

# Animation der Mimik und Gestik von Unity Charakteren auf Basis von Emotionen

Bachelorarbeit

Amir Safarha

Matrikelnummer: 4526084

09. Januar 2023



Fachbereich 3 — Mathematik und Informatik  
Studiengang Wirtschaftsinformatik

1. Gutachter: Prof. Dr. Rainer Koschke
2. Gutachterin: Dr. Hui Shi



# ERKLÄRUNG

---

Ich versichere, diese Arbeit — sofern dies nicht explizit anders gekennzeichnet wurde — ohne fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, sind als solche kenntlich gemacht.

*Bremen, den 09. Januar 2023*

---

Amir Safarha



## **GENDER-HINWEIS**

---

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.



# INHALTSVERZEICHNIS

---

1	Einführung	1
1.1	Formatierung . . . . .	1
1.2	Motivation . . . . .	1
1.3	Aufgabenstellung . . . . .	3
1.4	Fragestellung . . . . .	3
2	Konzept	5
2.1	Emotionstheorie . . . . .	5
2.2	Mimik . . . . .	7
2.3	3D-Modellierung . . . . .	8
2.3.1	Meshes und Materials . . . . .	8
2.3.2	BlendShapes . . . . .	9
2.4	Aktuelle Implementierung . . . . .	9
2.5	Zwischenfazit . . . . .	9
3	Implementierung	13
3.1	3D-Modell . . . . .	13
3.2	Animation . . . . .	14
3.3	Emotionsermittlung . . . . .	17
3.4	Zwischenfazit . . . . .	18
4	Evaluation	19
4.1	Hypothesen . . . . .	19
4.2	Aufbau . . . . .	20
4.3	Fragebogen . . . . .	21
4.4	Auswertung . . . . .	22
4.4.1	Demografische Auswertung . . . . .	22
4.4.2	Wiedererkennung von Mimiken . . . . .	23
4.4.3	Subjektive Wahrnehmung . . . . .	27
4.4.4	Erkennung dargestellter Emotionen . . . . .	28
4.5	Threats to Validity . . . . .	31
4.5.1	Interne Validität . . . . .	32
4.5.2	Externe Validität . . . . .	33
4.6	Zwischenfazit . . . . .	34
5	Fazit	35
A	Glossar	37
B	Akronyme	39
C	Abbildungsverzeichnis	41



# EINFÜHRUNG

---

Diese Bachelorarbeit befasst sich mit der Animation von spielbaren Charakteren in 3D-Umgebungen auf Basis menschlicher Emotionen. [Kapitel 1](#) erläutert den Aufbau dieser Arbeit, die Motivation dahinter und die zu untersuchende Fragestellung.

## 1.1 FORMATIERUNG

In dem folgenden Absatz wird beispielhaft dargestellt, wie der Text dieser Arbeit formatiert ist. In der digitalen Version sind alle nachfolgend genannten Objekte anklickbar und leiten den Leser zum Ursprungsort des Objektes.

Fachbegriffe und Akronyme, die erstmalig in dieser Arbeit genutzt werden, werden *braun* geschrieben und mit Randnotizen erklärt.<sup>1</sup> Bei jedem weiteren Vorkommen passt sich die Formatierung von Fachbegriffen dem übrigen Text an. Sie sind in der digitalen Version dennoch mit einem Link versehen. Zitiert wird nach dem *Deutsches Institut für Normung (DIN)*-Standard 1505. Ein Zitat wird stets in einer *purpurnen* Farbe abgebildet. Ein Beispiel dafür sähe wie folgt aus (Hays u. a., 1992). Referenzen auf andere Abschnitte dieser Arbeit werden in einer *blauen* Farbe dargestellt, wie es bereits am Anfang von [Kapitel 1](#) zu sehen ist.

*DIN: Eine unabhängige Plattform für Normung und Standardisierung.*

## 1.2 MOTIVATION

Die Implementierung von 3D-Umgebungen für die unterschiedlichsten Anwendungsfälle nimmt stetig zu. So wird die Virtualisierung und Visualisierung verschiedener Szenarien für ein risikoärmeres oder effizienteres Arbeiten genutzt. Ein gutes Beispiel hierfür ist die Luftfahrt, in der festgestellt wurde, dass Flugtrainings in Kombination mit Simulatoren effektiver sind als reine Flugstunden (Hays u. a., 1992), weswegen die Flugsimulation heutzutage ein nicht wegzudenkender Bestandteil einer jeden Flugausbildung ist.

Auch im Bereich der Software-Quellcodeanalyse wurde ein Ansatz der Visualisierung gefunden. Große Softwareprojekte und ihre Vielzahl an verschiedenen Komponenten werden nun in Form von Bauwerken einer Stadt dargestellt. Dieser Ansatz trägt den Namen *Code-City* (Wettel und Lanza, 2008). Die Merkmale dieser Gebäude definieren ausgesuchte Metriken der jeweiligen Softwarekomponenten. Infolgedessen könnte

---

<sup>1</sup> Sofern dies nicht bereits im Text geschehen ist.

die Höhe eines Bauwerkes beispielsweise die Anzahl der Codezeilen einer Komponente widerspiegeln.

*Software Engineering Experience (SEE)* ist eine an der Universität Bremen entwickelte Visualisierungssoftware, welche mittels der *Unity*-Engine implementiert wurde.

**Unity:** Eine der weltweit führenden Plattformen für die Erstellung interaktiver und plattformunabhängiger 3D-Umgebungen, basierend auf den Programmiersprachen C# und C++.

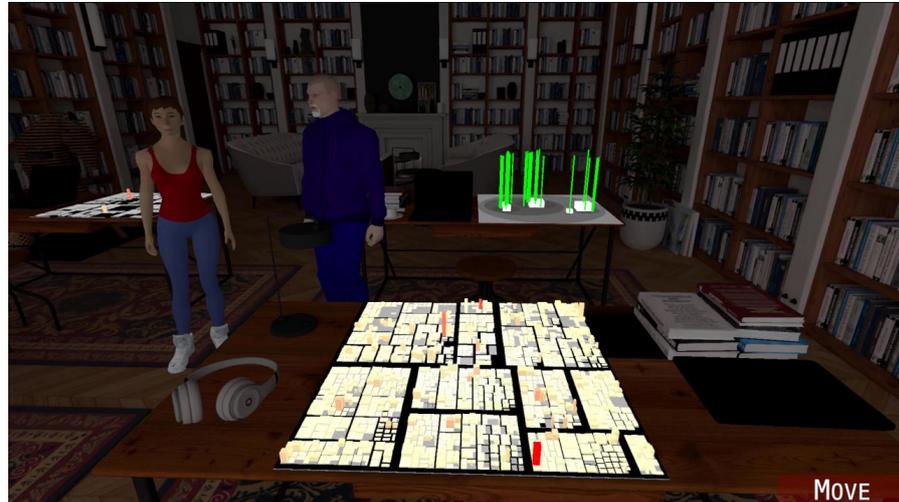


Abbildung 1.1: Darstellung einer Code-City in SEE (Koschke, 2021)

Dieses Projekt fasst ebenfalls den Gedanken der Code-Cities zur Softwareanalyse auf. [Abbildung 1.1](#) zeigt den Quelltext von SEE, dargestellt als Code-City, welche als dreidimensionale *Tree Map* zu interpretieren ist. Die einzelnen Komponenten werden als Blöcke dargestellt, die nach der jeweiligen Programmstruktur sortiert sind. Die Blöcke repräsentieren mit ihrer Form und ihrer Farbe verschiedene Metriken, die vom Nutzer selbst gewählt werden können. Dies gewährleistet einen schnellen Überblick über die Struktur des Softwareprojektes. Zudem bietet SEE noch einige zusätzliche Möglichkeiten der Softwareanalyse, welche auf der Homepage von SEE zu sehen sind.

Ein weiterer Grundgedanke von SEE ist die kollaborative und ortsunabhängige Softwareanalyse, weswegen SEE im Mehrspielermodus betrieben werden kann. Dabei ist es dem Nutzer überlassen, ob er an einem herkömmlichen Bildschirm oder mit einer *Virtual Reality (VR)*-Brille teilnimmt. Die Nutzer der Software werden, wie in [Abbildung 1.1](#) zu sehen, als menschliche Charaktere dargestellt. Sie begegnen sich hierbei in einem virtuellen Raum, in dem sie, neben der Softwareanalyse, untereinander kommunizieren und sich frei bewegen können. Der aktuelle Implementierungsstand nutzt die *Unity Multipurpose Avatar (UMA-2)*-Erweiterung für die Erstellung und Animation der Spielcharaktere. Diese Arbeit fokussiert die Animation ebendieser Figuren auf Basis menschlicher Emotionen.

**Tree Map:** Ein Kacheldiagramm, welches hierarchische Strukturen durch verschachtelte Rechtecke darstellt.

**VR:** Interaktive computergenerierte virtuelle Umgebung. Über spezielle Großleinwände, oder Brillen mit integrierten Bildschirmen übertragen.

**UMA-2:** Eine Unity-Erweiterung, mit der man anpassbare Charaktere, menschlicher Natur, erstellen kann.

## 1.3 AUFGABENSTELLUNG

Dieser Abschnitt ist die Einleitung für [Kapitel 2](#). Zunächst wird ein geeignetes psychologisches Modell benötigt, welches sich programmatisch abbilden lässt. Auf Basis dieses Modells werden die Gemütszustände des Spielcharakters definiert. Nachfolgend muss ermittelt werden, welche körperlichen Reaktionen ein gewisser Gemütszustand auslöst. Wie etwa, welche Muskelkontraktionen zu verzeichnen sind. In der psychologischen Forschung ist es vor allem die Mimik, welche menschliche Emotionen nach außen tragen ([Merten, 2016](#); [Bernhardt, 2019](#)). Diese gilt es, im letzten Schritt, naturgetreu zu animieren. Das Ziel dieser Arbeit wird somit sein, es den Spielcharakteren zu ermöglichen, menschliche Emotionen in Form von Mimik ausdrücken zu können.

## 1.4 FRAGESTELLUNG

Um das in [Abschnitt 1.3](#) genannte Ziel zu erfüllen, stellt sich die folgende Frage:

*Wie lassen sich Gesichtsausdrücke, ausgehend von einem Modell menschlicher Emotionen, in virtuellen 3D-Umgebungen am realistischsten darstellen?*

Hierbei handelt es sich um die zentrale Fragestellung dieser Arbeit. Um diese Frage zu beantworten, wird in [Kapitel 2](#) die entsprechende Literatur vorgestellt und mit der bestehenden Software abgeglichen. Es wird eine mögliche Implementierung konzipiert. [Kapitel 3](#) gibt Aufschluss über die genaue Implementierung, woraufhin [Kapitel 4](#) die vorangegangene Fragestellung mit einer empirischen Untersuchung beantworten soll. [Kapitel 5](#) beendet die Arbeit mit einem abschließenden Fazit und einem Ausblick auf die Zukunft.



## KONZEPT

---

In [Kapitel 2](#) werden die Grundlagen, auf denen diese Arbeit beruht, die Ziele, die dadurch entstehen, und das Konzept der Implementierung vorgestellt.

Wie bereits in [Kapitel 1](#) erwähnt, verfolgt SEE das Ziel, visualisierte Softwareanalysen auf kollaborativer Ebene durchführen zu können. Daher ist die Wahl der Unity-Engine für dieses Vorhaben, mit ihrer Plattformunabhängigkeit, leicht begründet. Nutzerinnen und Nutzer sollen, unabhängig von ihrem Endgerät und ihrem Standort, an der Analyse teilnehmen können. Kommuniziert wird größtenteils über einen internen Sprachchat. Wird eine Mikrofoneingabe registriert, so wird das Sprechen des Spielcharakters mit dem Öffnen und Schließen seines Mundes animiert. Weitere Gesichtsausdrücke gibt es nach dem aktuellen Implementierungsstand nicht. Zahlreiche psychologische Studien zeigen jedoch, dass eine nonverbale Kommunikation für das Übermitteln weiterer, nicht gesprochener Informationen maßgeblich ist ([Knapp u. a., 2013](#); [Wang, 2009](#)). Nonverbale Kommunikation kann über verschiedene Kanäle stattfinden. Hierzu gehören unter anderem Mimiken, Gestiken und die Stimmlage. Das Gesicht und seine Mimiken seien „eine der ausdrucksstärksten Kanäle der Kommunikation“ ([Ambady und Rosenthal, 1998](#)). Eine weitere interessante Beobachtung wurde von [Oh Kruzic u. a. \(2020\)](#) gemacht. Sie stellten fest, dass Mimiken, verglichen mit Gestiken, in virtuellen Umgebungen einen höheren Einfluss auf den zwischenmenschlichen Informationsfluss von Probanden haben. Die nächsten Abschnitte beschäftigen sich also mit der Frage, welche Emotionen existieren, welche Mimiken sie auslösen und wie man diese auf einen virtuellen 3D-Charakter übertragen kann.

### 2.1 EMOTIONSTHEORIE

Die Klassifizierung von Emotionen erfolgt größtenteils mittels zweier Theorien. Zum einen gibt es die Kategorisierung von Emotionen, in der angenommen wird, dass eine gewisse Anzahl an Basisemotionen existiert, über die jeder Mensch von Geburt an verfügt. So deklariert [Ekman \(1992\)](#) beispielsweise Angst, Ekel, Freude, Trauer, Wut und Überraschung als Basisemotionen, die allesamt individuelle Charakteristiken aufweisen, anhand derer man sie unterscheiden kann. Alle weiteren Emotionen werden den Basisemotionen kategorisch untergeordnet. Dieses Modell hat den Vorteil, dass die Emotionen in klar

definierte Kategorien eingeteilt werden, deren Titel leicht verständlich sind (PS und Mahalakshmi, 2017). Zum anderen existieren dimensionale Modelle, welche wiederum in zwei- oder dreidimensionale Modelle aufgeteilt werden. Die ersten zwei Dimensionen dieser beiden Ansätze sind identisch. Sie messen den Grad zwischen der Positivität oder Negativität des Gemütszustandes (meist *Valence* genannt) und den Grad der Aufregung (meist *Arousal* genannt). Dies misst, ob eine Person schläfrig oder äußerst erregt ist. Bei einem dreidimensionalen Modell, wie das *PAD*-Modell von Mehrabian (1996), wird noch die Dominanz (meist *Dominance* genannt) gemessen. Hierbei wird unterschieden zwischen Gefügigkeit und Dominanz. Emotionen der kategorischen Modelle sind hierbei als spezifische Punkte innerhalb der dimensionalen Modelle zu betrachten. [Abbildung 2.1](#) illustriert ein solches zweidimensionales Modell, wobei die x-Achse den Valence- und die y-Achse den Arousal-Grad repräsentiert.

*PAD: Pleasure-Arousal-Dominance Modell von Mehrabian (1996)*

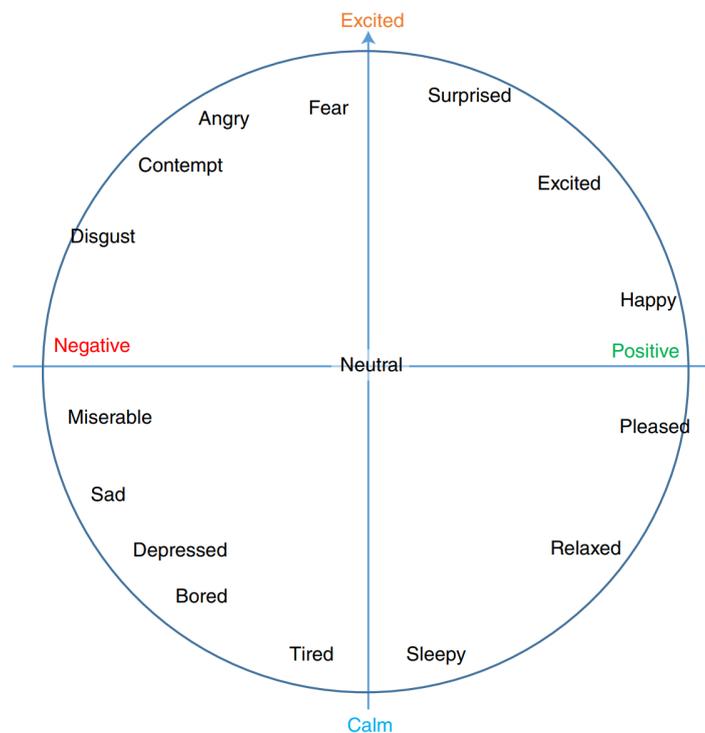


Abbildung 2.1: Zweidimensionales Modell der Emotionen nach Toisoul u. a. (2021)

Eine ängstliche Person hätte, ebenso wie eine deprimierte Person, eine negative Valence, während der Arousal-Wert bei der deprimierten Person im negativen und bei der ängstlichen Person im positiven Bereich läge (Barrett, 1998). Dieses Beispiel hebt den Vorteil von dimensionalen Modellen hervor, da man verschiedene Metriken besitzt, anhand derer man Emotionen festmachen kann. Diese Arbeit fokussiert die Bestimmung von Emotionen auf Basis eines Valence- und Arousal-Wertes.

## 2.2 MIMIK

In dem vorigen Abschnitt wurde dargestellt, dass erlebte Emotionen eine direkte Auswirkung auf die Mimiken einer Person haben. Nun muss ermittelt werden, welche Gesichtsausdrücke bei welcher Emotion zu verzeichnen sind und wie man diese beschreiben kann. In der Emotionstheorie sind Kodierungsverfahren für das Beschreiben von Mimiken maßgeblich. Sie spalten gesamte Gesichtsausdrücke in einzelne Muskelkontraktionen auf, sodass eine Mimik exakt beschrieben und nachgestellt werden kann. Das *Facial action coding system (FACS)* von Ekman und Friesen (1978) ist eines der am häufigsten genutzten Kodierungsverfahren in der Emotionstheorie. Es beinhaltet 44 sogenannte *Action Unit (AU)*s, die die Bewegung einzelner oder mehrerer Gesichtsmuskeln definieren. Eine Übersicht über alle AUs des FACS befindet sich im Anhang. Jede AU besitzt zudem eine Ausprägungsstärke auf einer fünfstufigen Skala. A, die kleinste Stufe, ist kaum, bis gar nicht, wahrnehmbar und E, die höchste Stufe, repräsentiert den höchstmöglichen Wert.



Abbildung 2.2: *Extended Cohn-Kanade-Datensatz (CK+)*: Abbildung einer Probandin (Lucey u. a., 2010)

Auf Basis zahlreicher Studien, wie dem „CK+“-Datensatz von Lucey u. a. (2010), lassen sich die Basisemotionen – glücklich, überrascht, ängstlich, wütend, verachtend, angeekelt und traurig – in ihre einzelnen AUs aufgliedern. [Abbildung 2.2](#) zeigt eine Probandin, die die zu der Trauer zugehörige, Mimik nachahmen soll. Hierbei sind die folgenden AUs zu verzeichnen: AU4<sup>1</sup>, AU15<sup>2</sup> und AU17<sup>3</sup>. Der Vorteil von kategorischen Emotionstheorien wird nun erneut deutlich. Befragte Personen können die Basisemotionen sofort einordnen und nachstellen. In diesen Studien wird zudem ersichtlich, dass ein Individuum eine individuelle Art besitzt, Emotionen mimisch zu äußern. Doch einige AUs, wie in diesem Beispiel AU15, sind maßgeblich für spezifische Emotionen, weswegen sie bei nahezu jedem Individuum auftreten.

1 AU4: Senken der Augenbrauen

2 AU15: Senken der Mundwinkel

3 AU17: Anheben des Kinns

Diese Arbeit fokussiert das Zusammenstellen von Mimiken auf Basis einzelner Muskelkontraktionen des Gesichtes nach dem Vorbild des FACS.

## 2.3 3D-MODELLIERUNG

Um jene Kodierungsverfahren auf das Gesicht eines 3D-Charakteres zu projizieren, bedarf es zunächst zu verstehen, wie 3D-Objekte aufgebaut sind. Fachbegriffe, die im Nachfolgenden genutzt werden, stammen aus der Unity-Umgebung. In anderen Umgebungen könnten die Techniken hinter diesen Begriffen unterschiedlich genannt werden.

### 2.3.1 *Meshes und Materials*

*Meshes*, was übersetzt Gitternetze bedeutet, bilden die Form eines 3D-Objektes. Sie bestehen aus einzelnen Punkten – Vertex genannt – in einem dreidimensionalen Koordinatensystem, deren Verbindungen Linien oder Flächen bilden. Die entstehenden polygonalen Flächen enthalten alle Informationen über die Höhe, Breite und Tiefe eines Objektes. Tatsächlich ist es in der 3D-Modellierung nicht möglich, Rundungen darzustellen. Auch diese werden durch Polygone dargestellt, die mit steigender Zahl der Kanten stets runder erscheinen. Doch der Rechenaufwand für die Darstellung eines Objektes steigt mit der Zunahme der Anzahl an Punkten, Kanten und Polygonen. Der linke Teil von [Abbildung 2.3](#) zeigt die Polygone, die für die Darstellung des Menschen benötigt werden.

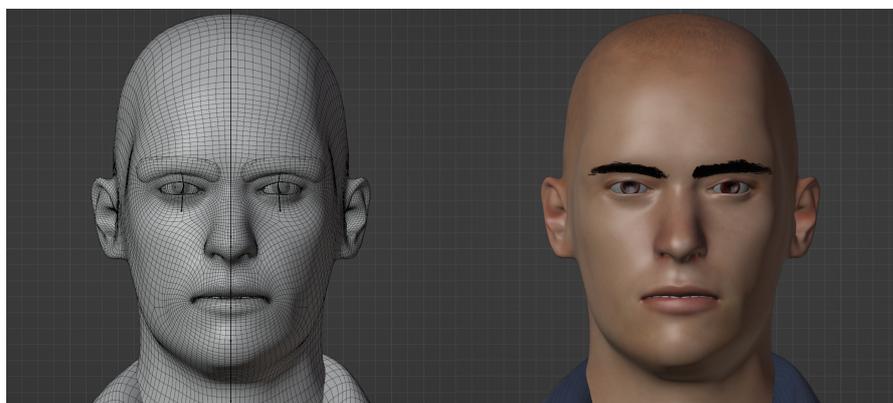


Abbildung 2.3: 3D-Modell eines Menschen

*Materials* beinhalten die Texturen eines jeweiligen 3D-Objektes. Diese Texturen werden über die Form des Gitternetzes gelegt, wodurch, wie in [Abbildung 2.3](#) rechts zu sehen ist, das Objekt sein eigentliches Aussehen erhält.

### 2.3.2 BlendShapes



Abbildung 2.4: Links: Ausdrucksloses Gesicht; Mitte: Abbildung zwei einzelner BlendShapes; Rechts: Kombination aus beiden BlendShapes

*BlendShapes* – auch *Shape Keys*, oder *Morph Targets* genannt – ermöglichen einem Mesh die Gestalt einer oder mehrerer vordefinierten Formen in Kombination anzunehmen. Ausgehend von einem Basiszustand, wie beispielsweise einem ausdruckslosen Gesicht, *morph*t man einen oder mehrere weitere Zustände, wie beispielsweise das Heben der Augenbrauen, hinzu. Durch das Morphen entsteht der sequenzielle Ablauf bis hin zur stärksten Ausprägung einer BlendShape. Somit entsteht die Möglichkeit, die Ausprägung jedes Zustandes prozentual vorzugeben. Diese Technik eignet sich somit hervorragend für das Vorhaben dieser Arbeit.

**Morphen:** Eine Technik, die fließende Übergänge zwischen zwei Bildsequenzen generiert.

## 2.4 AKTUELLE IMPLEMENTIERUNG

Die aktuelle Implementierung nutzt die UMA-2-Erweiterung für das Kreieren und Animieren von 3D-Charakteren. Die Animation der Gesichtsausdrücke erfolgt mittels eines Skriptes namens *UMA Expression Player*. Dieses Skript steuert 36 verschiedene, vom Entwickler vorgegebene, BlendShapes, welche zu verschiedenen Mimiken kombiniert werden können. Die 44 AUs des FACS lassen sich mit dieser Anzahl an BlendShapes jedoch nicht nachstellen, da das FACS einige Muskelkontraktionen definiert, welche der *UMA Expression Player*, aufgrund fehlender BlendShapes, nicht ansteuern kann.

## 2.5 ZWISCHENFAZIT

Durch die Kategorisierung von Basisemotionen und der eindeutigen Semantik, ist es in zahlreichen Studien gelungen, charakteristische Gesichtsausdrücke verschiedener Emotionen zu extrahieren. Dennoch ist das Bestimmen einer Emotion auf Basis verschiedener Metriken, wie der Valence und der Arousal, von Vorteil, da sie nicht nur die Basisemotionen klassifizieren können, sondern auch alle Gemütszustände, die zwischen ihnen liegen. Dies ist der Grund, warum diese Arbeit

beide Theorien miteinander verknüpft. Zum Ermitteln der erlebten Emotion dienen der Valence- und der Arousal-Wert als Eingabe. Diese werden in der vorliegenden Arbeit als x- und y-Koordinate eines kartesischen Koordinatensystems interpretiert. Die Basisemotionen, welche in [Abbildung 2.1](#) abgebildet sind, werden als Winkel interpretiert. Die Basisemotion, deren Winkel sich am dichtesten an dem Winkel der gebildeten Koordinate befindet, ist die ermittelte Emotion. Die Intensität der zugehörigen AUs besteht aus der euklidischen Distanz vom Nullpunkt bis zur Koordinate in Relation zu der euklidischen Distanz des vom Nullpunkt am weitesten entfernten Punktes mit demselben Winkel.

Es gibt bereits Ansätze in der Wissenschaft, die die besagten Valence- und Arousal-Werte anhand von Stimmanalysen extrahieren möchten ([Asgari u. a., 2014](#)). Ebenfalls existieren Ansätze, die anhand von Stimmanalysen Emotionen ermitteln können, wie beispielsweise das *openSMILE*-Projekt von [Eyben u. a. \(2010\)](#). Des Weiteren existieren VR-Brillen, welche die Bewegungen der unteren Gesichtshälfte registrieren und diese Daten an den Rechner übermitteln können. Die Einführung solcher Techniken ist für das Projekt SEE interessant, da das Projekt schon jetzt über die Schnittstellen – Sprachchat und VR-Unterstützung – verfügt, besagte Daten zu sammeln. Demnach zielt die Implementierung dieser Arbeit darauf ab, jedes genannte Anwendungsgebiet zu unterstützen.

[Abschnitt 2.2](#) zeigt auf, dass Mimiken sich in einzelne Muskelkontraktionen aufgliedern lassen, welche man mit Hilfe eines Kodierungsverfahrens beschreiben und nachahmen kann. Diese Arbeit bezieht sich auf das FACS, welches als eines der weltweit am weitesten verbreiteten Kodierungsverfahren für menschliche Gesichtsmuskelkontraktionen gilt. Die Animation der 3D-Charaktere erfolgt auf Basis der kategorischen Emotionen, deren AUs in zahlreichen Studien, wie dem „CK+“-Datensatz, erfasst werden.

Die Leitfrage dieser Arbeit, wie menschliche Gesichtsausdrücke in 3D-Umgebungen am realistischsten dargestellt werden können, lässt sich mit der folgenden Hypothese, welche in [Kapitel 4](#) untersucht werden soll, beantworten:

*Die realistischste Darstellung von Gesichtsausdrücken erhält man, wenn das Mesh eines 3D-Gesichtes dazu in der Lage ist, alle menschlichen Muskelkontraktionen nachahmen zu können.*

Die Ziele des nächsten Kapitels lassen sich stichpunktartig wie folgt zusammenfassen:

- Bereitstellung eines 3D-Modells mit allen AUs des FACS als BlendShapes

- Ermitteln einer Basisemotion und ihrer Intensität auf Basis der eingegebenen Valence- und Arousal-Werte
- Realistische Animation dieser Basisemotionen zusammengesetzt aus den AU-BlendShapes



## IMPLEMENTIERUNG

---

Die Kreation von BlendShapes auf Basis des FACS erfordert nicht nur die Erfassung von Veränderungsdaten basierend auf Ankerpunkten im Gesicht. Sie erfordert zudem fortschrittliche Fähigkeiten in der 3D-Modellierung. Somit übersteigt die Kreation von AU-BlendShapes für die aktuell genutzten UMA-2-Charaktere den Aufwand dieser Arbeit.

### 3.1 3D-MODELL

*MakeHuman* ist ein Open-Source-Werkzeug für das Kreieren menschlicher 3D-Modelle, welches stets von Nutzern erweitert wird. Dieses Werkzeug ist optimal für das Vorhaben dieser Arbeit, da verschiedene, von Nutzern kreierte Erweiterungen verfügbar sind, die das Importieren der 3D-Modelle in Unity und anderen Umgebungen erleichtern. So kann man dem 3D-Modell beispielsweise ein Skelett hinzufügen, welches nach Unity-Standards erstellt wurde. Dadurch erkennt Unity die Modelle als humanoid an, wodurch viele Funktionen menschlicher Charaktere auf diese Modelle anwendbar werden und somit das Ersetzen der UMA-2-Charaktere ermöglichen.

*FACSHuman* von [Gilbert u. a. \(2021\)](#), ist eine Software, welche entwickelt wurde, um Forschern das Erstellen von hypothetischen Gesichtsausdrücken in 3D-Umgebungen zu ermöglichen. Mittels dieser Software ist es möglich, nahezu alle AUs des FACS eines 3D-Gesichtes anzusteuern. Eine Übersicht über die implementierten AUs befindet sich im Anhang dieser Arbeit. Diese Software lässt sich als Erweiterung in das *MakeHuman*-Projekt einbinden. Somit ist es möglich, das Gesicht der in *MakeHuman* erstellten Charaktere auf Basis der FACS AUs zu verändern.

*Blender* ist eine Open-Source-3D-Modellierungssoftware, welche ebenfalls das Hinzufügen von Programmiererweiterungen unterstützt. Mit dieser Software lassen sich 3D-Modelle erstellen, welche in verschiedenen Dateiformaten exportiert werden können. Hierzu zählt auch das *.fbx*-Dateiformat, welches entwickelt wurde, um die Kompatibilität digitaler Inhalte in verschiedenen Umgebungen gewährleisten zu können. In einer *.fbx*-Datei sind somit beispielsweise die Informationen der Meshes, Materials, BlendShapes und des Skelettes enthalten, wodurch sie von Unity verarbeitet werden können. Zudem existiert in *Blender* eine Funktion mit dem Namen „Join as shapes“, welche die Möglichkeit

bietet, einem Objekt BlendShapes – in *Blender* „Shape Keys“ genannt – hinzuzufügen, die auf der Form anderer Meshes basieren.

Somit wird *MakeHuman* zunächst genutzt, um ein 3D-Modell eines Menschen zu erstellen. Hierbei lassen sich zahlreiche Variablen, welche das Aussehen des Menschen definieren, nach Belieben verändern. Das entstehende 3D-Modell zeigt einen Menschen mit einem ausdruckslosen Gesicht und dient als Basis für das finale Objekt. Es besteht aus mehreren Meshes und Materials für die verschiedenen Körperteile und dem vorher erwähnten Unity-Skelett. Nun wird die *FACSHuman*-Erweiterung von Gilbert u. a. (2021) genutzt, um die AUs des FACS auf das Mesh des Modells zu übertragen, indem die prozentuale Ausprägung einer jeden AU vorgegeben wird. Da jede AU eine separate BlendShape benötigt, wird für jede AU ein weiteres Modell desselben Menschen erstellt, welches die stärkste Ausprägung der jeweiligen AU darstellt. Somit entstehen, inklusive der Basisfigur, 98 verschiedene Meshes, welche als *.mhx2*-Dateien exportiert werden. Diese Dateien lassen sich mit Hilfe einer Programmiererweiterung in *Blender* importieren. Abschließend wird für die Erstellung der BlendShapes die vorher erwähnte Technik genutzt, mit welcher man BlendShapes auf Basis dritter Meshes erzeugen kann. Somit werden der Basisfigur die 97 Formen der anderen Meshes als BlendShapes hinzugefügt. Das Resultat ist ein einziges 3D-Modell eines Menschen, welches für jede AU des FACS eine BlendShape besitzt und als *.fbx*-Datei in Unity importiert werden kann. Alle Hyperlinks zu den erwähnten Programmen und Programmiererweiterungen befinden sich, ebenso wie detailliertere Anleitungen, im Anhang dieser Arbeit.

*.mhx2*: Dateiformat für das Importieren von *MakeHuman*-Daten nach *Blender*.

### 3.2 ANIMATION

Meshes, welche BlendShapes besitzen, nutzen in Unity die „*Skinned Mesh Renderer*“-Komponente für die Steuerung jener BlendShapes. Sie beinhaltet die Funktion `SetBlendshapeWeight(int index, float value)`, mit welcher die Ausprägung eines BlendShapes programmatisch verändert werden kann. Hierfür wird eine Fließkommazahl von 0 bis 100 übergeben, wobei ein Wert von 100 die stärkste Ausprägung herbeiführt. Führt man diese Funktion jedoch aus, kommt es zu einer sprunghaften Veränderung des Meshes, wodurch keine Animation stattfindet. Das Ziel dieses Abschnittes wird somit sein, dem *GameObject* eine oder mehrere AUs mitsamt Intensität übergeben zu können. Die Veränderung des Gesichtsausdruckes soll dabei stets animiert werden.

Hierfür wird dem *GameObject* das im Nachfolgenden erklärte Skript angefügt.

Dieses Skript erbt von der Klasse *MonoBehaviour*, welche als Basisklasse aller Unity-Skripte gesehen werden kann. Zunächst wird in dem Skript eine innere Klasse deklariert, welche die Entität einer Action Unit repräsentiert. Sie enthält zwei Variablen. Die erste string-

*GameObject*: Gruppierung von Komponenten in Unity, welche zu einem Objekt gehören.

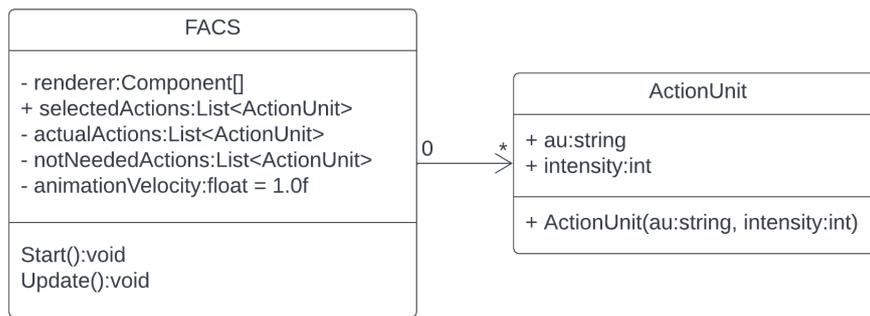


Abbildung 3.1: UML-Klassendiagramme der benötigten Klassen

Variable AU referenziert den eindeutigen Namen der BlendShapes. Die zweite Integer-Variable mit dem Namen Intensity, ist eine Ganzzahl, welche einen Wert zwischen eins und fünf annehmen kann. Sie definiert die fünfstufige Intensität der betroffenen Action Unit. Anschließend werden einige benötigte Variablen deklariert. Zum einen ein Array mit Component-Instanzen. In diesem Array werden sich die SkinnedMeshRenderer-Komponenten aller Meshes des GameObjects befinden. Diese werden benötigt, um die Ausprägung der BlendShapes programmatisch zu steuern. Zudem werden drei Listen des Typs ActionUnit benötigt. Die erste Liste mit dem Namen SelectedActions wird als public deklariert, sodass eine Modifizierung dieser Liste außerhalb des Skriptes möglich wird. Diese Liste enthält alle AUs, welche dargestellt werden sollen. Die Liste mit dem Namen ActualActions beinhaltet alle AUs, welche bereits aktiv sind. Die dritte und letzte Liste mit dem Namen NotNeededActions wird stets mit allen AUs befüllt, die zwar präsent sind, jedoch nicht weiter benötigt werden. Die letzten beiden Listen werden programmatisch befüllt, weswegen sie die private-Notation tragen. Die Fließkommazahl mit dem Namen animationVelocity enthält den Wert, mit welchem die Ausprägungsstärke eines Blendshapes pro Bild verändert wird.

Die Funktion Start() wird in Unity einmal bei der Initialisierung eines Skriptes ausgeführt. In dieser Funktion wird das eben vorgestellte Component-Array mit den Instanzen der SkinnedMeshRenderer befüllt. Die Klasse MonoBehaviour, von der dieses Skript erbt, enthält eine Variable namens gameObject. Diese Variable referenziert stets das Objekt, an welches ein Skript angehängt wird.

Die Update()-Funktion wird in Unity bei jedem Frame erneut ausgeführt. Die Logik befindet sich innerhalb dieser Funktion, in der zunächst durch die ActualActions-Liste iteriert wird, um diese mit der SelectedActions-Liste zu vergleichen. Existieren AUs, welche zwar in der ActualActions-Liste vorkommen, aber nicht in der SelectedActions-Liste, so handelt es sich um AUs, welche nicht weiter benötigt werden. Die Ausprägung dieser AUs muss wieder auf 0 sinken, weswegen sie erst einmal in die NotNeededActions-Liste aufgenommen werden.

Anschließend wird geprüft, ob die `SkinnedMeshRenderer`-Instanzen gefunden wurden. Ist dies der Fall, so wird durch das Array der `SkinnedMeshRenderer` iteriert, um die Veränderung aller Meshes zu gewährleisten.

Erst einmal wird durch die AUs, welche nicht weiter benötigt werden, iteriert. Da die Möglichkeit besteht, dass die Ausprägung einer AU in dieser Liste noch nicht den Wert 0 erreicht hat, wird geprüft, ob diese AU in der neuen Auswahl vorkommt. Ist dies der Fall, so wird diese AU wieder benötigt, weswegen sie aus der Liste entfernt wird. Die restliche Ausführung dieser Iteration wird nicht weiter benötigt, weswegen sie mit dem `break`-Statement gestoppt wird. Somit ist das nächste Objekt der Liste an der Reihe. `BlendShapes` eines Objektes sind in einem Array angeordnet, weswegen diese mittels ihres Index angesteuert werden. Die Funktion `GetBlendShapeIndex(string blendShapeName)` gibt den Index einer `BlendShape` anhand ihres Namens zurück. Dieser wird in der Variable `blendShapeIndex` aufbewahrt. Wird keine `BlendShape` mit dem angegebenen Namen gefunden, so gibt die Funktion `-1` zurück. Demnach wird geprüft, ob dies der Fall ist. Die Funktion `GetBlendShapeWeight(int index)` gibt die aktuelle Ausprägung einer `BlendShape` zurück. Dieser Wert wird ebenfalls in der Variable `actualBlendShapeWeight` aufbewahrt. Nun wird geprüft, ob die Ausprägung der jeweiligen `BlendShape` bereits den Wert 0 erreicht hat. Ist dies nicht der Fall, so wird die Ausprägung der `BlendShape` mit dem Wert der `animationVelocity`-Variable subtrahiert. Hat die Ausprägung bereits den Wert 0 erreicht, so wird diese AU aus der Liste entfernt. Das Ziel dieser Iteration ist es, die Liste `NotNeededActions` stets leer zu halten.

Nun wird durch die AUs iteriert, welche sichtbar sein sollen. Das Vorgehen dieser Iteration ist sehr ähnlich, wie bei der vorigen. Zunächst werden der Index und die aktuelle Ausprägung der `BlendShapes` erfasst. Da eine AU eine fünfstufige Ausprägungsstärke besitzt, welche nach dieser Implementierung durch eine Ganzzahl von 1-5 repräsentiert wird und `BlendShapes` eine prozentuale Ausprägungsstufe besitzen, wird die Ganzzahl mit 20 multipliziert. Im Anschluss wird geprüft, ob die aktuelle Ausprägung kleiner oder größer als die gewünschte Ausprägung ist. In diesen Fällen wird die Ausprägungsstufe entweder mit dem Wert der Variable `animationVelocity` addiert oder subtrahiert. Abschließend wird die `ActualActions`-Liste mit der `SelectedActions`-Liste gleichgesetzt.

Durch diese Logik wird erreicht, dass die Veränderung der Ausprägungen von `BlendShapes` nicht sprunghaft, sondern schrittweise, pro Ausführung der `Update()`-Funktion, geschieht. Hierdurch kommt eine Animation zustande.

## 3.3 EMOTIONSERMITTLUNG

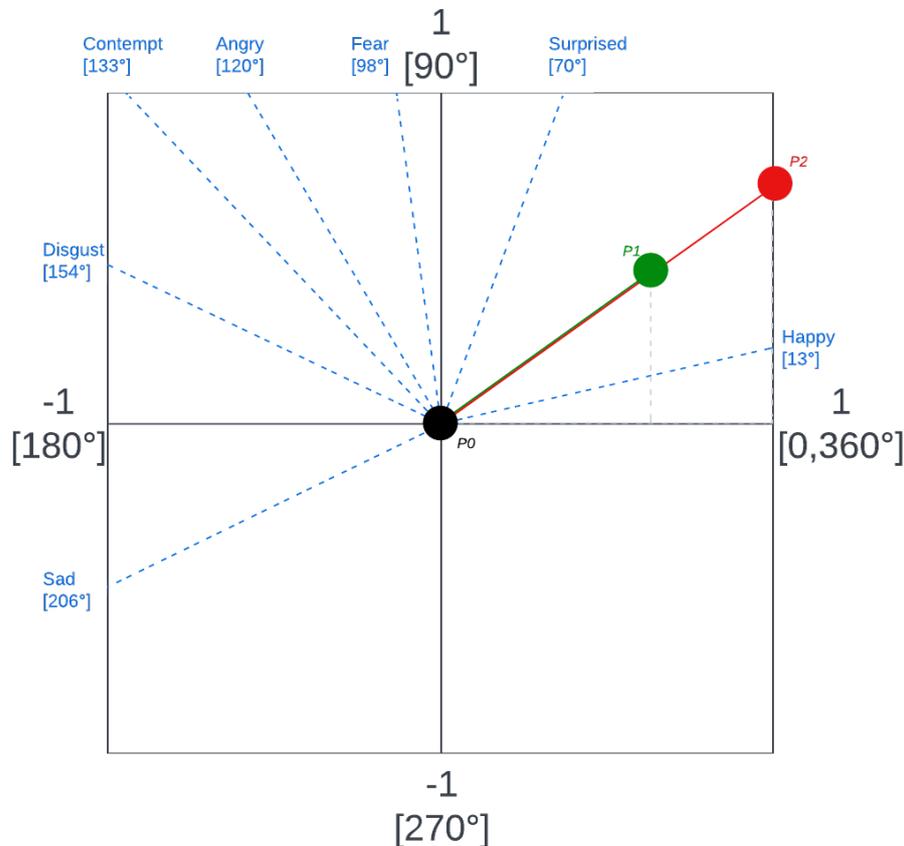


Abbildung 3.2: Eigeninterpretation des dimensionalen Modelles nach Toisoul u. a. (2021)

Abbildung 3.2 zeigt eine Eigeninterpretation eines dimensionalen Emotionsmodells. Hierbei werden diejenigen Basisemotionen aufgenommen, welche im „CK+“-Datensatz (Lucey u. a., 2010) dokumentiert sind. Zu jeder dieser Basisemotionen, deren Position sich an dem Modell von Toisoul u. a. (2021) orientiert, wird ein Winkel hinzugefügt. Die x-Achse dieses kartesischen Koordinatensystems repräsentiert den Valence-Grad und die y-Achse den Arousal-Grad. Hierbei gilt  $x \in [-1; 1]$  und  $y \in [-1; 1]$ , wobei diese Werte in Unity als Dezimalzahl mit zwei Nachkommastellen angegeben werden. Sie repräsentieren den prozentualen Grad der jeweiligen Achse. Der in Abbildung 3.2 gezeigte grüne Punkt  $P_1$  zeigt eine zufällig gewählte Koordinate  $P(\text{valence}, \text{arousal})$ . Mittels der *Math.Atan2*-Funktion kann das Bogenmaß ( $\alpha_{BM}$ ) dieser Koordinate berechnet werden, welches für die Bestimmung des Winkels benötigt wird. Eine Umrechnung in eine Gradzahl wird durch den Einsatz der folgenden Formel erreicht:

$$\alpha_{Grad} = \alpha_{BM} * 180^\circ / \pi$$

Nun kann ermittelt werden, welche Basisemotion sich am dichtesten an  $P_1$  befindet. In diesem Fall ist es das Glücklichein.

Für die Ermittlung der Intensität wird die Distanz ( $d$ ) – die **grüne Linie** in der Abbildung – vom Nullpunkt  $P_0$  bis hin zu der gewählten Koordinate  $P_1$  gemessen. Dies wird mit Hilfe des Satzes des Pythagoras erreicht. Es gilt:

$$d^2 = valence^2 + arousal^2$$

$$d = \sqrt{valence^2 + arousal^2}$$

Da jedoch jeder Winkel eine unterschiedliche Maximaldistanz vom Nullpunkt aus – die **rote Linie** in der Abbildung – besitzt, wird der soeben errechnete Wert mit der Maximaldistanz in Relation gesetzt. Um diesen Wert zu ermitteln, wird der zweite Strahlensatz genutzt, um festzustellen, welcher Punkt denselben Winkel hat, wie der Ausgangspunkt  $P_1$ , sich jedoch am äußersten Rand des Koordinatensystems befindet. Es gilt:

$$x_2 = \frac{x_1}{\max(|x_1|, |y_1|)}$$

$$y_2 = \frac{y_1}{\max(|x_1|, |y_1|)}$$

$x_1$  und  $y_1$  referenzieren hierbei den x- und y-Wert des Punktes  $P_1$ . Somit wird der gesuchte Punkt  $P_2$  ermittelt, dessen Distanz zum Nullpunkt  $P_0$  ebenfalls berechnet werden kann. Dividiert man nun die Distanz zum Punkt  $P_1$  durch die Distanz zum Punkt  $P_2$ , erhält man den Anteil der Distanz in Relation zur Maximaldistanz. Dieser Wert repräsentiert die Intensität einer ermittelten Emotion. Allgemein ist zu beachten, dass die vorangegangene Berechnung bei dem Punkt  $P(0,0)$  aufgrund der Division durch 0 nicht möglich ist. Der Punkt  $P(0,0)$  gilt in diesem Modell jedoch als emotionsloses Stadium, weswegen bei dem Punkt keine Berechnungen vorgenommen werden.

All diese Berechnungen werden in dem Skript `ValenceArousalSlider` durchgeführt. So wird anhand der zwei übergebenen Werte (Valence und Arousal) eine Emotion, ihre Intensität und die dazugehörigen AUs ermittelt. Diese Informationen werden an das bereits im vorigen Abschnitt erklärte Skript weitergeleitet, wodurch die Animation der ermittelten Emotion stattfindet.

### 3.4 ZWISCHENFAZIT

Die drei vorigen Abschnitte bieten als Ergebnis ein menschliches 3D-Modell, welches durch die Übergabe zweier Werte – dem Valence- und Arousal-Wert – eine Emotion ermittelt, deren AUs auf eine animierte Weise auf das Gesicht des Modells übertragen werden. Dies hat ein deutlich dynamischeres Auftreten als die vorher genutzten UMA-2-Charaktere. Zudem bestehen die neu erstellten Modelle aus deutlich mehr Polygonen, wodurch sie ein realistischeres Aussehen erhalten.

## EVALUATION

---

In diesem Kapitel wird die Darstellung von Mimiken und den daraus resultierenden menschlichen Emotionen in der Unity Umgebung durch eine Studie evaluiert. Dies geschieht, indem die Darstellung von Gesichtsausdrücken, basierend auf den sieben Basisemotionen des „CK+“-Datensatzes (Lucey u. a., 2010), bei den aktuell genutzten UMA-2-Charakteren mit den in [Abschnitt 3.1](#) erstellten *MakeHuman*-Charakteren verglichen wird. Zunächst werden die operationalisierten Hypothesen aufgeführt. Anschließend wird der dafür konzipierte Fragebogen vorgestellt. Abschließend werden die Ergebnisse der Studie ausgewertet.

### 4.1 HYPOTHESEN

Das Ziel dieser Studie ist, den Wahrheitsgehalt der in [Abschnitt 2.5](#) aufgestellten Hypothese zu überprüfen. Hierbei steht der Wiedererkennungsgrad, basierend auf einem Referenzbild eines Menschen des „CK+“-Datensatzes (Lucey u. a., 2010), im Vordergrund. Daraus ergeben sich die folgenden zwei Aspekte, welche überprüft werden sollen:

1. **Wiedererkennung einer Mimik:** Der Anteil korrekt zugeordneter Fotos von UMA-2-Charakteren gilt hierbei als  $W_U$ , während er bei den *MakeHuman*-Charakteren als  $W_M$  gilt.
  - **Nullhypothese**  $H_{W0}$ :  $W_M \leq W_U$ : Der Wiedererkennungsgrad  $W$  ist bei den *MakeHuman*-Charakteren kleiner als oder gleich dem Wiedererkennungsgrad der UMA-2-Charaktere.
  - **Hypothese**  $H_{W1}$ :  $W_M > W_U$ : Der Wiedererkennungsgrad  $W$  ist bei den *MakeHuman*-Charakteren größer als bei den UMA-2-Charakteren.
2. **Erkennung einer Emotion:** Der Anteil korrekt erkannter Emotionen bei UMA-2-Charakteren gilt hierbei als  $E_U$ , während er bei den *MakeHuman*-Charakteren als  $E_M$  gilt.
  - **Nullhypothese**  $H_{E0}$ :  $E_M \leq E_U$ : Der Erkennungsgrad einer Emotion  $E$  ist bei den *MakeHuman*-Charakteren kleiner als oder gleich dem Erkennungsgrad der UMA-2-Charaktere.
  - **Hypothese**  $H_{E1}$ :  $E_M > E_U$ : Der Erkennungsgrad einer Emotion  $E$  ist bei den *MakeHuman*-Charakteren größer als bei den UMA-2-Charakteren.

Ein weiterer Aspekt, welchen man hinzufügen könnte, wäre die Erkennung des Valence- und Arousal-Grades. Da dies jedoch Vorkenntnisse in der Psychologie voraussetzt, wird dies außer Acht gelassen.

## 4.2 AUFBAU

Um die im vorigen Abschnitt definierten Hypothesen zu untersuchen, werden 16 Bilder des „CK+“-Datensatzes (Lucey u. a., 2010) ausgewählt. Sie umfassen acht Bilder von weiblichen und acht Bilder von männlichen Probanden des Datensatzes. Sieben der acht Fotos zeigen die Basisemotionen, auf welche sich diese Arbeit bezieht. Das achte Bild zeigt zusätzlich ein neutrales Gesicht. Der Datensatz liefert zu jedem dieser Bilder die AUs nach dem FACS. Jedes dieser Bilder wird nun bei beiden 3D-Modellen nachgestellt. Bei den *MakeHuman*-Charakteren lassen sich die AUs übertragen, wobei das Nachstellen bei den UMA-2-Charakteren sich etwas schwieriger gestaltet, da hier eine begrenzte Anzahl an BlendShapes existiert. Hierbei werden diejenigen BlendShapes genutzt, welche den ausgewählten AUs am nächsten kommen.



Abbildung 4.1: Alle Gesichtsausdrücke der Evaluation. Die Bilder menschlicher Subjekte stammen aus dem „CK+“-Datensatz von Lucey u. a. (2010).

Abbildung 4.1 zeigt alle Objekte der Studie. Ausgehend von einem Foto eines Menschen, soll von den Probanden das zugehörige 3D-Modell ermittelt werden. Als Antwortmöglichkeiten sind alle acht Fotos eines 3D-Modells und die Option „Weiß ich nicht“ gegeben. Die gegebene Antwort kann entweder korrekt oder inkorrekt sein. Der Anteil korrek-

ter Antworten in Relation zu der Gesamtzahl an Antworten bildet den Wiedererkennungsgrad  $W$ . Zudem sollen die Probanden darlegen, welche Emotion abgebildet ist. Hierbei bestehen die Antwortmöglichkeiten aus den sieben Basisemotionen zuzüglich der Option „Weiß ich nicht“. Ein neutrales Gesicht wird hierbei nicht angezeigt. Auch hier kann die gegebene Antwort entweder korrekt oder inkorrekt sein. Der Anteil korrekter Antworten in Relation zu der Gesamtzahl an Antworten bildet den Erkennungsgrad von Emotionen  $E$ . Hierbei wird auch nach der abgebildeten Emotion bei den Fotos der realen Menschen gefragt. Somit lassen sich die Werte  $E_U$  und  $E_M$  mit dem Erkennungsgrad des menschlichen Vorbildes vergleichen. Da für die Beantwortung dieser Fragen keine Vorkenntnisse benötigt werden, kann jede Person an dieser Umfrage teilnehmen.

### 4.3 FRAGEBOGEN

Abbildung 4.1 zeigt, dass für jedes Foto des „CK+“-Datensatzes von Lucey u. a. (2010) jeweils ein zugehöriges Foto eines UMA-2-Charakters und eines des *MakeHuman*-Charakters existiert. Probanden der Studie füllen zunächst einen demografischen Fragebogen aus, worin nach dem Alter und Geschlecht der teilnehmenden Person gefragt wird. Anschließend beginnt die eigentliche Studie. Für die Ermittlung des Wiedererkennungsgrades  $W$  werden zunächst der Reihe nach die Fotos der realen Menschen, in Kombination mit der Frage, welches 3D-Modell denselben Gesichtsausdruck trägt, eingeblendet. Dies geschieht acht Mal je 3D-Modell. Als Antwortmöglichkeiten erhält die befragte Person alle acht Bilder des 3D-Modells und die Option „Weiß ich nicht“. In einer Pilotstudie sah man alle acht Fotos der realen Menschen gleichzeitig mit den acht Fotos eines 3D-Modells, welche einander zugeordnet werden sollten. Dieser Ansatz wurde verworfen, da somit die Beantwortung mittels des Ausschlussverfahrens ermöglicht wurde. In der gewählten Vorgehensweise ist es möglich, dass eine befragte Person dasselbe Foto des 3D-Modells mehrfach einem menschlichen Gesicht zuordnet. Somit hat jede befragte Person für die Ermittlung des Wiedererkennungsgrades  $W$  16 Fragen zu beantworten. Acht Mal bezüglich des UMA-2-Charakters und acht Mal bezüglich des *MakeHuman*-Charakters. Diese 16 Fragen beziehen sich auf ein einziges Geschlecht, welches abgebildet wird. Das Geschlecht ändert sich bei jeder abgeschlossenen Teilnahme, um eine Diversifikation zu ermöglichen.

Anschließend wird nach dem subjektiven Empfinden des Probanden gefragt. Es wird gefragt, bei welchem der beiden Modelle der Vergleich einfacher fiel und welches der Modelle ein natürlicheres Erscheinen hat.

Für die Ermittlung des Erkennungsgrades von Emotionen  $E$  werden nun randomisiert sieben Bildtriaten, mit der Frage, welche Emotion zu

sehen sei, angezeigt. Dabei kommt jede Basisemotion exakt dreimal vor. Einmal als Abbildung des realen Menschen und die zwei dazugehörigen 3D-Modelle. Als Antwortmöglichkeiten erhält die befragte Person alle sieben Basisemotionen und die Option „Weiß ich nicht“.

**ASP.NET Core:**  
Open-Source-  
Webframework für  
das Erstellen  
dynamischer  
Webanwendungen  
basierend auf der  
Programmiersprache  
C#.

**Dump:**  
Aufzeichnung der  
Tabellenstruktur  
einer relationalen  
Datenbank. Kann,  
aber muss nicht die  
Daten der Datenbank  
enthalten.

Dieser Fragebogen ist als *ASP.NET Core* Webapplikation implementiert, deren Antworten in eine relationale Datenbank geschrieben werden. Ein *Dump* dieser Datenbank und der Quelltext der Applikation befinden sich im Anhang. Der Fragebogen kann somit von jeder Person über den Webbrowser jedes Endgerätes durchgeführt werden. Die Dauer für die Bearbeitung beläuft sich auf ungefähr 15 Minuten.

## 4.4 AUSWERTUNG

Die soeben genannte Webapplikation enthält eine Funktion `GetStatistics()`, welche die gegebenen Antworten in drei `.csv`-Dateien ausgibt. Diese Dateien, sowie die Excel-Datei der gesamten Auswertung, befinden sich ebenfalls im Anhang. Mittels dieser Dateien lassen sich die Ergebnisse dieser Arbeit nachstellen. Teilgenommen an der Studie haben  $n = 40$  Probanden.

### 4.4.1 Demografische Auswertung

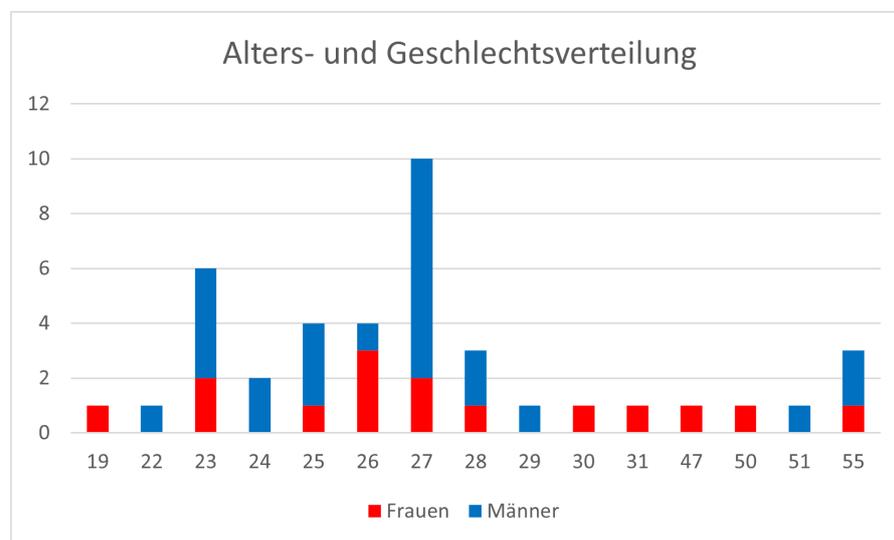


Abbildung 4.2: Alters- und Geschlechtsverteilung der Probandinnen und Probanden

Von den  $n = 40$  Teilnehmenden ist die jüngste Person eine 19-jährige Frau. Die ältesten Personen sind zwei Männer und eine Frau im Alter von 55 Jahren. Der Altersmedian liegt bei 27 Jahren, während der Mittelwert in der Nähe, bei 29,7 Jahren, liegt.

#### 4.4.2 Wiedererkennung von Mimiken

Die  $n = 40$  teilnehmenden Personen haben acht Antworten bezüglich der UMA-2-Charaktere und acht Antworten bezüglich der *MakeHuman*-Charaktere gegeben, wodurch insgesamt 640 gegebene Antworten entstehen, welche im Folgenden untersucht werden.

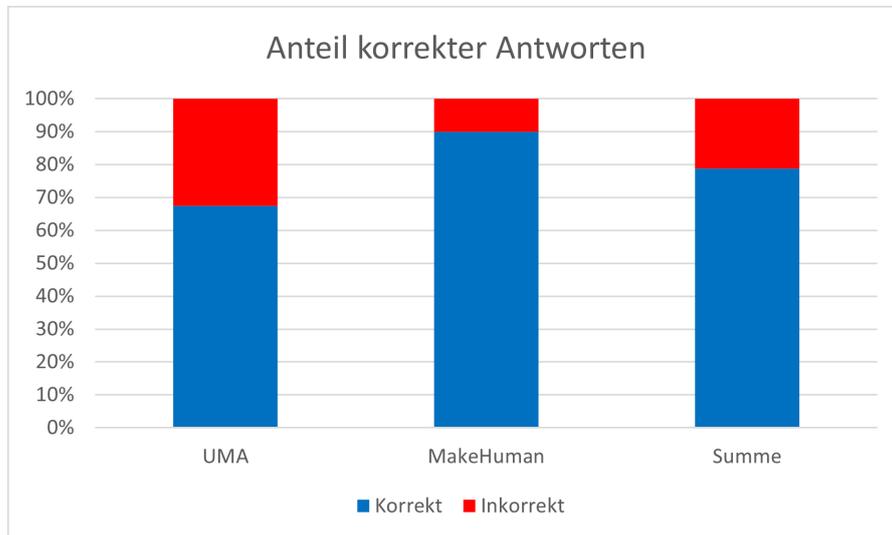


Abbildung 4.3: Anteil korrekter Antworten beim Vergleichen von Mimiken

Von den 640 Vergleichen wurden insgesamt 504 korrekt durchgeführt. Dies ergibt einen Anteil von 78,75%. Die Vergleiche bei den UMA-2-Charakteren erzielten 216 von 320 korrekte Antworten mit einem Anteil  $W_U = 67,5\%$ , während bei den *MakeHuman*-Charakteren 288 von 320 Vergleichen korrekt waren. Dies ergibt einen Anteil von  $W_M = 90\%$ .

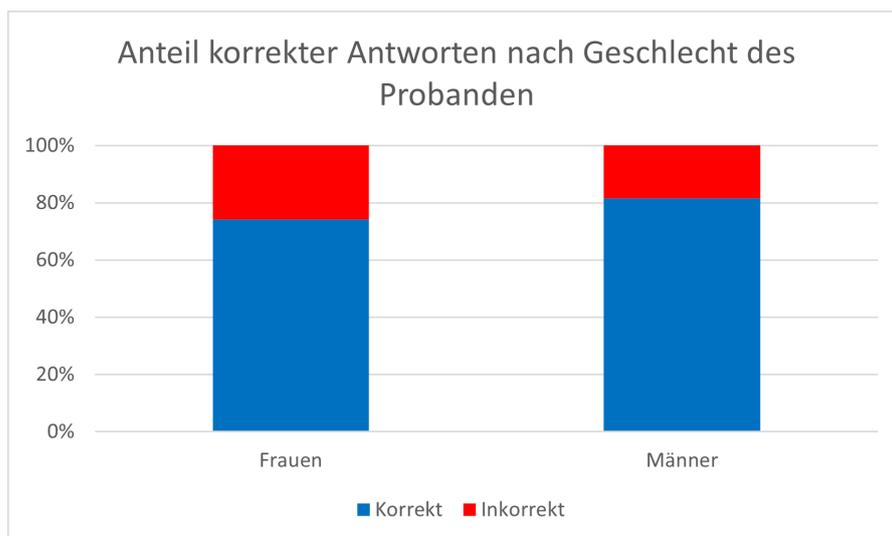


Abbildung 4.4: Anteil korrekter Antworten nach Geschlecht des Probanden

Hierbei spielt das Geschlecht der teilnehmenden Person kaum eine Rolle. Während bei den weiblichen Teilnehmenden 178 von 240 korrekte Antworten mit einem Anteil von rund 74,17% festzustellen sind, sind es bei den männlichen Teilnehmenden 326 von 400, mit einem Anteil von 81,5%.

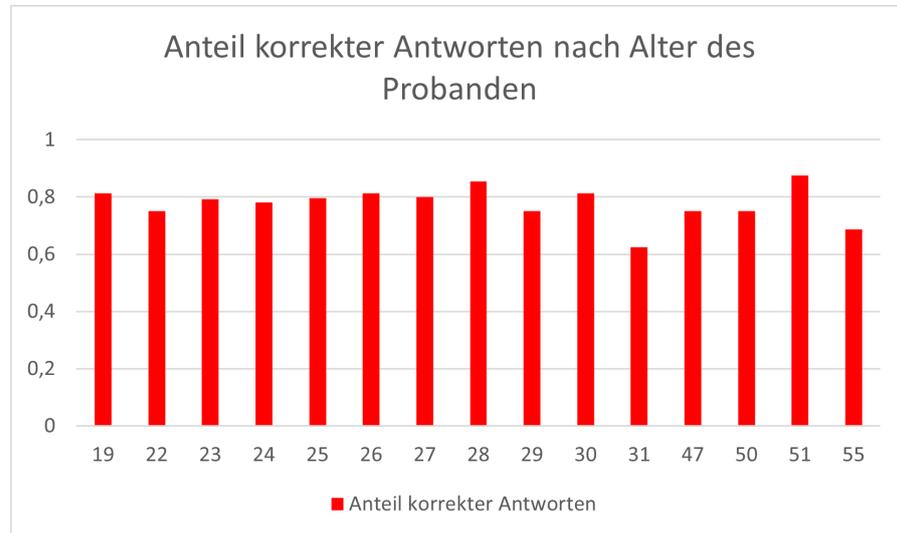


Abbildung 4.5: Anteil korrekter Antworten nach Alter des Probanden

Auch ist festzustellen, dass das Alter der teilnehmenden Personen keine signifikante Auswirkung auf die Erkennungsquote hat. Gruppiert nach dem Alter, beträgt die geringste Wiedererkennungquote 62,5%, während die höchste bei 87,5% liegt. Hierbei ist zu beachten, dass beide Werte sich auf jeweils zwei einzelne Subjekte der Studie beziehen. Andere Altersklassen, in denen mehrere Subjekte vertreten sind, pendeln sich bei ca. 80% ein.

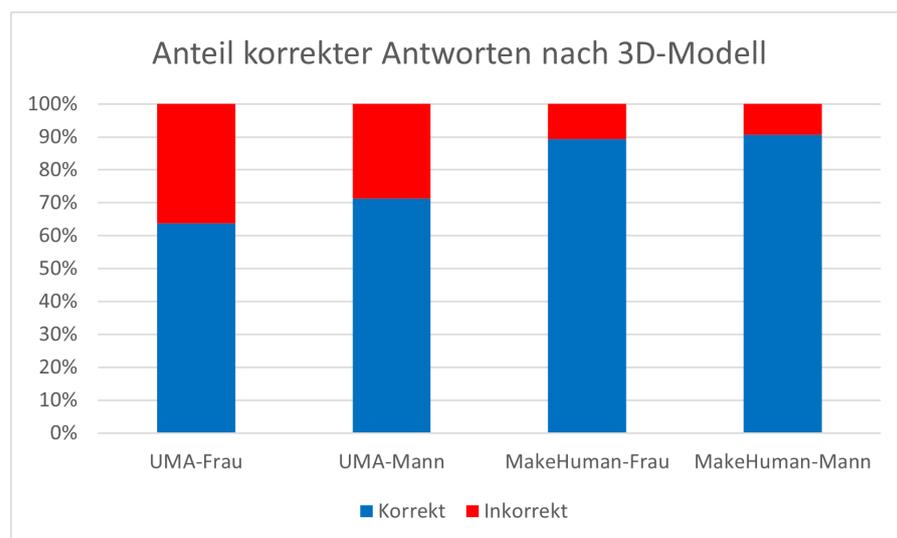


Abbildung 4.6: Anteil korrekter Antworten nach 3D-Modell

Wie bereits in [Abschnitt 4.3](#) erwähnt, ändert sich das Geschlecht des Studienobjektes bei jeder abgeschlossenen Durchführung. Somit existieren 160 vergleichende Antworten pro 3D-Modell. Die niedrigste Wiedererkennungquote hat das weibliche UMA-2-Modell mit 102 korrekten Antworten erreicht. Dies ergibt einen Anteil von 63,75%, während das männliche Modell mit 114 korrekten Zuordnungen einen Anteil von 71,25% erreicht. Bei den Geschlechtern der *MakeHuman*-Modelle ist der Wiedererkennungsgrad nahezu identisch. Mit 89,375%, also 143 korrekten Zuordnungen, schneidet das weibliche Modell ab, während das männliche Modell zwei weitere korrekte Zuordnungen erzielen kann. Hierbei beläuft sich der Anteil auf 90,625%. Der, bei den UMA-2-Charakteren festgestellte Unterschied lässt sich auf die Tatsache zurückführen, dass bei den männlichen und weiblichen Subjekten des „CK+“-Datensatzes unterschiedliche AUs aufgeführt sind, welche bei den 3D-Modellen nur durch die Kombination der beschränkten BlendShapes nachstellbar sind. Da bei den *MakeHuman*-Modellen exakt diejenigen AUs eingestellt werden können, welche auf dem zugehörigen „CK+“-Bild zu sehen sind, ist das Erkennen der daraus entstehenden Mimik deutlich einfacher, wodurch der nahezu identische Wiedererkennungsgrad entsteht.

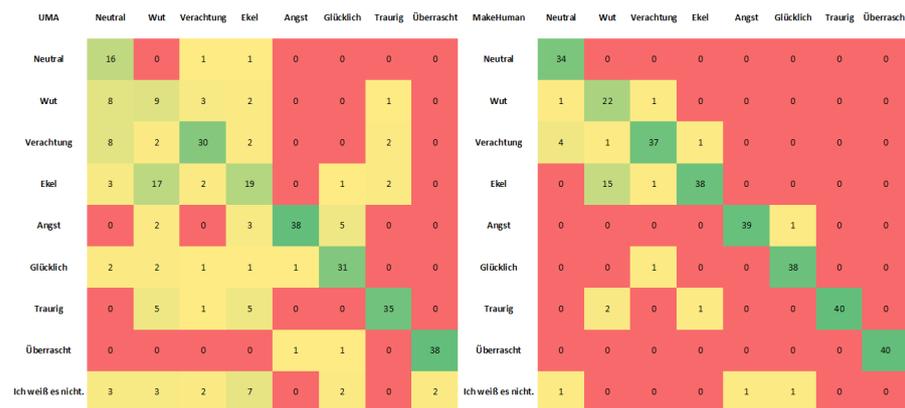


Abbildung 4.7: Heatmaps der abgebildeten Emotionen in Relation zu den erkannten Emotionen. Links UMA-2, rechts *MakeHuman*.

**Abbildung 4.7** zeigt für jede Art von 3D-Modell eine Heatmap. Die obere Leiste definiert die Emotion, welche in dem Fragebogen auf dem Bild des Menschen zu sehen ist, während die linke Leiste die Emotion des Gesichtes vom 3D-Modell, also die gegebene Antwort, definiert. Die linke Heatmap bezieht sich auf die UMA-2-Charaktere, während die rechte Heatmap sich auf die *MakeHuman*-Charaktere bezieht. Auf dem ersten Blick wird ersichtlich, dass es bei den *MakeHuman*-Modellen keine große Verteilung der Antworten gibt. Die Heatmap ist nahezu ideal verteilt. Es kommt dreimal vor, dass ein Proband ein Gesicht nicht zuordnen konnte. Bei den UMA-2-Charakteren kommt dies jedoch 19-mal vor. Bei beiden Modellen kam es häufig vor, dass ein wütendes menschliches Gesicht gezeigt und dennoch ein angeekeltes Gesicht

erkannt wird. Diese Erkenntnis deckt sich mit [Abbildung 3.2](#), auf der man erkennt, dass diese beiden Emotionen nicht weit voneinander entfernt liegen. Auch die Verachtung, welche sich in [Abbildung 3.2](#) zwischen der Wut und dem Ekel befindet, wurde ebenfalls bei beiden Modellen mindestens einmal mit den danebenliegenden Emotionen verwechselt. Ebenso kam es bei beiden Modellen vor, dass ein neutrales Gesicht mit dem Gesicht der Verachtung verwechselt wurde. Dies liegt daran, dass das Gesicht der Verachtung bei den Subjekten des „CK+“ Datensatzes nur eine AU besitzt. Diese scheint in den Fällen übersehen worden zu sein. Ebenfalls fällt auf, dass die Emotionen Angst, Glücklichsein, Trauer und Überraschung bei beiden Modellen überdurchschnittlich gute Werte erreichen. Dies ist darauf zurückzuführen, dass all diese Emotionen spezifische AUs besitzen, welche den Vergleich vereinfachen, wie beispielsweise das weite Aufreißen des Mundes bei der Überraschung.

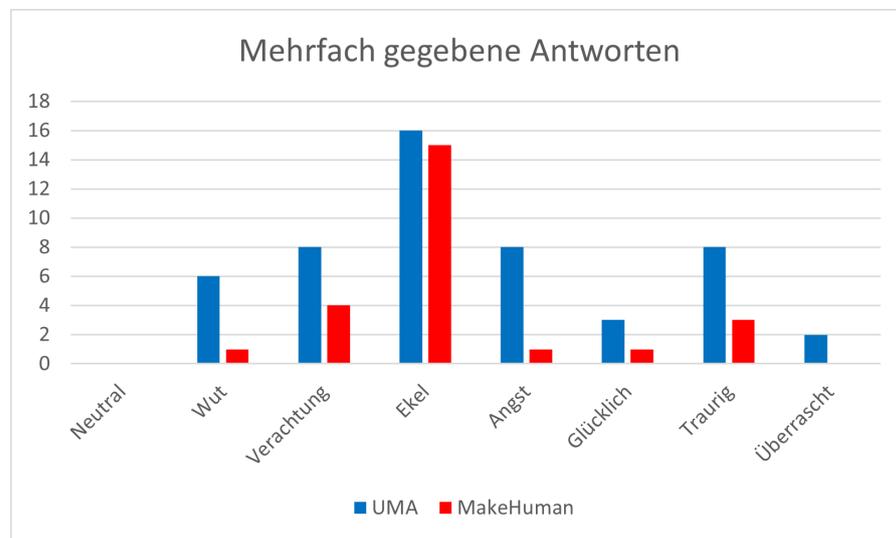


Abbildung 4.8: Anzahl mehrfach gegebener Antworten pro Proband

Der Fragebogen ist so konzipiert, dass die Gesichter der 3D-Modelle, pro Durchgang, in einer zufälligen Reihenfolge angeordnet werden. [Abbildung 4.8](#) zeigt die Anzahl mehrfach gegebener Antworten pro teilnehmender Person. Hierbei wird gezählt, ob ein Proband das Gesicht eines 3D-Modells mehr als einmal einem menschlichen Gesicht zugeordnet hat. Abgesehen von dem neutralen Gesicht, kommt es bei jedem anderen Gesicht mindestens einmal vor, dass ein und dieselbe Person mehreren menschlichen Bildern des „CK+“-Datensatzes dasselbe Gesicht des 3D-Modells zuordnet. Bei der Verachtung liegt es an dem bereits genannten Grund. Das Gesicht der Verachtung verzeichnet lediglich eine einzige AU, wodurch die Verwechslung zu einem neutralen Gesicht leicht gegeben ist. Die Wut und der Ekel verzeichnen beide einige identische AUs, wodurch auch diese Mimiken miteinander

verwechselt werden können.

Abschließend zu diesem Unterabschnitt ist festzustellen, dass die Hypothese  $H_{W1} = W_M > W_U$  mit Hilfe dieser Studie bekräftigt wird.

#### 4.4.3 Subjektive Wahrnehmung

Im Anschluss an die Vergleiche, wird die teilnehmende Person nach ihrem subjektiven Empfinden befragt. Da das Geschlecht des Studienobjektes bei jeder abgeschlossenen Durchführung wieder wechselt, erhalten wir von den  $n = 40$  Teilnehmenden 20 Antworten bezüglich der männlichen Modelle und 20 bezüglich der weiblichen.

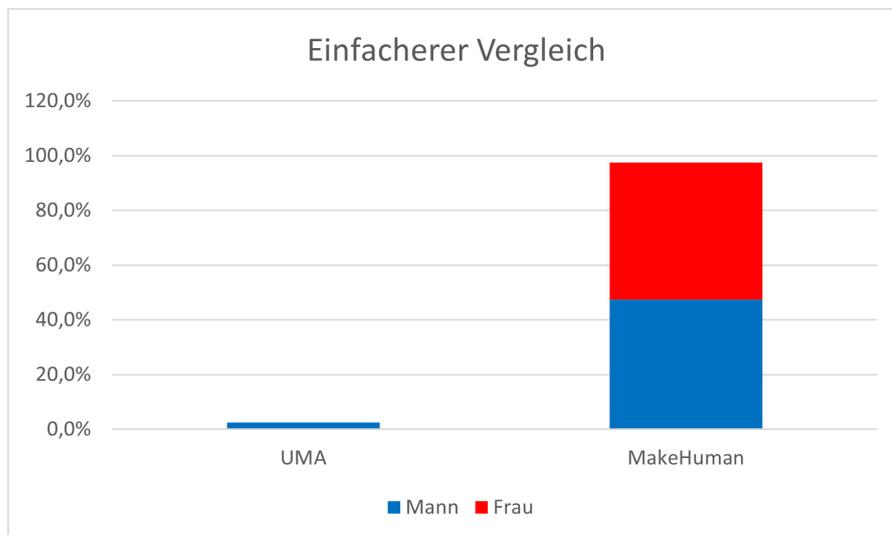


Abbildung 4.9: Anteil der Modelle, bei denen der Vergleich einfacher fällt.

Zunächst wird die teilnehmende Person gefragt, bei welchem der zwei gezeigten Modelle der Vergleich einfacher fällt. Hierbei hat nur eine einzige von allen  $n = 40$  Teilnehmenden für das UMA-2-Modell gestimmt. Somit finden 97,5% der Teilnehmenden der Vergleich bei *MakeHuman*-Modellen einfacher.

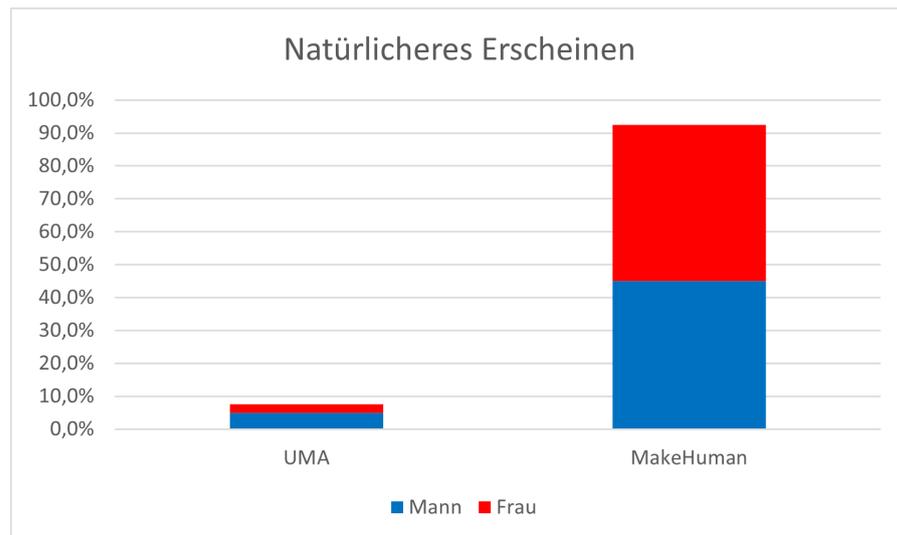


Abbildung 4.10: Anteil der Modelle, welche ein natürlicheres Erscheinen haben.

Anschließend wird die teilnehmende Person gefragt, welches der Modelle, ihrer Meinung nach, ein natürlicheres Erscheinen hat. Hierbei halten zwei Personen das männliche UMA-2-Modell und eine Person das weibliche Modell für natürlicher. Der Rest entscheidet sich für das *MakeHuman*-Modell. Hierbei entsteht eine Quote von 92,5% für die *MakeHuman*-Charaktere. Somit schließen diese in dem subjektiven Empfinden der Teilnehmenden deutlich besser ab, als die UMA-2-Charaktere. Dies ist jedoch auch darauf zurückzuführen, dass die *MakeHuman*-Charaktere aus weitaus mehr Polygonen bestehen und dadurch ein realistischeres Aussehen erhalten.

#### 4.4.4 Erkennung dargestellter Emotionen

Der Fragebogen ist so konzipiert, dass für die sieben Basisemotionen zufällig gewählte Bildtriaden erscheinen. Diese sind ebenfalls randomisiert angeordnet. Somit haben die  $n = 40$  Teilnehmenden insgesamt 840-mal einem Bild eine Emotion zugeordnet.

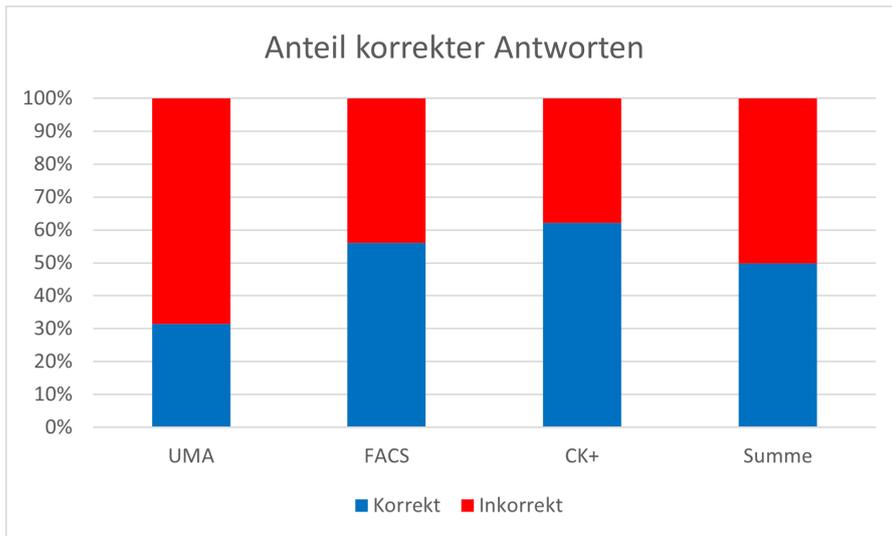


Abbildung 4.11: Anteil korrekter Antworten beim Zuordnen von Emotionen

Von den 840 gegebenen Antworten sind 419 korrekt. Dies ergibt einen Gesamtanteil von rund 49,88%. Das menschliche Vorbild des „CK+“-Datensatzes erreicht 174 von 280 korrekten Antworten mit einem Anteil von rund  $E_C \approx 62,14\%$ . Verglichen mit den anderen beiden Modellen ist dies der beste Wert. Die *MakeHuman*-Modelle erzielen 157 von 280 korrekten Antworten, mit einem Anteil von rund  $E_M \approx 56,07\%$ . Am schlechtesten schneiden die UMA-2-Modelle ab. Von 280 Antworten sind hier lediglich 88 korrekt. Dies ergibt einen Anteil von  $E_U \approx 31,43\%$ .

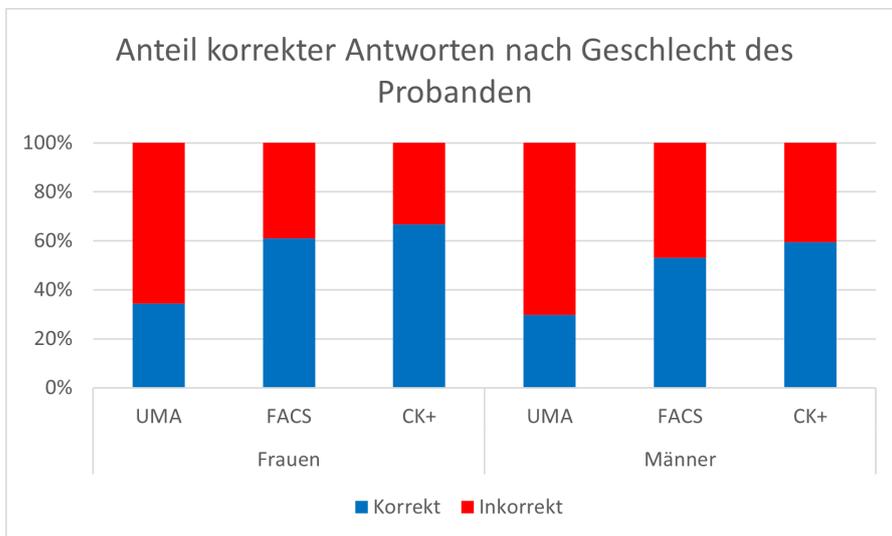


Abbildung 4.12: Anteil korrekter Antworten nach Geschlecht des Probanden und Modell

Auch hierbei wird deutlich, dass das Geschlecht der teilnehmenden Person keine große Auswirkung auf das Ergebnis hat. Dennoch hatten weibliche Teilnehmende bei dieser Fragestellung eine höhere Quote erreicht als die männlichen Teilnehmenden. Insgesamt liegt der Er-

kennungsgrad  $E$  bei den weiblichen Teilnehmenden bei rund 53,97%, wobei er bei den männlichen Teilnehmenden bei rund 47,43% liegt. Diese Erkenntnis machen auch [Abbruzzese u. a. \(2019\)](#), die das Erkennen von Emotionen anhand des Geschlechts und des Alters analysieren.

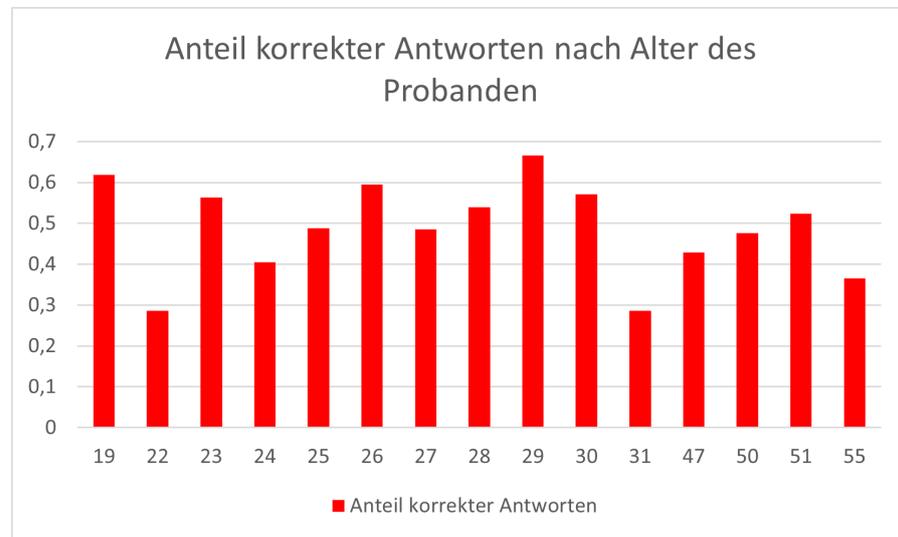


Abbildung 4.13: Anteil korrekter Antworten nach Alter des Probanden

Bei dem Anteil korrekter Antworten nach Altersgruppen sind deutliche Unterschiede festzustellen. Da diese Unterschiede jedoch oftmals bei Altersklassen vorkommen, in denen sich lediglich eine teilnehmende Person befindet, sind die Unterschiede nicht aussagekräftig. Dennoch ist mit steigendem Alter eine reduzierte Fähigkeit der Emotionserkennung festzustellen. Diese Erkenntnis machen auch [Abbruzzese u. a. \(2019\)](#).



Abbildung 4.14: Heatmap der abgebildeten Emotionen in Relation zu den erkannten Emotionen der CK+ Subjekte

Auch bei dieser Heatmap definiert die obere Leiste die gezeigten Emotionen des Fragebogens, während die linke Leiste die gegebenen Antworten referenziert. Hierbei wird deutlich, dass viele Teilnehmende

die Verachtung gar nicht zuordnen können. Ebenfalls ist festzustellen, dass die Emotionen des zweiten Quadranten von [Abbildung 3.2](#) oftmals miteinander verwechselt werden. So wurde die gezeigte Wut 16-mal der Verachtung zugeordnet, die Angst dem Ekel und der Ekel der Verachtung 15-mal. Auch dies ist auf die Tatsache zurückzuführen, dass all diese Emotionen einen hohen Arousal-Wert, aber einen niedrigen Valence-Wert besitzen.

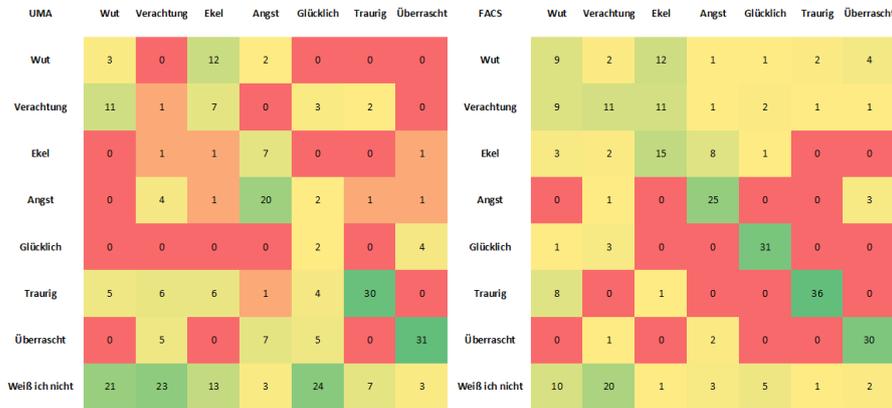


Abbildung 4.15: Heatmap der abgebildeten Emotionen in Relation zu den erkannten Emotionen. Links UMA-2, rechts *MakeHuman*.

[Abbildung 4.15](#) visualisiert noch einmal den Erkennungsgrad  $E_U \approx 31,43\%$ . Es wird deutlich, dass den meisten Gesichtern keine Emotion zugeordnet werden kann. Auch hier, sowie bei den *MakeHuman*-Charakteren, kommt es vor, dass die Emotionen des zweiten Quadranten von [Abbildung 3.2](#) – Wut, Angst, Verachtung und Ekel – gerne miteinander verwechselt werden. Bei beiden Modellen ist, genau wie bei den Subjekten des „CK+“-Datensatzes, die Verachtung oftmals nicht zuzuordnen. Emotionen, welche zuverlässig erkannt werden, sind die Angst, das Glücklichsein, die Trauer und die Überraschung.

Abschließend zu diesem Unterabschnitt lässt sich feststellen, dass auch die Hypothese  $H_{E1} = E_M > E_U$  durch diese Studie bekräftigt wird. Mit  $E_M \approx 56,07\%$  befindet sich der Erkennungsgrad bei *MakeHuman*-Charakteren sehr nah an dem menschlichen Vorbild des „CK+“-Datensatzes mit  $E_C \approx 62,14\%$ .

## 4.5 THREATS TO VALIDITY

In diesem Abschnitt geht es um die sogenannten „Threats to Validity“. Es wird unterschieden zwischen der internen und der externen Validität. Die interne Validität misst die Kausalität zwischen den unabhängigen Variablen der Studie – also den Objekten der Studie – und den abhängigen Variablen der Studie, welche die zu untersuchenden Metriken darstellen. Die externe Validität misst, ob die Ergebnisse der Studie auch außerhalb dieser anwendbar sind. Die „Threats to Validity“

stellen diejenigen Aspekte dar, welche die interne und externe Validität einer Studie bedrohen.

#### 4.5.1 *Interne Validität*

Zunächst gibt es die Störvariablen, welche in dieser Studie aus der Tatsache bestehen, dass jede Person an dieser Studie teilnehmen kann. Doch Personen, welche eine Vorkenntnis in der Psychologie besitzen, sind hierbei im Vorteil, weswegen zu erwarten wäre, dass diejenigen Personen deutlich mehr korrekte Antworten liefern. Da die Studie an einem Bildschirm stattfindet und die teilnehmende Person Bilder miteinander vergleichen muss, kann es zu einem Nachteil bei Probanden kommen, welche ein schlechteres Sehvermögen besitzen. Es ist jedoch davon auszugehen, dass diese Störvariablen keinen großen Einfluss auf die Studie haben, da in [Unterabschnitt 4.4.2](#) und [Unterabschnitt 4.4.4](#) festgestellt wurde, dass der Anteil korrekter Antworten bei den meisten Probanden ähnlich verteilt ist. Eine weitere Störvariable ist die Tatsache, dass der Vergleich mit zwei unterschiedlichen 3D-Modellen stattfindet. Die Anzahl an Polygonen ist bei den *MakeHuman*-Charakteren deutlich höher, wodurch eine realistischere Darstellung entsteht. Dieser Umstand führt nicht nur zu der Tatsache, dass die *MakeHuman*-Charaktere in [Unterabschnitt 4.4.3](#) deutlich besser abschneiden. Er könnte zudem für eine Sympathie der Probanden gegenüber der *MakeHuman*-Charaktere sorgen, was sich eventuell negativ auf die Zuordnung und Bewertung von UMA-2-Charakteren auswirken könnte. Die Tatsache, dass zwei verschiedene Modelle genutzt werden, ist darauf zurückzuführen, dass der Aufwand für das händische Kreieren von *BlendShapes* den Aufwand dieser Arbeit übersteigt. Um diesem Umstand dennoch entgegenzuwirken, wird in dem Fragebogen zunächst nur das UMA-2-Modell präsentiert, welches verglichen werden soll. Dennoch ist davon auszugehen, dass dieser Umstand einen hohen Einfluss auf die subjektive Wahrnehmung der Probanden in [Unterabschnitt 4.4.3](#) hat.

Die Zeit der Durchführung kann eine Rolle spielen, da Menschen dazu neigen, zu unterschiedlichen Zeiten am produktivsten und konzentriertesten zu sein. Da die Teilnahme an dieser Studie jedoch online erfolgt, ist es der teilnehmenden Person selbst überlassen, an welchem Wochentag und um welche Uhrzeit sie teilnimmt. Somit ist davon auszugehen, dass der zeitliche Aspekt in dieser Studie kaum bis gar keine Rolle spielt.

Die Sättigung bemisst die Ermüdungserscheinungen der Teilnehmer. Es ist davon auszugehen, dass Studien, welche zu zeitintensiv sind, Ermüdungserscheinungen der Teilnehmenden herbeiführen, wodurch das Ergebnis verfälscht werden kann. Der Fragebogen zu dieser Studie ist so konzipiert, dass eine konzentrierte Bearbeitung ca. 15 Minuten beansprucht. Des Weiteren sind die Aufgabenstellungen trivial und

erfordern kein übermäßig hohes Maß an Konzentration, wodurch davon auszugehen ist, dass die Ergebnisse dieser Studie kaum ein Problem bezüglich Ermüdungserscheinungen haben.

Wiederholtes Testen führt oftmals zu Lerneffekten bei Probanden. Dies wirkt sich auf das Ergebnis aus. Das Design dieses Fragebogens ermöglicht jedoch keinen Lerneffekt, da die teilnehmende Person nicht weiß, ob die gegebene Antwort korrekt oder inkorrekt ist. Somit hätte ein wiederholtes Durchführen dieser Studie keinen großen Effekt auf das Ergebnis. Zudem ist anhand von [Abbildung 4.2](#) davon auszugehen, dass eine wiederholte Durchführung kaum bis gar nicht vorgekommen ist.

Die Selektion bemisst, ob die Zuweisung in die Kontrollgruppen einer Studie zufällig geschieht. Hierbei kann es auch vorkommen, dass der Zufall die Gruppen ungünstig zusammensetzt. Die Kontrollgruppen dieser Studie – ob das gezeigte Modell weiblich oder männlich ist – werden automatisiert gleich groß gehalten. Somit ist davon auszugehen, dass die Selektion der Probanden keinen hohen Einfluss auf das Ergebnis hat.

#### 4.5.2 Externe Validität

Bei der externen Validität ist zunächst die Repräsentanz von Interesse. Hierbei wird bemessen, ob die Probanden die Gesamtpopulation widerspiegeln. An dieser Studie kann jede Person teilnehmen. Es werden keine Vorkenntnisse benötigt. [Abbildung 4.2](#) zeigt die Alters- und Geschlechtsverteilung der Probanden. Hier ist festzustellen, dass verschiedene Altersklassen von 19 bis 55 Jahren vertreten sind. Ebenfalls sind beide Geschlechter vertreten. Es wird im demografischen Fragebogen jedoch nicht nach der Berufsgruppe des Probanden gefragt. Die Repräsentanz ist somit nicht gegeben, da die Tätigkeiten der Probanden nicht bekannt sind. Ebenfalls ist festzustellen, dass die Probanden sich im direkten oder indirekten Umfeld des Autors befinden, was ebenfalls eine Reduzierung der Repräsentanz zur Folge hat.

Anschließend ist die Situation, in der die Studie durchgeführt wird, von Interesse. Da die Teilnahme von jedem mobilen Endgerät möglich ist, kann es sein, dass einige Teilnehmende mit dem Mobiltelefon teilgenommen haben. Dies führt zu einer Verschlechterung der Zuordenbarkeit von Gesichtern, da auf dem kleinen Bildschirm eines Telefons der Vergleich nur mittels intensivem Scrollen möglich ist. Um diesem Umstand entgegenzuwirken, wird die teilnehmende Person am Anfang der Studie darauf hingewiesen, dass sie die Studie am besten an einem Gerät mit einem größeren Bildschirm durchführen soll. Die Teilnahme mittels Mobiltelefon lässt sich jedoch nicht gänzlich verhindern.

## 4.6 ZWISCHENFAZIT

Die durchgeführte Studie liefert keine neuen wissenschaftlichen Erkenntnisse. Vielmehr sind die Ergebnisse der Studie als Untermauerung der bekannten Literatur anzusehen. Zudem lässt sich durch die Bekräftigung der Hypothesen feststellen, dass die vorangegangene Aufgabenstellung dieser Arbeit – das Animieren der 3D-Charaktere auf Basis menschlicher Emotionen – zufriedenstellend erfüllt wird. Die in [Abschnitt 2.5](#) geäußerte Hypothese, dass menschliche Mimiken in 3D-Umgebungen mittels Kodierungsverfahren von Muskelkontraktionen am realistischsten dargestellt werden können, kann mit der ersten Hypothese der Wiedererkennung von Mimiken bekräftigt werden. Zudem zeigt der Erkennungsgrad von Emotionen, dass, basierend auf den, mittels Kodierungsverfahren zusammengestellten Mimiken, Basisemotionen deutlich besser dargestellt werden können und somit vom Menschen als jene interpretiert werden.

## FAZIT

---

Dieses Kapitel befasst sich mit einem abschließenden Fazit, eventuellen Verbesserungsvorschlägen und einem Ausblick auf die Zukunft.

In [Kapitel 3](#) wird Aufschluss über die Implementierung dieser Arbeit gegeben. Das in [Abschnitt 3.1](#) erstellte 3D-Modell verfügt über mehrere Meshes. Vier der Meshes definieren die Körperteile des Modells. So besteht ein Mesh für den gesamten Körper, eines für die Augen, eines für die Zähne und eines für die Zunge. Die Separierung dieser Meshes ist sinnvoll, da jene Körperteile als eigenständig anzusehen sind. Dadurch entsteht jedoch der Umstand, dass jedes Mesh für jede AU des FACS eine BlendShape benötigt. Folglich wird in [Abschnitt 3.2](#) durch die *SkinnedMeshRenderer*-Komponenten iteriert, um die gewünschte AU für jedes Mesh zu setzen. Dies hat den Nachteil, dass jede Veränderung eines BlendShape-Wertes viermal erfolgen muss, was den Berechnungsaufwand unnötigerweise erhöht. Hierbei wäre eine Technik von Vorteil, die die BlendShapes mehrerer Meshes gleichzeitig ansteuern kann. Eine solche Technik ist dem Autor dieser Arbeit zum Zeitpunkt der Implementierung, aufgrund des Mangels an Erfahrung in der 3D-Modellierung, nicht bekannt.

Das Ziel, die 3D-Charaktere des Projektes SEE auf Basis menschlicher Emotionen zu animieren, wird jedoch erreicht. Das Ergebnis dieser Arbeit ist ein 3D-Modell, welches durch die höhere Anzahl an Polygonen schon im statischen Zustand eine natürlichere Darstellung besitzt. Zudem hat jenes Modell die Fähigkeit, nahezu alle Muskelkontraktionen im Gesicht vorzunehmen, wie es bei einem realen Menschen der Fall ist. Dadurch entsteht die Möglichkeit, alle Gesichtsausdrücke darzustellen, welche ein Mensch ausdrücken kann. Das Modell besitzt ebenfalls alle benötigten Eigenschaften, um die aktuell verwendeten UMA-2-Charaktere abzulösen. Hierfür ist Fleißarbeit gefragt, denn bereits bestehende Animationen und Aktionen müssen nun auf das neue Modell überführt werden. Ebenfalls lassen sich weitere nonverbale Informationen, wie beispielsweise Gestiken, durch das Hinzufügen des Skelettes implementieren. Auch hier gibt es in der Literatur einige Ansätze, welche die Veränderung der Körperhaltung bei verschiedenen Gemütszuständen analysieren, wie beispielsweise die Arbeit von [Unuma u. a. \(1995\)](#). Genannte Erweiterungen lassen sich hierbei gut mit dem in [Abschnitt 3.3](#) implementierten Skript verknüpfen. Ebenfalls ist es mit den neu erstellten Modellen und dem in [Abschnitt 3.2](#) beschriebenen Skript möglich, Gesichtsausdrücke zusammenzustellen.

Somit lassen sich Gesichtsausdrücke, welche durch Gesichtserkennung erfasst werden, unabhängig von der gelebten Emotion, auf das Gesicht des Avatares projizieren.



## GLOSSAR

---

- .mhx2** Dateiformat für das Importieren von *MakeHuman*-Daten nach *Blender*. [14](#)
- Arousal** Der Grad der Aufregung einer gelebten Emotion (schläfrig-äußerst erregt). [6](#), [9–11](#), [17](#), [18](#), [20](#), [31](#)
- ASP.NET Core** Open-Source-Webframework für das Erstellen dynamischer Webanwendungen basierend auf der Programmiersprache C#. [22](#)
- BlendShape** Eine Technik, die es ermöglicht ein Gitternetz zu einer oder mehreren vordefinierten Formen zu transformieren. [9–11](#), [13–16](#), [20](#), [25](#), [32](#), [35](#)
- Code-City** Ein Ansatz, große Softwareprojekte und ihre Vielzahl an verschiedenen Komponenten in Form von Wolkenkratzern einer Stadt zu visualisieren. Die Merkmale dieser Wolkenkratzer definieren ausgesuchte Metriken der Softwarekomponente. [1](#), [2](#)
- Dominance** Der Grad der Dominanz einer gelebten Emotion (dominant-gefügtig). [6](#)
- Dump** Aufzeichnung der Tabellenstruktur einer relationalen Datenbank. Kann, aber muss nicht die Daten der Datenbank enthalten. [22](#)
- GameObject** Gruppierung von Komponenten in Unity, welche zu einem Objekt gehören. [14](#), [15](#)
- Material** Texturen, die die Gitternetze eines 3D-Objektes bespannen. [8](#), [13](#), [14](#)
- Mesh** 3D-Gitternetz, bestehend aus Punkten (Vertex), die, miteinander verbunden, entweder Kanten (Edges) oder Flächen (Faces) bilden. Sie stellen die Form eines 3D-Objektes dar. [8–10](#), [13](#), [14](#), [16](#), [35](#)
- Morphen** Eine Technik, die fließende Übergänge zwischen zwei Bildsequenzen generiert. [9](#)
- PAD** Pleasure-Arousal-Dominance Modell von [Mehrabian \(1996\)](#) [6](#)
- Tree Map** Ein Kacheldiagramm, welches hierarchische Strukturen durch verschachtelte Rechtecke darstellt. [2](#)

**Unity** Eine der weltweit führenden Plattformen für die Erstellung interaktiver und plattformunabhängiger 3D-Umgebungen, basierend auf den Programmiersprachen C# und C++. [2](#), [5](#), [8](#), [13–15](#), [17](#), [19](#)

**Valence** Der Grad der Positivität oder Negativität einer gelebten Emotion (positiv-negativ). [6](#), [9–11](#), [17](#), [18](#), [20](#), [31](#)

## AKRONYME

- 
- AU** Bewegungseinheit des FACS. Definition der Bewegung einer oder mehrerer Gesichtsmuskeln. 7, 9–11, 13–16, 18, 20, 25, 26, 35
- CK+** Erweiterter Datensatz über Emotionen und den entsprechenden AUs auf Basis des Cohn-Kanade-Datensatzes von ? 7, 10, 17, 19–21, 25, 26, 29, 31, 41
- DIN** Eine unabhängige Plattform für Normung und Standardisierung. 1
- FACS** Kodierungsverfahren für das Beschreiben von Mimiken auf Basis von AUs (Ekman und Friesen, 1978). 7–10, 13, 14, 20, 35
- SEE** Eine interaktive Visualisierung von Software, welche die *Code-City*-Metapher verwendet und einen kollaborativen Multiplayer über verschiedene Plattformen<sup>1</sup> hinweg ermöglicht. 2, 5, 10, 35, 41
- UMA-2** Eine Unity-Erweiterung, mit der man anpassbare Charaktere, menschlicher Natur, erstellen kann. 2, 9, 13, 18–21, 23, 25, 27–29, 31, 32, 35, 41, 42
- VR** Interaktive computergenerierte virtuelle Umgebung. Über spezielle Großleinwände, oder Brillen mit integrierten Bildschirmen übertragen. 2, 10

---

<sup>1</sup> Neben Desktop- und Touch-Umgebungen noch Virtual Reality (z. B. *Valve Index*) und Augmented Reality (z. B. *Microsoft HoloLens*)



## ABBILDUNGSVERZEICHNIS

---

Abbildung 1.1	Darstellung einer Code-City in SEE (Koschke, 2021) . . . . .	2
Abbildung 2.1	Zweidimensionales Modell der Emotionen nach Toisoul u. a. (2021) . . . . .	6
Abbildung 2.2	CK+: Abbildung einer Probandin (Lucey u. a., 2010) . . . . .	7
Abbildung 2.3	3D-Modell eines Menschen . . . . .	8
Abbildung 2.4	Links: Ausdrucksloses Gesicht; Mitte: Abbildung zwei einzelner BlendShapes; Rechts: Kombination aus beiden BlendShapes . . . . .	9
Abbildung 3.1	UML-Klassendiagramme der benötigten Klassen	15
Abbildung 3.2	Eigeninterpretation des dimensionalen Modells nach Toisoul u. a. (2021) . . . . .	17
Abbildung 4.1	Alle Gesichtsausdrücke der Evaluation. Die Bilder menschlicher Subjekte stammen aus dem „CK+“-Datensatz von Lucey u. a. (2010). . . . .	20
Abbildung 4.2	Alters- und Geschlechtsverteilung der Probandinnen und Probanden . . . . .	22
Abbildung 4.3	Anteil korrekter Antworten beim Vergleichen von Mimiken . . . . .	23
Abbildung 4.4	Anteil korrekter Antworten nach Geschlecht des Probanden . . . . .	23
Abbildung 4.5	Anteil korrekter Antworten nach Alter des Probanden . . . . .	24
Abbildung 4.6	Anteil korrekter Antworten nach 3D-Modell . . . . .	24
Abbildung 4.7	Heatmaps der abgebildeten Emotionen in Relation zu den erkannten Emotionen. Links UMA-2, rechts <i>MakeHuman</i> . . . . .	25
Abbildung 4.8	Anzahl mehrfach gegebener Antworten pro Proband . . . . .	26
Abbildung 4.9	Anteil der Modelle, bei denen der Vergleich einfacher fällt. . . . .	27
Abbildung 4.10	Anteil der Modelle, welche ein natürlicheres Erscheinen haben. . . . .	28
Abbildung 4.11	Anteil korrekter Antworten beim Zuordnen von Emotionen . . . . .	29
Abbildung 4.12	Anteil korrekter Antworten nach Geschlecht des Probanden und Modell . . . . .	29

Abbildung 4.13	Anteil korrekter Antworten nach Alter des Probanden . . . . .	30
Abbildung 4.14	Heatmap der abgebildeten Emotionen in Relation zu den erkannten Emotionen der CK+ Subjekte . . . . .	30
Abbildung 4.15	Heatmap der abgebildeten Emotionen in Relation zu den erkannten Emotionen. Links UMA-2, rechts <i>MakeHuman</i> . . . . .	31

# LITERATURVERZEICHNIS

---

- [Abbruzzese u. a. 2019] ABBRUZZESE, Laura ; MAGNANI, Nadia ; ROBERTSON, Ian H. ; MANCUSO, Mauro: Age and gender differences in emotion recognition. In: *Frontiers in Psychology* 10 (2019)
- [Ambady und Rosenthal 1998] AMBADDY, Nalini ; ROSENTHAL, Robert: Nonverbal communication. In: *Encyclopedia of mental health* 2 (1998), S. 775–782
- [Asgari u. a. 2014] ASGARI, Meysam ; KISS, Géza ; SANTEN, Jan van ; SHAFRAN, Izhak ; SONG, Xubo: Automatic measurement of affective valence and arousal in speech. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, S. 965–969
- [Barrett 1998] BARRETT, Lisa F.: Discrete emotions or dimensions? The role of valence focus and arousal focus. In: *Cognition & Emotion* 12 (1998), Nr. 4, S. 579–599
- [Bernhardt 2019] BERNHARDT, Christian: Mimik und Emotionen. In: *Nonverbale Kommunikation im Recruiting*. Springer, 2019, S. 227–268
- [Ekman 1992] EKMAN, Paul: An argument for basic emotions. In: *Cognition & emotion* 6 (1992), Nr. 3-4, S. 169–200
- [Ekman und Friesen 1978] EKMAN, Paul ; FRIESEN, Wallace V.: Facial action coding system. In: *Environmental Psychology & Nonverbal Behavior* (1978)
- [Eyben u. a. 2010] EYBEN, Florian ; WÖLLMER, Martin ; SCHULLER, Björn: Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor. In: *Proceedings of the 18th ACM International Conference on Multimedia*. New York, NY, USA : Association for Computing Machinery, 2010 (MM '10), S. 1459–1462. – URL <https://doi.org/10.1145/1873951.1874246>. – ISBN 9781605589336
- [Gilbert u. a. 2021] GILBERT, Michaël ; DEMARCHI, Samuel ; URDAPILLETA, Isabel: FACSHuman, a software program for creating experimental material by modeling 3D facial expressions. In: *Behavior Research Methods* 53 (2021), Nr. 5, S. 2252–2272
- [Hays u. a. 1992] HAYS, Robert T. ; JACOBS, John W. ; PRINCE, Carolyn ; SALAS, Eduardo: Flight Simulator Training Effectiveness: A Meta-Analysis. In: *Military Psychology* 4 (1992), Nr. 2, S. 63–74. – URL [https://doi.org/10.1207/s15327876mp0402\\_1](https://doi.org/10.1207/s15327876mp0402_1)

- [Knapp u. a. 2013] KNAPP, Mark L. ; HALL, Judith A. ; HORGAN, Terrence G.: *Nonverbal communication in human interaction*. Cengage Learning, 2013
- [Koschke 2021] KOSCHKE, Rainer: *Eine kurze Einführung in SEE*. <https://www.youtube.com/watch?v=JsTzjJVNU3c&t=109s>. 2021. – [Online; Zuletzt aufgerufen am 12.12.2022]
- [Lucey u. a. 2010] LUCEY, Patrick ; COHN, Jeffrey F. ; KANADE, Takeo ; SARAGIH, JASON ; AMBADAR, Zara ; MATTHEWS, Iain: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: *2010 ieee computer society conference on computer vision and pattern recognition-workshops* IEEE (Veranst.), 2010, S. 94–101
- [Mehrabian 1996] MEHRABIAN, Albert: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. In: *Current Psychology* 14 (1996), Nr. 4, S. 261–292
- [Merten 2016] MERTEN, Jörg: Mimik und Emotion-Die Bedeutung der Gesichtsbewegungen. In: *Psychologie in Österreich* (2016), Nr. 5, S. 291–299
- [Oh Kruzic u. a. 2020] OH KRUZIC, Catherine ; KRUZIC, David ; HERRERA, Fernanda ; BAIENSON, Jeremy: Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. In: *Scientific Reports* 10 (2020), Nr. 1, S. 2045–2322
- [PS und Mahalakshmi 2017] PS, Sreeja ; MAHALAKSHMI, G: Emotion models: a review. In: *International Journal of Control Theory and Applications* 10 (2017), Nr. 8, S. 651–657
- [Toisoul u. a. 2021] TOISOUL, Antoine ; KOSSAIFI, Jean ; BULAT, Adrian ; TZIMIROPOULOS, Georgios ; PANTIC, Maja: Estimation of continuous valence and arousal levels from faces in naturalistic conditions. In: *Nature Machine Intelligence* 3 (2021), Nr. 1, S. 42–50
- [Unuma u. a. 1995] UNUMA, Munetoshi ; ANJYO, Ken ; TAKEUCHI, Ryoza: Fourier principles for emotion-based human figure animation. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, 1995, S. 91–96
- [Wang 2009] WANG, Haiyan: Nonverbal communication and the effect on interpersonal communication. In: *Asian Social Science* 5 (2009), Nr. 11, S. 155–159
- [Wettel und Lanza 2008] WETTEL, Richard ; LANZA, Michele: CodeCity: 3D Visualization of Large-Scale Software. In: *Companion of the 30th International Conference on Software Engineering*. New York, NY, USA :

Association for Computing Machinery, 2008 (ICSE Companion '08),  
S. 921–922. – URL <https://doi.org/10.1145/1370175.1370188>. –  
ISBN 9781605580791