen Wahrscheinlichkeit kann die Software dann bestimmen, wie sich ein Mensch gerade fühlt, und eine geeignete Reaktionsweise auswählen.

Allerdings haben Forschungen – wie oben erwähnt – festgestellt, dass gerade die automatisierte Gesichtserkennung äußerst anfällig für rassistische Vorurteile und diskriminierende Entscheidungen ist, wenn es zum Beispiel um die Auswahl von Bewerber:innen geht. Interessant ist für unseren Zusammenhang zu sehen, wie eng Möglichkeiten und Risiken der technologischen Entwicklung beisammen liegen. Einerseits hat die interaktive VR eine Reihe von Anwendungen ermöglicht, die sich mit den Themen Diversity, Rassismus und Diskriminierung auseinandersetzen. Andererseits birgt die Entwicklung interaktiver Technologien, die mit der automatisierten Erkennung von körperlichen Merkmalen und Ausdrucksweisen arbeiten, das Risiko von stereotypen und diskriminierenden Verhaltensweisen und Entscheidungen. Wie in Kapitel 2 schon diskutiert, ist es deshalb wichtig, den Sinn und Zweck und den Einsatzbereich solcher Anwendung gut zu durchdenken und möglichst breit zu diskutieren.

Empathiemaschinen oder empathische Maschinen?

Wie in Kapitel 1 erwähnt, war ein Ausgangspunkt für die Idee zum Virtual Skills Lab der Begriff der »Empathy Engine«, den Robert Praxmarer beim Ideen-Lab ins Spiel brachte. Die Verbindung von Virtual Reality und Empathie ist nicht neu. Sie wurde unter anderem durch Chris Milk verbreitet, der in einem vielbeachteten Vortrag von Virtual Reality als Empathy Machine sprach (Milk 2015). Als Filmemacher nutzte er die 360°-Video-Technologie, um das Leben eines syrischen Mädchens in einem jordanischen Flüchtlingslager zu dokumentieren. Der Film »Clouds over Sidra« ermöglicht es Zuseher:innen, die Lebensumstände dieses Mädchens und seiner Familie »hautnah« zu erfahren. Chris Milk präsentierte den Film beim Weltwirtschaftsforum in Davos, wo er hochrangige Entscheider:innen aus Wirtschaft und Politik mittels VR-Headset in die Realität des Flüchtlingsalltags eintauchen ließ. Sein Ziel war es, Betroffenheit und Interesse für die Situation der Millionen Geflüchteten zu wecken, damit sich diese verbessere. In seinem Vortrag ging er jedoch auch auf die Möglichkeiten dieses neuen Mediums ein. Dieses sollte in Zukunft als Mittel zur Erfahrbar-

keit von und damit zum Verständnis für andere, benachteiligte Lebensrealitäten genutzt werden.

REFLEXION

Wie ist Ihr Standpunkt zum Thema? Können solche Technologien Empathie fördern und sollen sie dazu genutzt werden? Was spricht für einen Einsatz von interaktiven Technologien zur Erweiterung der eigenen Perspektiven? Welche Argumente sprechen dagegen?

Die Frage ist, ob das tatsächlich funktionieren kann. In den letzten Jahren wurden daher einige Forschungen dazu angestellt. So entwickelte ein Team um den VR-Pionier Jeremy Bailenson und den Empathieforscher Jamil Zaki ein VR-Spiel mit dem Titel »*Becoming Homeless*« (Asher et al. 2018). In diesem Spiel schlüpft man in die Rolle einer Person, die zunächst ihren Job verliert und dann delegiert wird. In einer empirischen Studie fand das Team heraus, dass durch das Erleben der Szene aus der Perspektive des/r Betroffenen durchaus das empathische Empfinden gesteigert wird, sodass viele Teilnehmer:innen nach dem Spiel bereit waren, eine Petition für mehr leistbaren Wohnraum zu unterschreiben. Auch in anderen Prototypen wurden Reaktionen und Verhaltensweisen getestet, die als empathisch gedeutet werden können (Herrera et al. 2018, Van Loon et al. 2018, Troeger und Tümler 2020). Was Virtual Reality hier leisten kann, ist die Option eines körperlich erfahrbaren Perspektivenwechsels. Bis zu einem gewissen Grad ist es möglich, in die leibliche Erfahrung einer anderen Person hineinzuschlüpfen. Dadurch wird auch die kognitive Empathie angeregt, also das Vermögen, sich gedanklich auszumalen, was es zum Beispiel heißt, die Arbeit zu verlieren und was man in einer solchen Situation braucht. Dadurch entstehen auch Impulse im Sinne eines empathischen Handelns, wie aus der Studie »*Becoming Homeless*« hervorgeht.

Die Möglichkeit eines Perspektivenwechsels im Sinne der Erfahrung der Wirklichkeit anderer Menschen ist demnach ein Argument, das für den Einsatz von VR in Trainingsmaßnahmen spricht. Mit der Frage, inwiefern ein solches Hineinschlüpfen in die Rolle anderer zu einer Erweiterung der Perspektive in kognitiver Hinsicht und zu Handlungsimpulsen beitragen kann, beschäftigen sich die Arbeiten von Mel Slater bzw. des Kollektivs BeAnotherLab.⁶ Letzteres hat die »*Machine to be another*« entwickelt (Bertrand et al.

6 <http://beanotherlab.org/home/work/tmtba/> (abgerufen 25.4.2022)

2018). Diese besteht aus zwei VR-Headsets, auf denen jeweils Live-Kameras montiert sind. Die Spieler:innen bekommen dann das Live-Kamerabild der jeweils anderen Teilnehmer:in zugespielt, sodass sie im wörtlichen Sinn die Szene aus der Perspektive der anderen Person erleben und gleichzeitig miteinander interagieren können. Eine Forscher:innengruppe um Mel Slater nutzt die Möglichkeit des Aus-sich-Heraustretens durch die VR-Erfahrung für die Stärkung der Reflexionsfähigkeit (Osimo et al. 2015). In der VR-Erfahrung »Virtual Freud« haben Spieler:innen die Möglichkeit, sich selbst Ratschläge für Problemlagen zu geben, denen sie im Augenblick in ihrem Leben gegenüberstehen. Diese Art des Beratens mit sich selbst wird in der VR als Therapiesitzung inszeniert, in der die Ratschläge, die die Teilnehmer:innen für sich selbst eingesprochen haben, von einem Therapeuten vorgetragen werden, der äußerlich Sigmund Freud ähnelt. Diese eher spielerische Anwendung wurde dann von der Firma Bodyswaps weiterentwickelt.⁷ Die Firma bietet VR-Trainingsmaßnahmen an, bei denen etwa Führungskräfte üben können, ein Gespräch mit einem schwierigen Mitarbeiter zu führen. Dieser beschwert sich über seine Kollegen, doch die Führungskraft weiß, dass die Kollegen sich auch über seine mangelnde Kooperationsfähigkeit beklagen. Dann hat der/die Spieler:in die Möglichkeit, ein paar Worte an den Mitarbeiter zu richten. Diese Wortmeldung wird aufgenommen, und in der nächsten Szene betrachtet der/die Spieler:in sich selbst aus der Perspektive des Mitarbeiters. Darüber hinaus wird die Reaktion des/r Spieler/s/in mit verschiedenen Messinstrumenten ausgewertet. Diese Messung betrifft zum Beispiel den Blickkontakt, die Körpersprache und die Wortwahl.

Bei den beschriebenen Anwendungen steht die Frage im Vordergrund, inwiefern VR die Empathie von Menschen fördern und somit zu einem Mit-einander beitragen kann, das durch Offenheit und die Bereitschaft geprägt ist, die Perspektive der anderen einzunehmen. Über die verschiedenen technischen Möglichkeiten, menschliches Verhalten zu »lesen«, stellt sich aber auch die Frage, ob Maschinen selbst so etwas wie Empathie entwickeln können. Diese Frage hat eine erkenntnistheoretische und eine ethische Dimension. In erkenntnistheoretischer Hinsicht muss man diese Frage wohl verneinen. Wie eingangs schon hinsichtlich des Begriffs Intelligenz ausgeführt, ist auch Empathie an Lebendigkeit und lebensweltliche Vollzüge gebunden. Maschinen können menschlichen Ausdruck in Daten verwandeln, doch können sie sich nicht in dem Sinn in Menschen und andere Lebewesen einfühlen, wie

7 <https://bodyswaps.co/> (abgerufen am 25.4.2022)

Menschen und vielleicht manche Tiere dies tun. Wenn Einfühlung eine Art der Erfahrung ist, in der sich fremdes Bewusstsein kundtut (Stein 2008), dann kann Maschinen keine Empathie unterstellt werden, denn sie haben kein Bewusstsein und sind nicht lebendig. In moralisch-ethischer Hinsicht geht es darum, zu bestimmen, welche Funktionen eine simulierte Interaktion übernehmen kann, in der Menschen mit virtuellen Figuren interagieren, als handele es sich um Mitmenschen (Misselhorn 2021). Für den Bereich des Trainings sozialer Kompetenzen lässt sich dazu folgendes sagen: Wie aus den Interviews mit den Expert:innen hervorgeht, besteht die ethische Herausforderung bei solchen Anwendungen darin, die Lernenden einerseits aus der Reserve zu locken, ihnen also durchaus konfliktbeladene oder stressige Situationen zuzumuten. Andererseits darf die Eskalation nicht so weit getrieben werden, dass ethische Standards verletzt werden, wie zum Beispiel die Vermeidung von diskriminierendem Verhalten, Beleidigungen oder Beschimpfungen. Die emotionalen Reaktionen, die ein virtueller Agent zeigen kann, bzw. das Verhalten das diese Figur auslöst, dürfen nicht so gestaltet sein, dass allgemeine zwischenmenschliche Verhaltensstandards dadurch untergraben würden. Zwischen einer Lernerfahrung, die die Lernenden in Bezug auf ihre Sozialkompetenzen nicht fordert, und einer solchen, die völlig aus dem Ruder läuft, ist allerdings genug Spielraum für das Design sinnvoller Simulationen zum Training sozialer Kompetenzen. Dies kann durchaus Brüche in der Interaktion und vor allem Verläufe einschließen, die durch Mehrdeutigkeit und verschiedene mögliche Ausgänge geprägt sind, von denen keiner als der objektiv beste und adäquate bewertet werden kann.

Miras unsanfter Abgang und die Reflexion darüber

(HINTERGRUND)

Wie in Kapitel 2 skizziert, ist Empathie nicht die einzige soziale Kompetenz, die man trainieren kann und soll. Neben der Fähigkeit, die Perspektive anderer einzunehmen, zuzuhören und Kompromisse einzugehen, ist die komplementäre Kompetenz genauso wichtig, nämlich aus sich herauszugehen, also Entscheidungsfreude, Konflikt- und Durchsetzungsfähigkeit an den Tag zu legen. Darüber hinaus ist es wichtig, das eigene Verhalten in sozialen Situationen zu reflektieren, d.h. Gespräche vor- und nachzubereiten, sich präsentieren zu können und einschätzen zu lernen, wie die anderen einen wahrnehmen.

Aus dem Ko-Kreationsprozess ging als Thema das »wertschätzende Nein-Sagen« hervor. Damit war, wie erwähnt, der Ausgang der Geschichte in gewisser Weise festgelegt. In technischer Hinsicht hätte man die Geschichte dennoch so lösen können, dass man unterschiedliche Gesprächsverläufe anlegt und je nachdem, welche Antworten die Spieler:in gibt, die Situation mit einer gütlichen Einigung endet oder eskaliert. Wie der schwierige Mitarbeiter Barry hätte unser virtueller Agent auf manche Antworten eher aufbrausend und sehr enttäuscht reagiert und auf manche Antworten eher verständnisvoll. Aber wer legt fest, was in einem solchen Fall die richtige Reaktionsweise ist? Und was sollen sich die Spieler:innen aus der Erfahrung mitnehmen? Für den Prototyp entschlossen wir uns, die Einflussmöglichkeiten der Spieler:innen auf den Fortgang des Dialogs auf null zu begrenzen. Wir gaben zwar an jeder Stelle des Dialogs zwei oder drei alternative Antwortmöglichkeiten vor, doch führten diese nicht zu unterschiedlichen Reaktionen seitens der virtuellen Figur. Vielmehr sollten alle Personen, die den Prototyp testeten, dieselbe Erfahrung machen: Sie sollten ein unangenehmes Gespräch durchlaufen, das zunächst nicht so endet, wie sich der Spieler oder die Spielerin das wünscht. Die virtuelle Figur reagiert auf jeden Fall enttäuscht.

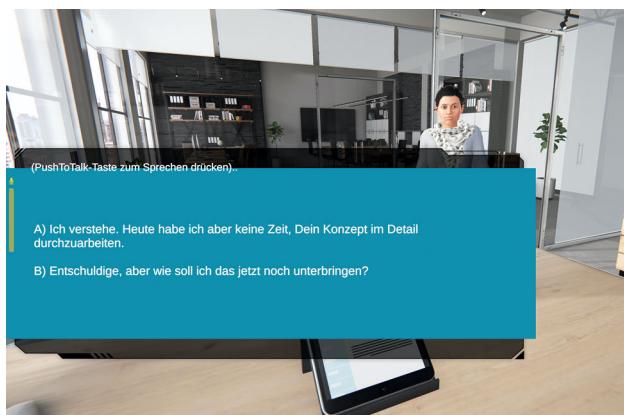


Abb. 6: Screenshot aus der interaktiven VR-Szene. Eingeblendet werden zwei Antwortmöglichkeiten.

Worum geht es in der prototypischen Szene? Als Spieler:in schlüpft man in die Rolle einer Führungskraft, die ein Team von ca. 10 Personen führt. Man sitzt im eigenen Büro, und gleich wird Mira Horvath, eine Mitarbeiterin, zur Tür reinkommen. Mira hat ein Template vorbereitet, das sie mit der Führungskraft durchgehen möchte, um es einer anderen Abteilung weiterleiten zu können. Das Meeting ist bereits zweimal verschoben worden, und nun erreicht die Führungskraft eine Textnachricht, die besagt, dass sie zu einer wichtigen Sitzung gehen muss. Es bleibt ihr also nichts weiter übrig, als das Meeting mit Mira erneut zu verschieben. Nach einem kurzen Dialog verlässt Mira verärgert das Büro.

Dabei bleibt es aber nicht. In der zweiten Sequenz erfährt die Führungskraft, dass Mira an diesem Tag früher nach Hause gegangen ist, und wird von einer anderen Mitarbeiterin gebeten, ihr etwas für den folgenden Tag auszurichten. Die Führungskraft hat nun die Aufgabe, eine Nachricht auf Miras Box zu sprechen. Dabei wird ihr freigelassen, ob sie zu dem schief gelaufenen Gespräch von vorhin Stellung nimmt oder nicht.

In der dritten und letzten Sequenz hat die Spieler:in dann die Möglichkeit, den Dialog Aussage für Aussage noch einmal durchzugehen und sich dazu zu äußern, wie sie an dieser Stelle reagiert hätte, wenn sie freie Wortwahl gehabt hätte. Darüber hinaus wird die Spieler:in immer wieder angeleitet, sich in Mira hineinzuversetzen und sich vorzustellen, wie sie sich in dieser Situation gefühlt hat.

Die Lernerfahrung besteht also aus zwei Komponenten. Zum einen wird Virtual Reality dazu genutzt, eine unangenehme Erfahrung aus der Sicht einer Führungskraft erfahrbar zu machen. Man bekommt die emotionalen Reaktionen der virtuellen Figur zu spüren und erfährt eine gewisse Hilf- oder Machtlosigkeit durch die vorgegebenen Antworten und die Eskalation, die man nicht verhindern kann. Zum anderen öffnet sich in der zweiten und dritten Sequenz der Reflexionsraum und ermöglicht den Spieler:innen, über soziale Kompetenzen allgemein bzw. über das eigene Verhalten in bestimmten Situationen nachzudenken.

»Was für eine Emotion können VR-Erlebnisse auslösen?«

(GESPRÄCH)

Die Entwicklung der Szene im engeren technischen Sinn stellte einen der zeitaufwendigsten Teile in unserem Projekt dar. Während des Ideen-Labs und über weite Strecken des Projektes prägte Robert Praxmarer die Arbeit an den technischen Komponenten der interaktiven Szene. Thomas Layer-Wagner war von Beginn an eng eingebunden und gestaltete nach Roberts Ableben die weitere Entwicklung zentral mit. Wie aus dem folgenden Gespräch ersichtlich wird, besteht der Prototyp nicht nur aus verschiedenen technischen Komponenten – diese verlangen ihrerseits hoch spezialisierte Kräfte, die in enger Kooperation ihren Beitrag zum Gesamtergebnis leisten. Jeder der Beteiligten kümmerte sich um eine bestimmte Dimension des interaktiven Geschehens, das die VR-Szene in ihrer prototypischen Form darstellt. Dass wir mit Mira sprechen können und dass sie uns versteht, verdanken wir dem Einsatz von Julian Watzinger, der tief in die Möglichkeiten und Grenzen der sprachlichen Verständigung von Mensch und Maschine eingedrungen ist. Denn dazu braucht es verschiedenste Ein- und Ausgabe-Technologien, die gesprochene Sprache in Text verwandeln können und die es umgekehrt der Maschine erlauben, adäquat auf menschliche Kommunikation zu reagieren. Daraüber hinaus muss es jemanden geben, der die Puppen zum Tanzen bringt. Für uns übernahm diese Aufgabe Wolfgang Stockinger, der Mira Horvath als 3-dimensionale animierte Figur designt hat.

Ideenfindungsphase

Klaus: Thomas, wie hast du die Idee aufgenommen, die am Ideen-Lab kreiert wurde?

Thomas: Während Robert am Ideen-Lab war, berichtete er mir in einem Telefonat, bei einem der möglichen Projekte gehe es darum, herauszufinden, ob man Verhandlungs- und Gesprächssituationen mit virtuellen Charakteren trainieren kann. Ich war gespannt, weil ich diese Möglichkeit in unterschiedlichen Kontexten unserer Entwicklungsarbeit bereits als sinnvoll wahrgenommen hatte. Wir orientierten uns zu diesem Zeitpunkt stärker in Richtung Ed-Tech mit dem Fokus Training und Bewusstseinsbildung. In diesem Bereich wollten wir angewandte Spieltechnologie und die damit verbundenen Konzepte nutzen. Was uns bei der Kombination aus VR-Erlebnissen und Kommunikationssituationen interessierte, war zunächst die Frage: Was bedeutet es, wenn man Dialoge in Spielen dynamischer gestalten kann, als dies bislang der

Fall ist? Was verändert sich, wenn gesprochene Sprache für die Eingabe zum Einsatz kommt? Die Frage nach der dynamischen Gestaltung von Dialogen in interaktiven Spielen hatte uns in anderen Projekten bereits beschäftigt, allerdings in einfacherer Form, über Texteingabe. Da hatten wir mit verschiedenen Entscheidungspfaden gearbeitet, über die man vom Gespräch aus mehr über den Hintergrund einer Geschichte erfahren konnte. Oder man kam über den Dialog zur Lösung eines Rätsels. Wir arbeiten vorzugsweise mit dem Spielkonzept des Escape Room. Bei diesem Konzept befinden sich Spieler:innen in einem (virtuellen oder realen) Raum, aus dem sie nur gelangen, wenn sie eine Aufgabe lösen. Der Wissensgewinn wird als Ausweg aus einer brenzligen Situation inszeniert. Insofern habe ich mir gedacht: Es interessiert mich, wie man die Ergebnisse dieses Ansatzes in ein Erlebnis integrieren kann wie dasjenige, das uns beim Virtual Skills Lab vorschwebte.

Klaus: Das Ergebnis sah dann so aus, dass es beim Dialog im Virtual Skills Lab zunächst keinen Ausweg gibt. Dieser eröffnet sich erst über die zweite und dritte Sequenz der Interaktion, wo man selbst Lösungen für den Konflikt formulieren kann bzw. über die Dynamik des schiefgelaufenen Gesprächs reflektieren kann.

Thomas: Für mich gab es noch einen zweiten Aspekt: Wenn wir in die Zukunft denken und schauen, was bei dem Vorhaben für uns rauskommt – es ist ja ein gemeinsames Ergebnis und betrifft nicht nur uns von Polycular, wobei die Zusammenarbeit mit so vielen Partnern für uns auch ein Experiment darstellte –, dann lautet die Frage, welche Vorteile VR hier bietet. Erste Hypothesen dazu haben wir im Antrag formuliert. Zum Beispiel: Es könnte sein, dass es für Leute, die eher schüchtern sind, eine einfachere und besser Möglichkeit bietet, Kommunikationssituationen zu trainieren, ohne auf Rollenspielpartner:innen angewiesen zu sein. Wir haben also von Beginn an Potenzial gesehen, etwa in Richtung der Ausbildung von Mitarbeiter:innen im Verkaufsbereich. Wir hatten klare Trainingsanwendungen vor Augen: zum Beispiel eine Situation, wo ich trainieren kann, was ich über ein Produkt erzählen muss, was den Leuten wichtig ist. Wie bringe ich das rüber? Was muss ich beachten, was soll ich vermeiden? Wir haben uns unvoreingenommen herangewagt und uns gefragt: Was kann dieses Virtual Skills Lab sein? Was könnte da rauskommen und wo könnten wir das dann auch einsetzen?

Julian: Ich bin erst in der Phase dazugekommen, als das Projekt schon in Planung war. Mich hat die Thematik der sozialen Kompetenzen angesprochen. Ich hatte da eine gewisse Vorerfahrung, weil ich mich mit der Methode des Neuro-Linguistischen Programmierens (NLP) beschäftigt hatte. Da geht es viel um nonverbale Kommunikation, beim Thema »mit anderen Menschen sprechen« zum Beispiel eher um die Frage: »Wie redest du?«, und weniger darum, was du sagst. Als ich diese Ausbildung machte, verwendete ich natürlich auch Bücher. Am besten jedoch war es im Seminar, weil man da mit Leuten zusammen ist und die Dinge üben kann, die man ansonsten nur liest. Die VR-Technologie bietet mir eine gute Möglichkeit, erfahrungsorientiert zu lernen, ohne ein Seminar zu besuchen. Denn wenn ich etwas lese und rausgehe und das ausprobieren, dann fragt mich mein Interaktionspartner vielleicht, ob ich auf den Kopf gefallen bin, weil ich gerade ausprobieren, wie es ist, wenn meine Stimme sich anders anhört. Die Möglichkeit, dies zunächst einmal für sich selbst in der virtuellen Welt auszuprobieren, habe ich am Virtual Skills Lab interessant gefunden.

Klaus: Das waren für dich mögliche Anwendungsfälle: zum Beispiel ein Seminar oder ein Buch, in dem Inhalte vermittelt werden, die dann in Form einer Simulation ausprobiert werden können.

Julian: Ja, ich bin beim Lernen eher Autodidakt, aber mit Soft Skills ist das schwierig. Die muss man ja in der Interaktion mit anderen ausprobieren und entwickeln. Wenn ich nun eine interaktive Technologie habe, dann kann ich das üben, ohne gleich rausgehen zu müssen.

Klaus: Im Sinne dessen, was Thomas vorhin sagte, würdest du tatsächlich in eine Zielgruppe für ein solches Training fallen. Wir mussten ja im Rahmen des Antrages Hypothesen aufstellen, was mögliche Vorteile dieser Technologie sein könnten. Eine Vermutung, die bereits in der Forschung behandelt worden war, ging dahin, dass Menschen, die sagen: »Ich möchte das mal für mich allein ausprobieren.«, einen Nutzen aus einer solchen Anwendung ziehen könnten. VR wäre dann eine gute Alternative zum Rollenspiel, weil man im virtuellen Raum eine zwar interaktive, aber sichere Umgebung vorfindet.

Julian: Genau. Rollenspiele fallen mir prinzipiell eher schwer. Aber in VR ist man unbefangener, da ist es einfacher, Dinge zu probieren. Da kann ich das Ganze in meiner eigenen Geschwindigkeit machen. So wie ich es mir vorstelle. Insofern würde ich tatsächlich in die Zielgruppe passen.

Klaus: Dein Hintergrundwissen aus den NLP-Kursen hat dich also motiviert. Da geht es stark um nonverbale Kommunikation, also um Gesten, Blickrichtungen usw. Das kann man in einer VR-Interaktion mit einer virtuellen Figur gut abbilden. Hat dich das dann auch weiter begleitet? Hast du den Eindruck, dass du dieses Vorwissen einbringen konntest?

Julian: In gewisser Weise ja. Ich habe zum Beispiel eine Funktion eingebaut, die sicherstellt, dass die virtuelle Figur dir direkt in die Augen schaut, dass die also »weiß«, wo dein Kopf ist. Bei den Tests hat man dann gemerkt, dass es einen Unterschied macht, ob diese Figur nur so dahinstarrt oder deinem Blick folgt. Wenn du zum Beispiel deinen Kopf zur Seite bewegst, dann schaut dich die virtuelle Figur immer noch an.

Anknüpfen an bestehende Erfahrungen

Klaus: Thomas, du hast ja schon angesprochen, dass ihr bei gewissen Vorerfahrungen angesetzt habt. Kannst du uns schildern, worin die bestanden?

Thomas: Was wir hatten, waren Chat-Systeme, über die die Nutzer:innen innerhalb des Dialogs mit den virtuellen Figuren an Hintergrundinformationen kommen können. Zudem hatten wir bereits VR-Projekte gemacht. Besonders Robert, der im Future-Lab an Projekten für den CAVE mitgewirkt hat.⁸ CAVE steht für Cave Automatic Virtual Environment. Anstatt ein Headset aufzusetzen, begibt man sich in einen Raum mit drei oder mehr Wänden, auf die eine 3-dimensionale virtuelle Umgebung projiziert wird. Man ist also in einer virtuellen Höhle und kann dort interagieren. Die Idee einen CAVE zu bauen wurde anfangs teils belächelt nach dem Motto: »Das werden sie eh nie schaffen.« Aber das Future-Lab, Robert war Teil des Teams, hat das dann doch geschafft und es entstand als erste immersive virtuelle Umgebung dieser Art in Europa. Allerdings stellt der CAVE über die Projektionen auf 5 Würfelseiten, in denen ich mich mit Shutter-Brillen befinde, eine andere Technologie dar als die heutigen HMDs.

Bei Polycular haben wir ebenfalls an einigen VR-Projekten gearbeitet, die recht unterschiedlicher Natur waren. Eines ging stärker in Richtung

⁸ <https://ars.electronica.art/newdigitaldeal/files/2021/08/futurelab25.pdf> (Seite 38). (abgerufen am 30.10.2022)

Visualisierung, Begehbarmachung von Architekturdaten. Ein eher künstlerisches, emotionales VR-Erlebnis, das wir designt haben, war ein Unterwasser-Schwimmausflug mit dem Titel »Swimming Pool«.⁹ Da tauchst du ab und hast das Gefühl, du bist im Paradies: Korallen, Fische. Wenn du dich aber weiterbewegst, siehst du stilisierte Wasserleichen herumschwimmen in dieser Idylle. Das war ein Erlebnis, das die Menschen mit der Situation von Flucht und Tod im Mittelmeer konfrontieren sollte. Dieses Thema erlangt leider immer wieder Aktualität. Bei diesem Projekt haben wir uns gefragt: »Wie sehr kann dich VR auf eine Reise schicken?« Das hat mich und Robert immer interessiert: Was für ein Erleben, was für eine Emotion können VR-Erlebnisse bei den Leuten auslösen? Können sie daraus etwas zehren, bringt das etwas auch aus künstlerischer Sicht? Kann man darüber etwas lernen? Das war damals bereits die Frage: Was löst das aus bei den Leuten? Wird das dann unmittelbarer, ist es dann für eine bestimmte Zeit ein Thema, mit dem sie sich stärker auseinandersetzen? Oder ist das völlig egal?

Komponenten des Prototyps

Klaus: Von dieser Erfahrungsbasis seid ausgegangen. Welche Herausforderungen habt ihr dann in der Phase der Erstellung des detaillierten Antrags gesehen? Wie habt ihr die Zusammenarbeit mit den anderen in dieser Phase erlebt?

Thomas: Auf unserer Seite ging es darum, zu konkretisieren, wie das Szenario aussehen könnte und was das in technischer Hinsicht impliziert. Was muss man dafür alles bauen, welche Teile der Software müssen an welcher Stelle miteinander interagieren? Julian, das war auch das erste Mal, als wir beide uns darüber unterhielten, welche Komponenten wir insgesamt benötigen. Wir versuchten dabei, nach den Sternen zu greifen, wie glaube ich alle im Konsortium. Was könnte man alles umsetzen und einbauen? Das begann bei der KI-basierten Emotionserkennung, also der Möglichkeit, maschinell zu erkennen, welche Emotionen der Nutzer oder die Nutzerin durchlebt. Wir waren uns relativ sicher, dass die Stimmanalyse und Spracherkennung gut funktionieren könnten. Diesbezüglich hatten wir bereits Erfahrungen gesammelt. Allerdings waren die Systeme damals noch nicht so weit wie jetzt. Und dann kamen noch die Fragen zur Darstellung der virtuellen Figur hinzu. Die muss sich

9 <https://www.polycular.com/portfolio/swimming-pool/> (abgerufen am 30.10.2022)

ja bewegen können, einen Gesichtsausdruck haben usw. Wir wussten, wir werden Audio brauchen, da war ebenfalls die Frage, wie weit KI da bereits entwickelt ist. Wie gut sind Tools zur Sprachverarbeitung nutzbar? Wir sahen Tools, von denen wir hofften, dass sich diese noch weiterentwickeln würden. In englischer Sprache ist das zum Teil auch passiert und die Sprachverarbeitung ist deutlich besser geworden. Im Deutschen ist das noch eingeschränkter.

Unsere Rolle haben wir in folgenden zwei Aufgaben gesehen: einerseits zu skizzieren, wie sich so ein Szenario anfühlen würde, was das überhaupt sein könnte, und andererseits zu überlegen, was für verschiedene Komponenten dazu gebraucht werden. Das war unser Rollenverständnis: einerseits Game-Konzepte zu designen und andererseits die Frage nach den Möglichkeiten der VR-Technologie zu bearbeiten. Für den Antrag entwarfen wir dazu ein Diagramm.

Technical Overview

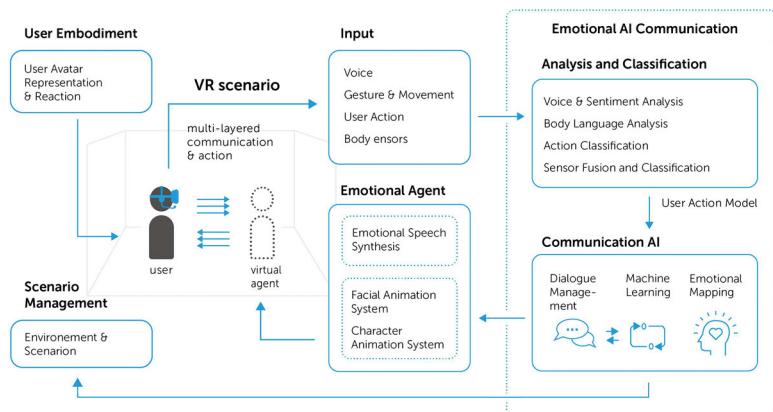


Abb. 7: Das ursprüngliche Diagramm zur Erklärung der Funktionsweise der technischen Komponenten.

Klaus: Ich glaube, dass diese Darstellung für alle wichtig war, zumindest habe ich das so erlebt. Ich hatte von den technischen Aspekten keine Ahnung.

So konnte ich erkennen, wo die neuralgischen Punkte und Herausforderungen liegen, die mit einer solchen Entwicklung verbunden sind. Man kann sagen, dass zwei Fragestellungen zusammengekommen sind. Einerseits die Frage nach dem Szenario: Welche Erkenntnisse liegen diesem aus organisationspsychologischer Sicht zugrunde? Andererseits die Frage der User Experience: Wie kann man ein solches Szenario mittels der VR-Technologie gestalten? An dieser Stelle berühren sich die Aspekte, die mit dem Training und mit der Interaktion zwischen Menschen zu tun haben, und die Aspekte, die man unter dem Begriff Mensch-Maschine-Interaktion zusammenfasst.

Thomas: In der Phase der Antragstellung war klar: Man wird einen Ablauf für die Nutzer:innen brauchen (Scenario Management). Damit ist die Art gemeint, wie die Nutzer:innen in die Szene eingeführt werden und was sie sehen. Welche Informationen brauchen sie? Welche Aktionen kann oder sollen sie in der virtuellen Umgebung durchführen können, um sich darin zurechtzufinden? Mit all diesen Fragen sind technische Problemstellungen und mögliche Lösungswege verbunden. Dann wird man verschiedene Formen von Input brauchen, also Systeme, die eine Eingabe mit der Stimme, mit Controllern oder mit einer Tastatur ermöglichen. Man wird einen Teil brauchen, der »versteht« und »analysiert«, was gerade passiert (Analysis and Classification). Der stimmliche oder textliche Input der Nutzer:innen muss von der Maschine seiner Bedeutung nach analysiert und eingeordnet werden. Dann muss ein sinnvolles Feedback zusammengestellt werden. Das Ganze sollte über Sprachausgabe, Gesichtsausdrücke und Animationssystem die Antwort des Avatars ergeben (Communication AI). Auf diese Weise kann ein Dialog zwischen einer realen Person und einer virtuellen Figur (Character, Avatar) entstehen.

Klaus: Wenn ihr diese Darstellung jetzt betrachtet, wie hat sie sich über den Entwicklungsprozess verändert?

Julian: In den Grundzügen entspricht das Resultat dem, was wir uns vorgenommen hatten. Wir haben daraus ein etwas zentrales System mit nur drei großen Komponenten entwickelt, nämlich Input, Output und Scenario Management. Diese kommunizieren dann miteinander, also letztlich ist es nicht so verschachtelt wie auf dem ursprünglichen Diagramm. Es ist simpler geworden, was auch damit zu tun hat, dass es ein längerfristiges Projekt ist. Viele dieser Komponenten bauen aufeinander auf. Am Anfang hatten wir noch kei-

ne Spracherkennung, sondern eine Texteingabe. Außerdem hatten wir am Anfang noch keine visuell gestaltete virtuelle Figur. Um zu sehen, dass das Szenario funktioniert, haben wir uns mal nur auf die Texteingabe und -ausgabe konzentriert. Ich gebe etwas ein, und die Maschine antwortet ebenfalls in Textform. Die erste Form des Szenarios war also ein Chatbot. Daraus hat sich langsam die jetzige Form entwickelt, nämlich aus den Komponenten Input, Output und dem Scenario Management, das mit beiden kommuniziert.

Klaus: Man muss es wohl im Rahmen des Prozesses noch einmal runterbrechen, weil man ja die einzelnen Komponenten für sich entwickeln muss. Daran kann ich mich erinnern. Ich habe die erste Form ja mal ausprobiert, bei der es nur um die sprachliche Kommunikation mittels Text ging.



Abb. 8: Unser Prototyp für ein interaktives Szenario¹⁰ begann mit einem Chatbot, der mit einer Spracherkennungssoftware verbunden wurde.

¹⁰ Auf diesem Foto sieht man Klaus Neundlinger, wie er eine frühe Version testet, indem er vorgegebene Sätze einspricht, auf die der Chatbot reagiert. Beobachtet wird er dabei von Julian Watzinger und Thomas Layer-Wagner.

Julian: Genau. Viele Dinge ergeben sich erst während der Entwicklung. Dann kann man erst einschätzen, was in der Praxis am besten funktioniert. Man muss viel ausprobieren und das Programm dann wieder umschreiben. Dann kommen noch Analyse und Debug-Features (Funktionen zum Finden und Beheben von Fehlern) dazu, wie zum Beispiel, wenn man direkt an bestimmte Stellen des Szenarios springen will, ohne die gesamte Interaktion durchspielen zu müssen. Man muss die gesamte Szenario-Struktur ändern, damit das funktioniert. Für die Entwicklungsarbeit ist das wichtig, weil man zu Probezwecken ja immer wieder einzelne Teile durchspielen oder anschauen muss. Und was dann noch dazu gekommen ist: Dieses Diagramm stellt ja nur die direkte Interaktion mit der virtuellen Figur dar, also das Hauptstück des Szenarios. Um dieses herum ist ja viel entwickelt worden, etwa die Einführung in die Szene am Anfang und dann vor allem der zweite Teil, die Reflexionsphase zum Szenario. Da ist also viel dazukommen, was wir in diesem Diagramm nicht vorausgesehen hatten, was dann aber für die didaktische Aufbereitung und die verschiedenen Studien im Rahmen der Evaluation wichtig war.

Thomas: Wir hatten zwar in der Antragsphase darüber nachgedacht, dass man die Szene erweitern kann. Zum Beispiel, indem man die Perspektive wechselt. Es sollte etwas passieren, das einen in irgendeiner Form über die Interaktion reflektieren lässt. Aber es war noch völlig offen, ob es sich um die App handelt, die später als Begleitung zum VR-Training entwickelt wurde, oder um etwas anderes. Da waren noch viele Wege offen.

Sackgassen

Thomas: Zu Beginn haben wir allerdings ein paar Wege eingeschlagen, die sich später als Sackgasse herausstellten. Bei der Sprachsynthese etwa mussten wir einsehen, dass wir im Rahmen dieses Projektes nicht weit genug kommen werden. Es war interessant, sich das anzuschauen, aber selbst im Englischen (wo man schon weiter ist als im Deutschen) ist das noch schwierig.

Ein weiterer Punkt, den wir zu Beginn dynamischer einschätzten, war die Möglichkeit, im Sinne einer Spielerfahrung Entscheidungen zu treffen, die zu unterschiedlichen Entwicklungen führen. Wir haben dann aber gesehen, dass es mit dem Budget schwierig wird. Wir hatten zwar eine Hardware für das Motion Capturing angeschafft, die hat aber nicht geleistet, was wir uns erwartet hatten. Beim Motion Capturing werden mit einem bestimmten Verfahren reale Menschen und ihre Bewegungen filmisch erfasst. Diese Aufnahmen bilden

die Grundlage für den Körper und die Bewegungen von computeranimierten Figuren. Wir mussten daher als Konsortium Entscheidungen treffen, wie man Teile des Szenarios einfacher lösen kann. Das war notwendig, um die Anzahl der Varianten nicht explodieren zu lassen, auf die der virtuelle Agent reagiert. Was hätte das dann für die Produktion bedeutet? Theoretisch wäre es möglich gewesen, und wir haben uns die Kompetenzen erarbeitet, um das auch machen zu können. Hinsichtlich des Budgets hätte das aber einen hohen Aufwand bedeutet, egal ob es mittels Computergraphik (CGI) umgesetzt worden wäre oder mittels eines 360°-Videos. Es hätte bedeutet, dass man viele unterschiedliche Varianten einer Szene produzieren muss. Bei 360°-Videos hätte es bedeutet, es sind noch mehr Takes aufzunehmen und auch bei der mittels Computergraphik animierten Variante hätte es ein Vielfaches an Gesichts-, Sprach- und Körperanimationen gebraucht. Andererseits hätte es den Vorteil eines dynamischeren Szenarios geboten. Jedenfalls hätte uns interessiert, wie die Reaktionen seitens der Nutzer:innen ausgefallen wären, wenn wir unterschiedliche Varianten der Geschichte ausformuliert und produziert hätten.

Julian: Eine weitere Sackgasse war die Sentiment-Analyse, also das Erkennen von Stimmungen seitens der Maschine. Da haben wir keine sinnvollen Lösungen gefunden. Da hätte man auch würfeln können, so zufällig waren die Ergebnisse, die da rausgekommen sind. Wir konnten ja keine eigene Software schreiben, sondern mussten bestehende Software-Lösungen nutzen. Wir hatten diesen Weg ja auch bei Speech-to-Text gewählt, aber was Speech-to-Sentiment angeht, waren die Lösungen sehr schlecht.

Klaus: Kannst du erklären, was eine Sentiment-Analyse ist?

Julian: Das ist ähnlich wie bei der Speech-to-Text-Analyse, doch geht es nicht darum, dass die Maschine erkennt, was die Person sagt. In diesem Fall muss die Maschine erkennen, in welchem Gefühlszustand die Person etwas sagt: Ist sie wütend? Freut sie sich? Eine solche Analyse erfolgt über bestimmte Parameter, sodass die Maschine eine bestimmte Wahrscheinlichkeit errechnen kann. Sie kann zum Beispiel eine 90 %ige Wahrscheinlichkeit dafür errechnen, dass die Person gerade wütend ist. Für diese Art von Analyse wurden bereits Softwarelösungen entwickelt. Für uns war klar, dass wir das nicht selbst entwickeln können. Die Lösungen, die wir geprüft haben, haben das allerdings nicht gut gemacht. Wir haben das mit vielen Leuten ausprobiert. Jeder hat mal probiert, reinzuschreien oder freundlich zu sein. Aber das, was zurückkam, war nie das,

was wir uns erwartet hatten. Aus diesem Grund war es nicht sinnvoll, das zu verwenden.

Thomas: Ich hatte das Gefühl, in unserem Fall war diese Software auch deshalb nicht gut einzusetzen, weil man im Szenario sehr kurze, knappe Antworten geben muss. Unter Umständen funktioniert die Erkennung besser, wenn man zwei, drei Sätze mehr spricht. Aber das hat mit unserem Szenario nicht zusammengepasst.

Julian: Doch selbst was Speech-to-Text angeht, mussten wir viel nachbearbeiten. Ich habe sehr viel Zeit darin investiert, die Spracherkennung noch einmal zu verbessern. Was da zurückkommt, entspricht nicht immer dem, was du haben willst. Im Szenario sind die Antworten vorgegeben, und basierend darauf kann man die Eingabe dann nachbearbeiten. Wenn wir nur die angekauften Lösungen verwendet hätten, ohne den Algorithmus, den ich draufgepackt habe, dann wäre die Qualität auch deutlich schlechter.

Klaus: Du hast einen Algorithmus entwickelt?

Julian: Ja. Da geht es darum, den Text, der aus dem System zurückkommt, mit den vorgegebenen Antworten zu vergleichen. Wenn der Nutzer oder die Nutzerin etwas sagt, was nichts mit den vorgegebenen Antworten zu tun hat, wird das nicht zugelassen. Wenn es hingegen ähnlich ist, soll es erkannt und zugelassen werden. Im Prinzip versucht der Algorithmus, möglichst viele Wörter in der richtigen Reihenfolge zu finden. Dann werden die Keywords noch mit reingerechnet, und so wird die Wahrscheinlichkeit kalkuliert, mit der es sich bei dem Text um eine der vorgegebenen Antworten handelt. Das ist die Grundlage dessen, was ich gemacht habe. Und das hat die Erkennung deutlich verbessert. Also wenn du zum Beispiel sagst: »Ich habe jetzt kein_Zeit.«, dann erkennt der Algorithmus, dass »kein« und »keine« vermutlich dieselbe Bedeutung haben. Um auf die Sentiment-Analyse zurückzukommen, das können wir hier nicht wirklich machen, weil wir nicht wissen, wie der Nutzer oder die Nutzerin sich verhält, welche Gefühlsintention sich hinter seiner oder ihrer Aussage verbirgt. Es gibt vielleicht eine Grammatik des Gefühlsausdrucks, aber diese ist wesentlich komplexer als diejenige der Sprache.

Visualisierung, Darstellung der virtuellen Figur

Klaus: Was die anderen Aspekte der virtuellen Figur betrifft, also die Bewegung, die Gesten und Mimik usw., wie ist es euch da ergangen?

Thomas: Wir hatten im Rahmen des Projekts früh Kontakt zu Wolfgang, der mittlerweile fix bei uns im Team arbeitet. Wolfgang ist bei uns 3D-Artist und Technical Artist. Er hat sich auf das Motion Capturing spezialisiert. In den anfänglichen Gesprächen und in der Zusammenarbeit mit ihm ging es um die Frage: Avatare, virtuelle Figuren, Bewegungen, was bedeutet das? Wir konnten da ebenfalls auf Vorerfahrungen zurückgreifen. Bereits in unserer Zeit an der Fachhochschule hatten Robert und ich mit Animationssystemen gearbeitet, mit computeranimierten Figuren, mit Bewegungs-Blending, aber auf einem abstrakteren Niveau. Das war nicht so auf einen Anwendungsfall zugeschnitten, sondern da haben wir aus verschiedenen Pools und Datenbanken (Librarys) an Animationen einen Ablauf zusammengestellt, um eine historische Situation zu schildern und zu zeigen. Konkret war das ein Projekt für ein Museum in Radstadt, bei dem die Geschichte der Bauernaufstände, der Bauernkriege erzählt wurde. Die Bewohner:innen von Radstadt waren unmittelbar von diesen Aufständen betroffen, sind auch belagert worden. Das Ziel dieser Animation war, Einblicke in die Zeit zu geben und zu vermitteln, dass es eine spannende Epoche war, weil da in ganz Europa viel passiert ist. Beispielsweise sollte eine Situation dargestellt werden, in der ein Anführer Leute einschüchtert. Also haben wir Animationen zusammengesucht, bei denen zuerst ein Charakter auf einen anderen einredet und gestikuliert und dann sein Gegenüber aber beim Kragen packt und mit der Hand eine drohende Geste macht. Da war auch schon die Fragestellung: Kann man die Bewegungs-Librarys irgendwie mischen, kann man damit eine Geschichte erzählen? Das waren immer schon Fragen, die Robert und ich uns gestellt hatten.

Von Wolfgang wussten wir, er hatte auch schon eine VR-Experience umgesetzt, bei der im Team ein aufwendiges Motion-Capturing gemacht worden war. Einen solchen Aufwand hätten wir in diesem Projekt allerdings nicht betreiben können: Nachbau von Charakteren anhand von Face-Scans, die gesamte Person in 3-D scannen und so weiter. Das wäre in unserem Projekt mit zu hohen Kosten verbunden gewesen. Aber wir hatten mit Wolfgang jemanden, der die Expertise im Bereich Motion Capture einbrachte.

Wir mussten uns vor dem Hintergrund des Budgets überlegen: In welchem Umfang produzieren wir Motion-Capture-Aufnahmen? Es ist nicht sinnvoll,

die Hälfte des Budgets in die Hard- und Software für ein Motion-Capture-System zu stecken. Denn das war ja nicht das einzige Problem, das wir angehen mussten. Damit waren Limitierungen verbunden. Von unserer Seite war die Frage zu lösen: Wie können wir mit diesen Einschränkungen möglichst gut umgehen? Was Gesicht und Oberkörper betrifft, sind wir recht weit gekommen. Was viel problematischer und schwieriger war, waren die Bewegungsabläufe. Wir hatten eigentlich gute Erfahrungen mit anderen Systemen diesbezüglich. Insofern haben wir das vielleicht ein wenig überschätzt, wie gut dieses System das liefern könnte.



Abb. 9: Die Schauspielerin Christine Winter, die Mira Horvath im 360°-Video verkörperte, bei den Motion-Capture-Aufnahmen.

Julian: Man darf sich das nicht so vorstellen, dass man jemanden in einen Anzug steckt, Aufnahmen macht und das ins Projekt reinwirft. Da muss man viel nachbearbeiten, das ist mit viel Handarbeit verbunden. Teilweise sind Sensordaten nicht richtig, dann kommen Posen raus, die nicht realistisch sind. Da ist auf einmal der Ellbogen im Oberkörper.

Thomas: Das ist sicher ein Bereich, der sich in den nächsten Jahren stark weiterentwickeln wird, wenn man an KI denkt. Über die Software, mit der man die Bewegungsabläufe aufnimmt, wird man dann das manuelle Bearbeiten wesentlich verbessern können. Man bekommt dann bessere Daten, mit denen man weiterarbeiten kann. Wir haben uns mit Leuten aus dem Filmbereich unterhalten, die so etwas zum Teil für Hollywood-Produktionen machen und in der Game-Engine-Entwicklung arbeiten. Die sagen übereinstimmend: Je höherwertiger und deshalb auch kostspieliger das System, umso besser wird es. Und selbst die machen zusätzlich zum Motion Capturing Videoaufnahmen, damit sie Fehler wie die beschriebenen mittels Nachbearbeitung korrigieren können.

Julian: Auch wenn die Aufnahmen perfekt sind, ist ein virtueller Agent nicht das 1:1-Ebenbild der Person, die das aufnimmt. Wenn die virtuelle Figur ein bisschen dicker oder dünner ist, muss man das entsprechend anpassen. Wir haben für die Vergleichsstudie vier verschiedene Avatare konstruiert. Das basiert immer noch auf denselben grundlegenden Motion-Capture-Aufzeichnungen, die mit einer Schauspielerin gemacht wurden. Die mussten dann auch bearbeitet werden, weil zum Beispiel eine deutlich korpulenter Person eine ganz andere Gangart und Skelettstruktur hat.

Thomas: Hier hat Wolfgang noch einmal viel Arbeit hineingesteckt. Da braucht man sehr viel Erfahrung.

Klaus: Die Frage ist, welches Ziel man mit der Darstellung erreichen will. Wie realistisch möchte man das gestalten? Ist das überhaupt erstrebenswert [siehe dazu Kapitel 3]? Ich fand es wichtig, im Rahmen des Projekts unterschiedliche Darstellungsweisen zu vergleichen, nach Alter, Geschlecht, Körperbau usw. Dies scheint ja große Auswirkungen darauf zu haben, wie das Szenario erlebt wird. Daraus muss dann abgeleitet werden, wie man die Möglichkeit, die Figur unterschiedlich aussehen zu lassen, in praktischen Anwendungen zum Einsatz bringt.

Thomas: Die Möglichkeit der unterschiedlichen Darstellung wird im Rahmen der Evaluation genutzt, um zu sehen, welche Reaktionen zum Beispiel ein anderes Geschlecht oder Alter bei den Spieler:innen hervorrufen, obwohl alles andere gleich bleibt, also der Dialog, die Umgebung, die Situation. Nur die Stimme ändert sich und damit verbunden der emotionale Ausdruck des Avatars. Es

ist eine spannende Erfahrung: Ich spiele exakt die gleiche Interaktion mit einer Frau durch und mit einem Mann durch – ändert das bei mir irgendwas [siehe dazu Kapitel 5 und 6]?

Zusammenarbeit im Konsortium

Klaus: Wie habt ihr den gemeinsamen Entwicklungsprozess im Konsortium erlebt?

Thomas: Diese Zusammenarbeit wurde lange Zeit von Robert gestaltet. Ich habe mich zwar viel mit ihm ausgetauscht, doch beschränke ich mich auf die Phasen und Aspekte, die ich selbst mitgestaltet habe. Natürlich gibt es innerhalb eines solchen Konsortiums unterschiedliche Zielsetzungen, Schwerpunkte und Interessenlagen. Wir waren durchaus daran interessiert, unterschiedliche Erlebnispfade und Ausgänge zu generieren. Aus wissenschaftlicher und psychologischer Sicht ist hingegen eine andere Herangehensweise gewählt worden. Es ist schwer zu entscheiden, welche der beiden besser ist. Es führt einfach zu sehr unterschiedlichen Ergebnissen. Aus meiner Sicht ist es spannend, mit Teams aus unterschiedlichen Fachbereichen zusammenzuarbeiten. Manchmal stellt die jeweilige Fachsprache ein Hindernis dar. Man braucht eine gewisse Zeit, bis man eine gemeinsame Sprache entwickelt hat. Was angenehm und hilfreich in der Zusammenarbeit war, war das Vertrauen der anderen in unsere Fähigkeit, in technischen Belangen die Dinge gut beurteilen zu können. Vor allem bezüglich dessen, was machbar ist und was weit über die Möglichkeiten im Projekt hinausschießt. Darüber hat es nie große Diskussionen gegeben, eher über Nuancen bestimmter Sachverhalte. Klar, es ist leichter, wenn man Dinge nur innerhalb eines eingespielten Teams ausmacht. Aushandlungsprozesse lassen sich dann schneller durchführen. In einer Konstellation wie dem Virtual-Skills-Lab-Konsortium, wo wir uns am Anfang noch gar nicht kannten, ist das anders. Beim Start des Projekts wussten wir auch nicht genau, wo die Reise hingehen wird. Es war ein Teil des Prozesses, das gemeinsam rauszufinden. Wir wussten noch nicht, wo die Schwerpunkte der anderen genau liegen, was sie herausfinden wollen. In diesem Spannungsfeld haben wir versucht, für die unterschiedlichen Seiten das zu bieten, was in ihrem Forschungsinteresse lag. Auf der einen Seite Usability (AIT), dann die Frage nach den psychologischen Wirkungen (KL Uni), dann die Frage, wie man so etwas in einem Workshop einsetzen kann (ice). Das sind doch drei unterschiedliche Interessenlagen.

Klaus: Dazu kommt noch die allgemein gesellschaftliche, gesellschaftspolitische Dimension (IHS). Da es sich beim Virtual Skills Lab um ein Forschungsprojekt handelt, ist der gemeinsame Prozess offen. Man formuliert eine Fragestellung und schaut, wo man hinkommt. Das hat es aus meiner Sicht für uns als Konsortium leichter gemacht. Es gab zwar immer wieder Punkte, an denen wir Entscheidungen treffen mussten. Aufgrund der Offenheit hat man wohl mehr Möglichkeiten, aufeinander einzugehen. Unser Unternehmenspartner hat wohl andere Vorstellungen gehabt. Die hätten vielleicht mit geschlossenen und konkreteren Fragestellungen mehr anfangen können.

Thomas: Wobei wir unserem Unternehmenspartner von Anfang an klar gemacht haben, dass das ein offener Prozess ist. Bei einem großen Konzern kommt es drauf an, wie das kommuniziert wird, wer dafür zuständig ist, wie die Leute sich verhalten, die in diesen Prozess mit einbezogen werden, aus welchen Ebenen die kommen. Für den Unternehmenspartner war das etwas Neues. Auch das Setting in der Ko-Kreation war offenbar etwas Neues für dieses Unternehmen. Es wurde dann relativ konkret, was einerseits gut, andererseits vielleicht nicht so gut war. Wenn man uns ein relativ enges Korsett anlegt, kann das gut sein. Andererseits kann auch eine große Offenheit Vorteile haben, weil man im Tun, im Experimentieren auf gewisse Dinge draufkommt und gewisse Dinge dynamisch oder anders macht. Von unserer Seite als Designer hätten wir es uns einfacher machen können, wenn wir einige Dinge klarer kommuniziert hätten. Dass man zum Beispiel die Szene so gestalten hätte sollen, dass Mira zu Beginn der Szene nicht zur Tür reinkommt, sondern bereits am Schreibtisch sitzt. So eine Szene hatten wir schon, mit einem Avatar, dessen Körper aufgrund der Sitzhaltung nur ab dem Rumpf sichtbar ist. Damit hätten wir in Bezug auf die Animation weniger Probleme gehabt. Da hat dann der Wunsch überwogen, es allen recht zu machen. Wir wollten uns auch selbst beweisen, dass wir das können. Hier könnte es hilfreich sein, einem technischen Entwicklungspartner Spielräume zuzugestehen, sodass dieser sagen kann: »Lasst uns das noch einmal ändern.« Denn der Aufwand, im Script noch einmal Änderungen vorzunehmen, ist wesentlich geringer als der technische Aufwand, das alles so zu realisieren.

Klaus: Ich fand es nicht schlecht, dass durch Miras Auf- und Abtritt Bewegung ins Szenario kommt. Aber ich habe auch mitbekommen, wie viel Aufwand das für euch bedeutete. Und ihr seid wohl bis heute noch nicht ganz zufrieden mit dem Ergebnis.

Julian: Nein, aber das ist immer so. Viele Dinge sind supercool, wenn sie gut funktionieren. In einem Projekt mit begrenztem Budget und begrenzter Zeit muss man da aber Abstriche machen.

Klaus: Man hätte noch einmal eine Schleife ziehen und fragen: Können wir das, was wir ausdrücken wollen, auch ohne den Auf- und Abtritt von Mira rüberbringen. Allerdings hat dieses Auf- und Abtreten seinen Sinn: Es transportiert viel von der Emotion, einerseits die Erwartung, mit der sie am Anfang reinkommt, andererseits die Enttäuschung, mit der sie am Ende rausgeht. Dramaturgisch ist das absolut schlüssig. Aber es ist leichter gefilmt, als es in der Computeranimation zu realisieren ist.

Thomas: Man hat das eh im 360°-Video gesehen. Das wurde ja gemacht, um einen Vergleich zu haben. Wir sind keine Filmcrew, sondern wollten das einfach mal ausprobieren, wie sich das übersetzen lässt, wo da die Hürden und Probleme liegen.

Klaus: Man muss sich immer wieder anschauen, was für den jeweiligen Zweck die schlüssigste Variante ist. So wie es jetzt ist, ist es eine klassische TheaterSzene: Auftritt, Abtritt.

Thomas: Es hat auf jeden Fall auch etwas Filmisches.

Potenzial

Klaus: Zum Schluss möchte ich euch noch fragen, wie ihr das Potenzial und den Nutzen dieser Technologie einschätzt. Wo wird die Reise hingehen?

Thomas: Es wird spannend sein zu sehen, wie sich die interaktive Szene in einen größeren Kontext einbetten lässt, in ein Trainingsprogramm, wo zu dem Thema auf unterschiedliche Weise gearbeitet wird. Für uns stellt sich dann die Frage: Was müsste man noch dazu bauen, um eine solche Einbettung zu unterstützen, sodass es auf der digitalen Ebene alleine funktioniert? Oder muss das eh gar nicht sein und man kann es im Idealfall in einer Workshop-Situation gut einsetzen. Und an welcher Stelle innerhalb eines Lernpfades wirkt das VR-Training am besten? Diese Fragen sind ja noch offen und die wollen wir gemeinsam bearbeiten. Das sind die Punkte, die für uns spannend sind. Wir haben uns in Bezug auf die technische Umsetzung viel an Kompetenz angeeig-

net und haben noch einige Dinge in der Hinterhand, die für den Prototyp nicht verwendet wurden. Vor allem könnten wir es in technischer Hinsicht dynamischer umsetzen. Wir sind auch gespannt, was bei dem Vergleich zwischen der immersiven Variante mit VR-Headset und der Desktop-Version ergeben wird. Ich würde mir erhoffen, dass das Eintauchen in die Szene, das durch die VR-Umgebung erzeugt wird, einen positiven Beitrag leistet. Aber selbst wenn es das nicht ist, sind es andere Erkenntnisse, die wir daraus mitnehmen können.

Julian: Da kann ich mich nur anschließen. Für mich persönlich ist auch die computergenerierte VR-Variante die interessanter, vom Technischen und von den Möglichkeiten her, wo es hingehen könnte. Die Video-Variante diente für mich eher zum Vergleich, obwohl bei den Evaluationen scheinbar das Video besser angekommen ist. Aber das heißt vielleicht nur, dass die computergenerierte Umsetzung noch nicht 100 % ausgereift ist.

Thomas: Wir haben ja auch vor, das gemeinsam weiterzuentwickeln, also gemeinsam in neue Gefilde vorzudringen.

