PAMOCAT: Kombination von qualitativen und quantitativen Methoden zur automatischen Analyse von menschlichen Verhaltensweisen in der Kommunikation basierend auf Bewegungsdaten

Dissertation

im Studiengang
Intelligente Systeme

vorgelegt von

Bernhard-Andreas Brüning
Matr.-Nr.: 1659625

am 28.08.2014
an der Universität Bielefeld

Erstprüfer/in:	habil. Dr. Sven Wachsmuth
Zweitprüfer/in:	Prof. Dr. Philipp Cimiano

Danksagung

Ich möchte mich bei allen bedanken, die dazu beigetragen haben, dass ich diese Dissertation verfassen konnte. Zunächst geht mein Dank an meine Eltern, die mir den Weg in die Wissenschaft gewiesen haben. Dann geht mein Dank an meine Kollegen des Central Labs Team, insbesondere an Holger Dierker, der mich in verschiedener Weise unterstützt hat. Meinen Betreuern Professor Philipp Cimiano und Doktor Sven Wachsmuth gebührt spezieller Dank für ihre inspirierenden wissenschaftlichen Anleitungen und wiederholten Ermutigungen. Mein besonderer Dank geht auch an die Verantwortlichen des Exzellenzclusters der Universität Bielefeld, der sich mit dem Verstehen von kognitiver Interaktion befasst, für die finanzielle Unterstützung während meiner Arbeit an dieser Dissertation.

Kurzfassung

In der Biologie, Linguistik, Psychologie und Soziologie wird versucht, (menschliches) Interaktionsverhalten zu verstehen und zu beschreiben. In der Robotik ist ein Schwerpunkt, dieses (menschliche) Interaktionsverhalten zu modellieren, damit eine natürliche Interaktion mit Robotern möglich ist. Ein Bestandteil der natürlichen Interaktion ist unter anderem, zu erkennen, wann ein Interaktionspartner die Sprecherrolle übernehmen darf, ohne unfreundlich zu wirken und den anderen Interaktionspartner zu unterbrechen. Ein weiterer Schwerpunkt ist die Analyse, wie verschiedene Menschen beim Sprechen gestikulieren, um z. B. gleiche Sachinhalte mittels Sprache und sprachbezogener Gesten zu beschreiben. Sind aus solchen Analysen Verhaltensmuster erkannt worden und wurden diese Interaktionsverhaltensweisen implementiert, muss verifiziert werden, ob Menschen das z. B. von einem Roboter oder sozialen Agenten ausgeführte Verhalten als natürlich empfinden. Eine gängige Methode, ein solches Verhalten zu analysieren, ist die Aufzeichnung in verschiedenen multimedialen Daten wie Audio und Video, sodass diese anschließend im Detail analysiert werden können. Leider ist dieser Videoanalyseprozess sehr zeitintensiv, da er manuell durch Menschen durchgeführt werden muss. Um eine Bewegung in einem Video analysieren zu können, muss diese erst aus dem Video extrahiert werden, wobei dieses nicht immer genau durchgeführt werden kann. Dieses kann der Fall sein, wenn Gelenk- und andere Körperteilepositionen nicht genau bekannt sind, da diese Körperteile verdeckt sein können. Da diese Analyse ein zeitintensiver Prozess ist, der durch viele Arbeitsstunden teuer wird, gibt es Bemühungen, möglichst Mechanismen zu finden, durch die diese Arbeiten automatisch durchgeführt werden können. Als erstes Problem muss bei einer Analyse von Videodaten ermittelt werden, was Personen sind und in welcher Körperhaltung sie sich befinden. Allgemein funktioniert dieses, ist allerdings fehleranfällig. Um genauere Daten der Interaktionen zu erhalten und um auch automatische Analysen durchführen zu können, geht ein Trend dazu über, weitere modale Daten wie Motion-Capture-Daten zusätzlich aufzuzeichnen. Dadurch kann die Bewegung der interagierenden Personen viel genauer in räumlicher Relation zueinander analysiert werden. Um dieses durchführen zu können, stellen sich die Fragen, „wie die Motion-Capture-Daten sinnvoll mit angemessenem Arbeitsaufwand für die Untersuchungen genutzt werden können“ und „wie die Interaktionen mehrerer Personen über eine längere Zeitspanne robust aufgezeichnet werden können“. Beim Motion-Capturing ist eine lange Aufnahme mit einem Vielfachen dieser Zeit als Nachbearbeitungsphase verbunden. In dieser Nachbearbeitungsphase werden die Daten aufgearbeitet, damit einzelne Marker immer den zugehörigen Körperteilen zugeordnet werden können. Um einen deutlichen Nutzen aus dem Motion-Capturing ziehen zu können, darf die Zeit, die für das zusätzliche Motion-Capturing aufgewendet wird, nicht höher sein als die Zeit, die für das Annotieren der Video-Analyse aufgewendet würde. In dieser Arbeit wird gezeigt, wie das Motion-Capturing mit einem angemessenen Zeiteinsatz verwendet werden kann, um automatische Analysemöglichkeiten nutzbringend durchführen zu können. Dabei wird auf die Frage-

stellung eingegangen, „was die Motion-Capture-Daten für Möglichkeiten bei der Verhaltensforschung bei Interaktionen bieten“. Dazu wird gezeigt, dass diese neuen Möglichkeiten in einer automatischen detaillierten Analyse liegen, die eine standardisierte Basis für Analysen mit einer immer gleichbleibend guten Qualität liefern.

Um die Nützlichkeit der Motion-Capture-Daten hervorzuheben, wird gezeigt, wie diese im Forschungsalltag eingesetzt werden können. Die hierbei gesammelten Erfahrungen sind in die Entwicklung eines Annotationstools „PAMOCAT“ eingegangen, bei dem verschiedene elementare Verhaltensbestandteile als abstrakte Kategorien (wie z. B. Bewegung in elementaren Gelenken, etwas angucken, Handbewegungen oder Posen) automatisch annotiert werden können. Dabei haben sich verschiedene elementare Kategorien herauskristallisiert, die ein breites Spektrum von möglichen Einsatzbereichen in der Verhaltensforschung bieten. Dazu wird eine Basis von elementaren Interaktionsphänomenen bereitgestellt, die durch Kombinationen mit anderen Interaktionsphänomenen als Suche nach Zeitpunkten, bei denen diese zusammen auftreten, angesetzt werden kann. Dadurch ist eine detailliertere Analyse komplexen Verhaltens einfacher und schneller möglich, als es zuvor möglich war. Um diese Analysefunktionalität einem möglichst großen Anwenderkreis bereitzustellen, ist ein Graphical User Interface - GUI entwickelt worden, welches in Zusammenarbeit mit Endnutzern optimiert wurde. Damit ergeben sich neue Möglichkeiten bei der Analyse großer Korpora und es kann viel Zeit eingespart werden, sodass die Aufmerksamkeit auf eine detaillierte Analyse fokussiert werden kann.

Schlagwörter: PAMOCAT, Annotation, Bewegungsanalyse, Elementarbewegung, Bewegungssegmentation, Posturerkennung, Multi Personen-Motion-Capturing, Verhaltensanalyse, Konversation-Analyse.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

GUI	Graphical-User-Interface
KA	Konversationsanalyse
HRI	Human-Robot-Interaktion
HHI	Human-Human-Interaktion
MMI	Mensch-Maschine-Interaktion
DOF	Degree of Freedom
Mocap	Motion-Capturing - Bewegungserfassung

1 Einleitung

In der auf den Menschen bezogenen Verhaltensforschung mit dem Schwerpunkt Mensch-Maschine-Interaktion MMI wird daran gearbeitet, verschiedene Verhaltensweisen zu verstehen, um die Interaktion mit Robotern oder Maschinen einfacher und natürlicher gestalten zu können. Dabei grenzt dieses sehr stark an die Forschungsbereiche der Soziologie, der Linguistik, der Psychologie und der Biologie, bei denen allgemein versucht wird, das menschliche Verhalten zu verstehen und zu beschreiben [1]. In diesen Forschungsbereichen werden meistens ähnliche Vorgehensweisen und gleiche Werkzeuge genutzt, um den Arbeitsablauf zu unterstützen. Genau an dieser Stelle setzt diese Arbeit an, die Verhaltensforschung zu unterstützen, um menschliches Interaktionsverhalten zu erforschen und dieses bei der MMI zu nutzen. Dazu wird im späteren Verlauf dieser Arbeit gezeigt, wie Teile der menschlichen Bewegung und grundlegende Interaktionsbestandteile automatisch erkannt und die entsprechenden Zeitpunkte genau markiert werden können. Diese markierten und mit elementaren Verhaltensweisen (z. B. Handgelenk bewegen oder jemanden angucken) bezeichneten Sequenzen werden von den Verhaltensforschern genutzt, um komplexere Verhaltensweisen zu analysieren. Der Fokus dieser Arbeit beinhaltet einmal die Ermittlung von elementaren, auf die Bewegung bezogenen Verhaltensbestandteilen, das Bereitstellen einer Suchfunktionalität nach Kombinationen dieser Bewegungsbestandteile und die Bereitstellung dieser Funktionalität in einer Weise, dass so gut wie jede Person diese nutzen kann.

Im Folgenden werden die Hintergründe dieser Arbeit aus Sicht der MMI und der Verhaltensforschung betrachtet. Am Ende dieses Kapitels werden eine Zielsetzung und ein Überblick über die gesamte Arbeit gegeben.

1.1 Hintergrund

Seit der Konstruktion der ersten Computer wird die Interaktion mit diesen Maschinen ständig weiterentwickelt. Diese Interaktion der Menschen mit den Maschinen wird immer mehr auf Bewegungselemente oder Bewegungsgesten erweitert, von denen die Ursprünge in der natürlichen Mensch-Mensch-Interaktion zu finden sind. Kerngedanke ist es, die Benutzung oder Bedienung der Maschinen zu erleichtern und an die natürliche Interaktion von Menschen miteinander anzulehnen. Bei der Mensch-Mensch-Kommunikation spielt das Zeigen eine große Rolle, welches dem Mitmenschen auf natürliche Weise symbolisiert, was er z. B. meint oder haben will. Das Resultat dieser Mensch-Maschine-Interaktion ist, dass heutzutage immer mehr Geräte mit Touchscreens ausgestattet werden, bei denen der Benutzer auf das zeigen kann, was er haben oder benutzen will.

Im Bereich der Human-Roboter-Interaktion - HRI ist es das Ziel, Roboter zu bauen, mit denen natürlich interagiert werden kann. Dazu wird gezielt das „Mensch-zu-Mensch“ Kommunika-

tionsverhalten analysiert und versucht, dieses im Detail zu verstehen. Die so gewonnenen Erkenntnisse können dann in ein Reaktionsmodell eines Roboters oder einer Maschine integriert werden, um die Interaktion angenehmer, leichter und natürlicher zu gestalten [2] [3] [4].

Aber um dieses realisieren zu können, muss diese Interaktion als zwischenmenschliche Kommunikation im Detail analysiert werden [5]. Dieses kann als Interaktion auf verschiedenen Ebenen aufgefasst werden, in einer Ebene der Sprache, einer der körperlichen Bewegung und einer der Gesichtsmimik. In der Prosodie (sprachlichen Ebene) wird die genaue Ausdrucksweise der gesprochenen Sprache analysiert, welche dazu meistens erst in eine schriftliche Form überführt wird, um den genauen Satzbau analysieren zu können. In der Ebene der körperlichen Bewegung wird die Bewegung in kleinere Bewegungssequenzen zerlegt, um diese Darstellungen mit einer zeitlichen Abfolge von textuellen Beschreibungen in Relation setzen zu können. In der Ebene der Gesichtsmimik werden die jeweiligen Gesichtsausdrücke ermittelt und durch Annotationen (eine textuelle Darstellung zum zeitlichen Geschehen) zur späteren Analyse aufbereitet. Diese Arbeit wird sich hauptsächlich auf die Bewegungen des Körpers konzentrieren und die mit diesen verbundenen möglichen Posen und Gesten. Im Folgenden werden Begriffe in ihrer Bedeutung beschrieben, die für diese Arbeit wichtig sind.

Pose:
Beschreibt die Position oder Stellung von den Gelenken eines Menschen zu einem bestimmten Zeitpunkt.

Geste:
Beschreibt die Änderung von Gelenken über einen ausgedehnten Zeitraum, bei der eine oder mehrere unterschiedliche Posen eingenommen werden können, um Gedanken oder Gefühle auszudrücken [6]. *Sie beschreibt ein kommunikatives Bewegen der Hände und Arme, um wie mit der Sprache Gedanken, Gefühle und Intentionen auszudrücken* [7].

Die Begriffe Posen und Gesten spielen eine zentrale Rolle in dieser Arbeit, sie werden im Verlauf dieser gesamten Arbeit nicht nur in direktem Zusammenhang mit Gesprächs- und Gestenanalyse benutzt. Die Themengebiete, die in dieser Arbeit vertieft werden, sind „turn taking“ (Wechsel der aktiven sprechenden Person) und sprachbezogene Gesten. Als Untersuchungsrahmen der sprachbezogenen Gesten werden Probanden in einem Szenario zusammengeführt, bei dem verschiedene Personen gleiche Sachverhalte in geometrischer Anordnung (Beschreiben eines Weges) durch sprachbezogene Gesten den anderen beschreiben sollen.

Allgemein ergibt sich die Fragestellung, wie Bewegung verwendet werden kann, um verschiedene Aspekte von Verhalten zu analysieren. Generell ist das menschliche Interaktionsverhalten sehr komplex und verbindet viele verschiedene Merkmale zu bestimmten Zeitpunkten. In Korpora werden diejenigen Merkmale gesucht, die auf eine bestimmte Verhaltensweise hindeuten. Dabei ist das Finden dieser verschiedenen Merkmale, die bei einer bestimmten Verhaltensweise zusammenkommen, aber auch das Wiederfinden der einzelnen Merkmale in

Kombinationen schwierig. Dazu wird die Verhaltensweise basierend auf den kinetischen Bewegungsgesten nach elementaren Phänomenen untersucht.

> ***Phänomen:***
> *Allgemein beschreibt Phänomen etwas Wahrnehmbares, ein Ereignis und auch etwas Besonderes* [8]. *In dieser Arbeit wird unter Phänomen ein elementarer Bewegungsbestandteil verstanden, der sich auf verschiedene abstrakte Kategorien bezüglich der Bewegung in Interaktionen bezieht.*

Dieses kann ein statisches oder dynamisches Phänomen aus einer Bewegung sein. Ein Beispiel hierfür ist eine dynamische Bewegung eines bestimmten Gelenks oder die Bewegung einer Hand. Ein statisches Phänomen ist „auf etwas Orientieren“ oder eine einzelne statische Pose. Weiterhin werden Phänomene unterschieden, die sich auf einzelne Personen beziehen und personenübergreifend sind. Mit personenübergreifendem Phänomen ist gemeint, dass nicht nur eine Person daran beteiligt ist, z. B. „es orientieren sich zwei Personen zueinander“ oder „eine Person kommt mit der Hand einer anderen nah“. Durch das Finden dieser elementaren Phänomene kann ein Korpus analysiert werden, indem die Zeitpunkte gefunden werden, bei denen die Phänomene in einer bestimmten Konstellation zusammen vorkommen.

> ***Konstellation:***
> *Eine Konstellation beschreibt das zeitliche Zusammentreffen von verschiedenen Phänomenen.*

Ein Beispiel hierfür wäre das Auffinden einer Zeigegeste, die durch zwei verschiedene elementare Phänomena gefunden werden kann, einmal eine Pose des Körpers, bei der ein Arm vom Körper weg gerichtet ist, und eine Bewegungsaktivität der Hand. Mithilfe dieser Konstellationssuche können Korpora von verschiedenen Studien gezielt auf verschiedene Verhaltensbestandteile durchsucht werden. Allgemein ist das Durchführen von Studien ein wichtiger Bestandteil der Verhaltensforschung in der Mensch-Mensch-Kommunikation. Bei diesen Studien überprüft man durch Experimente, ob eine Hypothese bezüglich einer Verhaltensweise richtig ist, und kreiert aus den Analyseergebnissen neue Theorien. Nachdem das Experiment durchgeführt wurde, werden die Daten für eine spätere genaue Analyse durch eine Annotation aufbereitet. Dazu werden unterschiedliche abstrakte Kategorien gewählt, nach denen annotiert wird. Die Bezeichnung des Annotierens stammt aus der Linguistik und beschreibt das Hinzufügen von Zusatzinformationen zu Rohdaten. Dieses ist ein sehr zeitintensiver Prozess, der durch diese Arbeit mit automatischem Annotieren von verschiedenen Bewegungsbestandteilen unterstützt werden soll. Allgemein können die Grundlagen für das Annotieren der Rohdaten aus geschriebenem Text, Bildern oder auch aus Videos bestehen. Angefangen hat das Annotieren bei Texten, wodurch nachträglich eine Analyse der genauen Struktur möglich wurde. Das Annotieren von Körperbewegungen (auch in Echtzeit) und die körperliche Interaktion mit anderen Menschen [9] wurden durch verschiedene spezielle Gestik-Notation-Schemata oder Coding-Schemata eingeführt [10] [9]. Der Begriff „Coding“ bezeichnet das aktive Erstellen

von Annotationen. Durch Coding-Schemata werden Bewegungs- und Interaktionsbestandteile auf eine einheitliche Weise durch spezifische Vorgaben möglicher Kombinationen verschiedener Bestandteile dieser Gesten beschrieben.

Coding-Schema:

Beschreibt ein Vorgehen, wie einheitlich annotiert werden sollte. Dazu werden Vorgaben für die Bewegungsbestandteile definiert.

Beispiele eines Coding-Schemas für die geometrische Bewegung ohne analytische Bestandteile können Handform, Handorientierung, Handposition und Bewegungsart sein [11]. Solche Coding-Schemata sollen die individuelle Auffassung von Situationen einzelner Individuen reduzieren und ein einheitliches Vorgehen für das Annotieren definieren. Damit soll erreicht werden, dass nicht fehlerhaft Schlussfolgerungen aus Annotationen gezogen werden, die nur auf einer unterschiedlichen Auffassung einer Situation beruhen.

Die Entwicklung und Verbreitung von Tonaufnahmegeräten ermöglichte es, gesprochene Texte aus einer Unterhaltung im Nachhinein detailliert zu analysieren. Dadurch wurde es möglich, nicht nur genaue Wortreihenfolgen zu analysieren, sondern auch, wie die Wörter betont wurden. Zum Beispiel könnte dieses eine ängstliche zittrige Stimme sein oder das Hervorheben einzelner Wörter, um Andeutungen zu machen. Dieses Verfahren konnte mit Hilfe der neuen Technik in die Analyse mit einbezogen werden. Mit der Verfügbarkeit von Videokameras konnten später zusätzlich Video-Daten für die spätere Analyse mit aufgezeichnet werden. Dadurch ergaben sich die zusätzlichen Möglichkeiten, im Nachhinein die Körpersprache, Mimik und die Bewegung im Kontext zur Umgebung mit in die Analyse einzubeziehen. Unter anderem konnten so anschließend Rückschlüsse auf den emotionalen Zustand der Versuchsperson gezogen werden (Mimik und Körperhaltungen). Zudem wurde es möglich, komplexere Verhaltensweisen wie die Interaktion in einer Gruppe später detaillierter zu analysieren. Darüber hinaus konnte erstmals die körperliche Gestik bei verschiedenen verbalen Äußerungen auf Basis der Videodaten analysiert werden.

Allerdings werden solche Analysen mit weiteren Medien immer komplexer. Daher ist viel manueller Annotationsaufwand nötig, um die körperliche Gestik, mit der z. B. eine sprachliche Aussage untermauert wird, in die Analyse einzubeziehen, und z. B. einen Widerspruch zwischen verbaler Aussage und einer körperlichen Geste zu ermitteln.

1.2 Motivation

In der heutigen Verhaltens- und der Gestenforschung wird das Annotieren der Rohdaten genutzt, um zu analysieren, wie gesprochene Sprache in Bezug auf die körperlichen Gesten verwendet wird. Speziell der genaue Zusammenhang zwischen diesen ist von Interesse. Beispielsweise werden Zeigegesten genutzt, um die Rolle des Sprechers zu übernehmen oder andere Handbewegungsgesten von hinten nach vorn, um das „entlang eines Weges gehen“ zu symbolisieren. Andere Gesten untermauern das Gesprochene direkt und untermalen bestimm-

te Wörter. Um diese Verhaltensweisen zu untersuchen, wird in der Verhaltensforschung, der Gesprächsanalyse oder der Konversationsanalyse die Standard Herangehensweise genutzt, um Hypothesen zu evaluieren (in der Gesprächsanalyse) oder neue Hypothesen (in der Konversationsanalyse) zu erzeugen. Bei der Evaluierung der Hypothesen oder der Erzeugung neuer Hypothesen werden die annotierten Daten genutzt und bilden dazu die Basis. Verhaltensforscher nehmen die Interaktionen ihrer Versuchspersonen auf, um an diesen später detaillierte qualitative Analysen durchführen zu können. Durch diese verschiedenen hervorgehobenen Zeitpunkte (Annotationen) können sie Verhaltensweisen von verschiedenen Personen miteinander vergleichen und analysieren. Dabei werden die Rohdaten in Bezug auf verschiedene Forschungshypothesen annotiert, die meist Grundlage für die spätere Überprüfung von Hypothesen, das Belegen von Hypothesen oder aber das Aufstellen neuer Hypothesen bilden.

Die annotierenden Personen sind „nur“ Menschen, und es kommt vor, dass diese Fehler machen oder aber auch einfach Sachverhalte anders wahrnehmen [12]. Die einzelnen Personen haben unterschiedliche Kenntnisse, die sie in ihre Annotationen stecken können, oder auch einen anderen Auffassungssinn. Dabei ist die sich ändernde Qualität ein Problem für die Analyse, da diese später zu Fehlinterpretationen führt könnte. Um diese Fehlerquelle zu vermeiden, werden diese Annotationen meist nicht nur von einer Person, sondern gleich von mehreren Personen durchgeführt, damit nachher die einzelnen Annotationen zu einer qualitativ hochwertigen Annotation zusammengeführt werden können. Zudem ist es schwierig, eine gleichbleibend hohe Qualität über den gesamten Korpus aufrechtzuerhalten, wenn mehrere Leute mit unterschiedlichen Qualifikationen einen Korpus bearbeiten, da dadurch nicht einheitliche Annotationen erzeugt werden. Macht eine Person in einer Situation immer genau den gleichen Fehler, hat man eine gute Chance, diesen nachträglich zu beseitigen. Wünschenswert ist hier eine einheitliche Qualität, die gegebenenfalls ein wenig schlechter sein kann als die manuellen Annotationen, aber mit einer einheitlichen Qualität.

Der gesamte Ablauf des Annotierens ist ein sehr zeitaufwendiger Prozess. Die Zeit für das Annotieren von Sprache kann ungefähr das 35-fache der Aufnahmezeit betragen, die Übersetzung der Annotationen in eine andere Sprache kann noch einmal die 25-fache Zeit der Aufnahmezeit erfordern, und bei der Annotation von Gesten kann die gesamte Annotationszeit sogar mehr als das 100-fache der Aufzeichnungszeit kosten[1] [13]. Dieser Zeitaufwand wird noch höher, wenn nicht nur einzelne Personen annotiert werden, sondern eine Interaktion von mehreren Personen. Dabei erhöhen nicht nur die Anzahl der beteiligten Personen die Zeit zum Annotieren[2], sondern auch die Interaktionen in der Gruppe, da eine größere Anzahl von Kategorien bearbeitet werden müssen.

Um diesen Prozess des Annotierens zu vereinfachen, gibt es eine Reihe von Tools, die es einer annotierenden Person ermöglichen, Zusatzinformationen wie Beschreibungen und Analy-

[1] DOBES Project www.mpi.nl/dobes

[2] Je nach Aufgabenstellung sprechen und interagieren nicht alle Personen gleichzeitig, daher ist die Zeit nicht direkt proportional zur Personenanzahl.

seelemente in exakte zeitliche Verbindung zu den Aufnahmen zu bringen. Dabei haben diese einzelnen Tools verschiedene Zusatzfunktionen und auch eingeschränkte automatische Annotationsfunktionen, die das Annotieren erleichtern und teilweise übernehmen. Die am meisten verbreiteten Softwareprogramme im Bereich der multimodalen Annotation sind aktuell: ELAN [13], ANVIL [14], EXMARaLDA [15], TASX [16] und Praat [17][3]. Auf diese Tools wird im Kapitel 3.1 näher eingegangen werden. Leider bieten diese Tools nur wenige Möglichkeiten, automatische Annotationen durchzuführen.

Ein Versuch vieler Forscher der letzten Jahre besteht darin, neben den heute üblichen multimedialen Datenquellen wie Audio und Video eine weitere modale Datenquelle, nämlich das Motion-Capturing, mit in die Analyse einzubeziehen. Die automatische Annotation oder die Visualisierung einer Interaktion zwischen mehreren Leuten kann leider noch keines dieser Tools durchführen. An dieser Stelle setzt diese Dissertation an. Um Motion-Capturing für die Verhaltensforschung nutzen zu können, muss man sich jedoch zunächst eine Reihe von Fragen stellen.

- Wie können multiple Personen über eine längere Zeitspanne robust aufgezeichnet werden?
- Wie kann die Gesamtzeit die durch die zusätzliche Datenquelle (Motion Capture) entstehende Vor- und Nachbereitungs-Zeit gegenüber dem Mehrgewinn, der aus diesen Daten gewonnen werden kann, in einem angemessenen Verhältnis halten?
- Was kann aus diesen Daten an nützlichen Zusatzinformationen gewonnen werden?
- Wie können diese Zusatzinformationen praktikabel in den Forschungsarbeitsablauf integriert werden, sodass diese schnell und einfach verwendet werden können?
- Welche technischen Systeme (Marker basiert, rein optisch, magnetisch usw.) sind für das Analysieren von Gruppeninteraktionen nutzbar?
- In welcher Form können automatische Annotationen auf Basis von Motion-Capture-Daten durchgeführt werden?
- Wie sehen elementare Bestandteile basierend auf Gruppeninteraktion aus?
- Können verschiedene Verhaltensweisen automatisch erkannt werden und wie gut funktioniert diese Erkennung?
- Wie können die Motion-Capture-Daten bestmöglich visualisiert werden und welche Bestandteile müssen für eine gute Analyse der Daten hervorgehoben werden?
- Wie kann diese gesamte Funktionalität für das Annotieren von Verhaltensweisen genutzt werden?

Die hierbei gesammelten Erfahrungen liegen der Entwicklung des Annotationstools „PAMOCAT“ zugrunde, bei dem verschiedene abstrakte Kategorien oder elementare Phänomene automatisch annotiert werden können. Dabei haben sich mehrere elementare Phänomene herauskristallisiert, die in dieser Arbeit entwickelt wurden und ein breites Spektrum an

[3] Praat ist eigentlich nicht multimodal, sondern nur audiobasiert, wird aber sehr stark in Kombination mit den anderen Tools eingesetzt.

Einsatzbereichen ermöglichen. Dafür steht ein Katalog von elementaren Phänomenen zur Verfügung, mit dem durch Kombination dieser Phänomene komplexere Verhaltensweisen wiedergefunden werden können. Um diese Funktionalität einem möglichst großen Anwenderkreis bereitzustellen, wird diese mit einer Graphical User Interface - GUI zusammen bereitgestellt. Damit werden neue Möglichkeiten der Analyse durch automatische Annotation von großen Korpora durch Zeitersparnis und die Lenkung der Aufmerksamkeit auf die Interpretierten ermöglicht.

1.3 Zielsetzung

Annotation im Allgemeinen und insbesondere die von menschlicher Bewegung verlangt ein gewisses Maß an Interpretation. Die annotierenden Personen erfassen menschliche Bewegungen je nach Charaktertyp, Bildung und Gemütszustand mal genauer und mal ungenauer in abstrakten Kategorien. Um die Annotationen für die Auswertung in Analysen nutzen zu können, müssen diese ein gewisses Maß an Qualität aufweisen. Daher ist es nötig, mehrere Annotationen anzufertigen, die zu hochwertigeren Annotationen zusammengeführt werden können [12].

Speziell im Bereich der Erforschung von Gesten mit Kameras ist es schwierig, auch unter optimalen Bedingungen die verschiedenen Phasen der Bewegungen zu finden; hinzu kommt noch, dass die zugrundeliegenden Posen der Versuchspersonen nicht immer eindeutig aus einem bestimmten Blickwinkel, oder auch wegen ungenügender Auflösung oder Verdeckung, gesehen werden können. Daher sind automatische Annotationen basierend auf Videodaten im diesen Bereich schwierig und gegebenenfalls ungenau. Bei manuellen Annotationen von mehreren Personen spielt der menschliche Faktor eine große Rolle, da auch im Falle sehr guter Vorbereitung die Ergebnisse immer noch unterschiedlich sein können. Menschen würden nicht exakt gleiche Kriterien oder Merkmale zur Annotation oder Kategorisierung der durchgeführten Bewegung bei der Transkribierung verwenden bzw. diese mehr oder weniger gleich interpretieren.

Grundgedanke dieser Arbeit ist es, zu erarbeiten, wie sich Motion-Capturing als weitere Modalität zur Annotation eignet, und was, basierend auf den Motion-Capture-Daten, an Annotationen automatisch erkannt und durchgeführt werden kann. Die Grundlage dafür bieten die Motion-Capture-Daten, welche eine hohe Präzision von menschlichen Bewegungsdaten für eine empirische Analyse ermöglichen [18].

Dabei sollen Muster in der Bewegung auf einer Ebene gesucht werden, die das elementare Analysieren von Bewegung in allgemeinen Situationen ermöglichen, auch wenn diese Bewegungen der gleichen Gesten unterschiedlich aufgebaut sind. Das Annotationstool PAMOCAT - Pre Annotation Motion Capture Tool wird vorgestellt, das im Rahmen dieser Dissertation entwickelt wurde. Der Kern dieser Arbeit besteht in der Ermittlung der unterschiedlichen Weisen, in denen das Motion-Capture-Daten Tool für den Verhaltensforschungszyklus für

eine Vielzahl an Personen eingesetzt werden kann. Dazu werden verschiedene elementare Bestandteile und Kategorien bei der Interaktion in einer Gruppe erarbeitet, die die Grundlage bilden, komplexere Interaktionssituationen zu analysieren. Diese abstrakten Kategorien sollen automatisch annotiert werden können. Die anschließende Auswertung der Daten soll ebenfalls durch eine Suche nach Kombinationen dieser Kategorien unterstützt werden, ebenso wie viele verschiedene multimodale Visualisierungen wie Motion-Capture, Videos und Geschwindigkeitsplots. Diese Visualisierungen sollen verschiedene Kategorien hervorheben und so dem Forscher eine leichtere und schnellere Analyse ermöglichen. Am Rande wird ein Einblick in verschiedene Hardwaresetups gegeben, um zu vermitteln, welche technischen Systeme für welche Art von Verhaltensforschung einsetzbar sind. Außerdem wird untersucht, wie mit Fehlern in der Aufzeichnung von Motion-Capture-Daten umgegangen werden kann, und es werden verschiedene Analysefunktionen zum Ermitteln elementarer Bestandteile von Verhaltensweisen basierend auf der Bewegung bereitgestellt.

1.4 Entstehungsumgebung

Diese Arbeit ist im Rahmen einer Anstellung als wissenschaftlicher Mitarbeiter beim CITEC im „Central Lab“ entstanden. Das CITEC wurde als Exzellenzcluster für Kognitive Interaktionstechnologien 2007 durch die deutsche Bundesregierung finanziert. Ziel dieser Institution ist es, Interaktive Intelligente Systeme in vier zentralen Forschungsbereichen zu entwickeln. Diese sind: Bewegungsintelligenz, Systeme mit Aufmerksamkeit, Situierte Kommunikation sowie Gedächtnis und Lernen [19]. Im CITEC selber sind mehrere Forschungsgruppen von verschiedenen Arbeitsgruppen und Fakultäten vorhanden, die es ermöglichen, interdisziplinär zu forschen. Diese Arbeitsgruppen sind von der Biologie, Linguistik, Mathematik, Psychologie, Sport und der Technischen Fakultät. Das „Central Lab“ ist als ein zentrales Labor für Experimente und Demonstrationen gedacht. Darüber hinaus stellt das „Zentral Labor“ Infrastruktur des CITEC bereit und gibt in verschiedenen Bereichen technische Unterstützung. Ein Aufgabenbereich des Zentral Labors ist die Unterstützung bei Motion-Capture-Systemen und virtuellen Visualisierungen. In diesem Rahmen wurden mehrere Studien durchgeführt, die als Grundlage dieser Arbeit dienen.

1.5 Überblick

Im folgenden Kapitel 2 wird kurz auf die Grundlagen der biologischen Merkmale des menschlichen Bewegungsapparates eingegangen, und es wird erklärt, wie diese mathematisch beschrieben werden können. Anschließend werden Grundlagen in der Verhaltensforschung vorgestellt. In Kapitel 3 wird der Stand der Forschung und Technik in den diesbezüglich relevanten Bereichen vorgestellt. Dazu gehören aktuelle Motion-Capture-Systeme, aktuelle Annotationstools, und verschiedene einzelne Arbeiten im Bereich von Motion-Capturing und des automatisches Annotierens. Die hier erarbeitete zugrundeliegende Technik des Motion-Capturing, die in der Verhaltensforschung für die Analyse von Gruppeninteraktion eingesetzt

werden kann, wird in Kapitel 4 vorgestellt. Anschließend werden die im Rahmen dieser Arbeit erstellten Korpora in Kapitel 5 vorgestellt.

In Kapitel 6 wird die praktische Anwendung des Annotationstools „PAMOCAT“ konzeptuell aufgezeigt. Dazu zählen elementare Bestandteile der Bewegung, die es ermöglichen, komplexere Verhaltensweisen zu analysieren. Anschließend, in Kapitel 7, folgt ein Überblick über die implementierte Software mit den darin erstellten Bibliotheken und den zugrunde liegenden Abhängigkeiten. Das Tool PAMOCAT wird selber in Kapitel 8 mit verschiedenen Anwendungsfällen vorgestellt. Um den praktischen Nutzen von PAMOCAT darzulegen, wurden zwei Studien durchgeführt, in der das Tool PAMOCAT in Kapitel 9 validiert und die Nützlichkeit der Funktionalität in Bezug auf andere Tools ermittelt wird. Darauf folgt ein Schlusswort in Kapitel 10 und ein Ausblick auf Möglichkeiten der Erweiterungen, um die Analysen noch besser durchführen zu können.

2 Grundlagen

Die Mensch-Maschine-Interaktion wird von der Mensch-Mensch-Interaktion inspiriert. Die Grundlage dieser Interaktionen bildet der menschliche Bewegungsapparat. Um die menschliche Bewegung im Detail analysieren zu können, wird diese in einer mathematischen Darstellungsform beschrieben. Die mathematische Beschreibung bildet die Basis der später beschriebenen automatischen Annotationen. Eine weitere Grundlage ist die computergrafische Darstellung von Bewegungen, welche zum einen verwendet wird, um die Bewegungen zu visualisieren, und zum anderen, eine Grundlage für die automatischen Annotationen darstellen. Ziel dieser Arbeit ist die Verhaltensforschung zu unterstützen, welche in verschiedenen Forschungsdisziplinen angesiedelt ist. Die Verhaltensforschung, die in den verschiedenen Bereichen wie Soziologie, Psychologie, Linguistik und der Biologie durchgeführt wird, wird hier aus allgemeiner Sichtweise der Linguistik betrachtet. Dazu werden zunächst Grundlagen der Linguistik vorgestellt. Die technischen Systeme hinter dem Motion-Capturing werden am Ende dieses Kapitels vorgestellt, um einen Einblick zu erhalten, welche Systeme sich für welche Einsatzbereiche eignen.

2.1 Mathematische Beschreibung von menschlicher Bewegung

Die Kinematik beschreibt die Bewegung von Körpern im Raum. Die menschliche Bewegung kann mit der Kinematik mathematisch beschrieben werden. Dazu müssen zunächst die Bewegungseigenschaften des menschlichen Skeletts betrachtet werden. Mit deren Hilfe kann das menschliche Skelett als eine Reihe von kinematischen Ketten durch eine Folge von verschiedenen Transformationen darstellt werden, welches eine mathematische Darstellungsform ergibt.

2.1.1 Biologische Bewegungsfreiheiten des menschlichen Skelettes

Die biologische Grundlage der Beweglichkeit des Menschen bildet das Skelett, dessen Gelenke, Muskeln und Sehnen. Die Gelenke ermöglichen die Bewegung entsprechend verschiedener Freiheiten und Bewegungsmöglichkeiten. Es gibt sechs Typen von Gelenken. Diese sind das Scharniergelenk, das Zapfengelenk (Radgelenk), das Sattelgelenk, das Kugelgelenk, Planesgelenk und das Eigelenk [20] siehe **Abbildung 1**. Das Planesgelenk ist ein Wirbelgelenk, welches hier nicht von Interesse ist, da die Krümmung der einzelnen Gelenke im Rücken mit der später vorgestellten Technik (Kapitel 4) nicht erfasst werden kann. Diese verschiedenen Typen von Gelenken unterscheiden sich in den unterschiedlichen Bewegungsmöglichkeiten oder auch Bewegungsfreiheitsgraden, zu Englisch „Degree of Freedom“ - DOF. Diese DOFs bei den menschlichen Gelenken entsprechen der Anzahl und Orientierung der Achsen eines Gelenkes, um die rotiert werden kann. Sie sind in der **Abbildung 1** mit ihren DOFs darge-

stellt. Die Rotationsachsen sind mit einem kreisförmigen Pfeil in einem Koordinatensystem in den Gelenken eingezeichnet.

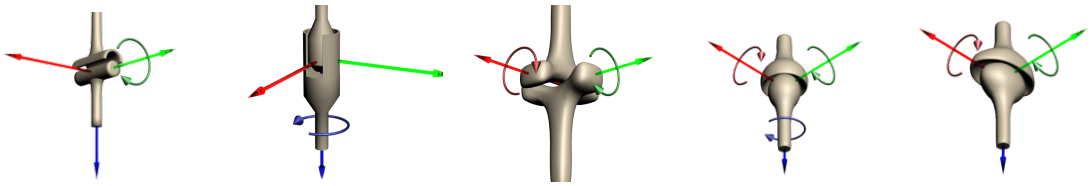

Abbildung 1 Fünf der sechs grundlegenden Gelenktypen, nämlich: Scharnier-, Zapfen-, Sattel-, Kugel- und Eigelenk (Inspiration durch [21])

Im Folgenden sind diese einzelnen Gelenktypen [21] mit ihren Positionen im Skelett und ihren DOFs aufgeführt:

- Das Scharniergelenk im Ellenbogen hat 1 DOF, die Achse der Bewegungsfreiheit (oder die Rotationsachse) ist ein Vektor, der senkrecht auf dem Oberarm und dem Unterarm liegt.
- Das Zapfengelenk z. B. im Ellenbogen ermöglicht es, den Unterarm um eine Rotationsachse zu drehen, die vom Ellenbogengelenk zur Hand geht.
- Im Daumen ist das Sattelgelenk mit 2 DOFs. Es ermöglicht, den Daumen seitlich und aufrecht zu bewegen.
- Das Kugelgelenk in der Hüfte und in der Schulter hat drei Rotationsachsen, die orthogonal zueinander stehen; dieses entspricht einem Bewegungsfreiheitsgrad von 3 DOFs.
- In der Hand liegt das Eigelenk mit 2 DOFs; dieses erlaubt eine Bewegung der Hand seitlich und aufrecht bezüglich des Unterarmes.

Diese 5 verschiedenen Gelenktypen sind in der folgenden **Abbildung 2** entsprechend der Position im Skelett dargestellt. Der Übersichtlichkeit halber sind nicht alle Gelenke des gesamten Skeletts hervorgehoben.

2.1.2 Mathematische Repräsentation von Gelenken

Um das gesamte Skelett mathematisch darstellen zu können, müssen erst einmal die einzelnen Gelenke beschrieben werden. Zu diesem Zweck werden zunächst einzelne Gelenktypen mathematisch beschrieben; anschließend wird ein Verfahren vorgestellt, mit dem es möglich ist, ein gesamtes menschliches Skelett mathematisch zu beschreiben. Ein 1 DOF Gelenk kann durch eine Rotationsmatrix dargestellt werden. Ein 2 DOF Gelenk kann durch die Multiplikation zweier Rotationsmatrizen mathematisch beschrieben werden, bei dem die Rotationsachsen sich unterscheiden. Bei der mathematischen Beschreibung eines Gelenks mit 3 DOFs gibt es unterschiedliche Darstellungsmöglichkeiten.

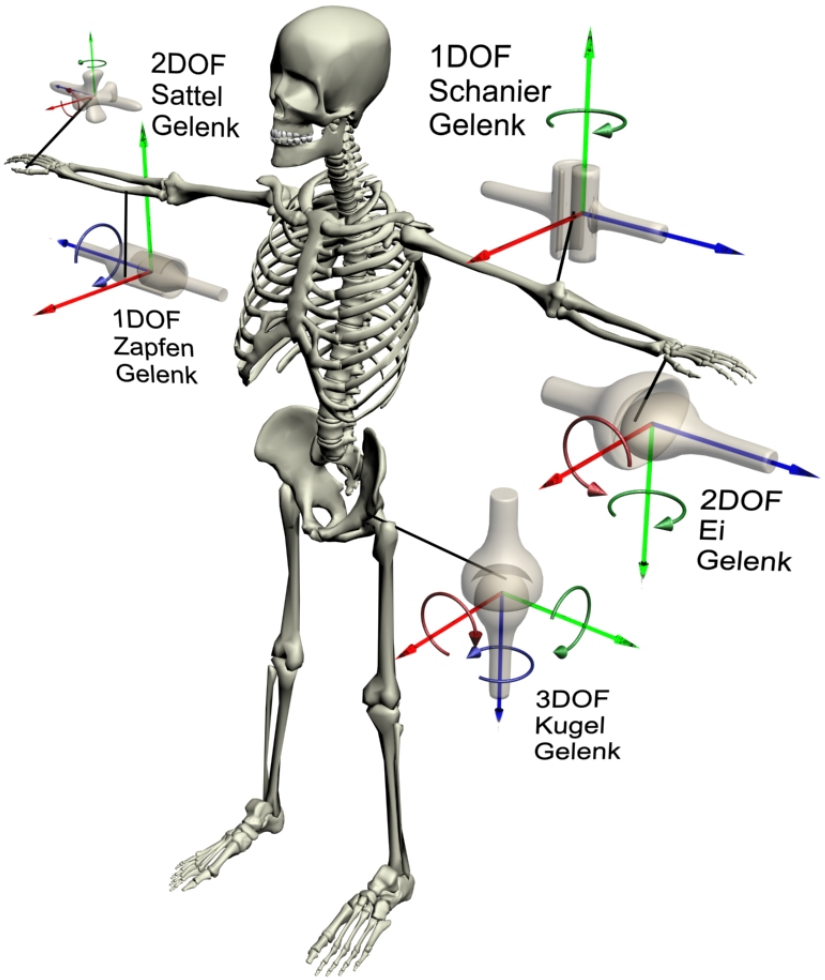

Abbildung 2 Positionen der Gelenke im Skelett

Die sogenannten Euler- und Roll-Pitch-Yaw-Winkel können sowohl zur Darstellung einer Orientierung im dreidimensionalen Raum verwendet werden als auch zur Beschreibung eines Gelenks mit 3 DOFs [22]. Die Eulerwinkeldarstellung kann folgendermaßen aus drei Rotationsmatrizen aufgebaut werden. Dabei bezeichnet $R_{z,\Phi}$ die Rotation um den Winkel Φ um die z-Achse, entsprechend für die anderen Rotationen:

$$R_{Euler} = R_{z,\Phi} \times R_{y,\Theta} \times R_{z,\Psi} =$$

$$\begin{pmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos(\Theta) & 0 & \sin(\Theta) \\ 0 & 1 & 0 \\ -\sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \times \begin{pmatrix} \cos(\Psi) & -\sin(\Psi) & 0 \\ \sin(\Psi) & \cos(\Psi) & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \mathrm{c}(\Phi)\,\mathrm{c}(\Theta)\,\mathrm{c}(\Psi) - \mathrm{s}(\Phi)\,\mathrm{s}(\Psi) & -\mathrm{c}(\Phi)\,\mathrm{c}(\Psi) - \mathrm{c}(\Phi)\,\mathrm{c}(\Theta)\,\mathrm{s}(\Psi) & \mathrm{c}(\Phi)\,\mathrm{s}(\Theta) \\ \mathrm{s}(\Phi)\,\mathrm{c}(\Theta)\,\mathrm{c}(\Psi) - \mathrm{c}(\Phi)\,\mathrm{s}(\Psi) & \mathrm{c}(\Phi)\,\mathrm{c}(\Psi) - \mathrm{s}(\Phi)\,\mathrm{c}(\Theta)\,\mathrm{s}(\Psi) & \mathrm{s}(\Phi)\,\mathrm{s}(\Theta) \\ -\mathrm{c}(\Theta)\,\mathrm{c}(\Psi) & \mathrm{s}(\Theta)\,\mathrm{s}(\Psi) & \mathrm{c}(\Theta) \end{pmatrix} \quad \textbf{(1)}$$

Die Roll-Pitch-Yaw Winkeldarstellung wird durch drei Rotationsmatrizen aufgebaut. Der Unterschied zu der Eulerwinkeldarstellung liegt in der Achse der letzten Rotationsmatrix, bei der in der letzten Rotation anstelle um die z-Achse um die x-Achse rotiert wird.

$$R_{RollPitchYaw} = R_{z,\Phi} \times R_{y,\Theta} \times R_{x,\Psi} =$$

$$\begin{pmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos(\Theta) & 0 & \sin(\Theta) \\ 0 & 1 & 0 \\ -\sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Psi) & -\sin(\Psi) \\ 0 & \sin(\Psi)\ 0 & \cos(\Psi) \end{pmatrix} =$$

$$\begin{pmatrix} c(\Phi)\,c(\Theta) & -s(\Phi)\,c(\Psi) + c(\Phi)\,s(\Theta)\,s(\Psi) & s(\Phi)\,c(\Psi) + c(\Phi)\,s(\Theta)\,s(\Psi) \\ s(\Phi)\,c(\Theta) & c(\Phi)\,c(\Psi) + s(\Phi)\,c(\Theta)\,s(\Psi) & -c(\Phi)\,s(\Psi) + s(\Phi)\,s(\Theta)\,c(\Psi) \\ -s(\Theta) & c(\Theta)\,s(\Psi) & c(\Theta)\,c(\Psi) \end{pmatrix} \quad \textbf{(2)}$$

Bei Euler-Winkeln wird als Letztes um die z-Achse rotiert und bei Roll-Pitch-Yaw-Winkeln wird als Letztes um die x-Achse rotiert (siehe dazu die **Abbildung 3**). Die mathematische Beschreibung eines Gelenkes mit 3 DOFs kann durch die Multiplikation von drei Rotationsmatrizen mit den Variablen Φ, Θ und Ψ als Winkel beschrieben werden. Je nachdem, welcher Winkel für die einzelnen Variablen eingesetzt wird, kann durch Ausrechnung und die Multiplikation der einzelnen Rotationsmatrizen die entsprechende Endposition des Gelenkes bestimmt werden. Das gesamte Skelett besteht nicht nur aus Gelenken, sondern auch aus Verbindungen zwischen diesen Gelenken. Diese Verbindungen, die in Englisch „links“ genannt werden und die einem Gelenk zugeordnet werden können, führen in der mathematischen Beschreibung dazu, dass die Rotationsmatrizen zusätzlich einen Verschiebungsanteil hinzugeführt bekommen.

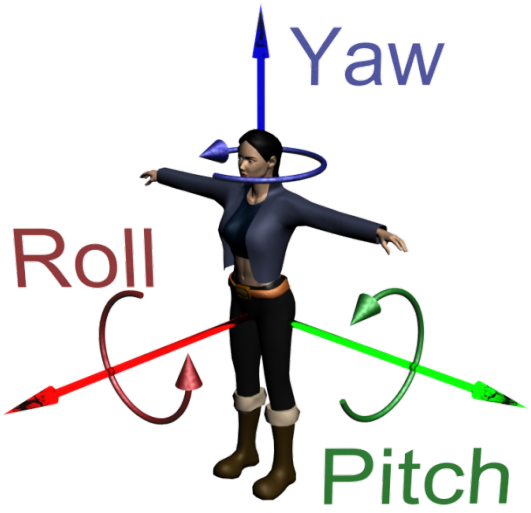

Abbildung 3 Namen der Rotationsachsen (rote Achse ist z, grüne Achse ist y und blaue Achse ist x)

2.1.3 Die Denavit-Hartenberg-Konvention

Die Denavit-Hartenberg-Konvention ist ein Verfahren aus der Robotik, das verwendet wird, um die Vorwärtskinematik eines Roboters mathematisch zu beschreiben. Bei diesem Verfahren wird eine Folge von homogenen Transformationen verwendet, um eine Transformations-

matrix zwischen zwei in einer Hierarchie aufeinander folgenden Koordinatensystemen zu bestimmen. Diese Transformation entspricht dem Link (der Verbindung) zwischen zwei Gelenken. Die einzelnen Transformationen der Links können verwendet werden, um die gesamte Transformation durch Multiplikationen bis zu dem letzten Gelenk auszurechnen. Dies ist wichtig für die Modellierung und mathematische Beschreibung eines Skelettes in einer hierarchischen Struktur aus mehreren Gelenken und entspricht der Anwendung der Kinematik zur Berechnung jedes Gelenkeinflusses auf die folgenden Gelenke. In der Robotik wird aus der gesamten Transformation der Gelenke auf die Lage und Ausrichtung des Endeffektor[4] Vektors geschlossen[5] [22]. Mit diesem Verfahren kann das gesamte menschliche Skelett in jeder möglichen Pose mathematisch beschrieben werden.

2.1.4 Vorgehensweise zur mathematischen Beschreibung eines Skelettes

Zunächst werden nur zwei einzelne Gelenke betrachtet, um einen Teil der gesamten Transformation zu bestimmen. Die Transformation, die den Übergang von einem Koordinatensystem in ein anderes beschreibt bzw. auch den Link i zwischen dem Gelenk $i-1$ und dem Gelenk i beschreibt, kann aus den vier homogenen Transformationen zusammengesetzt werden [23]:

1. Eine Rotation um den Winkel θ_i bezogen auf die z_{i-1}-Achse.

2. Eine Verschiebung um d_i entlang der z_{i-1}-Achse.

3. Eine Verschiebung um a_i entlang der x_i-Achse.

4. Eine Rotation um den Winkel α_i bezogen auf die x_i-Achse.

In der folgenden **Abbildung 4** wird diese Abfolge von Transformationen dargestellt, wie diese sich zwischen zwei Gelenken zusammensetzen. Dazu sind in der Abbildung zwei windschiefe schwarze Geraden (diese stellen die jeweilige z-Achse der Gelenke dar) dargestellt, für die eine Transformation gefunden werden muss, um das vorherige Koordinatensystem KS_{i-1} in das folgende Koordinatensystem KS_i zu überführen. Die eigentliche Ausrichtung der Koordinatensysteme steht zunächst noch nicht fest, nur die Ausrichtung der z_{i-1}-Achse, die den zwei windschiefen schwarz dargestellten Geraden entspricht. Zunächst muss die x_{i-1} in die x_i um die z_{i-1} gedreht werden. Die Ausrichtung der x_i-Achse ist durch die Tatsache gegeben, dass sie auf den beiden Achsen z_{i-1} und z_i, an dem Punkt der kleinsten Distanz zwischen ihnen, senkrecht steht. Damit kann der Parameter θ_i ermittelt werden. Anschließend wird entlang der z_{i-1}-Achse vom Ursprung θ_{i-1} des $i-1$-ten Koordinatensystems die Distanz d_i zu dem Schnittpunkt der x_i und der z_{i-1}-Achse ermittelt. Daraufhin wird von der z_{i-1}-Achse bis zur z_i-Achse entlang der x_i-Achse die Distanz α_i festgelegt. Zum Schluss

[4] Werkzeug an der Spitze des Roboters wie z. B. ein Bohrer.

[5] Oft ist aber auch das Gegenteil von Interesse, da man die Lage des Endeffektors vorgegeben hat und wissen will, wie die Gelenkstellungen der kinematischen Kette aussehen müssen, um eine spezifische Position und Orientierung im Raum zu erreichen. Dieses wird inverse Kinematik genannt.

wird die z_{i-1}-Achse in die z_i-Achse um x_i rotiert. Dies wird durch den Parameter α_i dargestellt. Die Matrixmultiplikation dieser vier Transformationen ergibt die gesamte Transformation A_i

$$A_i = R_{z_{i-1},\theta} \times T_{z_{i-1},d_i} \times R_{x_i,a_i} \times R_{x_i,\alpha} \quad \textbf{(3)}$$

des i-ten Links. Die Folge von Transformationen der Gelenke, ausgehend von dem Wurzelgelenk (Gelenk, das durch kein anderes Gelenk beeinflusst wird) bis zum Blattgelenk

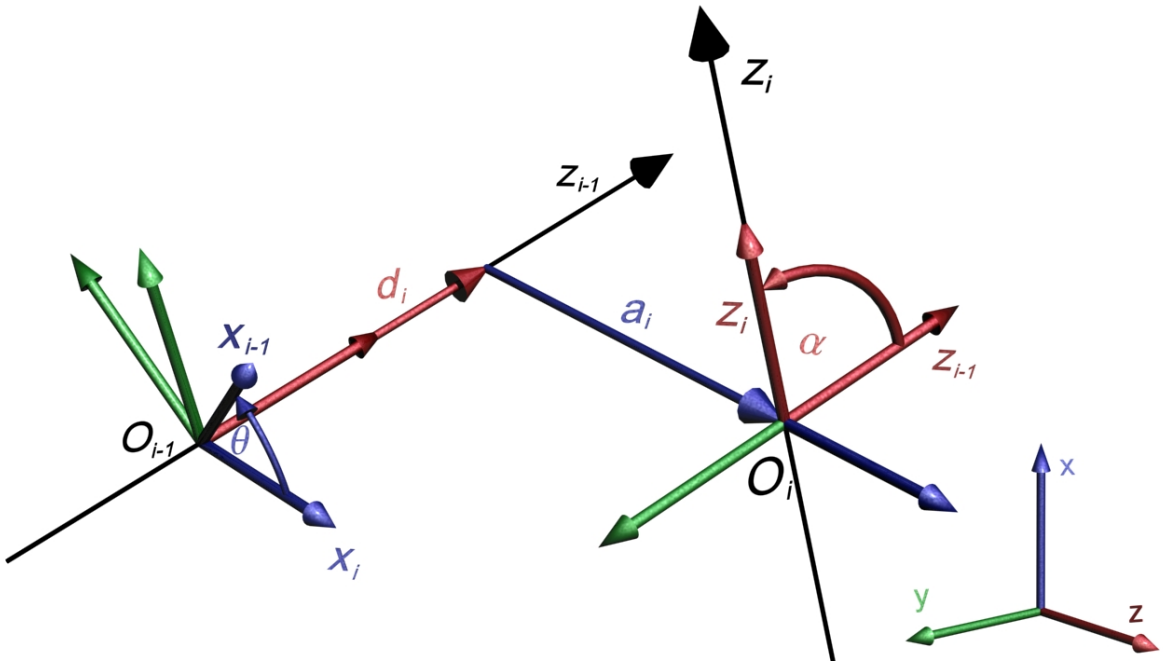

Abbildung 4 DH-Transformationen zwischen zwei windschiefen Geraden

(dasjenige Gelenk, das keine weiteren Gelenke beeinflusst) in der Gelenkhierarchie wird als kinematische Kette bezeichnet. Nach dem letzten Gelenk folgt der Endeffektor, kurz EE. Dabei kann ein Manipulator[6] auch mehrere EE haben und damit auch mehrere kinematische Ketten beinhalten. Um diese Gesamttransformationen des Manipulators auszurechnen, müssen alle Gelenk-Transformationen A_i von der Wurzel ausgehend bis zu dem jeweiligen EE ausmultipliziert werden. Es sind nicht immer nur die Transformationen bis zum EE gefragt, manchmal ist auch wichtig, in welcher Lage sich ein Gelenk innerhalb einer kinematischen Kette befindet. Bei der folgenden Formel steht n für die Tiefe in einer Hierarchie, bis zu der die Transformationen ausgerechnet werden soll [22].

$$T_0^n = A_0 \times A_1 \times \ldots \times A_n \;; n \in N \quad \textbf{(4)}$$

Die Gesamttransformation T_0^n beschreibt die Lage und Orientierung des Endeffektors. Als Beispiel könnte man sich einen Manipulator vorstellen, der einer Hand nachempfunden wurde und aus 23 Gelenken besteht (3 Gelenke bestimmen die Ausrichtung der gesamten Hand und jeweils 4 Gelenke sind nötig, um die Bewegung eines Fingers nachzubilden). In diesem Falle

[6] Im Falle dieser Arbeit wäre ein Manipulator die mathematische Beschreibung eines Menschen.

ist die Tiefe in der Hierarchie 7, und es gibt insgesamt 5 kinematische Ketten, die die Transformation zu den Fingerspitzen darstellen.

2.2 Charakter-Animations-Techniken

Um Bewegung analysierbar zu machen, ist eine geeignete Darstellungsform zu finden. Alternativ kann Bewegung ohne Visualisierung analysiert werden, basierend auf den kinematischen Bewegungsdaten. Im Verlauf dieser Arbeit wird eine Analysemethode basierend auf Verfahren und Konzepten aus der Charakteranimation vorgestellt. Dazu betrachten wir zunächst verschiedene Charakter-Animations-Techniken. Diese sind Key-Frame-Animation, Algorithmische Animation und die direkte Darstellung der Motion-Capturing-Daten. Diese verschiedenen Techniken finden alle Anwendung in dem Tool PAMOCAT.

2.2.1 Key-Frame-Animation

Zur Darstellung eines virtuellen und künstlichen Charakters wird eine sogenannte Key-Frame-Animation (Hunger 1974) oder zu Deutsch Schlüsselbilddarstellung verwendet. Die Darstellung einer Bewegung wird durch gezielte Veränderungen von Gelenken jeweils bezüglich Startzeitpunkt und Endzeitpunkt für die zu animierende Figur per Hand definiert. Ein Zeitpunkt entspricht einem Frame, in einer Sekunde können z. B. 60 Frames dargestellt werden. Je mehr Frames verwendet werden, desto flüssiger kann die Animation dargestellt werden[7]. Wird eine Bewegung durch mehrere Zeitpunkte bezüglich aller Gelenke im Körper animiert, sieht eine Bewegung natürlicher aus. Der Startzeitpunkt und der Endzeitpunkt bilden zusammen mit der Änderung zwischen diesen einen sogenannten Key-Frame (oder Schlüssel-Zeitpunkt), dieses Key-Frame bezieht sich auch auf einzelne Gelenke oder Körperteile. Zwischen diesen Key-Frames können die Gelenkwinkel interpoliert werden, wie in **Abbildung 5** gezeigt wird. Dieses wird als Key-Frame-Animation bezeichnet [24]. Dabei ist nicht zwingend die gesamte Änderung zu einem Zeitpunkt durchzuführen, es können auch einzelne Gelenke an verschiedenen Zeitpunkten geändert werden. Um eine möglichst realistische Darstellung der Bewegung zu erzeugen, müssen möglichst alle einzelnen Gelenke (oder auch noch elementarer jeder einzelne DOF) zu vielen unterschiedlichen Zeitpunkten geändert werden. Um darzustellen, wie aufwendig die Erstellung einer möglichst realistischen Key-Animation sein kann, ist es nötig, zu wissen, wie viele Gelenke manuell animiert werden können. Das gesamte Skelett des Menschen kann durch 104 DOFs[8] dargestellt werden, denn gemäß dem Beispiel aus Unterkapitel 2.1.4 wird eine einzelne Hand durch 24 DOFs dargestellt, und entsprechend gilt für die übrigen beweglichen Skelettelemente:

$$2 \times Hand + 2 \times Arm + Körperorientierung + Kopf +$$

$$2 \times Beine + 2 \times Fuß =$$

[7] Das menschliche Auge kann allerdings nur durchschnittlich 25 Frames pro Sekunde wahrnehmen.

[8] Abhängig von den gewünschten Freiheiten der Animation.

$$2 \times 21 + 2 \times 7 + 3 + 3 + 2 \times 6 + 2 \times 15 = 104 \quad \textbf{(5)}$$

Würde man die Wirbelsäule mit allen ihren Freiheitsgraden mit berücksichtigen und dazu noch die Muskeln des Gesichtes als DOF ansehen, wäre diese Zahl noch erheblich höher. Daraus ist ersichtlich, dass dies ein komplizierter und aufwendiger Vorgang ist, der je nach gewünschtem Natürlichkeitsgrad einen entsprechend hohen Zeitaufwand erfordert. Erfahrene Personen nehmen allerdings solche vereinfachten Key-Frame-Animationen immer noch als unnatürlich war. In **Abbildung 5** wird zwischen zwei Schlüssel-Positionen interpoliert, dieses ist verteilt auf vier Zeitpunkte. Die interpolierten Posen sind leicht durchsichtig dargestellt. Dabei finden eine seitliche Bewegung des Kopfes von rechts nach links und des linken Armes von der Körpermitte nach links außen statt[9]. Um solche Animationen noch echter wirken zu lassen, können nichtlineare Interpolationstechniken (höhergradige Interpolationen) verwendet werden. Bei diesen wirken die Beschleunigungs- und Abbremsphasen realistischer, da es keine eckigen Übergänge in der Geschwindigkeit einer Bewegung gibt.

Abbildung 5 Interpolation zwischen zwei Key Frames

Diese kurze Key-Animation bestehend aus 6 einzelnen Zeitpunkten in der **Abbildung 5** wird wahrscheinlich von den meisten Menschen als nicht natürliche Bewegung wahrgenommen. Dies liegt hier an der Anzahl der verwendeten Key-Intervalle und damit der beteiligten Gelenke. Um eine natürlicher wirkende Bewegung zu erstellen, müssen viele verschiedene Start- und Endzeitpunkte für die verschiedenen DOFs ausgewählt werden, mit denen die Gelenkänderungen durchgeführt werden sollen. Dabei sind die einzelnen Start- und Endzeitpunkte voneinander unabhängig. Um einen Überblick zu erhalten, welche möglichen Kombinationen maximal zur Erstellung verfügbar wären, kann die Anzahl der DOFs mit einer Zeitspanne multipliziert werden:

[9] Eigentlich sieht man auch eine Bewegung des Charakters von links nach rechts, dieser ist aber nicht im Fokus und nur der Darstellung halber enthalten, da sonst die einzelnen Posen übereinander lägen und nicht mehr unterscheidbar wären.

$$\frac{DOF\ eines\ Skeletts \times Zeit}{MinimaleAnzahlAnFrame} = \frac{104 \times 60}{2} = 3120 \quad \textbf{(6)}$$

Das Resultat von 3120 ist die Anzahl an maximal möglichen Key-Frames für alle DOFs bei einer Sekunde, wenn eine Framerate von 60 Frames verwendet wird. Die Teilung durch 2 ergibt sich aus der Tatsache, dass ein Key-Frame einen unterschiedlichen Anfangs- und End-Frame hat. Dieses ist ein unrealistischer Wert, der aber die maximal mögliche Anzahl darstellt, die beim Motion-Capturing verfügbar ist. Um Bewegungen möglichst real aussehen zu lassen, muss in der Animation jedes der einzelnen Gelenke zu unterschiedlichen Zeitpunkten angepasst werden; dieses kostet viel Zeit und damit auch viel Geld.

2.2.2 Algorithmische Animationen

Die algorithmische Animation [25] wurden entwickelt, um schnellere und günstigere Animationen zu erstellen. Regelmäßige Bewegungen wie das Schwingen der Flügel eines Schmetterlings kann automatisch als eine Sinusschwingung vereinfacht animiert werden. Es können aber auch physikalische Gesetze die Grundlage für eine Animation sein, zum Beispiel das Gravitationsgesetz bei einem Partikelsystem für die Darstellung eines Springbrunnens. Es können verschiedene Bewegungen durch Algorithmen animiert werden, wenn diese sich mit mathematischen Funktionen oder Gesetzmäßigkeiten beschreiben lassen. Trotzdem müssen die entsprechenden Algorithmen entwickelt werden, was bedeutet, dass erst nach der Entwicklung viel manuelle Animationszeit eingespart werden kann. Eine weitere sehr praktische Einsatzmöglichkeit ist die Beschreibung von zielgerichteter Bewegung, wie es etwa der Fall beim Greifen einer Hand ist, die sich dabei entlang einer Trajektorie bewegt. Dazu wird zwischen den einzelnen Positionen von der Start- bis zur Endposition der Trajektorie interpoliert und mittels einer inversen Kinematik wird für jeden Zeitpunkt die Gelenkstellung der Manipulators (z. B. Arm) ausgerechnet. Dieses mit Key-Frame-Animationen zu realisieren, würde sehr viel Zeit in Anspruch nehmen. Der Grund dafür ist, dass immer wieder Bewegungen von anderen Gelenken zu einem nicht sofort ersichtlichen Teil kompensiert werden müssen. Dabei ist die Schwierigkeit, die gesamte Vorwärtsbewegung als eine flüssige und natürliche Bewegung aussehen zu lassen. Auf diese Weise lassen sich viele Animationen leicht und kostengünstig realisieren, allerdings nicht alle Arten von Animationen wie die der komplexeren Bewegungen virtueller Menschen. Eingesetzt werden solche Animationen z. B. bei virtuellen Menschen [26], bei denen nicht die gesamte Bewegung vordefiniert werden kann. Virtuelle Menschen müssen sich auf eine flexible Art und Weise bewegen können, die nicht vordefiniert werden kann. Zum Beispiel müssen sie aus einer beliebigen Körperhaltung auf ein beliebiges Objekt zeigen können.

2.2.3 Motion-Capturing

Der Begriff Motion-Capture (zu Deutsch Bewegungserfassung) bezeichnet eine Technik, die es ermöglicht, Bewegungen (meist von Menschen) aufzuzeichnen und in einem computerleserlichen Format zu speichern. Dazu werden die Positionen und die Ausrichtung der Gelenke

des Skeletts in 3D erfasst (mehr hierzu ist in Kapitel 4 zu finden). Diese Bewegungen können dann verwendet werden, um sie auf 3D-Modelle eines virtuellen Charakters oder auch auf einen Roboter zu übertragen. Eine neu aufkommende Verwendung des Motion-Capturings ist es, dieses für Analysen in der Verhaltensforschung einzusetzen. Am häufigsten wird Motion-Capturing in der Filmindustrie und Computerspielindustrie eingesetzt. Weitere Anwendungsgebiete von Motion-Capturing sind in der medizinischen Analyse des Ganges durch die Orthopäden, im Bereich des Sportes zur Leistungssteigerung und auch in der Strafverfolgung zur Rekonstruktion von Handlungsabläufen [27] zu finden. Allgemein ermöglicht das Motion-Capturing, schnell natürlich aussehende Bewegungen festzuhalten, die von einem Schauspieler, Artisten oder Stuntman (bzw. auch Patienten) ausgeführt werden. Die dargestellte Bewegung muss nicht immer die eines Menschen sein, z. B. könnten die Bewegungen eines Menschen für die Animation eines Frosches oder Vogels verwendet werden, wobei tierische Bewegungen vom Menschen gespielt werden. Es werden aber auch Bewegungen von Tieren verwendet, um diese zu analysieren. Ein bekanntes Beispiel hierfür sind die Aufnahmen eines galoppierenden Pferdes[10]. Eadweard Muybridge [28] hat 1872 dabei mit Hilfe von 12, 24 und 36 sukzessiv auslösenden Fotoapparaten eine Serie von Fotos eines galoppierenden Pferdes erstellt, um jeweils die exakte Beinstellung zu ermitteln. Durch diese Aufnahmen hatte er einen sichtbaren Beweis, dass beim Galoppieren zeitweise alle vier Hufe des Pferdes in der Luft sind. Dies ist als Vorläufer des heutigen Motion-Capturings zu sehen. In der folgenden **Abbildung 6** sind die so ermittelten Beinstellungen dargestellt.

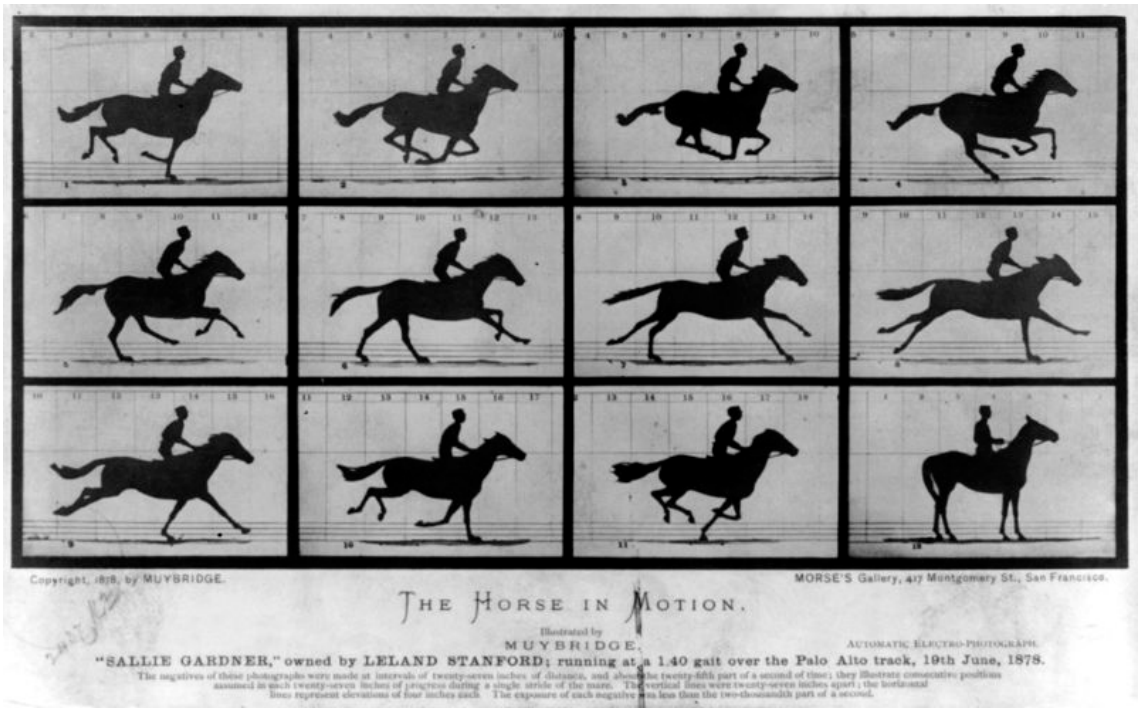

Abbildung 6 Ein galoppierendes Pferd, aufgenommen [28]

[10] In der Biologie werden Bewegungen von Tieren mit Hilfe von Motion-Capturing untersucht.

Eine Erweiterung der Motion-Capture-Technik wird als Performance-Capture bezeichnet, bei der zusätzlich noch die Bewegungen des Gesichts mit aufgezeichnet werden. Wird nur die Bewegung des Gesichts aufgenommen, bezeichnet man dieses als Facial-Motion-Capture. Ein Gesicht mit den Grundpositionen der Marker, die meistens verwendet werden, um verschiedene Geschichtsausdrücke aufzuzeichnen, ist in der **Abbildung 7** dargestellt. Das direkte Übertragen der Motion-Capture-Daten auf einen virtuellen Charakter wird als Performance-Animation bezeichnet [29].

Der Hauptnachteil des Motion-Capturings sind die sehr hohen Kosten für die erforderlichen Aufnahmesysteme, sodass sich meistens nur große Entwicklerstudios solche Systeme leisten können. Es kommen allerdings auch immer mehr günstige Alternativen auf den Markt, bei denen aber noch nicht die gewünschte Genauigkeit vorhanden ist. Die Vor- und Nachteile der verschiedenen Systeme werden genauer im folgenden Abschnitt 2.3 vorgestellt.

2.3 Motion-Capture-Systeme

In der Industrie bei Film- und Spielproduktionen ersparen Motion-Capture-Systeme Animationszeit, um natürliche Bewegungen zu erhalten. Aus der Sicht von Verhaltensforschern ist die Analysemöglichkeit von Bedeutung, die mit der automatischen Berechnung auf der Grundlage der kinematischen Daten bzw. der Motion-Capture-Daten erstellt werden kann. Die automatisch berechneten Annotationen haben den Vorteil, dass sie eine immer gleich bleibende Qualität liefern, welche auf immer denselben Rahmenbedingungen basiert und nicht gegebenenfalls von einer annotierenden Person und deren Stimmung beeinflusst wird. Um solche Motion-Capture-Aufnahmen zu erstellen, benötigt man entsprechende Systeme, die auf verschiedenen Techniken basieren und von verschiedenen Firmen angeboten werden. Im Folgenden werden diese Motion-Capture-Systeme vorgestellt. Viele dieser Systeme basieren auf optischen Kameras, es gibt aber auch magnetisch-, mechanisch- und Schall-basierte Systeme.

Abbildung 7 Gesicht mit Grundpositionen der Marker, wie es beim Performance-Capturing [30] oder Facial-Motion-Capturing verwendet wird

Diese einzelnen Techniken eignen sich für verschiedene Anwendungen jeweils individuell gut. Im Folgenden werden die unterschiedlichen Verfahren und Techniken erläutert und deren Stärken und Schwächen diskutiert.

2.3.1 Optische Trackingsysteme

Es gibt viele verschiedene optische Trackingsysteme, die auf Kameras zur Aufnahme der Bewegung basieren. Diese optischen Trackingsysteme unterscheiden sich bei der Verwendung von Markern. Es gibt passive, aktive und markerlos basierte Trackingsysteme.

2.3.1.1 Passive Marker

Passive Marker sind meistens rund und mit einer speziellen reflektierenden Beschichtung versehen, die gut infrarotes Licht reflektiert[11]. Die Beschichtung der Marker besteht aus vielen kleinen Kugeln, die das Licht in die Richtung der Lichtquelle reflektieren. Bei den Systemen, die passive Marker verwenden, wird die Infrarotkamera mit Infrarotstrahlern kombiniert, um die Marker optimal auszuleuchten. Die normalen passiven Marker sind nicht eindeutig zu identifizieren. Eine Erweiterung dieser Marker sind Rigidbodys, die aus mindestens vier einzelnen dieser Marker bestehen und in einer eindeutigen räumlichen Anordnung zueinander stehen, wodurch diese durch das speziell auf die Rigidbodys abgestimmte Trackingsystem (z. B. der Firma ART, Vicon oder OptiTrack) eine eindeutige Orientierung und Identifizierung zulassen (siehe **Abbildung 8**). Optische Trackingsysteme bestehen aus mindestens zwei Kameras[12], die im Raum verteilt und auf die Aufnahmefläche ausgerichtet sind. Die Kameras werden relativ zueinander kalibriert, wodurch dem System die Position und Orientierung aller Kameras zueinander bekannt gemacht wird. Zur Laufzeit kann das System aus den jeweiligen 2D Kamerakoordinaten mit Hilfe von Schnittpunkten der optischen Strahlen in 3D die Positionen der einzelnen Marker errechnen. Daraus können einzelne Markergruppen als Rigidbodys identifiziert und die Position und Orientierung können errechnet werden (siehe dazu die **Abbildung 9**).

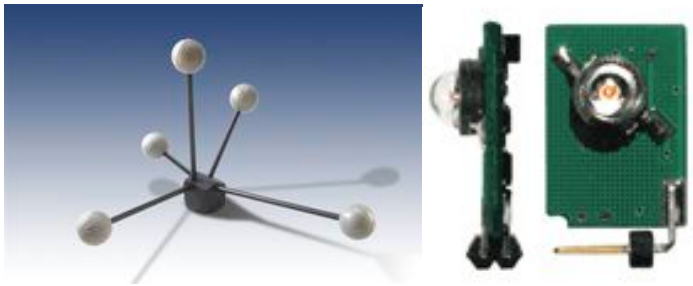

Abbildung 8 Rigidbody, bestehend aus einzelnen passiven Markern [31] und einem aktiven Marker von zwei Seiten [32]

[11] Es gibt aber auch Systeme, die mit verschiedenen Farben arbeiten, Systeme mit Infrarot sind generell robuster, da sie lichtunabhängig arbeiten.

[12] Motion-Capturing-Systeme für das Erfassen von ganzen Menschen sollten 8 oder mehr Kameras haben.

2.3.1.2 Aktive Marker

Aktive Marker sind selbstleuchtend und bestehen aus Infrarot-LEDs, wodurch die Marker allerdings etwas eingeschränkt sind, da sie mit Energie versorgt werden müssen. Daher muss der Proband, dessen Bewegung aufgezeichnet wird, eine zentrale Energieversorgung tragen, die mit den einzelnen Markern verbunden ist. Die einzelnen LEDs haben einen gewissen Abstrahlwinkel, wodurch sie nicht optimal in alle Richtungen leuchten. Ein sehr positiver Aspekt bei aktiven Markern ist, dass bei einigen dieser Systeme die Marker eindeutig zu identifizieren sind, da jeder in seiner eigenen Frequenz

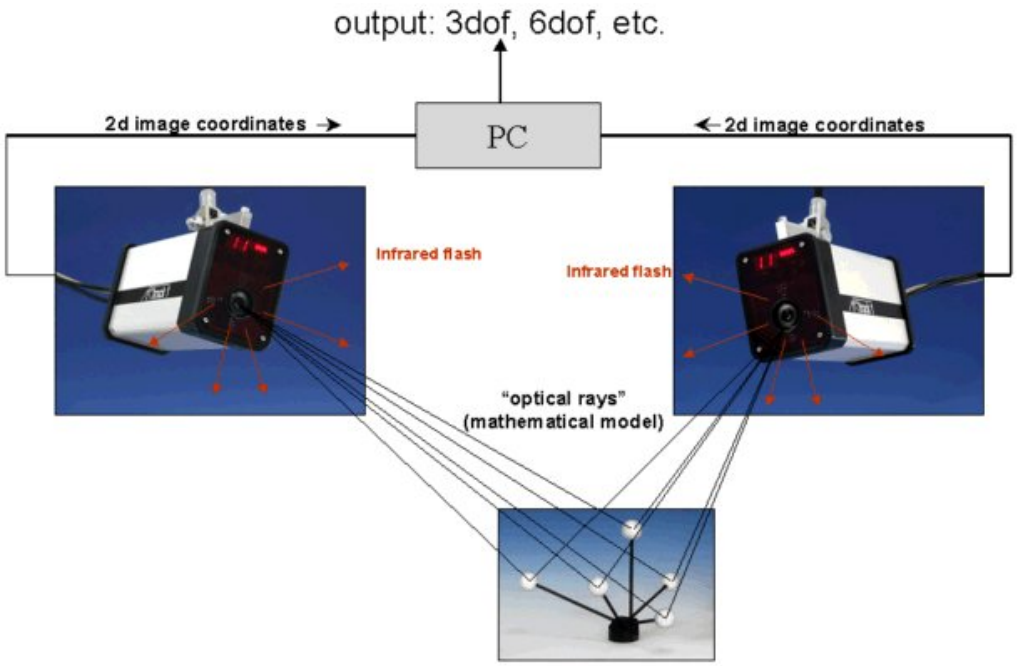

Abbildung 9 Optische Trackingsysteme [31]

aufleuchtet[13]. Leider gibt es auch Nachteile durch diese eindeutige Identifizierung der einzelnen Marker, da eine Limitierung der Anzahl der zu verwendenden Marker durch das Frequenzspektrum im infraroten Wellenlängenbereich berücksichtigt werden muss. Damit ist es nicht möglich, mehrere Personen gleichzeitig aufzunehmen. In der **Abbildung 8** sind die beiden Sorten von Markern abgebildet. Die jeweiligen Marker werden am Körper verteilt angebracht. Das geschieht so, dass die Marker möglichst wenig verdeckt werden können und dass aus den Positionen der Marker auf die Körperhaltung des Akteurs geschlossen werden kann.

2.3.1.3 Markerloses Tracking

Eine Alternative zu Systemen, bei denen der Akteur nicht erst mit Markern ausgerüstet werden muss, sind Systeme, bei denen die Bewegung direkt aus den Videobildern berechnet wird. Solche Systeme gibt es von der Firma Polhemus, Organic Motion oder Vicon. Allerdings ist

[13] Nicht alle aktiven Marker arbeiten mit verschiedenen Frequenzen und lassen sich daher nicht immer eindeutig unterscheiden.

die Genauigkeit deutlich schlechter als bei Systemen, die mit Markern arbeiten, und sie sind empfindlicher gegenüber Lichtänderungen. Diese Systeme arbeiten normalerweise mit einem speziellen farbigen Hintergrund, und die Personen müssen entsprechend andersfarbig gekleidet sein. Bei solchen Systemen wird, um gute Aufnahmeergebnisse zu erzielen, nur mit einer Person im andersfarbigen Aufnahmebereich gearbeitet, um an die Qualität der anderen Systeme heranzukommen. Zudem ist die Genauigkeit bei der Ausrichtung von Endgliedmaßen wie den Händen und dem Kopf nicht sehr genau.

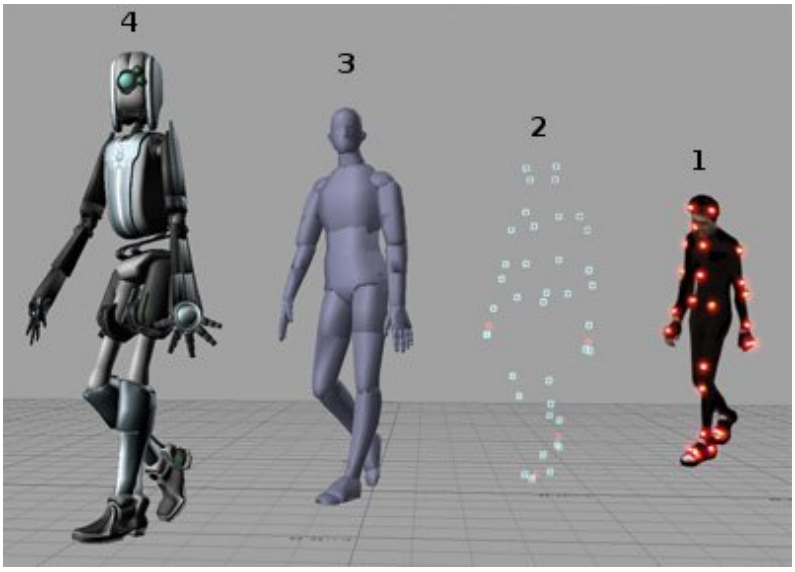

Abbildung 10 Bewegungserfassung mit Markern [30].

2.3.2 Magnetische Tracking-Systeme

Magnetische Motion-Capture-Systeme arbeiten nach dem Induktionsprinzip, d. h. fließt ein Strom durch eine Spule, baut sich in ihr ein Magnetfeld auf. Wenn ein Leiter, z. B. ein Metalldraht, im Magnetfeld einer Spule bewegt wird, wird in ihm ein Strom induziert. Dadurch erzeugt ein sich zeitlich änderndes Magnetfeld in einer Spule einen Stromfluss, der abhängig von der Orientierung zum Magnetfeld verschieden stark ist. Magnetische Tracking-Systeme bestehen aus drei Transmittern, einem Steuerrechner und mehreren Sensoren. Die Transmitter sind fest im Raum installiert. Sie stellen drei einzelne orthogonal ausgerichtete Spulen dar, die jeweils zeitversetzt ein Magnetfeld aufbauen. Die gesamte Folge bildet einen Zeitschritt in dem System, welches durch das Steuergerät kontrolliert wird, dargestellt in der **Abbildung 11**. Die Sensoren bestehen ebenfalls jeweils aus drei orthogonal zueinander ausgerichteten Spulen, in der **Abbildung 11** durch eine Box mit drei Spulen dargestellt. Bei diesen Spulen werden die durch die Transmitter induzierten Ströme gemessen. Aus den drei Messwerten eines Sensors lassen sich Position und Orientierung der Sensoren ermitteln. Dies basiert darauf, dass aus der Stärke des induzierten Stromes auf die Entfernung geschlossen werden kann und ebenso aus der Form des Magnetfeldes, welches eine Torus ähnliche Form besitzt. Diese Sensoren werden in kompletten Anzügen eingesetzt, die wie in der **Abbildung 12** dargestellt aussehen. Nachteilig ist die große Störanfälligkeit der magnetischen Felder. Diese werden

durch Metall, aber auch durch andere elektrische Geräte beeinflusst, sodass ein solches System nicht in jedem Gebäude ohne Weiteres aufgestellt werden kann. Außerdem sind die Genauigkeit und der Sensorbereich im Vergleich zu optischen oder mechanischen Trackingsystemen relativ klein. Hinzu kommt noch, dass die exakte Stellung des Skelettes nicht genau bekannt ist, sondern nur die Positionen der einzelnen Marker in Bezug auf ihre vorherige Position zu einem früheren Zeitpunkt. Daher können auch Personen nicht in Relation zueinander im Detail erkannt werden, bei denen die Kopforientierung als Indiz für die Blickrichtung von Interesse wäre.

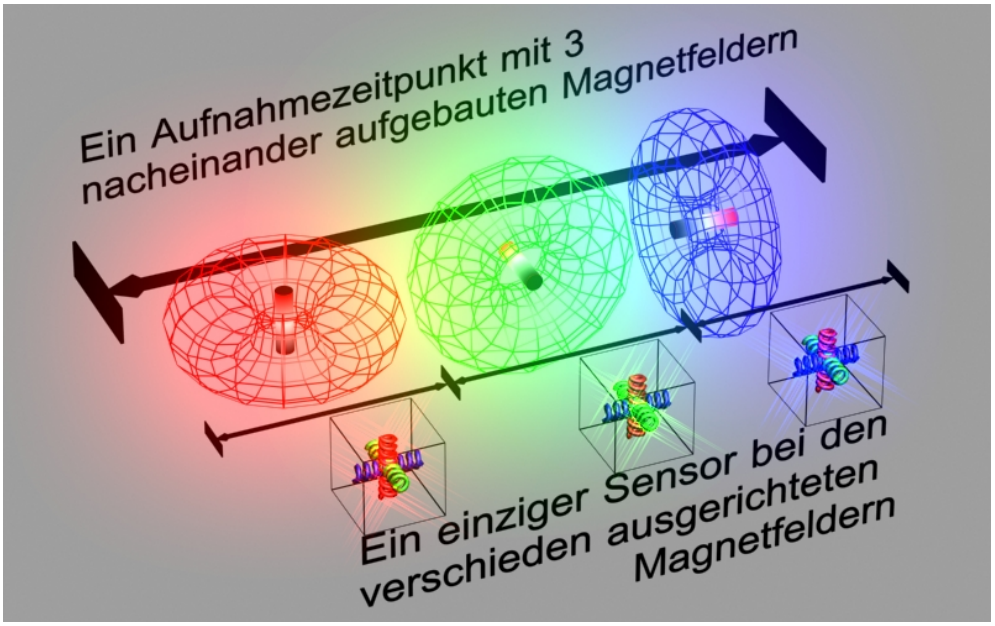

Abbildung 11 Drei zeitversetzte Magnetfelder, die hintereinander erzeugt werden, und ein Sensor mit drei Spulen, in denen jeweils ein Stromfluss induziert wird

Abbildung 12 Magnetische Tracking-Anzüge [33], [34] und [35]

2.3.3 Schall- und Trägheitssensor basierte Tracking-Systeme

Das Tracking-System von Vlasic und Adelsberg [36] arbeitet mit einer Kombination von Ultraschall- und Trägheitssensoren zusammen. Der Ultraschall wird von mehreren Quellen ausgestrahlt, die am Körper befestigt sind. Dieses geschieht mit einer Wiederholungsrate von 40 kHz. Das erzeugte Signal wird von mehreren Sensoren mit Mikrophonen, die wie in der **Abbildung 13** am Körper befestigt sind, wahrgenommen. Aus der vergangenen Zeit von der Ausstrahlung eines Signals von der Quelle bis zum Empfang durch einen Sensor kann auf die Entfernung geschlossen werden. Um die Position und Orientierung der Sensoren ermitteln zu können, besitzen die Sensoren ein Gyroskop und einen Accelerometer. Ein Gyroskop misst die Änderung der Orientierung und das Accelerometer misst die Beschleunigung. Durch die verschiedenen Sensordaten aus Mikrophonen, Gyroskopen und Accelerometern wird die wahrscheinlichste Position und Orientierung, ausgehend von der letzten Lage, durch Verwendung des Kalman-Filters errechnet. Dieses Verfahren wird in dem Paper [36],„Practical Motion Capture in Everyday Surroundings“ näher beschrieben. Die Besonderheit dieses Systems ist, dass es in jeder Umgebung eingesetzt werden kann. Die Bewegung des Probanden ist allerdings einschränkt durch einen Anzug und einen Rucksack, um die Daten zu messen und aufzuzeichnen. Leider können bei diesem System nicht die Bewegungen mehrerer Personen erfasst werden.

2.3.4 Tiefensensor Tracking-Systeme

Im Jahre 2010 wurde die Kinect von Microsoft auf den Markt gebracht. Es ist eine Tiefenkamera kombiniert mit einer Farbkamera. Um die Tiefe zu berechnen, besitzt die Kamera eine Infrarotprojektionsmaske und eine Infrarotkamera, die die entsprechende Verzerrung der Projektionsmaske ermittelt und daraus die Tiefe eines Gegenstandes berechnen kann. Mit der Bibliothek OpenNI [37] lassen sich die Posen der Menschen auslesen. Es ist auch möglich,

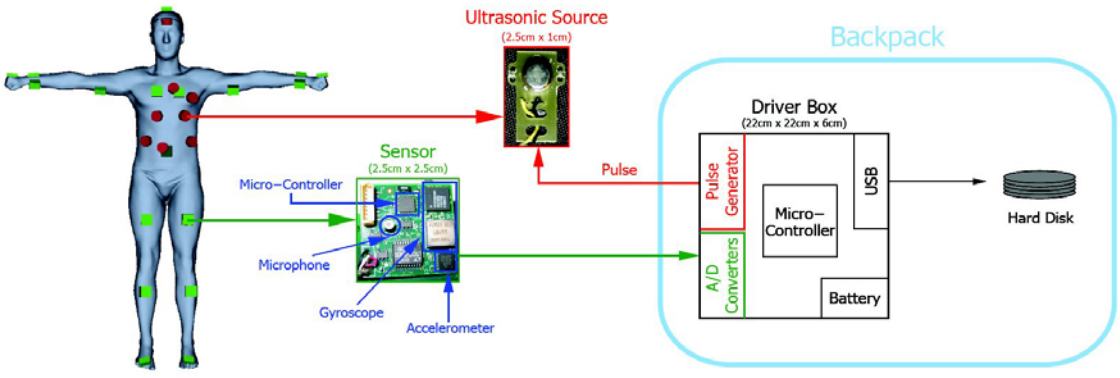

Abbildung 13: Funktionsweise eines auf Ultraschall- und Trägheitssensoren basierenden Motion-Capture-Systems [36]

mehrere Personen aufzuzeichnen, allerdings kann die Orientierung einzelner Körperteile (Ausrichtung des Kopfes oder der Hände) nicht oder nur instabil erfasst werden. Außerdem

kann das System nicht ohne Weiteres mit mehreren Kameras erweitert werden, um die Verdeckung von anderen Personen bei der Aufzeichnung von Gruppeninteraktion zu vermeiden.

2.3.5 Mechanische Systeme

Bei mechanischen Motion-Capture-Systemen gibt es zwei verschiedene Techniken; diese sind einmal exoskelettbasierte Systeme und zum anderen Systeme, die auf Verformung basieren. Bei mechanischen Systemen werden die Winkel durch ein Exoskelett ermittelt. Ein solches Exoskelett ist in der **Abbildung 14** dargestellt. Dieses ist ein zweites Skelett, welches außerhalb des Körpers angebracht wird und sich parallel zum eigentlichen Skelett des Akteurs bewegt. Durch Winkelenkoder, die in der Lage sind, die Winkel der Gelenke zu messen, ist die aktuelle Pose des Akteurs immer eindeutig messbar. Die zweite Technik, die auf Verformung basiert, arbeitet mit sogenannten Shapetapes. Dies sind Sensoren, die aus verschiedenen fiberoptischen Materialien bestehen und es ermöglichen, Verbiegungen und Verdrehungen im dreidimensionalen Raum wahrzunehmen. Damit sind die Position jedes einzelnen

Abbildung 14 (a) Gypsy5 Exoskelett und (b) ShapeTape, (c) ShapeHand [38]

Zwischenelementes und die Orientierung in Echtzeit wahrnehmbar. Die Zustände der Sensoren und damit die Pose der Probanden können mit einer Rate von 10 kHz aufgezeichnet werden. Dieser Biegesensor und ein ganzer Anzug, der dieAufnahme von menschlichen Bewegungen ermöglicht, sind in der **Abbildung 14** dargestellt. Vorteilhaft an mechanischen Trackingsystemen ist die Tatsache, dass immer und mit hoher Zeitgenauigkeit ein korrekter Winkel für alle Gelenke ermittelbar ist. Der Sensorbereich ist nahezu unbegrenzt. Nachteil ist, dass das Exoskelett die Bewegungen beeinflusst, sodass die Bewegungen durch die zusätzliche Masse etwas schwerfälliger wirken können. Die auf Shapetapes basierende Technik ist relativ unabhängig von der Größe im Gegensatz zum Exoskelett und vom Gewicht her leichter. Diese Systeme sind nicht für die Analyse von Gruppeninteraktionen geeignet, da die Positionen der Personen und ihrer Körperteile zueinander nicht bekannt sind.

2.3.6 Einsatzgebiete der verschiedenen Motion-Capture-Systeme

Die verschiedenen zuvor vorgestellten Motion-Capture-Systeme eignen sich unterschiedlich gut für verschiedene Einsatzgebiete und Analysen. Für die Analyse von Gruppeninteraktionen eignen sich die meisten Systeme nicht. Um zu klären, welche Systeme sich für das Analysieren von Interaktionen eigenen, sind die gängigsten Motion-Capture-Systeme mit ihren technischen Daten in der folgenden **Tabelle 1**[14] aufgeführt. Man sieht in dieser Tabelle, dass es verschiedene optische Systeme gibt, die sich in ihrer Auflösung und Aufzeichnungsrate unterscheiden. Die mechanischen Systeme haben insgesamt eine sehr hohe Genauigkeit, wiegen aber zum Teil viel. Insgesamt eignen sich optische Trackingsysteme am besten für die Analyse von Gruppeninteraktionen. Die Vorteile der optischen Tracking-Systeme sind einmal, dass nur kleine Sensoren (oder keine) ohne Kabel verwendet werden können[15], die die Bewegungsfreiheit nicht oder nur wenig einschränken. Hierdurch wird die Beweglichkeit nicht verfälscht, wie es eventuell bei Aufnahmen mit einem Exoskelett basierten System der Fall wäre. Der größte Vorteil von optischen Systemen mit mehreren Kameras ist die Möglichkeit, verschiedene Personen auf einmal in Relation zueinander erfassen zu können. Außerdem können optische Trackingsysteme, die auf Infrarotsensoren basieren, auch an Orten verwendet werden, in denen die Lichtverhältnisse sich sehr stark ändern. Diese Systeme sind damit unabhängig von den Lichtverhältnissen. Der Sensorbereich kann je nach Belieben mittels zusätzlicher Kameras vergrößert werden. Die zeitliche Auflösung solcher Systeme ist sehr exakt in Bezug auf die Gelenkstellungen. Magnetische Systeme hingegen sind störanfällig gegenüber Metall, das eventuell in der Gebäudestruktur verwendet wurde. Gyroskop Systeme kennen nur die Lage und Ausrichtungsänderungen relativ zum Ausgangspunkt. Mit diesen Systemen ist eine Analyse von mehreren Personen in Relation zueinander nicht möglich, wie es bei optischen Motion-Capture-Systemen der Fall ist. Mit optischen Kameras liegt die zeitliche Auflösung zwischen 60 und 1000 fps (Frames pro Sekunde) und die bildliche Auflösung zwischen 640x480 und 4000x4000 Pixeln. Die räumliche Genauigkeit ist über die Auflösung und Anzahl der Kameras gegeben. Dieses hängt meistens jedoch eher von der räumlichen Anordnung der Kameras des Systems ab. Allerdings sind Exoskelett basierte und das Ultraschallsystem ortsunabhängig, haben kein Verdeckungsproblem und haben eine extrem hohe Abtastrate (bzw. fps). Bei Marker basierten Trackingsystemen ist es auch möglich, die Bewegung der Finger neben der Körperbewegung zu erfassen. Denkbar sind auch Aufnahmen von mehreren Personen und der Gesichtsmimik, wie es bei Motion-Capture Aufnahmen für das

[14] Nicht alle Hersteller haben die Auflösung für ihre Produkte aufgeführt.

[15] Bei aktiven Marker basierten Tracking-Systemen sind Kabel nötig, um die einzelnen LEDs mit Strom zu versorgen, wenn die einzelnen Marker mit verschiedenen Frequenzen arbeiten.

Firma	**Art**	**Auflösung**[16]	**Genauigkeit**	**Anmerkung**
ART-Advance Realtime Tracking	Optisch passiv/aktiv	640x480	15 - 60 Hz	Fingertracking möglich
OptiTrack V120 SLIM /Prime 41	Optisch passiv	640x480 - 4 Megapixel	120 - 250 Hz	-
VICON MX (T10 bis T160) [39] /VICON Motus	Optisch passiv/ markerloses Tracking	1 - 16 MPixel	50 - 1000 Hz	Schnelle Frameraten
PhaseSpace Impulse X2	Optisch aktiv Markers	12 Megapixel	960 Hz	Fingertracking, bis zu 8 Menschen, ca. 150 Gramm, 8 Stunden Batteriebetrieb
PTI-Phoenix Visualeyez VZ 4000	Optisch aktiv Markers	-	Ca. 4000 Hz	Sensor Bar mit mehreren Kameras in einem Gerät
MotionAnalysis Raptor 12	Optisch passiv	2 - 12 Megapixel	150 - 900 Hz	außen einsetzbar
Organic motion openStage2	Optisch ohne Marker	-	60 - 120 Hz	25 bis 100 ms Latenz
Xsens MTi	Gyroskop Anzug	-	400 Hz	Sensor 11 Gram,
Ascension Motion Tracking	Magnetisch	0.25 cm Position 0.1 Grad Orientierung	100 Hz	-
Meta Motion Gypsy7 [38]	Exoskelett	0.125 Grad	30 - 120 Hz	4 kg 14 Sensoren
Measurand ShapeTape	Exoskelett	0.5 Grad	110Hz	-
MS Kinect	Tiefenkamera	640x480	30 Hz	-

Tabelle 1 Motion-Capture-Systeme Übersicht

[16] Ist auch ein indirektes Maß für die räumliche Genauigkeit.

Kino oder die Spieleindustrie der Fall ist. Allerdings ist dabei auch von einer erhöhten Nachbearbeitungszeit auszugehen [18]. Markerlose optische Tracking-Systeme sind leider noch nicht präzise genug, speziell in Bezug auf die Ausrichtung einzelner Körperteile wie Kopf und Hände. Gegen optische Trackingsysteme spricht generell, dass Marker der Körperteile durch bestimmte Bewegungen verdeckt werden können, sodass nicht garantiert werden kann, dass immer alle Stellungen der menschlichen Gelenke richtig ermittelt werden können[17]. Die auf optischen Markern basierten Trackingsysteme eignen sich am besten für die Aufzeichnung von menschlicher Interaktion in Gruppen, sind aber leider auch am teuersten. Diese Informationen sind im Detail in **Tabelle 2** noch einmal zusammengefasst dargestellt worden.

Merkmale	**Kameras**	**Marker**	**Magnetisch**	**Exoskelett**	**Kinect**
Zeitliche Auflösung	Langsam (30 Hz)	Schnell (60 - 4000Hz)	Durchschnittlich (100 - 400)	Durchschnittlich (30 - 120)	Langsam (30 Hz)
Räumliche/ Auflösung	Gut (1 – 16 MP)	Gut (1 - 16 MP)	-	0.125 Grad	Niedrig (1 MP)
Multiple Personen	Daten nicht in Relation	Sehr gut	Daten nicht in Relation	Daten nicht in Relation	Möglich mit Einschränkung
Bewegungseinschränkung	Keine	Gering durch Marker	Stark durch verkabelte Sensoren	Stark durch Exoskelett	Keine
Robustheit	Schlecht	Gut	Durchschnittlich	Sehr robust	Durchschnittlich
Nachbearbeitung	Möglich	Möglich	Nicht möglich	Nicht nötig	Nicht möglich
Störungseinflüsse	Durch Verdeckung	Durch Verdeckung	Durch Metall und elektrische Geräte	keine	Durch Verdeckung
Preis	20.000 USD	10.000 – 100.000 USD	9.000 USD	10.000 USD	200 USD

Tabelle 2 Eignung der verschiedenen Motion-Capture Techniken für den Forschungsalltag

[17] Theoretisch ist diesem mit mehreren Kameras entgegenzuwirken.

2.4 Linguistische Grundlagen

Um die Anforderungen an ein Annotationstool, mit dem in der linguistischen Verhaltensforschung gearbeitet wird, zu verstehen, ist es wichtig, den allgemeinen Arbeitsablauf bzw. Research-Cycle zu kennen. Anschließend werden verschiedene linguistisch relevante Aspekte vorgestellt, die für diese Arbeit bezüglich der auf Gesten basierenden Verhaltensforschung von Bedeutung sind.

2.4.1 Ein Einblick in den Research-Cycle

In dieser Arbeit geht es um interaktives Verhalten in Gruppen, bei dem das Handeln der einzelnen Menschen, ausgehend von Aktionen und Reaktionen, von Interesse ist. Allgemein ist die Analyse von Verhalten ein Bestandteil vieler unterschiedlicher Forschungsrichtungen wie Soziologie, Psychologie, Linguistik, aber auch der Biologie, bei denen es aus der Sicht einer Entwicklung für ein Annotationstool nur wenig Unterschiede gibt; in der Soziologie, bei der es um die Aktion und Reaktion von verschiedenen Gruppenmitgliedern in bestimmte Situation geht; in der Psychologie wird die Reaktion auf das Erlebte und das daraus resultierende Verhalten über längere Zeitperioden betrachtet. Die linguistische Verhaltensforschung ist der soziologischen Verhaltensforschung insofern ähnlich, als beide ihren methodischen Ansatz in der sogenannten Konversationsanalyse haben, bei der es um die Analyse der auf verbalen Äußerungen basierenden menschlichen Interaktion geht. Konversationsanalyse beschäftigt sich damit, wie Gespräche funktionieren und wie diese Gespräche strukturiert werden. Bei den verschiedenen Forschungsrichtungen der Analyse des Verhaltens wird das Verhalten stets zunächst aufgezeichnet und anschließend im Detail untersucht[18]. Ein sehr wichtiger Aspekt der Verhaltensanalyse ist es, möglichst alle Bestandteil einer Interaktion festzuhalten. Der Schwerpunkt dieser Arbeit liegt in der Analyse des Verhaltens von Menschen. Die ethnomethodologische Konversationsanalyse geht im Ansatz zurück auf die Soziologie [40] und stellt eine Verbindung von Soziologie und Linguistik dar [41]. Die Konversationsanalyse selber basiert nicht darauf, Hypothesen am Aufnahmematerial zu prüfen, sondern Hypothesen anhand von Daten zu entwickeln. Der Forschungsablauf kann allgemein in vier Phasen eingeteilt werden, und zwar in das Planen eines Versuches, die Aufnahme der Studie, die Aufbereitung der aufgenommenen Daten und schließlich die eigentliche Analyse des Versuches mit der Evaluation einer Annahme (oder einer Hypothese in der Gesprächsanalyse).

2.4.1.1 Planungsphase

Nach einer genauen Zielsetzung der Studie wird überlegt, wie die entsprechenden Verhaltensweisen untersucht werden können. Zunächst muss eine Möglichkeit gefunden werden, eine bestimmte Verhaltensweise künstlich zu provozieren. Außerdem muss ausgeschlossen werden, dass das Verhalten durch Nebeneinflüsse beeinträchtigt wird und so die Daten ver-

[18] Ausgenommen in der Informatik, in der das Verhalten reproduzierbar ist und zusätzlich direkt analysiert werden kann.

fälscht werden. Eine Fragestellung, die in dieser Arbeit eine große Rolle gespielt hat, ist die Klärung der technischen Realisierbarkeit. Dazu muss untersucht werden, was in welcher Form technisch mit den verfügbaren Mitteln möglich ist. Darüber hinaus muss eine genaue Planung des technischen Aufbaues sowie die zeitliche Koordinierung des den Versuch begleitenden Personals und der Probanden durchgeführt werden. Neben der Klärung der technischen Möglichkeiten sind aber auch die Möglichkeiten bezüglich des Annotierens mit zu beachten, die durch verschiedene Softwaretools begrenzt sind. Hierbei müssen Einschränkungen bezüglich der unterstützten Datenformate, Annotationsfunktionalitäten und der Unterstützung bei verschiedenen Modalitäten (z. B. Audioannotationsmöglichkeiten und die Unterstützung von Motion-Capture) mit berücksichtigt werden. Je nach gewünschtem Ziel und damit verbundenen Annotationen kann viel Zeit durch die Wahl des richtigen Annotationswerkzeuges eingespart werden. Um frühzeitig feststellen zu können, ob das gewünschte Verhalten so provoziert werden kann, dass es anschließend analysierbar ist, empfiehlt es sich, Testaufnahmen durchzuführen. Wird eine Gruppe in der Interaktion so aufgenommen, dass jedes Gruppenmitglied frontal gefilmt wird, steigt der technische Aufwand erheblich an. Zum Beispiel müssen die Versuchsleiter rechtzeitig instruiert werden, wie die technische Ausrüstung zu bedienen ist und dass sie dafür sorgen müssen, dass die Kameras synchronisiert werden (z. B. mit einer Filmklappe). Bei einem komplexen Versuchsaufbau mit vielen technischen Aufnahmegeräten für gegebenenfalls mehrere Modalitäten empfiehlt es sich, eine Checkliste anzufertigen, die bei jedem Versuchsdurchgang neu abgearbeitet wird. Um die Hypothese möglichst gründlich prüfen zu können, hilft ein Fragebogen, der die Ausschlusskriterien nochmals abfragt, um so eine eventuelle Verfälschung der Studie zu vermeiden, aber auch, um noch verschiedene Aspekte der Auswertung des Themas zu formulieren. Zudem muss natürlich immer auch die Privatsphäre der Versuchspersonen gewahrt werden. Dazu müssen die Fragebögen anonymisiert werden, aber es muss auch festgehalten werden, welcher der Fragebögen zu welcher Versuchsperson gehört. Die Daten aus den ausgefüllten Fragebögen, den Audio- oder Videoaufnahmen müssen zur Wahrung der Privatsphäre unter Verschluss gehalten werden.

2.4.1.2 Die Studiendurchführung

Unter Berücksichtigung der vorher definierten Checklisten sowie einer ausführlichen Instruktion des Hilfspersonals und der Versuchspersonen werden die Aufnahmen durchgeführt. Zusätzlich wird festgehalten und dokumentiert, wie der Versuch durchgeführt wurde, um spätere Unstimmigkeiten gegebenenfalls ausräumen zu können. Typischerweise werden Fotos vom Versuchsaufbau und eine detaillierte Skizze gemacht. Um die aufgezeichneten Daten überhaupt für wissenschaftliche Zwecke nutzen zu dürfen, müssen Einverständniserklärungen der Probanden eingeholt werden.

2.4.1.3 Aufbereitung der Daten

Je nach Studiendesign und der Art des zu untersuchenden Verhaltens ist die Aufbereitung der Aufnahmen die zeitintensivste Arbeitsphase. In dieser Phase werden die Daten für ein Anno-

tationstool vorbereitet (z. B. müssen multiple Kameraaufnahmen synchronisiert werden) und in den Annotationstools entsprechend der zu untersuchenden Unterhaltungssituation spezifisch annotiert werden. Dieses können zum Beispiel Annotationen der gesprochenen Sprache sein, bei der jedes gesprochene Wort mit annotiert wird. Hierbei ist es aber auch wichtig, z. B. Verzögerungswortlaute wie „äh“, „öh“ oder „mhh“ mit festzuhalten, damit die genaue Situation bei der späteren Analyse berücksichtigt werden kann. Hinzu können je nach Hintergrund der Studie noch verschiedene weitere Kriterien bezogen auf die gesprochene Sprache mit annotiert werden. Dazu sind im Folgenden einzelne mögliche Merkmale und Kategorien aufgeführt, um zu verdeutlichen, wie aufwendig dieser Annotationsprozess sein kann [41]:

- gleichzeitiges Sprechen
- Abbrüche
- Wiederholungen
- unvollständige Äußerungen
- Versprecher
- äh, öhm, hm usw.
- Stimmhebung
- Stimmsenkung
- Betonungen
- Dehnungen
- Pausen

Dieser Vorgang der Verschriftlichung gesprochener Sprache wird als Transkribieren bezeichnet. Körperliche Gestik und Gesichtsmimik spielen bei der Interaktion auch eine wichtige Rolle und müssen je nach Studienziel mit annotiert werden. Kategorien für körperliche Ausdrucksformen können zum Beispiel folgende sein:

- Gesichtsausrichtung (Aufmerksamkeitsfokus)
- Zeigegesten
- Handaktivität allgemein
- Hand in bestimmte Richtungen bewegen
- Hände symmetrisch bewegen
- bestimmte Körperposen einnehmen (Körperhaltung)
- Fingerbewegungen
- Bewegungsabläufe
- Bewegungsgeschwindigkeiten
- Bewegung einzelner DOFs von einem Gelenk
- Simultanbewegungen
- Bewegungen in Bezug zu anderen Personen (sich nähern)
- Interaktion mit Objekten

- Bestimmung der verschiedenen Bewegungsphasen[19] (Vorbereitungs-, Haupt-, Zurücksetzungs-Phase)

Macht man eine entsprechende Annotation bezüglich einzelner oder aller aufgeführten Punkte, steigt die Annotationszeit immens, besonders, da diese Punkte für mehrere Probanden in der Interaktion annotiert werden müssen[20]. Um die gesamte Komplexität zu berücksichtigen, muss man bedenken, dass es immer verschiedene weitere Unterpunkte zu den einzelnen Merkmalen gibt. Zum Beispiel kann das Gesicht folgende Ausdrücke einnehmen:

- fröhlich
- ängstlich
- traurig
- neutral
- zornig
- gelangweilt
- genervt
- überrascht
- entsetzt
- erschrocken
- enttäuscht

Anhand dieser Merkmale und Kategorien kann man sich vielleicht vorstellen, dass für jede der Kategorien spezifische Unterkategorien und Merkmale definiert werden können, die das Annotieren entsprechend komplizierter und zeitintensiver machen. Ein Katalog dieser Kategorien mit den Zuständen wird als Coding-Schema bezeichnet [42]. Dieses beinhaltet die Informationen, die durch Annotation hervorgehoben werden sollen, je nachdem, welcher Schwerpunkt von Interesse ist. Zu jedem dieser Kategorien wird sich in der Regel jedes Video einmal genau angeschaut (oder angehört bei sprachbezogenen Annotationen), um sich auf die jeweilige Kategorie und die Merkmale konzentrieren zu können. Das Annotieren von Sprache dauert ca. das 35-fache, die Übersetzung in eine andere Sprache dauert noch einmal das 25-fache, und die Annotationszeit mit zusätzlichen Gesten sogar mehr als das 100-fache der Aufzeichnungszeit[21] [13]. Dazu kommt, dass Menschen Sachverhalte unterschiedlich wahrnehmen und Fehler machen, sodass normalerweise die gleichen Annotationen von mehreren Personen annotiert werden, die später wiederum zu einer qualitativ höherwertigeren Annotation zusammengefasst werden.

[19] Eine detailliertere Beschreibung der Bewegungsphasen wird im Verlauf dieses Kapitels im Abschnitt 2.4.2 gegeben.

[20] Im Gegensatz zur Transkription, bei der meistens nur einer redet und es höchstens kurzzeitig zu Überlappungen kommen kann.

[21] DOBES Project www.mpi.nl/dobes

2.4.1.4 Die eigentliche Analyse

Nach erheblichem Aufwand können die aufbereiteten Daten anhand der annotierten Kategorien entsprechend einer Forschungsfragestellung analysiert werden. Diese Analyse wird innerhalb eines oder mehrerer Annotationstools durchgeführt, welche die Aufnahmen und synchron dazu die Annotationen darstellen. Dabei unterstützen die meisten der relevanten Tools[22] auch das Handhaben mehrerer Datentypen, die bestenfalls sogar synchron zur aktuellen Abspielzeit des betrachteten Videos genutzt werden können (z. B. eine Darstellung der Frequenzen der Tonspur). Mit diesen Tools werden die Hypothesen geprüft, indem die einzelnen zeitlich aufbereiteten Zusatzinformationen in Kombination mit anderen Annotationen in Zusammenhang analysiert werden.

2.4.1.5 Allgemeiner Research-Cycle

Es gibt es auch viele Gemeinsamkeiten beim Vorgehen der verschiedenen Fachrichtungen [43]. Bei allen Fachrichtungen kommen die hier vorgestellten vier Phasen (Planung, Durchführung, Aufbereitung und Analyse) vor und können als Verallgemeinerung dieser aufgefasst werden. Mit der Betrachtung des Research-Cycles im Allgemeinen wurde nur der grobe Arbeitsablauf beschrieben, in dem viele einzelne Arbeitsschritte und Analysen durchgeführt werden. Die bei Weitem aufwendigste Phase ist die Aufbereitungsphase der Daten, bei der verschiedene Bestandteile des Verhaltens aufbereitet werden, basierend auf den verwendeten Datenquellen.

2.4.2 Bestandteile von Gesten

Gesten sind ein wichtiger Bestandteil in der Mensch-Mensch-Kommunikation. Um diese genauer zu analysieren, wird die Bewegung in kleinere Elemente (welche als Phasen bezeichnet werden) unterteilt. Dadurch kann ermittelt werden, welcher Teil einer Bewegung von Bedeutung ist. Diese Phasen können Aufschluss auf verschiedene Typen von Gesten geben, und sie erlauben, zusätzliche Informationen zu ermitteln, die es ermöglichen, die Gesten unterschiedlich zu deuten. Kita unterteilt Gesten in verschiedene Phasen; bei einer genaueren Betrachtung wird die Bewegung [44] oder auch die Phase in kleinere Phasen (engl. phases) unterteilt. Dabei gibt es einzelne Phasen in einer Geste, die eine komplexere Struktur aufweisen können, die durch eine Grammatik (Strukturdefinition) aufgebaut werden kann. Im folgenden Unterabschnitt wird zunächst angeschaut, wie diese Unterteilung gemacht werden kann; dann, wie die Bewegungseinheiten (Phasen) identifiziert werden können, und anschließend, wie deren Typen identifiziert werden können [44].

2.4.2.1 Segmentierung der Bewegung in Phasen

Eine Bewegungseinheit beginnt, wenn sich eine Hand von einer Ruheposition anfängt zu bewegen. Eine Ruheposition ist meistens durch ein stützendes Objekt bestimmt, wie zum Bei-

[22] Eine genaue Vorstellung dieser Tools wird in der Sektion 3.1 mit einem Vergleich gegeben.

spiel eine Stuhllehne oder der eigene Körper. Es können aber auch die Finger mit den Haaren spielen oder die Hände an der Kleidung positioniert sein, wenn sie sich zum Beispiel in einer Hosen- oder Pullovertasche befinden. Die Hände können sich auch gegenseitig halten, wenn die Arme ineinander verschränkt sind, oder aber an einem Objekt wie einer Kaffeetasse fixiert werden. Aus analytischer und mathematischer Sicht sind die Hände lange an einer Stelle und bewegen sich nicht mit großer Geschwindigkeit. Eine Bewegungseinheit lässt sich in mehrere Phasen unterteilen. Die Struktur der einzelnen Phasen in einer Bewegung kann wie folgt aufgebaut sein:

$$Bewegungseinheit = Bewegungsausdruck *$$

$$Bewegungsausdruck = (Vorbereitung) \Rightarrow Ausdrucksphase \Rightarrow (Rückzug)$$

$$Ausdrucksphase = einzelne\ Haltung\ (engl. hold)$$

$$Ausdrucksphase = (abhängige)\ Haltung \Rightarrow Bewegungszug\ (engl. stroke)$$

$$\Rightarrow (abhängige) Haltung$$

$$Vorbereitung = (befreiende\ Bewegung) \Rightarrow lokale\ Vorbereitung$$

$$\rangle\rangle\ Hand\ interne\ Vorbereitung$$

$$Rückzug\ (wenn\ sie\ von\ einem\ weiteren\ Bewegungsausdruck\ gefolgt\ wird) = partieller\ (unvollständiger)\ Rückzug$$

Tabelle 3 Grammatik (Strukturdefinition) von Bewegungsphasen bei Handgesten [44]

Die Notation für diese Grammatik ist:

X = Y	X besteht aus Y
*	eins oder mehrere Elemente
⇨	diskreter Übergang
( )	optional
>>	gemischter mitunter diskreter Übergang

Diese Definition nach Kita entspricht einer Grammatik, durch die eine Bewegungseinheit aus beliebig vielen Bewegungsausdrücken aufgebaut sein kann. Ein Bewegungsausdruck hat immer eine Vorbereitungsphase, eine Ausdrucksphase und eine Rückzugsphase. Es gibt insgesamt fünf verschiedene Typen von Bewegungsphasen: Bewegungszug (engl. stroke), Haltung (engl. hold), Vorbereitung (engl. preparation), Rückzug (engl. retraction) und unvollständiger Rückzug (engl. partial retraction), die entsprechend der Grammatik in der Tabelle 3 aufgebaut sein können. Die Bewegungseinheiten werden in weitere Einheiten unterteilt, wenn bei der Bewegung eine Richtungsänderung eintritt und eine Unterbrechung der Geschwindigkeit

(bzw. die Beschleunigung gleich Null wird) auftritt. Daraufhin wird die Bewegungseinheit in zwei Bewegungsausdrücke unterteilt und als eine Einzelsegmentphase bezeichnet; ist keine Geschwindigkeitsunterbrechung vorhanden, wird sie als Multisegmentphase bezeichnet. Eine sich wiederholende Bewegung wie das Klopfen mit einem Finger wird als Wiederholungsphase definiert.

2.4.2.2 Identifikation von Phasentypen

Es gibt verschiedene Identifikatoren für die verschiedenen Phasentypen [44]:

- Ein Phasensegment, das mehr Kraft[23] beinhaltet als die umliegenden Phasen, ist eine Stroke-Phase. Dabei spielt die Richtung eine Rolle. Wenn die Bewegungsrichtung nach unten gerichtet ist, hat die Gravitation einen Einfluss. Eine Multisegmentphase und eine Wiederholungsphase werden als Stroke-Phasen bezeichnet.
- Ein Phasensegment, in dem die Hand ruht und nicht in einer Ruheposition ist, wird als Hold-Phase bezeichnet, wobei das Ruhighalten der Hand relativ zu den umliegenden Phasen zu betrachten ist.
- Ein Phasensegment, welches bei einer Ruheposition beginnt und keine Stroke-Phase darstellt, ist eine Präparations-Phase, und ein Phasensegment, welches bei einer Ruheposition endet, ist eine Retraktions-Phase.
- Ein Phasensegment zwischen zwei Stroke-Phasen ist eine Präparations-Phase. Manchmal bewegt sich die Hand in Richtung einer möglichen Ruheposition hin, bereitet dann aber doch noch eine weitere Stroke-Phase vor[24]. Dieses wird als Partial-Retraktions-Phase bezeichnet.
- Eine Bewegung aus der Ruheposition, bei der die Hand z. B. aus einer Tasche genommen wird, um einen Ausdruck durchzuführen, wird als Vorbereitungsphase bezeichnet.

2.5 Zusammenfassung

In diesem Kapitel wurde der biologische menschliche Bewegungsmechanismus mathematisch beschrieben, um diese mathematische Beschreibung für automatische Analysen verwenden zu können. Anschließend wurden verschiedene Animationstechniken vorgestellt, die wieder im späteren Konzeptteil aufgegriffen und bei den automatischen Analysen benötigt werden. Danach wurde ein Überblick über verschiedene Motion-Capture-Systeme gegeben und deren Vor- und Nachteile diskutiert. Zum Schluss wurden linguistische Aspekte der Arbeit und das Aussehen der verschiedenen Grundlagen der Gestenforschung eingeführt. Dazu wurde auf den allgemeinen Research-Cycle eingegangen, gefolgt von der Vorstellung eines Konzepts zur Beschreibung von Gesten mittels einzelner Bestandteile.

[23] Große Beschleunigung ist das Resultat von großer Kraft.

[24] Beispielsweise, da der interagierenden Person plötzlich eine Idee gekommen ist.

2.6 Fazit

Um die gesetzten Ziele erfüllen zu können, müssen Wege gefunden werden, die Bewegung des Menschen zu erfassen und in ein für Annotationen nützliches Format zu überführen. Außerdem wird eine Möglichkeit benötigt, den gesamten Zusatzaufwand des Motion-Capturens im Verhältnis zu positiven Ergebnissen abschätzen zu können, die auf diesem Zusatzaufwand basieren.

3 Stand der Forschung und Technik

Forschung im Bereich des menschlichen Interaktionsverhaltens wird in den verschiedenen Forschungsdisziplinen wie Psychologie, Linguistik und Soziologie durchgeführt. Dabei spielen Annotationstools für multimodale Daten (Audio und Video) eine zentrale Rolle. Im folgenden Kapitel wird der Stand der Forschung und Technik bezüglich Annotieren und Motion-Capturing vorgestellt. Im Rahmen dieser Arbeit wird ein neues Tool vorgestellt, das Motion-Capturing-Daten von mehreren Personen als Grundlage für automatische Annotationen verwendet. Um darzulegen, dass es nötig war, ein neues Tool zu entwickeln, werden die existierenden Tools im Bereich Annotation vorgestellt. Es existiert eine Reihe von multimodalen Annotationstools, die einem breiten Kreis von Personen zur Verfügung stehen. Im folgenden Kapitel werden ihre Stärken und Schwächen diskutiert. Außer diesen Annotationstools für multimodale Daten gibt es noch eine Vielzahl von Arbeiten über Algorithmen, die zur automatischen Detektion von Ereignissen benutzt werden können. Leider sind diese meistens nicht direkt nutzbar, da sie nur schwer zugänglich und für einen breiteren Personenkreis zu kompliziert zu benutzen sind. Die dabei benutzten Algorithmen sind oft nur für eine konkrete Problemstellung und nicht für den Allgemeinfall anwendbar. Sie stellen eine interessante Entwicklung dar und weisen auf mögliche Funktionalitäten in kommenden Annotationstools (oder neue Versionen der existierenden Tools) hin. Diese werden am Ende dieses Kapitels vorgestellt, um einen Überblick zu geben, welche automatischen Annotationsmöglichkeiten im Ansatz oder als Weiterentwicklung bereits existieren. Dazu wird ein Überblick über die für diese Arbeit funktional relevanten Algorithmen in Verbindung zum Motion-Capturing gegeben.

3.1 Multimodale Annotationssoftware

Im Folgenden werden die verschiedenen Annotationstools vorgestellt. Dabei wird der Fokus zunächst auf die am weitesten verbreiteten Tools gelegt, die, basierend auf einer Untersuchung von Rohlfing et al. mit der Community im Jahre 2006 [17], auf den heutigen Stand gebracht wurden. Diese werden mit ihren Schwächen und Stärken vorgestellt. Anschließend werden diese Tools gegenübergestellt, um einen schnellen Überblick über die Unterschiede und mögliche Einsatzgebiete zu geben. Als Maßstab zur Verbreitung und Wichtigkeit der Tools wurde die Community[25] selber gewählt, um festzustellen, welche Tools aktuell benutzt und aktiv weiterentwickelt werden [17] [42]. Anschließend werden alle bekannten Tools gegenübergestellt, um die Unterschiede dieser Tools hervorzuheben und um zu beleuchten, welche weiteren Fähigkeiten wünschenswert sind.

[25] Die Community besteht aus aktiven Entwicklern der Tools sowie Anwendern selber, die sich zusammengetan haben, um die Schwächen und Stärken der Tools herauszufinden.

3.1.1 Allgemeine Mediaspieler und Texteditoren

Man muss keine spezielle Software zum Annotieren nutzen, da es eine Vielzahl von Programmen zum Abspielen verschiedener Medien und Texteditoren gibt. Zum einen sind diese für jedes beliebige System verfügbar, und zum anderen ist die Einarbeitung meist nicht mehr erforderlich oder minimal, da die Software durch ihren alltäglichen Einsatz schon bekannt ist. Mediaabspiel-Software sind normale Video-Betrachtungs-Programme oder Video-Bearbeitungs-Tools wie zum Beispiel „MS Media Player“, „Virtual Dub“, „Quick Time“, „VLZ“, (K)MPlayer, „Adobe Premiere“, „Apple Final Cut“ usw. Diese Tools sind normalerweise auch unabhängig von Video-Codecs, Video-Datei-Formaten und haben keine Längenbegrenzung für die Mediafiles. Zum Annotieren wird die Medienabspiel-Software zusammen mit einem Texteditor gestartet. In diesen kann der Zeitpunkt des Auftretens und des Endens eines Ereignisses festgehalten werden. Beinahe in jedem dieser Tools gibt es eine einfache Text Suchfunktion, die bei einer späteren Suche nach speziellen Ereignisses (oder auch Phänomentypen und Beschreibungselementen) genutzt werden kann. Die Synchronisation beschreibender Textelemente (Annotationen) zu Videoausschnitten geschieht hier ausschließlich über festgehaltene Zeitangaben, manche Mediaplayer unterstützen das direkte Einstellen eines spezifizierten Zeitpunktes. Einzelne Mediaspiel-Softwares unterstützen auch Slow-Motion Abspielfunktionalität, um sich spezielle Sequenzen im Detail anschauen zu können. Die größte Schwäche ist, dass die spätere Analyse der Annotionsdaten mühsam wird, da die annotierten Daten nicht automatisch synchron zu den zugehörigen Videodaten gehalten werden.

3.1.2 PRAAT

Praat[26] ist ein auf Audiodaten spezialisiertes Annotationstool, dessen Entwicklung von Paul Boersma und David Weenink im Jahr 1996 an der Universität Amsterdam begonnen wurde. Es bietet eine graphische Darstellung der Tonspur und unterstützt durch verschiedene Funktionalitäten das Audioannotieren. In der **Abbildung 15** ist die Benutzeroberfläche dargestellt. Die Software bietet die Möglichkeit, neben der Synchronisation der zu Audiodaten annotierten Daten eine Grundfrequenzanalyse der gesprochenen Sprache zu erstellen. Außerdem können Bilder von Oszillogrammen, Spektrogrammen, Transkriptionen und Kombinationen daraus erstellt werden [45].

[26] Streng genommen ist „Praat“ kein multimodales Annotationstool, da es nur die Modalität „Audio“ unterstützt. Aber es bietet verschiedene automatische Audio Annotations Funktionen und zum anderen ist es auf unterschiedliche Weise mit anderen, im Folgenden beschriebenen Tools kombinierbar.

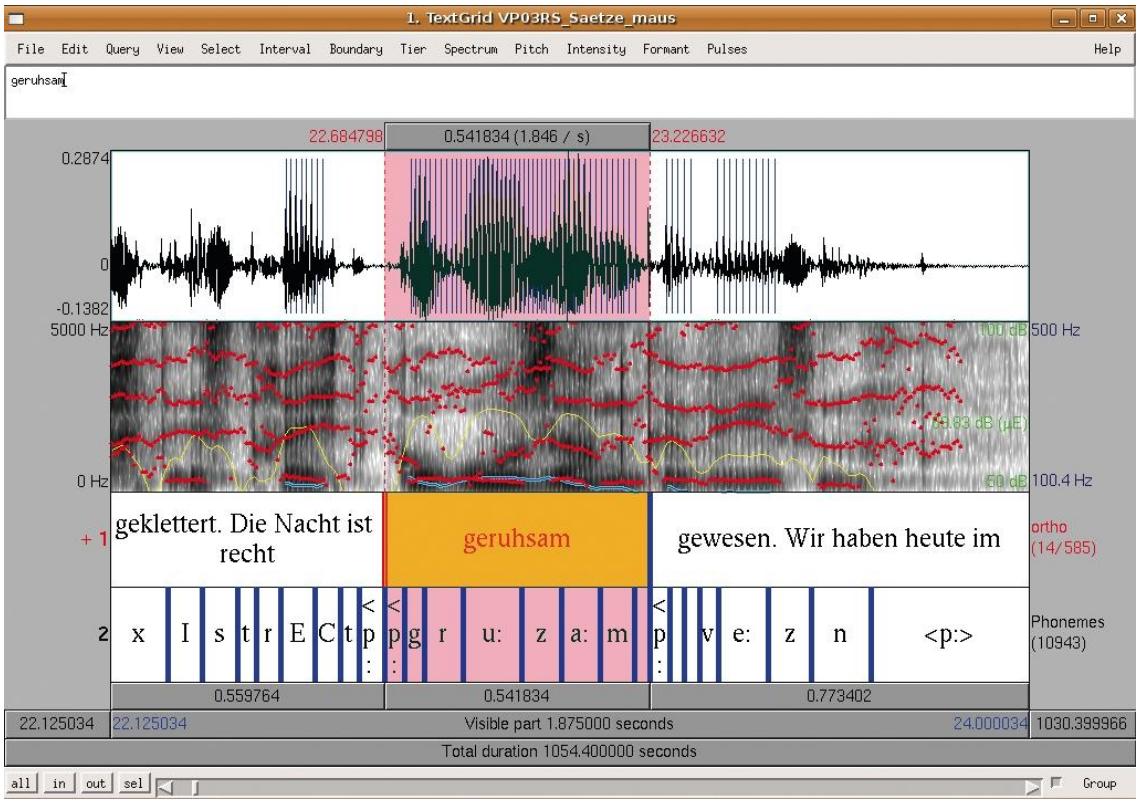

Abbildung 15 Praat-Benutzeroberfläche zum Annotieren von Audioaufnahmen mit Audiosegmentierungsfunktionalität [45]

3.1.3 TASX

„Time Aligned Signal data eXchange“ - TASX ist ein multimodales Annotationstool, welches auf XML[27] Datenstruktur arbeitet und multiple Tier[28] Annotationen ermöglicht. Es wurde an der Universität Bielefeld im Jahr 2002 von Jan Torsten Milde entwickelt [16]. Es ist eines der ersten Annotationstools, bei dem Audio und Video annotiert werden konnten. Ein Tier ist eine Annotationskategorie, welche einer Person, einem Phänomen oder einer Funktionalität zugewiesen werden kann (z. B. Handbewegungen, Blickrichtungen, Transkriptionen von Gesprochenem, Übersetzung in andere Sprachen). TASX hat zwei verschiedene Fenster, eines davon ist ein funktionaler Medienspieler, das andere bietet verschiedene Möglichkeiten zum Annotieren. Dazu gehören eine visuelle Darstellung des Audiosignals als Frequenzplot und verschiedene Darstellungsformen der Daten. Es kann in die Annotationsdaten gezoomt werden, und es kann auf bestimmte Zeitpunkte direkt zugegriffen werden. In der folgenden **Abbildung 16** ist die Benutzeroberfläche dargestellt.

[27] XML steht für Extensible Markup Language und wird für die Darstellung von hierarchischen Textstrukturen verwendet.

[28] Tier ist Englisch und bedeuten im Deutschen Ebene, gemeint sind verschiedene Ebenen für unterschiedliche Kategorien.

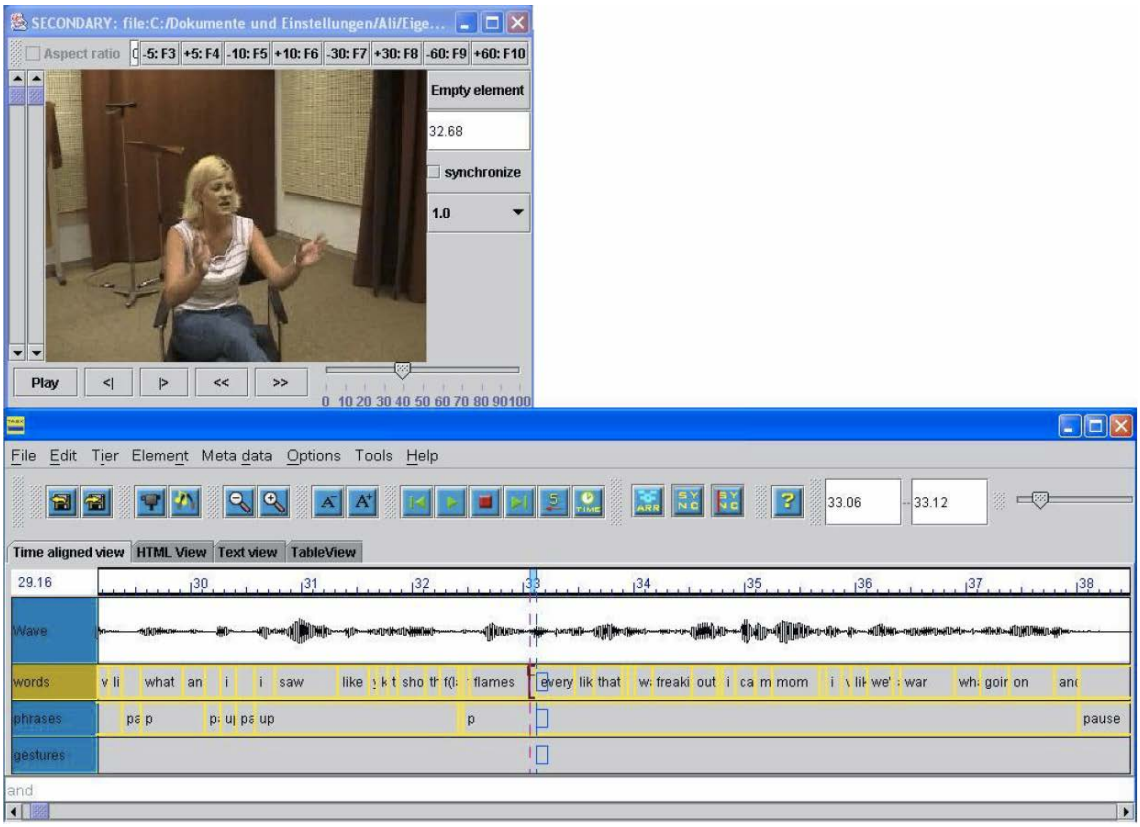

Abbildung 16 Benutzeroberfläche des Annotationstools TASX [16]

3.1.4 ANVIL

Das Annotationstool ANVIL wurde für Video-Annotation von Michael Kipp im Jahre 2001 an der Universität des Saarlandes entwickelt und bietet zusätzlich zu der Funktionalität von TASX eine hierarchische Anordnung von Tiers. Zur Hervorhebung bestimmter Annotationstypen in den verschiedenen Tiers können diese eingefärbt werden, damit Kategorien leichter voneinander zu unterscheiden sind (siehe **Abbildung 17**). Durch die Möglichkeit, hierarchisch Tiers anzuordnen, kann eine Annotation erst einmal grob durchgeführt werden, um z. B. alle Zeitspannen mit Bewegung hervorzuheben, und anschließend kann eine verfeinerte Analyse der Einzelheiten einer Bewegung durchgeführt werden. Darüber hinaus bietet es die Möglichkeit, ein Praat-File mit zu visualisieren. Eine neuere Funktionalität ist die Darstellung von Motion-Capturing-Daten einer einzelnen Person (siehe **Abbildung 18**). Damit kann die Bewegung einer einzelnen Person von allen Blickwinkeln im Detail angeschaut werden und eine genaue Bestimmung der einzelnen Körperteile zur manuellen Annotation genutzt werden. Zudem werden in dieser 3D Ansicht der Motion-Capture-Daten auch die Trajektorien („color coded motion trails“)[29] entsprechend den Annotationen eingefärbt, was eine genaue Analyse der gesprochenen Sprache im Bezug auf die Bewegungen ermöglicht. Der Schwerpunkt des Tools liegt in der globalen Analyse der Bewegung, wodurch die aktive Bewegung leider nicht den einzelnen Gelenken zugeordnet werden kann. Um dennoch lokale

[29] „Coding“, zu Deutsch Kodieren, beschreibt das eigentliche Annotieren oder Transkribieren. Das Coding-Schema beschreibt, was in welcher Form annotiert werden soll mit allen möglichen Werten.

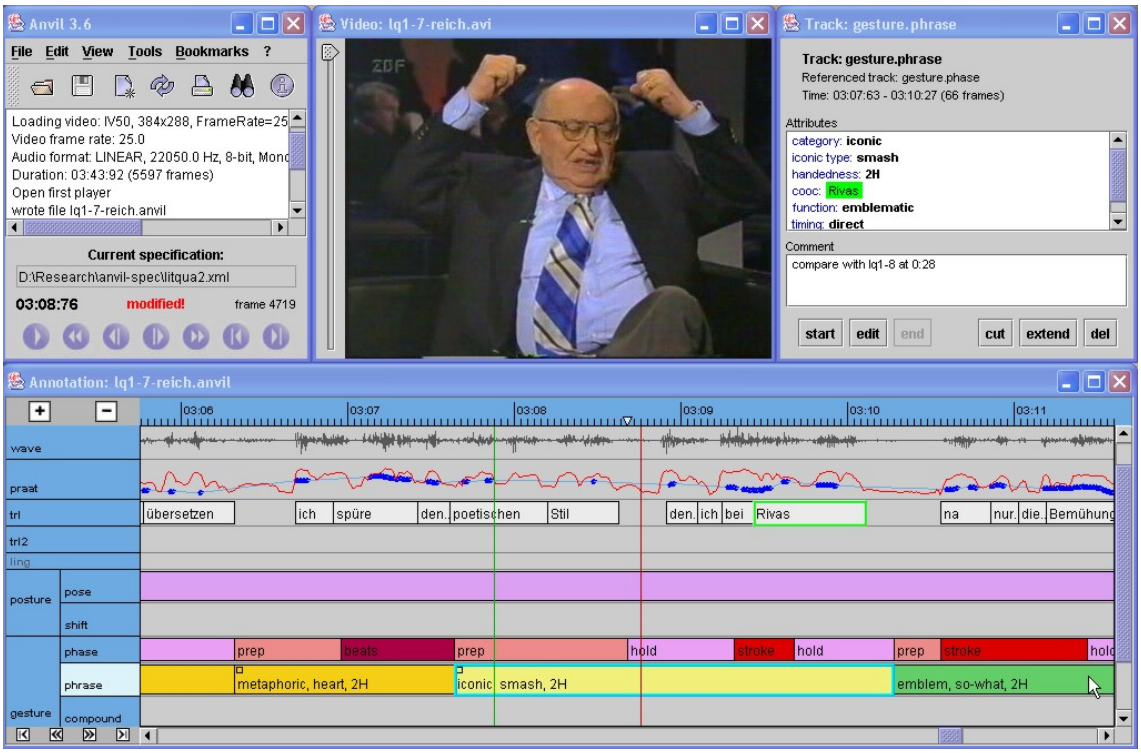

Abbildung 17 ANVIL Benutzeroberfläche mit Stimmenintensitätsanzeige [12]

Bewegungen bestimmter Gelenke betrachten zu können, gibt es die Möglichkeit, ein Gelenk des Skelettes in globalen Koordinaten festzustellen, wodurch sich nur die in der Gelenkhierarchie nachfolgenden Gelenke bewegen. Macht man dieses zum Beispiel mit der Schulter, kann die Bewegung des Unterarmes genauestens analysiert werden, ohne dass die Bewegungen des Oberkörpers Einfluss auf diese haben. Eine weitere Visualisierung stellt einen Plot von den

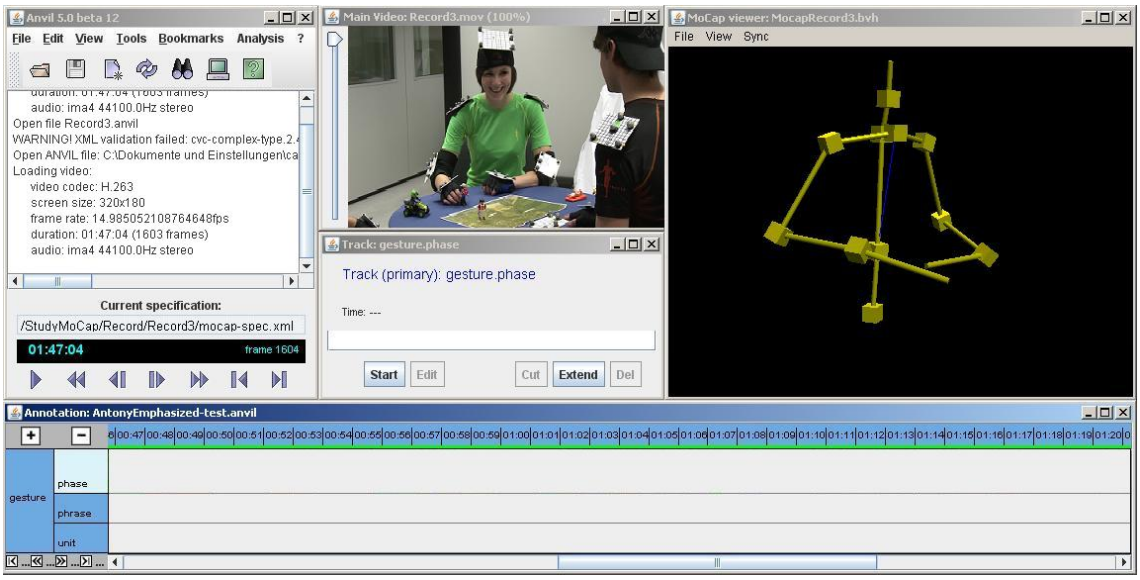

Abbildung 18 ANVIL mit dem Einzelpersonen Motion-Capture-View, bei der aus PAMOCAT die Bewegung einer einzelnen Person exportiert wurde[30]

[30] Die dargestellte Szene stammt aus dem sogenannten „Obersee“ Korpus [71].

Handpositionen in globalen Koordinaten dar. Zudem wurden Funktionen zur Erkennung von Handbewegungen entwickelt, die unterscheiden können, ob die linke, rechte oder beide Hände aktiv sind [12].

3.1.5 EXMARaLDA: Extensible Markup Language for Discourse Annotation

EXMARaLDA steht für „Extensible Markup Language for Discourse Annotation". Es wurde am Hamburger Zentrum für Sprachkorpora (HZSK) von Thomas Schmidt, Kai Wörner und Hanna Hedeland entwickelt [15]. Es ist mit dem Ziel der Computer assistierenden Unterstützung bei der Analyse gesprochener Sprachen für Korpora entwickelt worden. Es stellt eine Kollektion von Datenformaten und Software Tools dar, die das Erstellen, Analysieren und den Zugriff auf Sprachkorpora unterstützen. Zu den Softwarefunktionalitäten zählen Tools zur Transkription, zur Korporaverwaltung und Unterstützung bei der Durchsuchung von Korpora. Ein wichtiger Punkt ist dabei der leichte Zugriff auf verschiedene archivierte Korpora und die Möglichkeit, eine Suche über verschiedene Korpora hinweg durchführen zu können. Die Datenformate sind in XML darauf ausgelegt, ein Standardformat für Korpora zu etablieren. In der **Abbildung 19** wird die Benutzeroberfläche des Partitur-Editors gezeigt, welcher zum Annotieren bzw. Kodieren entwickelt wurde.

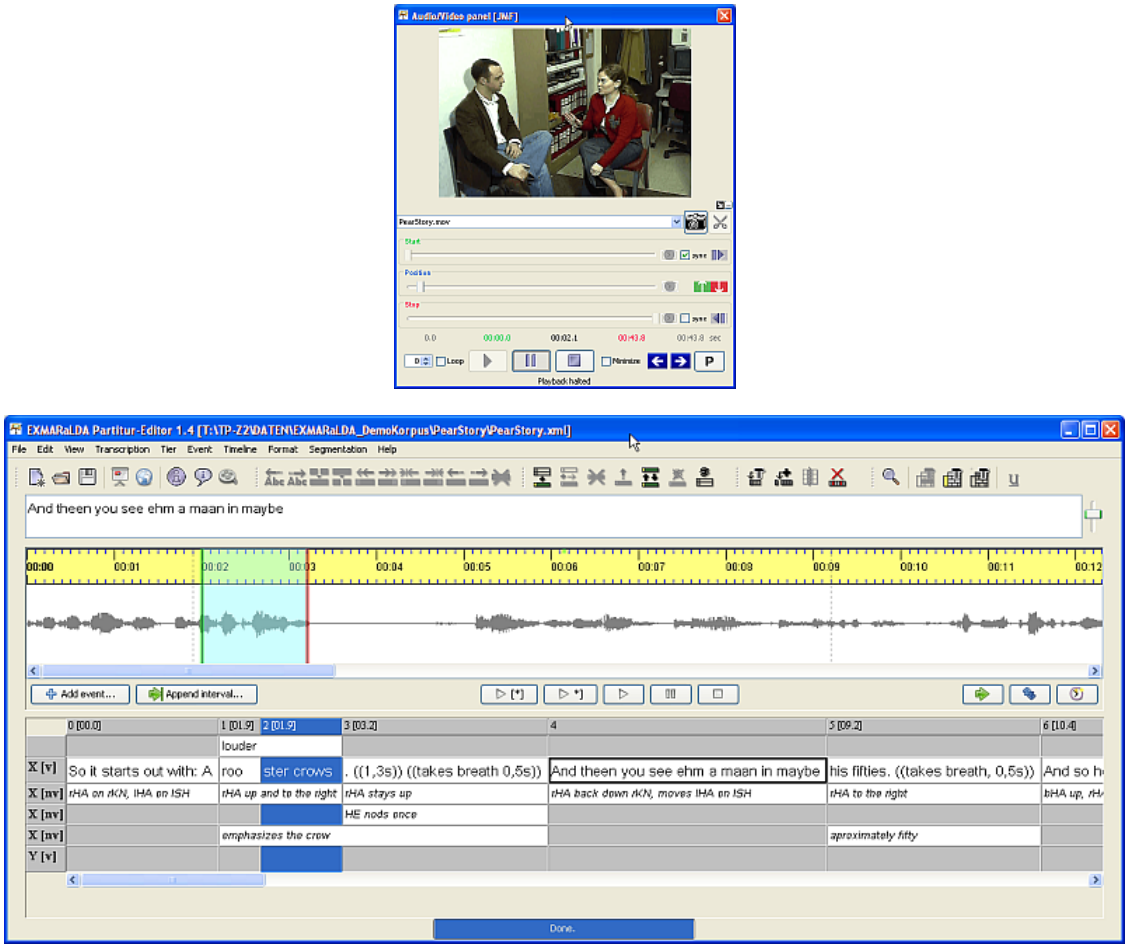

Abbildung 19 Benutzeroberfläche des Partitur-Editors von EXMARaLDA [15]

3.1.6 ELAN

ELAN wurde von Han Sloetjes am Max-Planck Institut 2002 in Nijmegen in den Niederlanden entwickelt [13]. Derzeit wird es von zwei Max-Planck Instituten und einem Fraunhofer Institut weiterentwickelt. Dabei wird ein großer Schwerpunkt auf das semi-automatische Annotieren von Audio- und Videodaten gelegt. Dazu hat es verschiedene Filter, zum Beispiel, um Hintergrundgeräusche herauszufiltern, damit eine Mustererkennung bei einem Audiosignal durchgeführt werden kann. Aber auch einfache Funktionen, lange Redepausen automatisch zu erkennen, helfen bei dem mühsamen Annotationsprozess und reduzieren die Annotationszeit. Neben Funktionen zur Segmentierung von Sprache werden mit ELAN aber auch Tests zur Spracherkennung, Sprechergruppierung und Geschlechtserkennung basierend auf Standardalgorithmen durchgeführt. Darüber hinaus lässt sich auch die Audio Annotationssoftware Praat aus ELAN heraus starten und für eine detaillierte akustische Analyse verwenden. Es gibt ähnlich wie bei EXMARaLDA eine Suchfunktionalität namens TROVA, welche reguläre Ausdrücke bei der Suche zulässt. Weitere interessante Möglichkeiten sind verschiedene Implementierungen von Filtern basierend auf Videodaten, mit denen einzelne Gebiete markiert und Aktivitäten in diesen detektiert werden können. Außerdem ist eine Schnittstelle für Plug-Ins in ELAN integriert [46], damit die Community gewünschte Fähigkeiten selbst integrieren kann.

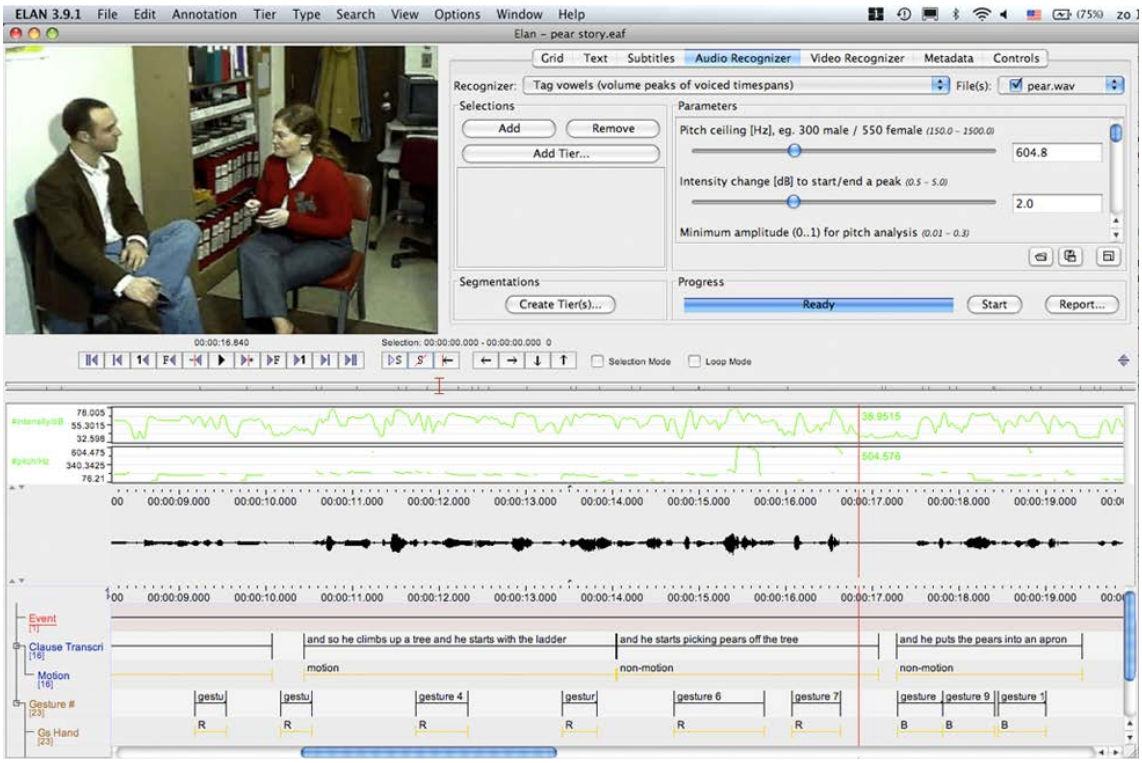

Abbildung 20 Benutzeroberfläche von ELAN mit Stimmenintensitätsanzeige [13]

3.1.7 Weitere Annotationstools

Neben den zentralen Annotationstools gibt es noch eine Reihe von nicht so verbreiteten Tools, die hier vollständigkeitshalber kurz erwähnt werden sollen. MacVis ist ein Tool spezi-

ell zur Annotation von strukturellen Zusammenhängen für Audio- und Video-Dateien (nur für MacOS). C-Bas ist ein Tool zum Annotieren von Audio- und Video-Gesten basierend auf einer XML Datenstruktur[31]. Theme ist eine kommerzielle Software zur Detektion von Mustern bei aufeinanderfolgenden Ereignissen basierend auf Annotationen. Transformer ist ein Tool zum Austauschen von Daten zwischen den bekanntesten Annotationstools. Folker [47], Transcriber und Clan sind Tools für die Annotation von Audiodaten. Annex ist ein webbasiertes Videoannotationstool [48]. Software, die beim Aufzeichnen von Motion-Capture-Daten verwendet werden kann wie zum Beispiel „Vicon NX“, reichert zwar auch die Rohdaten mit Zusatzinformationen an (wie z. B. welcher Marker gehört zu welchem Körperteil), wird aber

Name	**Merkmale / Fähigkeiten**	**Nachteile für Interaktionsanalyse**
Praat	Annotationstool für Sprache	Kein Video
TASX	Tool zum Kodieren von Audio- und Video-Aufnahmen	Nicht mehr aktuell verfügbar
ANVIL	Einzelne Person Motion-Capture-Support, hierarchische Tier Strukturierung möglich, farblich hervorgehobene Trajektorien in Bezug zur Annotation	Keine Gruppeninteraktion -analysierbar
EXMA-RaLDA	Erstellung von sekundengenauen Annotationen ohne zeitliche Rahmenbedingungen wie Frameanzahl	Umständlicher Wechsel zwischen Videos
ELAN	Verbreitetes Annotationstool für das Annotieren von Video und Sound mit Praat-Integration. Schnittstelle für Video- und Audioplugin.	Kein Motion-Capturing
MacVisSTA	Speziell zur Annotation von Zeitfolgen struktureller Zusammenhänge	Nur für Mac OSX verfügbar und letzte Version war 2008. Keine online Dokumentation verfügbar
C-Bas	Audio und Video Annotationstool von Gesten basierend auf einem XML Format mit Coding-Schema Unterstützung.	Aktuell nicht mehr verfügbar.
Theme	Software zum Detektieren von Mustern bei aufeinanderfolgenden Ereignissen in den Annotationsdaten.	Kommerzielle Software, macht keine eigene Datenauswertung
Transformer	Ermöglicht es, Annotationen-Daten zwischen Annotationstools auszutauschen und erstellt druckfähige Dokumente aus diesen [49]	Nur für Windows 7 verfügbar
FOLKER	Speziell für Audio Annotationen der Sprache Annotationsschema.	Keine Import und Export Funktionalität
Transcriber	Speziell für Audio Annotationen.	Keine Import und Export Funktionalität
Clan	Speziell für Audio Annotationen.	Keine Import und Export Funktionalität

Tabelle 4 Annotationstool-Übersicht basierend auf [17], [42] und [50].

[31] Die meisten aktuell verwendeten Tools setzen auf XML als Datenstruktur.

an dieser Stelle nicht als Annotation gesehen. Dabei wird ein Fehler ausgeglichen, der durch die Motion-Capturing-Technik hervorgerufen wird und somit in die Nachbereitung des Motion-Capturings fällt.

3.1.8 Direkter Vergleich von Annotationstools

Um einen Überblick über die zuvor vorgestellten multimodalen Annotationstools zu geben, werden diese Tools nun direkt gegenübergestellt. In **Tabelle 4** sind alle multimodalen bekannten Annotationstools aufgeführt, im weiteren Verlauf wird auf wichtigeEinzelheiten eingegangen. Wir beginnen mit einer groben Einordnung der verschiedenen multimodalen Annotationstools basierend auf den verschiedenen Fähigkeiten in Relation zueinander. Dazu werden die Besonderheiten der jeweiligen Tools (siehe **Tabelle 4**[32]) kurz aufgeführt. Anschließend werden die Annotationstools im Detail miteinander verglichen, dazu zählt die Unterstützung allgemeiner Funktionalitäten, die durch andere Tools bekannt sind (siehe dazu **Tabelle 5**). Einzelne Tools hatten in einer früheren Version Probleme bei längeren Aufzeichnungen gehabt. Für die spätere Analyse müssen die Aufnahmedaten kodiert (annotiert) werden. Wie dieses geschehen kann, und wie die Tiers aufgebaut werden können, unterscheidet sich durch

Name	Aufnahmelimit	Abspielgeschwindigkeit änderbar	Suchfunktionen	Auto. Analyse
Praat	Nein	Nein	Nein	Ja
TASX	Unübersichtlich bei + 4h	Ja	Basic	Ja
ANVIL	Probleme > 30 min	Ja	Ja	Ja
EXMA-RaLDA	Nein	Nein	Ja	Ja
ELAN	Nein	Ja	Komplex	Ja
MacVisSTA	Framedrops > 40min	Ja	Nein	Nein
C-Bas	Unbekannt	Nein	Unbekannt	Nein
Theme	Unbekannt	Nein	Ja	Nein
Transformer	Keine Betrachtung möglich	Nein	Ja	Nein
FOLKER	Nein	Nein	Ja	Nein
Transcriber	Nein	Nein	Ja	Nein
Clan	Nein	Nein	Ja	Nein

Tabelle 5 Eigenschaften der Annotationstools in einer Übersicht basierend auf [17], [42] und [50]

[32] Diese Tabellen basieren auf den Tabellen in Kapitel 4 aus der Arbeit [16].

die Möglichkeit, Hierarchien für die Anordnung der Tiers zu verwenden. Bei Analysen von miteinander interagierenden Personen werden meist mehrere Kameras benutzt, um alle Details festhalten zu können. Diese beinhalten immer auch identische Informationen, sodass in zwei ähnlichen Aufnahmen ähnliche Vorkommnisse annotiert werden. Ein Tool, das mit mehreren Videos arbeiten kann, erleichtert hierbei die Arbeit. Außerdem ist die Import- und Exportfunktionalität ein wichtiges Merkmal, da nicht jedes Tool die gleichen Fähigkeiten besitzt. Es ist aber durchaus gewünscht, alle Fähigkeiten oder auch automatische Analysen durch die verschiedenen Annotationstools bei der finalen Auswertung der Aufzeichnungen benutzen zu können. Die Erweiterbarkeit der Tools macht es Softwareentwicklern möglich, zusätzliche Funktionalitäten in das Tool zu integrieren, die für bestimmte Experimente benötigt werden. Eine Software, die als „OpenSource“ bezeichnet wird, kann prinzipiell erweitert werden, wobei zu berücksichtigen ist, wie deren Softwarearchitektur aufgebaut ist (z. B. kann eine Software wie ELAN, das ein Plugin-Interface besitzt, leicht erweitert werden). Diese Informationen sind in der **Tabelle 6** zusammengeführt. Ein wichtiger Punkt für den Erfolg von Software ist die Benutzbarkeit. Eine Software wird nicht erfolgreich sein, wenn sie zwar viele Funktionen bietet, aber schwierig zu bedienen ist. Daher sind weitere wichtige Punkte zu berücksichtigen wie die Benutzbarkeit der Tools [51]. Diese kann durch die Geschwindigkeit, mit der sich eine Person einarbeiten kann, gemessen

Name	Hierarchien	Multiple Videos	Import	Export	Open Source
Praat	Nein	Nein	Nein	Tabellen, Graphen	Ja
TASX	Nein	Ja	Praat, ANVIL	HTML, Praat	Ja
ANVIL	Ja	Nein	Praat	Tabelle, SPSS	Nein
EXMARaLDA	Nein	(Nein)	Praat, TASX, ELAN	TASX, ELAN, Praat	Nein
ELAN	Ja	Bis zu 4	Praat benutzbar	Text, Chat	Ja
MacVisSTA	Nein	Ja, mit Einschränkungen	Praat	MYSQL	Ja
C-Bas	Unbekannt	Nein	XML	XML	Nein
Theme	Ja	Nein	Observer XT 7.0-9.0	Observer XT 7.0-9.0	Nein
Transformer	Nein	Nein	Praat, TASX, ELAN	Praat, TASX, ELAN	Nein
FOLKER	Nein	Nein	GAT2	GAT2	Nein
Transcriber	Nein	Nein	Keine	Keine	Ja
Clan	Nein	Nein	Keine	Keine	Ja

Tabelle 6 Zusatzfunktionalität von Annotationstools in einer Übersicht

werden. Dazu wurden die Tools einer Usability Studie von anderen Entwicklern und Nutzern unterzogen. Dabei wurden vier zentrale Fragen untersucht wurden: Lerngeschwindigkeit, Annotationsgeschwindigkeit, Komplexität und Fehlerhäufigkeit [17]. Ein weiteres wichtiges Merkmal ist die Möglichkeit, Support von den Entwicklern bei Fragen oder auch Fehlern zu erhalten. Dabei spielt auch die Größe der Community um das Tool herum, die diese Hilfe leisten kann, eine wichtige Rolle. Ein weiterer Punkt ist die Weiterentwicklung der Tools und die damit verbundene Beseitigung von z. B. aufkommenden Problemen bei Inkompatibilitäten. Diese Fragen werden in der **Tabelle 7** beantwortet [17]. Die meisten Datenformate der

Name	**Benutzeranzahl**	**Lernschwierigkeit**	**Positiv bezüglich MMI Analyse**	**Negative bezüglich MMI Analyse**
Praat	ca. Xx100	Schwierig, Dokumentation	Audiosegmentierungsfunktionen	Ständig sich ändernde Benutzeroberfläche
TASX	Klein	Schwierig	Verschiedene Datenvisualisierung, viele unterstützte externe Datenformate	Keine Weiterentwickelung, keine Korpusverwaltung
ANVIL	ca. Xx100	Leicht - Normal, Dokumentation	Motion-Capturing, Sprachfrequenzen, Intensitäten	Keine Darstellung von Kopf- und Handorientierungen
EXMA-RaLDA	ca. Xx100	Leicht - Normal, Dokumentation	Transkriptionssysteme, Korpusmanagement	Umständlicher Wechsel zwischen Videos
ELAN	ca. Xx100	Leicht - Normal, Dokumentation	Sprachfrequenzen, Intensitäten, Video Annotation, reguläre Ausdrücke zum Suchen.	Keine Motion-Capturing-Unterstützung
MacVisSTA	Klein	Schwierig	Grafische Elemente „Motion traces“	Komplexe Kategorien, keine Möglichkeit zum Drucken
C-Bas	Klein	Unbekannt	Audio und Video Annotationstool basierend auf einem XML File-Format	Nicht mehr verfügbar
Theme	Klein	Normal	Mustererkennung auf Events	Kommerziell, keine automatischen Annotationen
Transformer	Klein	Normal	Gute Suche und Austausch mit anderen Annotationsprogrammen	Keine manuelle oder automatische Annotationsmöglichkeit
FOLKER	Klein	Leicht mit Dokumentation	Für Sprachanalysen der deutschen Sprache mittels Coding Schema	Keine Video Darstellung
Transcriber	Klein	Normal	Audio und multiple Video	Datenformat nicht XML
Clan [52]	Klein	Normal	Audio	Kein Video

Tabelle 7 Benutzung und Einflüsse basierend auf [17], [42] und [50]

Tools basieren auf einer XML-Struktur und dem „Annotation Graph Toolkit“ [53]. Die Wichtigkeit des Austauschs von Annotationsdaten zwischen verschiedenen Tools haben auch die verschiedenen Toolentwickler erkannt und diesbezüglich Maßnahmen ergriffen. Dazu wurde in den am meisten genutzten Tools, die am Anfang dieses Teilkapitels vorgestellt wurden, ein einheitliches Datenformat entwickelt, basierend auf dem Annotation-Graph-Framework. Die beteiligten Toolentwickler (ELAN, ANVIL, EXMARaLDA und Transformer) haben ihren Tools die Möglichkeit zum Datenaustausch durch Import- und Export-Funktionen gegeben [42].

3.2 Management von multimodalen Datenkollektionen

Neben dem Erstellen der Korpora, die bisher betrachtet wurde, spielt das Management von multimodalen Daten eine wichtige Rolle, um Analysen durchführen zu können. Dazu gehört, dass ein Korpus erstellt und zu einem späteren Zeitpunkt Informationen in diesem gefunden werden können, um Hypothesen zu prüfen. Unter diesem Gesichtspunkt werden daher EXMARaLDA und MexiCo noch einmal näher betrachtet.

3.2.1 EXMARaLDA

EXMARaLDA [15] beinhaltet ein Korpus-Management-Tool mit dem Namen „CoMa“. Es dient zur Verwaltung mehrerer Transkriptionen und ermöglicht es, Metadaten zu den jeweiligen Transkriptionen zu speichern. Damit ist es leicht möglich, Personen gebundene Metadaten, die in mehreren Transkriptionen auftreten, in Beziehung zueinander zu setzen. CoMa ermöglicht es, einen Korpus zu strukturieren, z. B. nach Interaktionen und beteiligten Sprechern, welche dann als eigenständige Einheiten einander zugeordnet werden können. Alle Einheiten können Attribute als Metadaten zugewiesen bekommen. Basierend auf diesen Attributen und den Transkriptionen kann eine Wortsuche mit Regular Expression[33] durchgeführt werden, um den gesamten Korpus nach bestimmten Vorkommnissen oder Eigenschaften zu durchsuchen.

3.2.2 MExiCo

MExiCo: „A Library for Managing Multimodal Data Collections“ wurde an der Universität Bielefeld von Peter Menke und Philipp Cimiano entwickelt [54]. Die Motivation für diese Entwicklung war es, ein Werkzeug zu schaffen, damit Daten aus verschiedenen interdisziplinären Forschungsprojekten einheitlich verwendet werden können. Die Daten, die aus verschiedenen Arbeitsgruppen unterschiedlicher Fakultäten kommen, sind meistens Audio-, Video- und Text-Annotationen [43]. Die Schwierigkeit ist, die Daten einheitlich zu behandeln, um Suchen über mehrere teils verschiedene Korpora zu ermöglichen. Da meist unterschiedli-

[33] Unter anderem Verknüpfungsausdrücke wie „and“ und „or“.

che Ziele[34] untersucht werden, sind die erstellten Daten (Annotationen) je nach Zielsetzung in verschiedenen Programmen und den dazugehörigen Datenformaten abgelegt. MExiCo ermöglicht es, Daten mit anderen Forschern und anderen Projekten auszutauschen. Dazu werden die Datenformate der gängigsten Programme als Import und Export Möglichkeit der Daten genutzt, um diesen Austausch zu ermöglichen[35]. MExiCo behandelt die Probleme, Daten von einem Format in ein anderes zu überführen, das z. B. weniger Strukturelemente (z. B. Hierarchien) besitzt. Die aktuell unterstützten Formate sind die Formate der Tools Praat, ELAN, ANVIL und eingeschränkt EXMARaLDA. Zusätzlich ist MExiCo eine ein Projekt begleitende Software, die bei der Planung, Organisation und Erstellung hilft. Dieses geschieht z. B. durch Verwaltung von Probanden, Variablen und Ressourcen, aber auch durch Funktionalitäten wie Checklisten. In der Nachbereitungsphase können fehlende Tiers in den Annotationsdaten ermittelt werden. Darüber hinaus ermöglicht MExiCo es, Metadaten zu speichern und Publikationen zu verwalten, die für den Korpus relevant sind. Eine ausführlichere Beschreibung ist in der Dissertation von Peter Menke zu finden [55].

3.3 Bewegungsklassifikation

Im Bereich des automatischen Annotierens im Allgemeinen, aber auch im Bezug zur MMI Verhaltensforschung gibt es eine Vielzahl von unterschiedlichen Forschungsarbeiten. Meistens sind diese für Spezialfälle geschrieben und daher nicht für allgemeinere Fälle verwendbar. Außerdem sind diese für normale Benutzer ohne große Spezialkenntnisse in der Softwareentwicklung schwierig zu nutzen. Trotzdem zeigen diese Arbeiten interessante und praktische Funktionalitäten, auch wenn sie teilweise noch nicht ganz ausgereift sind oder nur unter bestimmten Bedingungen funktionieren. Diese Arbeiten zeigen, dass diese entsprechenden Funktionalitäten prinzipiell funktionieren, aber noch nicht alltagstauglich sind. Die Hoffnung ist, dass diese Funktionalitäten bald in Standard Annotationstools, wie die im vorigen Teilkapitel genannten, verfügbar sein könnten. Hier wird ein kleiner Überblick gegeben, was heutzutage und später einmal machbar sein könnte. Allerdings zählen zu diesem Gebiet auch sehr viele Arbeiten, die hier nicht alle vorgestellt werden können. Daher geht der Überblick nur auf Arbeiten ein, die für das Thema Bewegungserkennung relevant sind. Auch wenn es in dieser Arbeit nicht direkt um Bewegungsklassifikation geht, wäre die Möglichkeit, bekannte Bewegungen zu finden, für ein Annotationstool sehr hilfreich, denn es sollen auch noch nicht bekannte Bewegungsphänomene gefunden werden. Wir beginnen mit einem kurzen Überblick über mögliche Verfahren, anschließend werden zwei diesbezügliche Arbeiten vorgestellt.

[34] Bei manchen Studien geht es nur um die gesprochene Sprache, bei anderen um Sprache in Kombination mit Gesten.

[35] Dazu wurden auch von der Community und deren Entwicklern der einzelnen Annotationstools Bemühungen getroffen, und es wurde ein gemeinsames Datenformat gewählt, welches durch Import und Export Funktionalität unterstützt wird [41].

3.3.1 Allgemein

Die meisten Arbeiten bezüglich Bewegungserkennung basieren auf Videodaten. Es gibt auch Arbeiten, bei denen Motion-Capture-Daten zur Wiedererkennung von Bewegung eingesetzt werden. Hintergrund zu diesem Thema ist die Detektion von Bewegungen, um Videos klassifizieren zu können und um den Zugriff auf diese leichter zu gestalten. Andere Einsatzgebiete sind, die Bewegung als Interaktion und zur Interaktionsanalyse bei der Verhaltensforschung zu nutzen. Im Prinzip geht es bei allen Arbeiten um eine ähnliche Funktionalität; bei den auf Videos basierten Daten besteht allerdings erst noch das Problem, eine Motion-Capture-Repräsentation der Videos zu berechnen [56]. Da die Videos, die klassifiziert oder indiziert werden sollen, meist nur aus einem Sichtwinkel aufgenommen wurden, muss zudem das Problem der Nicht-Eindeutigkeit mancher Posen gelöst werden [57]. Nicht alle Bewegungsklassifizierungssysteme arbeiten auf extrahierten Motion-Capture-Daten. Einige bildbasierte Bewegungserkennungsmethoden arbeiten mit Formen, Farben und auch Textur. Allgemein werden verschiedene „Maschine Learning“ Verfahren verwendet, um Bewegungen zu klassifizieren. Einige davon sind im Folgenden aufgelistet [58]:

- Probabilistische Klassifier wie z. B. das von Bayes, das die Wahrscheinlichkeit[36] $P(C_i|X)$ für jede Klassifikationsklasse C_i bei den gegebenem Featurevector[37] X berechnet [59].
- Artificial Neural Networks werden benutzt, um ein neuronales Netz mit Eingangs- und den dazugehörigen gewünschten Ausgabewerten zu trainieren [60].
- Support Vector Machines – SVM werden genutzt, um mehrere Klassen klar voneinander mittels einer K-Mean-Kernel-Funktion durch eine Hyperebene zu trennen [61].
- Decision Trees - das Problem wird in kleinere Teilprobleme zerlegt, z. B. in Torso, Arme und Beine [62].
- Template Matching kann verwendet werden, um die bildlichen Bewegungsänderungen einer Bewegung zu vergleichen [63].
- Nearest Neighbor – Suche nach der ähnlichsten Bewegung, basierend auf Bildänderungen [64].

Bei diesen Arbeiten werden vorher bekannte Standardbewegungen wie zum Beispiel Gehen, Rennen, Winken, Fangen, Springen, Sitzen und Stehen detektiert. Allerdings sind diese Verfahren nur dazu geeignet, bekannte Bewegungen wiederzufinden, nicht aber, wie es in der Verhaltensforschung üblich ist, nach unbekannten Verhaltensweisen zu suchen, die sich durch nicht bekannte Bewegungsmuster identifizieren lassen.

[36] Die Wahrscheinlichkeit, dass eine Bewegung auftritt unter verschiedenen Bedingungen, muss aus einer Studie oder Datenbank diesbezüglich erst extrahiert werden.

[37] Feature-Vektor stellt verschiedene Eingangsattribute dar.

3.3.2 Automatisches Annotieren von Alltagsbewegungen

Ein System, das Bewegungen automatisch auf der Basis von Video-Daten annotieren kann, würde viele manuelle Annotationen überflüssig machen. Deva Ramanan und D.A. Forsyth stellen ein solches System vor. Dieses System besteht aus 3 Kernkomponenten, die sie miteinander kombiniert haben, um brauchbare Ergebnisse zu erzielen. Die erste Komponente beinhaltet 3D-Motion-Capture-Daten, in der die zu erkennende Alltagsbewegung entsprechend annotiert vorliegt. Die zweite Komponente ist eine Softwarebibliothek, die aus Video-Daten zweidimensionale Bewegungsdaten erzeugt. Die dritte Komponente vergleicht die Alltagsbewegungen, die als 3D-Motion-Capture-Daten vorliegen, mit den zweidimensionalen Bewegungsdaten, die aus den Videodaten extrahiert worden sind. Die 3D-Motion-Capture-Daten der Alltagsbewegungsdaten sind 7 Minuten lang und bestehen aus Rennen, Gehen, Winken, Springen, Rechtsdrehen, Linksdrehen, Fangen, Ankommen, Tragen, Rückwärtsgehen, Hocken, Stehen und Aufheben. Dabei sind sogar mögliche Kombinationen wie Gehen und Fangen zu detektieren. Intern wird eine Bibliothek mit SVM Support-Vector-Maschinen [65] verwendet, um die Gelenkpositionen mit den Trajektorien über die Zeit als Erkennungsmerkmal zu nutzen [66]. In der **Abbildung 21** sind die Resultate zu sehen. Dabei gibt es zu jeder enthaltenen Bewegung einen zeitlichen Verlauf, der symbolisiert, ob das System diese als aktiv erachtet hat.

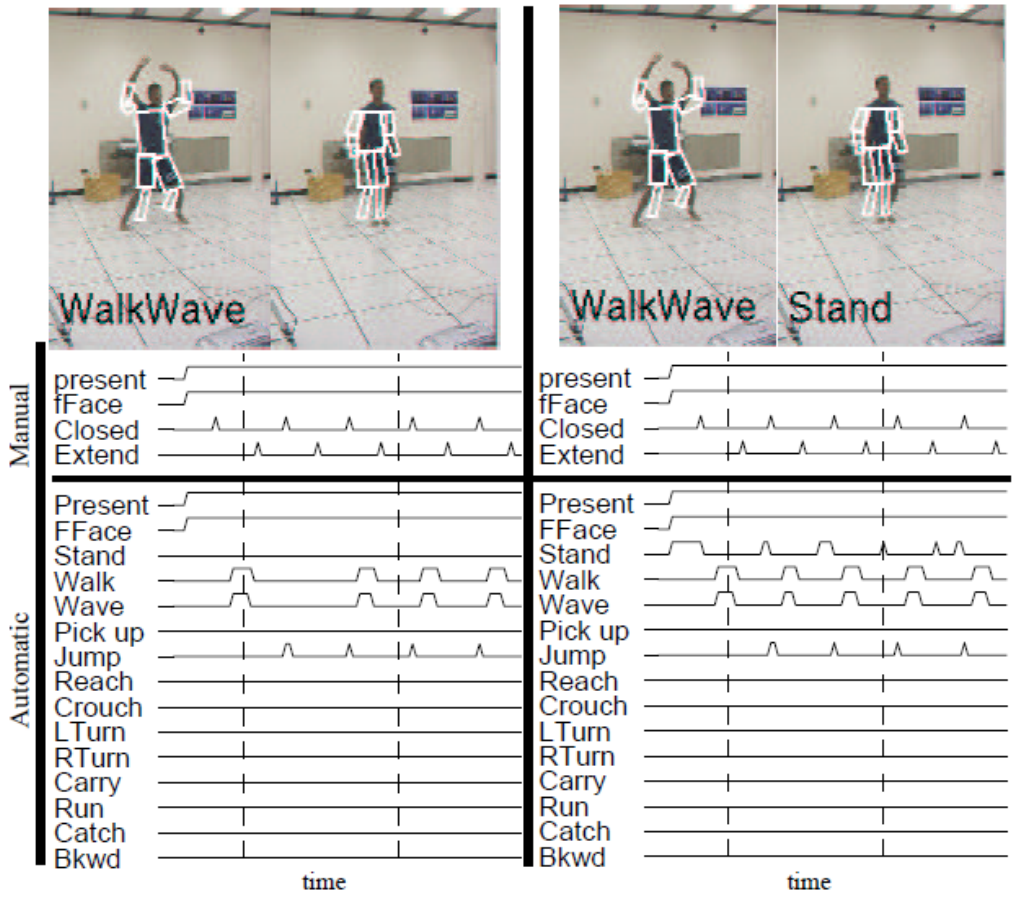

Abbildung 21 Resultat der automatischen Annotation von Alltagsbewegungen [66]

Um die Korrektheit des Systems zu ermitteln, sind im oberen Bereich der **Abbildung 21** manuelle Annotationen dargestellt, die den automatischen gegenübergestellt werden. Bezogen auf die Mensch-Maschine-Interaktion soll das System auch mehrere Menschen annotieren können.

3.3.3 Bewegungswiedererkennung

Körperbewegung wiederzuerkennen ist ein praktisches Einsatzgebiet zum Annotieren großer Mengen von Bewegungsdaten. Dabei gibt es verschiedene auf reinen Videodaten oder auf Motion-Capture-Daten basierende Versuche. Dabei können die auf Video basierenden Verfahren sehr praktisch auf verschiedene existierende Daten angewendet werden. Das Verfahren aus dem Paper „Efficient and Robust Annotation of Motion Capture Data“ [67] basiert auf einem „Motion Template“, das mit einer speziellen „Feature Vector“ Erkennung arbeitet[38] [68]. Dieser verwendete „Feature Vector“ hat 39 verschiedene Merkmale, dazu zählen Eigenschaften wie „z. B. Hände nach vorne bewegen“, „die räumliche Position von Körperteilen“, „schnelle Bewegungen einzelner Körperteile“, aber auch „Gelenkwinkel“. Die gesamte Liste ist unter [68] Tabelle 6 auf Seite 10 einzusehen. Das Verfahren verwendet positive und negative Beispiele, um eine Merkmale-Matrix zu erstellen (siehe **Abbildung 22**). Schwarze Felder beschreiben Merkmale, die verschiedene positive Beispiele gemeinsam haben, graue, die nur bei manchen gemeinsam auftreten und schließlich weiße, die keine Gemeinsamkeiten aufweisen. Anhand dieser Feature-Matrix können Schlüssel-Körper-Posen bestimmt werden, die als Basis für die Bewegungswiedererkennung dienen. Damit scheint es eine robuste Möglichkeit zu geben, um spezielle Bewegungsfolgen in Motion-Capture-Daten wiederzufinden.

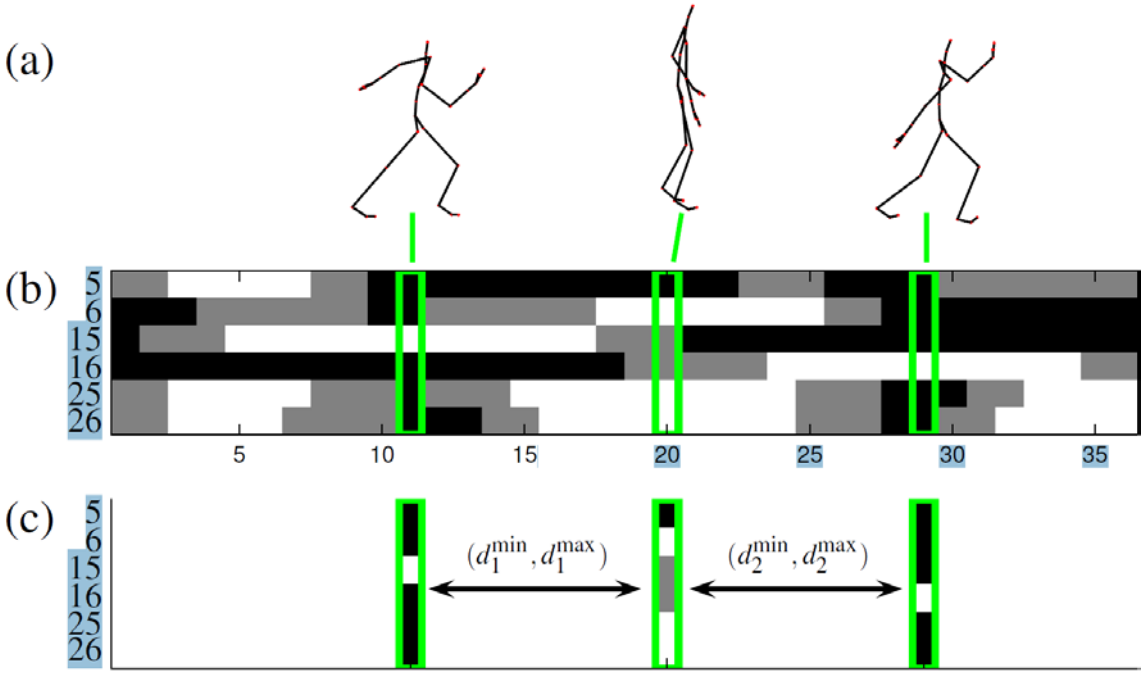

Abbildung 22 Bewegungsvergleich (a) Geh Bewegung (b) Merkmale-Matrix (c) Berechnete Key Frames als Pose zum Wiederfinden [67].

[38] Dabei stellt ein „Feature Vector“ eine Liste verschiedener Merkmale da.

3.4 Motion-Capturing basierte Forschung

In diesem Teilkapitel wird der relevante Teil der aktuellen Forschung bezogen auf das Motion-Capturing vorgestellt werden. Dazu zählt einmal die Richtung, Motion-Capture-Daten als Interaktionsinterface zu benutzen, und das Skeleton Fitting, in dem Marker zu einzelnen Körperteilen in Bezug gebracht werden sollen. Langfristing gesehen ist es Wünschenswert automatische Analysen in Echtzeit basierend auf Bewegungsdaten verwenden zu können um z.B. Roboter in Interaktionsszenarien Gesten richtig deuten zu lassen.

3.4.1 Motion als Interaktions-Interface

In der virtuellen Realität wird Bewegung nicht nur aufgezeichnet (um diese später darzustellen), sondern auch direkt als Interaktion mit der Umgebung verwendet und dient damit als Eingabegerät. Damit kann z. B. in einer virtuellen Werkstatt die Karosserie eines Autos entworfen und die Veränderungen an seiner Struktur können direkt in einem virtuellen Windkanal untersuchen werden. Eine andere Anwendung, die in dieses Feld der Interaktion fällt und möglichst viele Interaktionsparameter benötigt, ist die Interaktion mit virtuellen Agenten wie z. B. Max [26]. Diese Eingabe durch Posen und Bewegungen könnte durch das Wissen über die Körperhaltung auf verschiedene Gemütszustände des Interaktionspartners schließen lassen (Körpersprache). Je nach aktueller Gesprächssituation kann aus einer Körperhaltung, bei der die Arme verschlossen sind, darauf geschlossen werden, dass sich diese Person auch im Gespräch verschlossen gibt. Darüber hinaus lassen sich aus körperlichen Gesten weitere Informationen erschließen. Gesten alleine können verwendet werden, um mit anderen Menschen zu interagieren, die nicht dieselbe Sprache beherrschen, sodass oft grundlegende Bedeutungen verständlich gemacht werden können (z. B. auf etwas zeigen als „das da“). In dem Paper „Motion-Capture-Based Avatar Control Framework in Third-Person View Virtual Environments“ [69] wird ein Verfahren vorgestellt, um in einer Cave zu interagieren. Dieses ist ein begrenzter Raum, in dem nicht beliebig weit gegangen werden kann. Daher wird versucht, eine vorher definierte Folge von elementaren Bewegungen zu erkennen, um aus diesen auf komplexere Bewegungen schließen zu können. Diese komplexeren Bewegungen können anschließend als Interaktion in der virtuellen Umgebung dienen. Ein Beispiel solch einer Folge von elementaren Bewegungen ist die Gehbewegung, die in dieser Applikation definiert ist als abwechselnde Auf- und Abwärtsbewegung der Füße in Kombination mit einer jeweils gegenläufigen abwechselnden Vorwärts- und Rückwärtsbewegung der Arme. Dabei bewegt sich der Akteur auf der Stelle, und um zu lenken, verändert dieser kurzzeitig seine Orientierung in die gewünschte Richtung.

3.4.2 Skeleton-Fitting

Im Bereich der Filmindustrie werden beim Motion-Capturing mit optischen Tracking-Systemen viele einzelne Marker an einem Menschen befestigt, um dessen Bewegungen aufzuzeichnen. Diese erscheinen im System als Datenwolke von mehreren 3D-Positionen im

Raum, von denen die Zugehörigkeit zum Skelett des Akteurs unklar ist. Die Zuordnung zu einem Skelett wird als „Skeleton-Fitting“ bezeichnet. Normalerweise wird diese Zuordnung per Hand von einem Menschen durchgeführt. In dem Paper „Mapping optical motion capture data to skeletal motion using a physical model“ [70] wird ein Verfahren vorgestellt, dieses zu automatisieren. Dazu wird ein physisches Modell mit den Längen des Akteurs verwendet, um die Zuordnung der Marker zum Skelett durchzuführen. Dieses ist eine sehr nützliche Entwicklung, da heutzutage sehr viel Zeit aufgebracht werden muss, um nach der eigentlichen Aufnahme die Daten nachzubearbeiten, welche ca. 10-mal der Aufnahmezeit entsprechen kann. Bei der Film- und Spiele-Industrie ist dieses nicht so gravierend, da meist nur kurze Bewegungssequenzen aufgezeichnet werden und auf diese Weise viel Geld gespart werden kann gegenüber einer nicht so echt aussehenden Key-Frame-Animation. Allerdings kann bei der Analyse von Interaktionsverhalten diese Zeit nicht aufgebracht werden, da viele Personen stundenlang aufgezeichnet werden müssen. Problematisch ist nur, wie diese Algorithmen reagieren, wenn einzelne Marker fehlen, weil sich Interaktionspartner z. B. sehr nah kommen und dadurch Marker verdeckt oder zusammengefasst werden.

3.5 Zusammenfassung

In diesem Kapitel wurde ein Einblick in die aktuellen Probleme im Forschungsalltag bezogen auf Mensch-Mensch-Interaktionen gegeben und die damit verwandten Probleme des Annotierens betrachtet. Dazu wurden die aktuell existierenden Annotationstools mit ihren Möglichkeiten und Unterschieden betrachtet. Die aktuellen multimodalen Annotationstools wurden im Detail mit ihren Besonderheiten und Schwächen näher betrachtet, um anschließend diese untereinander zu vergleichen. Danach wurden einzelne automatische Annotationsmethoden vorgestellt. Diese Funktionalitäten sind leider noch nicht in den Annotationstools zu finden, da sie meist nur für einige spezielle Bedingungen entwickelt wurden und noch nicht im Allgemeinen funktionieren. Am Ende dieses Kapitels wurden kurz relevante Arbeiten aus dem Bereich des Motion-Capturings betrachtet.

3.6 Fazit

Die Grenzen der bisher existierenden Annotationstools sind die nicht oder nur sehr beschränkte Unterstützung des Motion-Capturings und der damit verbundenen Fähigkeiten zur automatischen Annotation. Existierende Tools bieten nicht die Möglichkeit zur Integration von Motion-Capturing verbunden mit automatischen Annotationen. Zum anderen, ausgehend von der Vielzahl allgemeiner Annotationen für spezifische Problemsituationen, wäre auch die Integration in eine Softwareumgebung wünschenswert, deren Funktionalität jedermann benutzen kann. Ebenso wäre eine Klärung des Umstandes wünschenswert, wie genau das Motion-Capturing zur Unterstützung des Annotierens eingesetzt werden kann. Gegenwärtig arbeiten die „Skeleton-Fitting“ Algorithmen noch nicht in zufriedenstellender robuster Weise, und sie haben Probleme mit dem temporären Verlust von Markern. Darüber hinaus stellt sich die Fra-

ge, in welcher Form das Motion-Capturing am effektivsten zur automatischen Annotation und zur Analyse von menschlichem Interaktionsverhalten genutzt werden kann.

4 Robustes Motion-Capturing mehrerer Personen über einen längeren Zeitraum

Ein grundlegender Bestandteil des Tools PAMOCAT ist die Möglichkeit, Motion-Capture-Daten über einen längeren Zeitraum mit wenig Arbeitsaufwand live aufzunehmen und zu visualisieren. Dabei stehen drei Dinge im Vordergrund. Zuerst die Möglichkeit, mehrere Personen aufzuzeichnen, sodass das Gruppenverhalten analysiert und die genaue Bewegung der einzelnen Personen zueinander betrachtet werden kann. Zweitens muss das Motion-Capturing robust sein, es dürfen möglichst keine Marker durch Verdeckung verschwinden. Als drittes muss die Motion-Capture-Aufnahme mit minimaler Vor- und Nachbearbeitung durchgeführt werden können, sodass nicht ein Vielfaches der Aufnahmezeit aufgewendet werden muss, um die Daten nutzbar zu machen.

4.1 Rigidbody basiertes Motion-Capturing

Normales optisches Motion-Capturing wird mit einzelnen Markern durchgeführt, die am Anfang der Aufnahme einzelnen Körperpositionen manuell zugeordnet werden müssen. Bei mehreren Personen ist dieser Vorgang etwas schwieriger, da alle Personen gleichzeitig im Aufnahmebereich sein müssen und die Marker von verschiedenen Personen falsch zugeordnet werden können. Der Prozess des „Labellings“ (das Zuordnen der Marker zu Körperteilen) bedeutet, dass Zeit für die Vorbereitung eingeplant werden muss, da er bei jeder Person einzeln durchgeführt werden muss. Um die einzelnen Marker überhaupt einzelnen Körperteilen zuordnen zu können, müssen die Akteure am Anfang der Aufnahmen die sogenannte T-Pose (Arme ausgestreckt) einnehmen. Das Einnehmen einer bestimmten Pose lenkt die Personen von einer natürlichen Interaktion ab und lenkt die Gedanken der Probanden darauf, dass sie sich in einem Experiment befinden. In der Praxis der orientierten Anwendungen des Motion-Capturings durch die Industrie werden mit dieser Technik einzelne kurze Aufnahmen durchgeführt, die später nachbearbeitet werden müssen, da die einzelnen Marker oft verloren gehen. Die Nachbearbeitung beträgt ca. das Zehnfache der Aufnahmezeit, wenn eine Person aufgenommen wird, da verloren gegangene Marker einzelner Personen und Körperteile jedes Mal neu zugeordnet werden müssen. Dabei ist die Nachbearbeitungszeit kein Problem, da sie relativ kurz im Verhältnis zur Arbeitszeit für das Erstellen einer entsprechenden Key-Frameanimation ist. Außerdem sieht die resultierende Animation mit echten Bewegungen realistischer aus. Bei Aufnahmen mit mehreren Personen wird dieses Wiederzuordnen der Marker zu den entsprechenden Körperteilen deutlich länger dauern, da zum einen mehrere Ghostmarker (nicht real existierende Marker) auftreten und von den realen Markern unterschieden werden müssen. Zum anderen aber auch, weil Verdeckung von Markern bei mehreren Personen häufiger auftritt. Da eine deutlich längere Vorbereitungs- und Nachbearbei-

tungszeit notwendig ist (geschätzt 50-mal die Aufnahmezeit), und da für die Verhaltensforschung viele Sequenzen in mehreren Gruppen mit längerer Aufnahmezeit benötigt werden, wurde in der vorliegenden Arbeit das Motion-Capturing mit Rigidbodys durchgeführt. Dafür existieren inzwischen kommerzielle Ganzkörper-Trackingsysteme, die allerdings nur bis zu zwei Personen gleichzeitig aufnehmen können. Andere kommerzielle Systeme sind nicht dafür ausgelegt, Personen mit Rigidbodys aufzunehmen, können aber viele Rigidbodys unterscheiden. Daher wird ein kommerzielles System verwendet, das mit einer eigenen Berechnung der Skelettwinkel viele Rigidbodys unterscheiden kann, um die Posen der jeweiligen Probanden zu bestimmen. Dieses hat den Vorteil, dass einmal drei Personen aufgezeichnet werden können, die kinematischen Beschreibungen der Posen vorliegen und darauf basierend auch verschiedene Analysen einfach integriert werden können. Die Entscheidung, die Berechnung der Skelettwinkel selber durchzuführen, basiert darauf, dass eine eigene Motion-Capturing-Software mit Bewegungsvisualisierung für eine Person eingebettet in einer GUI bereits zur Verfügung stand [71], wodurch eine erhebliche Zeitersparnis bei der Integration von Motion-Capture-Daten und der Berechnung der inversen Kinematik mit einer Darstellung der Bewegung von mehreren Personen möglich war. Zusätzlich konnte die Kinematik den gewünschten Freiheitsgraden angepasst werden, und es konnte eine leichte, auf der inversen Kinematik basierende Anbindung verschiedener Analysen implementiert werden.

4.2 Rigidbodys

Die Technik der Rigidbodys stammt aus der Echtzeit-Interaktion, wie sie in großen virtuellen Reality-Anlagen wie einer Cave[39] oder einer Powerwall eingesetzt wird. Zum Zeitpunkt der Datenerhebung existierten keine Systeme mit Rigidbodys, die drei Personen gleichzeitig tracken[40] konnten. Daher musste eine eigene Konstruktion von Rigidbodys zur Aufzeichnung verwendet werden. Ein Rigidbody ist ein Muster von einzelnen Markern (infrarot reflektierenden Kugeln) im dreidimensionalen Raum. Durch die Rigidbodys sind die eindeutigen Positionen und Ausrichtungen einzelner Körperteile der verschiedenen Personen erkennbar. Der Vorteil gegenüber einzelnen Markern ist einerseits, dass die Rigidbodys fest einem Körperteil und einer Person zugeordnet werden können. Dadurch müssen nicht erst einzelne Marker zu Personen und Körperteilen manuell zugeordnet werden, auch nach einem möglichen Verschwinden und Wiederauftauchen ist die Position am Körper der zugehörigen Person des Rigidbodys automatisch bekannt. Dieses spart Zeit, da keine Nachbearbeitung der Aufnahmen nötig ist. Zusätzlich hat es den Vorteil, dass ein Rigidbody bestehend aus mehreren Markern nicht so leicht verdeckt wird bzw. verloren geht, da meist nicht alle Marker, aus denen ein Rigidbody besteht, gleichzeitig verdeckt sind. Die Position ist zumindest immer noch ermittelbar, wenn nur ein Marker und die Orientierung, wenn 3 Marker sichtbar sind. Außerdem

[39] Eine Cave besteht aus 3 bis 6 Wänden (4x Seiten, Boden und Decke), auf die ein dreidimensionales Bild projiziert wird.

[40] Bewegung der Personen aufzeichnen.

müssen die Probanden keine spezielle Körperpose einnehmen, die ihnen noch mehr bewusst macht, dass sie gefilmt werden. Diese Rigidbodys müssen eine hohe Anzahl an Variationen zulassen, und die einzelnen Muster durch Markeranordnungen müssen sich stark voneinander unterscheiden, damit sie stabil detektiert werden können. Dadurch wird verhindert, dass gegebenenfalls einzelne Körperteile vertauscht werden. Nach verschiedenen Versuchen wurde ein Rigidbody-Modell gebaut, das aus fünf einzelnen Markern besteht. Um eine große Anzahl von Variationen zu erhalten, sind die Kugeln auf verschieden langen Stäben (1, 3, 5, 7 cm) fixiert, die wiederum auf einer Platte mit einem Rastermuster befestigt werden. Dabei gibt es zwei verschiedene Größen, einmal 7x7 cm und 10x10 cm, damit die Rigidbodys möglichst wenig stören. Da das verwendete Trackingsystem der Firma Vicon das Ursprungskoordinatensystem und die Ausrichtung eines Rigidbodys im 3D-Raum durch die ersten drei Markerpositionen definiert, sind alle Rigidbodys spezifisch aufgebaut. Dazu sind die ersten drei Marker immer auf derselben Höhe und stehen in einer rechtwinkligen Anordnung[41] zueinander. Durch diese Aufbauweise wird erreicht, dass das Koordinatensystem rechtwinklig zur Basisplatte ausgerichtet ist. Es gibt so nur eine translatorische Verschiebungskomponente des ersten Markers zum Mittelpunkt auf der Oberfläche der Basisplatte, der leicht durch Abmessung einmalig definiert werden kann (siehe **Abbildung 23**). Um die Anordnung der einzelnen Kugeln zueinander möglichst unterschiedlich gestalten zu können, wurde eine Software geschrieben, welche verschiedene Kombinationen von Marker-Mustern mit einer möglichst großen Variation unter Einhaltung verschiedener Kriterien berechnet. Die berücksichtigten Regeln sind:

- Rechtwinklige Anordnung der ersten drei Marker auf gleicher Höhe.
- Die rechtwinklige Anordnung darf nicht von der Höhe 1 cm sein, da die einzelnen Marker dann schlecht von der Seite zu sehen wären.
- Die rechtwinklige Anordnung darf nicht auf der maximalen Höhe von 7 cm liegen, da der Rigidbody nicht zu ausladend wirken soll, um die Versuchspersonen nicht abzulenken.
- Der seitliche Abstand der Marker zueinander soll mindestens eine freie Rasterposition betragen.
- Es können keine zwei Stangen an derselben Position auf dem Raster platziert werden.
- Die Marker, die nicht zu der quadratischen Anordnung gehören, müssen auf einer unterschiedlichen Höhe zueinander und zur quadratischen Anordnung sein, damit eine möglichst große Variation im Dreidimensionalen entsteht.
- Die Position aller Kugeln zueinander darf nicht durch Rotation und Translation einer anderen Anordnung von Kugeln zueinander entstehen.

Ein Beispiel-Rigidbody, der nach diesem Schema und nach den beschriebenen Regeln gebaut wurde, ist in der **Abbildung 23** dargestellt. In dieser Abbildung ist ein kleiner Rigidbody mit

[41] Zwei Vektoren zu einem Marker x und y, die vom Ursprung O Marker ausgehen, stehen in einem Winkel von 90° zueinander.

den Ausmaßen 7x7 cm abgebildet. Die quadratische Anordnung befindet sich auf einer Höhe von 3 cm, ein Variationsmarker liegt in der Höhe 1 cm und einer in der Höhe 5 cm. Dabei ist in der Abbildung (a) das Koordinatensystem im ersten Marker, und die X- und Y-Achsen sind von diesem aus durch den zweiten und dritten Marker definiert. Die Position des ersten Markers zum Mittelpunkt des Rigidbodys, bezogen auf alle anderen Rigidbodys, ist unterschiedlich. Daher muss jeweils der Abstand des ersten Markers entsprechend der X-, Y- und Z-Achse zum Mittelpunkt des Rigidbodys festgehalten werden. Damit kann die mittlere Position des jeweiligen Rigidbodys durch diesen Offset (Versatz) berechnet werden.

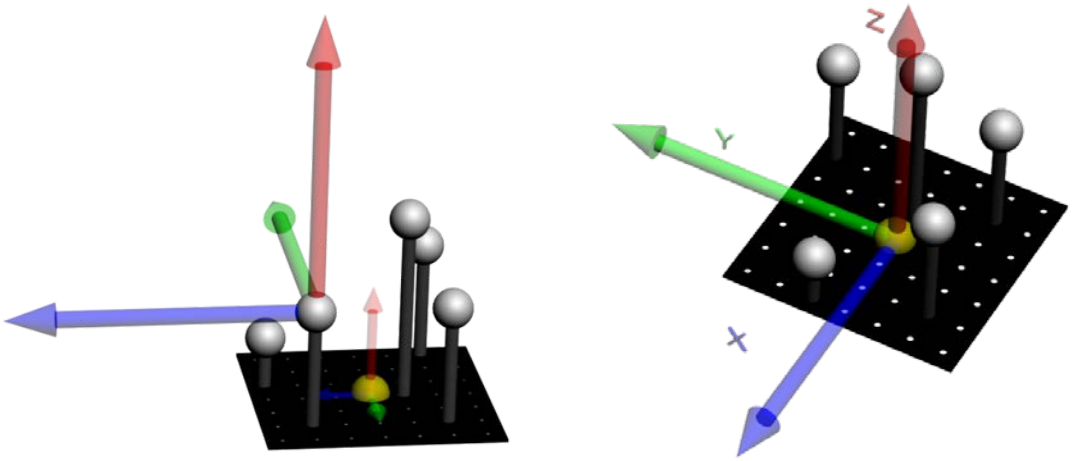

Abbildung 23 Beispiel eines Rigidbody-Designs für eine möglichst große Variation (a) Koordinatensystem im ersten Marker (b) Koordinatensystem im Mittelpunkt des Rigidbodys.

4.3 Positionierung der Rigidbodys am Körper

In den Szenarien für die Erstellung der Korpora (siehe dazu Kapitel 5), bei denen die Aufnahmen gemacht wurden, sitzen die Probanden auf Stühlen, währenddessen sie reden und gestikulieren. Bei dem später im Detail vorgestellten Korpus „Obersee“ [72] sitzen drei Leute um einen Tisch herum, sodass es schwierig ist, die Bewegung des gesamten Körpers aufzunehmen, da der Tisch den unteren Körperteil verdeckt. Bei dem zweiten Korpus „Sagaland“ sitzen drei Probanden in einem Kreis. Generell ist es kein Problem, die Bewegung des gesamten Körpers aufzunehmen, nur sind dazu mehr Rigidbodys nötig. Und eine höhere Anzahl an Rigidbodys bedeutet auch, dass mehrere ähnliche Muster auftreten, die sich entsprechend wenig unterscheiden. Daher ist es ratsam, immer nur die wirklich benötigte Anzahl an Rigidbodys zu verwenden, damit diese möglichst robust und stabil erkannt werden können. Die Positionen, an denen die Rigidbodys befestigt werden, sind unter den Merkmalen der guten Sichtbarkeit und der möglichst geringen Verdeckbarkeit (durch den eigenen Körper) gewählt worden. Das Wichtigste allerdings ist, dass die Körperposen berechnet werden können und der damit verbundene Arbeitsaufwand möglichst gering ist. Durch dieses Design ist eine schnelle Abmessung der Rigidbodys zu den Gelenken möglichst entlang einer einzigen Achse mög-

lich[42]. Um die Bewegung des Oberkörpers mittels der Rigidbodys aufzeichnen zu können, müssen diese daran befestigt werden. Dazu werden flexible T-Shirts aus dem Laufsport verwendet, an denen Klettverschlussauflagen an der Rückseite der Rigidbodys angenäht bzw. angeklebt wurden. Um die Rigidbodys an den Ellenbogen zu befestigen, wurden Ellenbogenschoner (bei denen die Plastikpanzerung entfernt wurde, um nicht die Bewegungsfreiheit einzuschränken) ebenfalls mit einem Klettverschluss versehen. An den Händen werden Fahrradhandschuhe ohne Fingerspitzen mit einem Klettverschluss auf den Handrücken zur Befestigung des Rigidbodys verwendet. Diese Arbeiten wurden im Rahmen des Papers von Karola Pitch et al. „Linking Conversation Analysis and Motion Capturing“ [72] durchgeführt. Die Befestigung des Rigidbodys am Kopf wurde zuerst durch einen Hut mit Klettverschluss bewerkstelligt (siehe dazu **Abbildung 24**), es hat sich aber als besser erwiesen, einen Haarreifen dafür zu verwenden, da dieser genauer an den Kopf ausgerichtet werden kann und stabiler sitzt. Auf Wunsch der Probanden werden die älteren Hüte verwendet, da es unangenehm sein kann, diese Haarreifen mit keinen oder mit wenig Haaren zu tragen. Mit dieser Konfiguration aus Kopf-, Schulter-, Rücken-, Ellenbogen- und Hand-Rigidbodys sind zusätzlich zu früheren Arbeiten [73] auch die Bewegungen der Schultern mit ermittelbar. Das hier vorgestellte System ist auf Robustheit ausgelegt, daher sind die Rigidbodys an Körperstellen platziert worden, die möglichst immer gut sichtbar sind, im Vergleich zum kommerziellen System der Firma ART[43], bei dem die Kugeln sehr nahe am Körper und an den Körpergliedern angebracht werden. Außerdem ist eine Berechnung der Skelettposen, die möglichst ohne Orientierung einzelner schlechter sichtbarer Rigidbodys auskommt, viel stabiler als ein System, welches bei Verlust der Orientierung komplett falsche Körperposen berechnet.

Abbildung 24 (a) Positionierung der alten 2D Rigidbodys am Körper (b) überarbeitete 3D-Rigidbodys am Körper (c) Rigidbodys mit T-Shirt, Handschuhen, Ellenbogenbefestigung und Hut (wurde ersetzt durch Haarreifen)

[42] Es muss kein Rotationsoffset mitbestimmt werden zu den Gelenken.

[43] Das kommerzielle Trackingsystem der Firma ART kann aktuell nur die Bewegung von 2 Personen gleichzeitig aufzeichnen und war zur damaligen Zeit nicht als Motion-Capture-System verfügbar.

4.4 Aufbau des Studiensetups

Ziel ist es, Motion-Capture-Aufnahmen in Verbindung mit Videoaufnahmen aller einzelnen Personen zu erhalten, um das Interaktionsverhalten analysieren zu können. Daher werden alle Details (Fingerbewegung, Gesichtsmimik usw.) der Probanden festgehalten, die nicht in den Motion-Capture-Aufnahmen sichtbar sind. Damit die Motion-Capture-Aufnahmen durchgeführt werden können, muss ein komplexes Setup entsprechend den aktuellen Anforderungen aufgebaut werden. Das hier verwendete Motion-Capture System ist von der Firma Vicon und wird mit mindestens 10 oder mehr Vicon T20 Infrarotkameras und einem bzw. zwei Vicon MX Gigant-Servern[44] verwendet. Diese sind in einem Kreis möglichst weit (entsprechend den Raumgegebenheiten) vom Aufnahmebereich höhenvariiert montiert. Die Vicon MX Gigant-Server sind wiederum mit einem Windows-Rechner mit acht Kernen verbunden, über den das gesamte Motion-Capture-System gesteuert wird. Dieser ist zudem für die Mustererkennung (die Anordnung der einzelnen Marker) bezüglich der Rigidbodys verantwortlich. Die erkannten Positionen und Ausrichtungen der Rigidbodys werden von diesem Windows Rechner über ein Netzwerk zu einem anderen Rechner[45] verschickt. Dieser Rechner ist für das Anzeigen und Aufzeichnen der Daten verantwortlich. Um die Videodaten der einzelnen Personen mit Gesichtsausdrücken aufnehmen zu können, wurden zusätzlich drei einzelne HD Kameras aufgestellt, die frontal auf jede Person ausgerichtet sind.

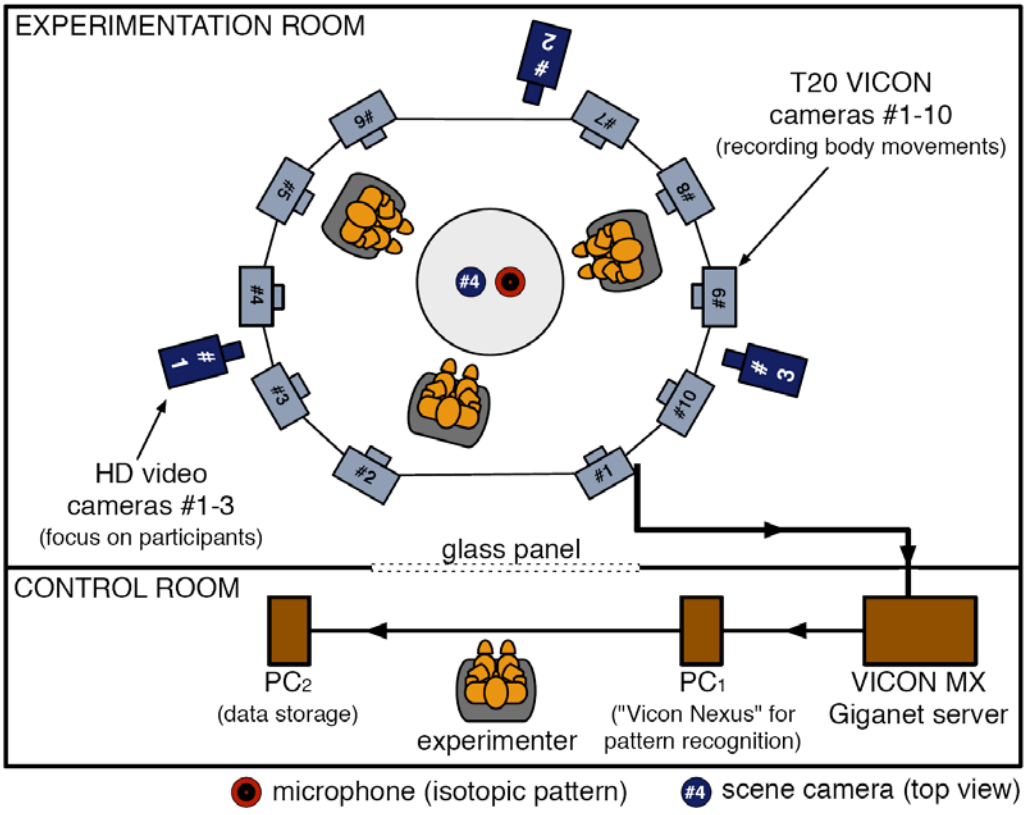

Abbildung 25 Studiensetup bei der Erstellung des Obersee Korpus [72]

[44] Ab 10 verwendeten Kameras muss ein weiterer Vicon MX Giganet Server verwendet werden.

[45] Auf diesem Rechner, der für das Speichern der Daten zuständig ist, arbeitet die Software namens PAMOCAT, die hier später detaillierter vorgestellt wird.

Zusätzlich ist eine vierte HD Kamera von der Decke herab auf den Aufnahmebereich ausgerichtet, um das Interaktionsverhalten zwischen den verschiedenen Probanden mit ihren Gesprächspartnern festzuhalten. Bei den neueren Korpusaufzeichnungen „Sagaland“ wurden miteinander synchronisierte Kameras verwendet, wodurch das spätere Zusammenführen der gesamten Daten erleichtert wird. Um ein gutes Audiosignal aufzuzeichnen zu können, wurde von der Decke ein zusätzliches Mikrofon zu den vier Mikrofonen der HD Kameras installiert (bei der Erstellung des „Sagaland“ Korpus wurden Headsetmikrophone verwendet). Die genaue Positionierung der einzelnen Geräte ist in der **Abbildung 25** dargestellt. Die Blickrichtung der Vicon Kameras ist von schräg oben herab, damit bei Bewegung der Probanden möglichst wenig durch andere Probanden verdeckt werden kann. Die teilweise laute Aufnahmehardware liegt hinter einer Wand mit einem Glasfenster, damit die Audioaufnahmen nicht beeinträchtigt werden.

4.5 Aufnahmevorbereitung und Nachbereitungen

Da sehr viel technisches Equipment bedient werden muss und dabei kein Zwischenschritt vergessen werden darf, ist es nötig, eine Liste mit Aufgaben abzuarbeiten und abzuhaken. Eine einzelne Person ist dabei mit der Bedienung überfordert, da nicht nur das technische

Nr.	Aufgabe	Erledigt
0	Motion-Capture-System kalibrieren.	
1	Probanden mit T-Shirts versehen.	
2	Probanden mit Rigidbodys bestücken.	
3	Experimentiererläuterung geben.	
4	Einverständniserklärung einholen.	
5	Abmessungen der Rigidbodys zu den Gelenkpositionen durchführen.	
6	Alle Probanden nacheinander in den Aufnahmebereich bringen.	
7	Im Tool PAMOCAT mittels einer Skelettdarstellung prüfen, ob die Rigidbodys korrekt platziert wurden.	
8	Kameras 1 bis 4 in den Aufnahmemodus bringen.	
9	Separate Audioaufnahme starten.	
10	Motion-Capture-Aufnahme im Tool PAMOCAT starten.	
11	Motion-Capture-Video-Klappe im Mittelpunkt des Aufnahmebereiches zusammenklappen, sodass sie von allen vier Kameras sichtbar ist.	
12	Rigidbodys und T-Shirts von den Probanden entfernen.	
13	Probanden Fragebogen zum Experiment ausfüllen lassen.	
14	Sichern der Daten auf einem externen Speichermedium.	

Tabelle 8 Arbeitsschritte zur Durchführung einer Motion-Capture-Aufnahme

Equipment bedient werden muss, sondern auch die Versuchspersonen vorbereitet und deren Fragen beantwortet werden müssen. Daher wurde eine Checkliste (siehe **Tabelle 8**) erstellt, bei der alle Schritte nacheinander abgearbeitet werden sollen, um sicherzustellen, dass kein Punkt der Checkliste vergessen wird. Um die Motion-Capture-Aufnahmen mit den multiplen Video- und Audio-Aufnahmen synchronisieren zu können, wird eine Filmklappe verwendet, die zusätzlich mit Markern versehen ist. Wenn die Klappe zusammengeklappt ist, repräsentiert dies eine Markeranordnung eines definierten Rigidbodys, sodass in den Motion-Capture-Daten automatisch die Startposition der Aufnahme ermittelbar ist, wenn das erste Mal dieser Rigidbody auftaucht. Die modifizierte Filmklappe mit Markern ist in der **Abbildung 26** dargestellt. Die Kalibrierung der Motion-Capture-Kameras zueinander

Abbildung 26 Motion-Capture-Video-Synchronisationsklappe (a) offen (b) zugeklappt Markerklappe, die in dieser Anordnung einen Rigidbody definiert.

muss nicht jedes Mal neu durchgeführt werden, falls sichergestellt ist, dass keiner der Probanden gegen eine Kamera gestoßen ist und diese somit bewegt wurde. Nach der Aufnahme müssen die Daten von den verschiedenen Geräten zusammenkopiert und synchronisiert werden. Um die Daten zu synchronisieren, müssen alle Videos den gleichen Anfangszeitpunkt haben[46]. Dazu müssen die Videos so zurechtgeschnitten werden, dass der exakte Zeitpunkt, bei dem die Klappe zusammengeklappt war, am Anfang liegt. Diese Synchronisation ist durch das Audio- und Videosignal zu dem Zeitpunkt, wann die Klappe zusammenklappt, möglich. Da man aus linguistischer Sicht am liebsten auch alle Videoaufnahmen zur späteren Analyse verwenden können möchte, kann ein gemeinsamer Zeitpunkt davor ausgewählt werden. Dabei muss die Zeitdifferenz vom Anfang bis zum Zeitpunkt, wann die Klappe zugeklappt wurde, bei allen Videos gleich sein und im Projektkonfigurationsmodus von PAMOCAT gespeichert werden. Mit dieser Referenzenzeit können die Motion-Capture-Daten synchron zu den Videos bzw. Audiodaten gehalten werden.

[46] Bei den späteren Aufnahmen muss nur eine zusätzliche Webcam mit den zueinander synchronen Kameras angeglichen werden.

4.6 Berechnung der Skelettposen durch die Durchführung der inversen Kinematik

Um später die Posen der verschiedenen Probanden durch ein Skelett darstellen zu können, müssen die Winkel aller Gelenke berechnet werden. Mittels dieser Winkel kann auch später die Bewegung im Detail analysiert werden. Dazu muss zunächst ein Skelett definiert werden [71].

4.6.1 Beschreibung des Skeletts

Die Denavit-Hartenberg-Konvention [22], vorgestellt in Abschnitt 2.1.3, beschreibt die Beziehung zwischen zwei Gelenken. Mit dieser DH-Konvention und den daraus resultierenden DH-Parametern können Skelette für komplexe Roboter und Probanden mathematisch beschrieben werden. Zu diesem Zweck wird zunächst ein einzelner Arm, der in symbolischer Gelenkdarstellungsform in **Abbildung 27** dargestellt ist, mit den DH-Parametern beschrieben. Später folgt die gesamte Darstellung eines kompletten Skeletts in dieser Form.

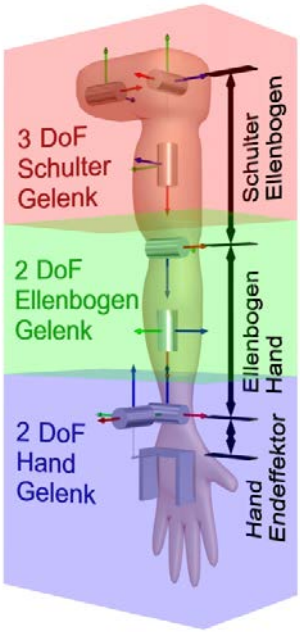

Abbildung 27 Beschreibung eines Armes in der DH-Konvention

In der Abbildung wird das Schultergelenk durch drei einzelne Gelenke, die jeweils um 90° um die x-Achsen gedreht sind, mathematisch beschrieben und alle zugehörigen z-Achsen schneiden sich in einem Punkt. Nur das letzte Schultergelenk hat eine Länge, nämlich die Länge des Oberarms von der Schulter zum Ellenbogen. Das Ellenbogengelenk, welches zwei Freiheitsgrade hat, wird durch einen Freiheitsgrad als Ellenbogengelenk und einen weiteren

Freiheitsgrad zusammen in der Hand beschrieben. Dadurch wird die Hand durch ein Gelenk mit drei Freiheitsgeraden beschrieben[47].

Analog zum einzelnen Arm lässt sich der gesamte Körper beschreiben. In der Hierarchie vor dem Arm liegen drei Gelenke, welche die gesamte Ausrichtung des Körpers definieren. Zudem liegen vor dem Schultergelenk zwei weitere Gelenke, welche es erlauben, die Schulter anzuheben und nach vorne zu bewegen. Diese sind der Übersichtlichkeit halber nicht in der **Abbildung 28** dargestellt. Die Beine sind fast identisch zu den Armen aufgebaut, haben allerdings einen Freiheitsgrad weniger. Im Kniegelenk gibt es keine zwei Freiheitsgerade wie beim Ellenbogengelenk. Die Ausrichtung des Kopfes kann durch drei einzelne Gelenke ähnlich zu der Anordnung des Schultergelenkes beschrieben werden. Die einzelnen DH-Parameter für einen Arm sind in der folgenden **Abbildung 28** aufgeführt. Dabei sind dort

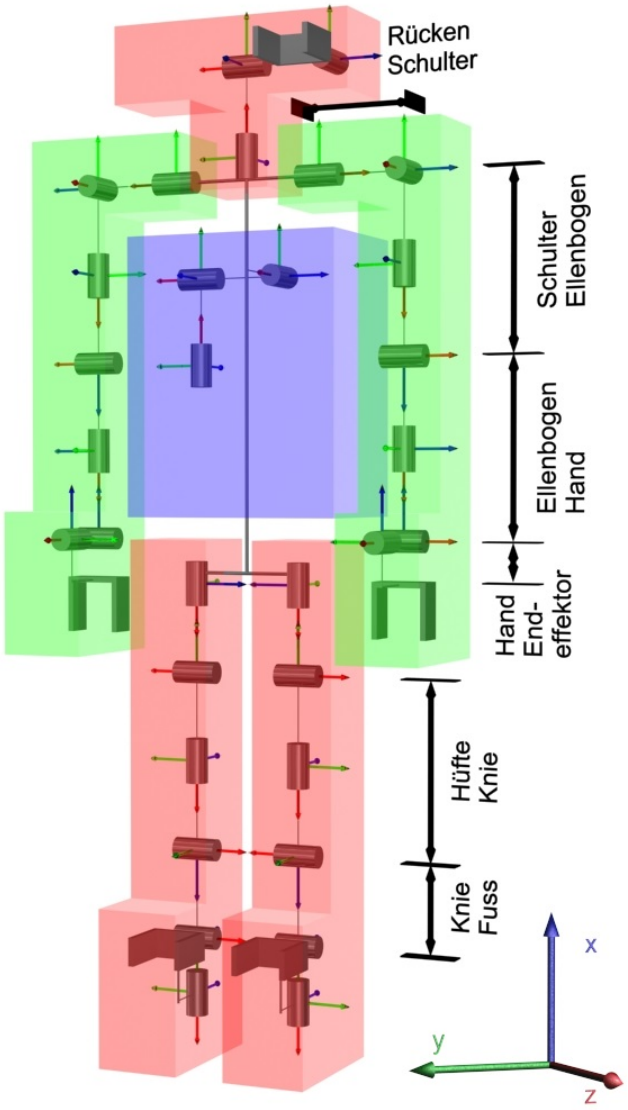

Abbildung 28 Beschreibung eines kompletten Skeletts in der DH-Konvention

[47] Trotzdem hat das Handgelenk eigentlich aus Sicht der Gelenke nur zwei Freiheitsgrade, das dritte Gelenk ist am Ellenbogen, welches die Hand um sich selbst drehen lässt.

Gelenkname	**übergeordnetes Gelenk**	α_i	θ_i	d_i	a_i
Orientierung (Yaw)	Root	0	0	0	0
seitlich (Roll)	Orientierung	90	-90	0	0
aufrecht (Pitch)	seitlich	90	90	0	0
Blattgelenk1	aufrecht	-90	-90	40*	0
Blattgelenk2	Blattgelenk1	-90	0	140*	0
Schultergelenk1	Blattgelenk2	90	90	0	0
Schultergelenk2	Schultergelenk1	90	0	0	0
Schultergelenk3	Schultergelenk2	-90	180	0	300*
Ellenbogengelenk	Schultergelenk3	90	90	320*	0
Handgelenk1	Ellenbogengelenk	0	0	0	0
Handgelenk2	Handgelenk1	90	-90	0	0
Handgelenk3	Handgelenk2	-90	-90	0	0

Tabelle 9 Auszug der DH-Parameter für die Beschreibung eines Armes aus den 27 Gelenken in der Oberkörperkonfiguration (von 41 in der Ganzkörperkonfiguration), dabei sind Winkel in Grad und Distanzen in mm angegeben.

nicht alle DH-Parameter für das gesamte Skelett aufgeführt, da sich beide Arme zueinander und auch zu den Beinen nicht viel unterscheiden. Der Unterschied besteht in einem Vorzeichen im Schultergelenk oder im Hüftgelenk. In der **Tabelle 9** sind Längenwerte nur als Beispiel eingetragen und hängen im Einzelfall von den jeweiligen Probanden ab.

4.6.2 Berechnung der Winkel

Auf Grund der mathematischen Beschreibung des Skeletts kann die Vorwärtskinematik verwendet werden, um die Winkelstellungen der Gelenke zu berechnen (Inverse Kinematik). Dieses ist ein Vorgang, der bei dem Wurzelgelenk anfängt und immer weiter bis zu den letzten Gelenken, den sogenannten Blattgelenken, durchgeführt wird. Dabei werden die zuvor berechneten Winkel auf das Skelett übertragen und genutzt, um die nächsten Winkel (iterativ) auszurechnen.

Durch die Orientierung des Rigidbodys am Rücken (siehe **Abbildung 24**) und mit Hilfe der Winkelextraktion im Anhang Anhang Akönnen die ersten 3 Gelenkstellungen ausgerechnet und auf das Skelett übertragen werden. Anschließend folgt der iterative Vorgang, der ausgehend von der Wurzel[48] alle Winkel der Gelenke nacheinander bis zu den Endeffektoren (Hände und Füße) ausrechnet. Dieser Vorgang wird im Detail nun am Beispiel des Schultergelenks verdeutlicht; dabei sind die Positionen der Schulter, des Ellenbogens und die Orientierung im Gelenk vor dem Schultergelenk bekannt. Dieser Vorgang wurde dabei schon für die beiden Gelenke am Schulterblatt, die vor dem Schultergelenk liegen, durchgeführt.

Im ersten Schritt müssen die Gelenkpositionen in globalen Koordinaten berechnet werden. Positionen wie die des Rückenzentrums, Schultergelenks, Ellenbogengelenks und Handgelenks werden durch eine Translation bezogen auf die Orientierung und Position der einzelnen

[48] Erstes Gelenk in einer Hierarchie.

Rigidbodys ermittelt. Diese müssen manuell im Vorfeld für jede Person ausgemessen werden, z. B. die Distanz des Rigidbodys am Ellenbogen zum Ellenbogengelenkzentrum.

Explizit wird dieses durch die Multiplikation der homogenen Matrix des Rigidbody am Ellenbogen und des Offsets berechnet, welche die Abmessungen entsprechend den Achsen enthalten. Anstatt dieser manuell abgemessenen Offsets können Durchschnittswerte verwendet werden um den Arbeitsaufwand zu reduzieren. Eine bestmögliche Genauigkeit ist allerdings nur mit diese manuellen abgemessen zu erziehlen.

$$Matrix_{Rigidbody} \times P_{Offset} = P_{Gelenk} \tag{7}$$

Das Resultat ist die Position des Gelenks in globalen Koordinaten des Ellenbogens ausgehend vom Rigidbody des Ellenbogens und den manuellen Abmessungen siehe **Abbildung 29**.

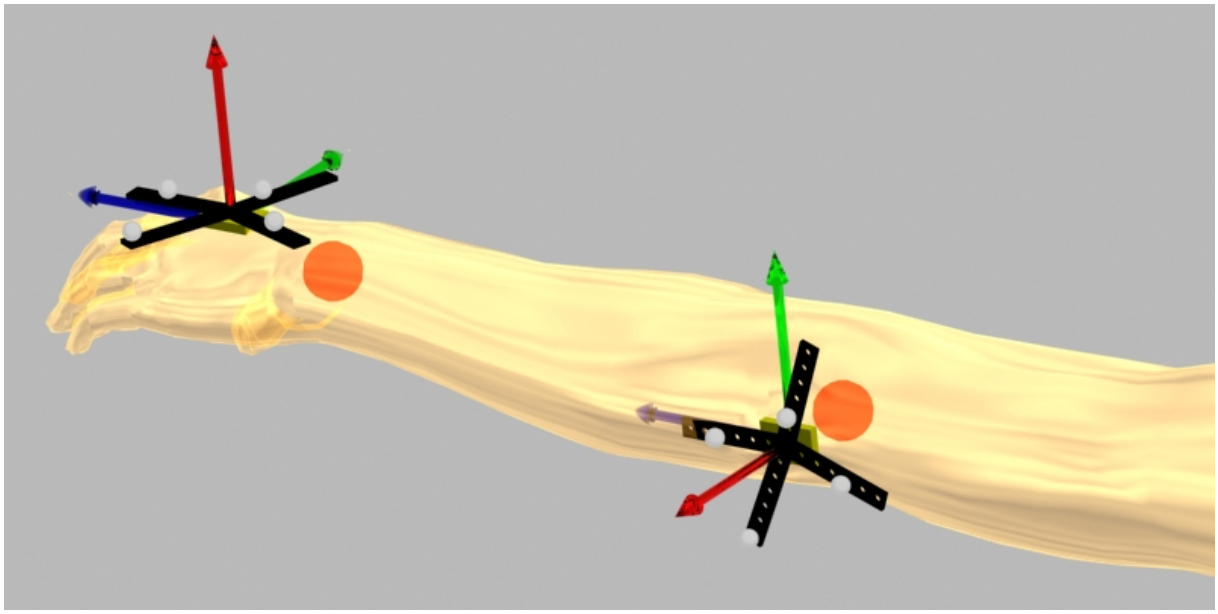

Abbildung 29 Gelenkpositionen dargestellt durch rote Kugeln im Inneren des Arms im Verhältnis zu den Rigidbodys (alte Darstellung nach Vorlage von ART [31])

Der zweite Schritt ist die Berechnung der lokalen Positionen aus Sicht des Skelett-Koordinatensystems im ersten Schultergelenk. Dazu muss die Transformation des Skeletts, ausgehend von der Wurzel bis zu dem aktuellen Schultergelenk, aufmultipliziert werden, um die genaue Position und Ausrichtung des Schultergelenks zu bestimmen. Wird die Transformation invertiert und mit der Gelenkposition multipliziert, resultiert die Position des Ellenbogengelenks in lokalen Koordinaten aus Sicht des Schultergelenks.

$$M_{Schultergelenk}{}^{-1} \times P_{Gelenk} = M_{Schultergelenk}{}^{-1} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{pmatrix} = \begin{pmatrix} x_{l0} \\ y_{l0} \\ z_{l0} \\ 1 \end{pmatrix}$$

$$Mit\ M_n = T_0^n = A_0 \times A_1 \times \dots \times A_n\ ; \tag{8}$$

Das Ergebnis dieser Multiplikation ist ein Vektor mit der lokalen Position x_{l0}, y_{l0} und z_{l0} mit l0 als lokale Koordinaten der Position P_0. In diesen lokalen Koordinaten kann ein rechtwinkliges Dreieck definiert werden, welches durch die Projektion der Positionen des Ursprunges P_0, x_{l0} und y_{l0} auf die zweidimensionale Ebene, die durch x_{l0} und y_{l0} aufgespannt wird, dargestellt wird. Der Zusammenhang ist in der **Abbildung 30** dargestellt.

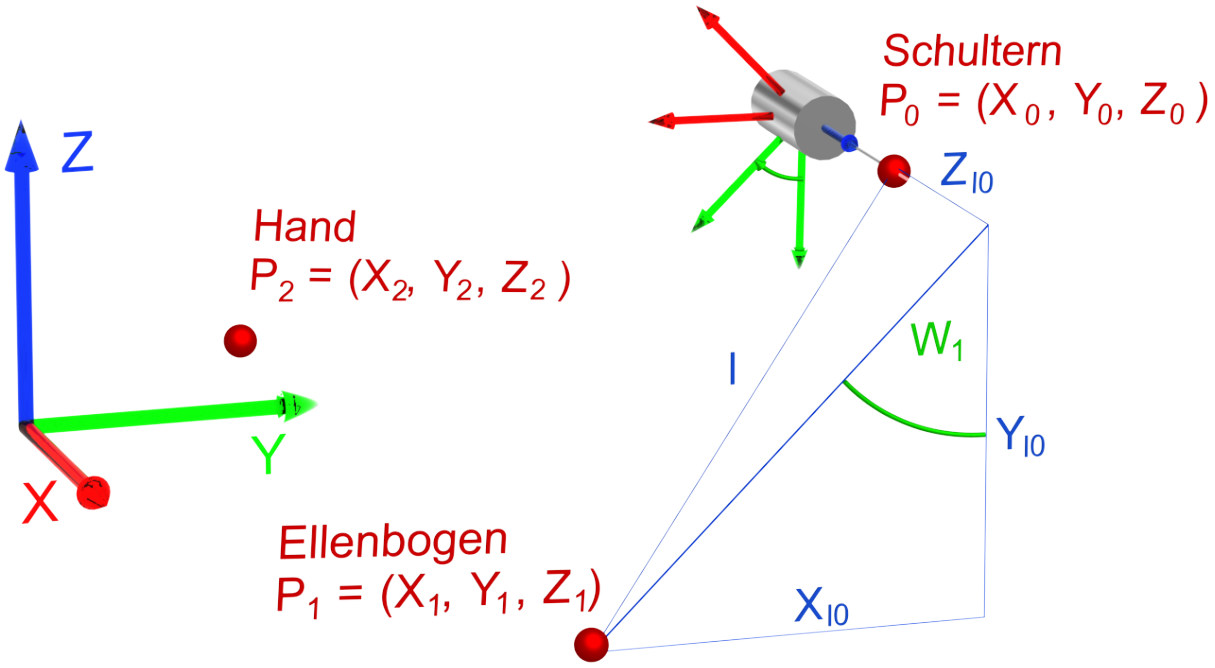

Abbildung 30 Iteratives Vorgehen bei der Berechnung der Gelenkstellungen am Beispiel des ersten Schultergelenkes.

An diesem Dreieck kann mittels des Sinus und des Kosinus der eigentliche Winkel W_1 ausgerechnet werden. Dieses geschieht durch die Anwendung des Tangens:

$$Tan(w_1) = \frac{\text{Sin}(w_1)}{\text{Cos}(w_1)} = \frac{Gegenkathete}{Ankathete} = \frac{x_{l_0}}{y_{l_0}}$$

$$\Rightarrow w_1 = \text{arcTan}\left(\frac{x_{l0}}{y_{l0}}\right) \quad \textbf{(9)}$$

Um das iterative Vorgehen zu verdeutlichen, wird kurz die Berechnung des zweiten Schultergelenkes w_2 gezeigt. Dazu werden die Schultergelenkposition, die Ellenbogengelenkposition und die aktuelle Orientierung des Gelenkes vor dem zweiten Schultergelenk genutzt.

Durch das Übertragen des ausgerechneten Gelenkwinkels w_1, welches die zuvor ausgerechnete Orientierung beinhaltet, auf das Skelett wird die Orientierung dieses Schultergelenks aktualisiert. Das Koordinatensystem des zweiten Schultergelenkes ist einmal um die x-Achse und einmal um die z-Achse in der Skelettdefinition, jeweils um 90°, gedreht. Daher muss nun der gleiche Vorgang wie im vorigen Fall durchgeführt werden. Es muss die lokale Position des Ellenbogengelenkes, diesmal bezogen auf die Orientierung des ersten Schultergelenks, berechnet werden. Nach der Multiplikation der Ellenbogengelenkposition mit der invertierten

homogenen Matrix der Position und Orientierung des ersten Schultergelenkes kann der Winkel durch die atan2 Funktion berechnet werden (siehe **Abbildung 31**). Beim dritten Schultergelenk ist die Kenntnis der Position der Schulter, des Ellenbogens und nun auch der Hand nötig, die nun in die lokalen Koordinaten des zweiten Schultergelenks transformiert werden müssen.

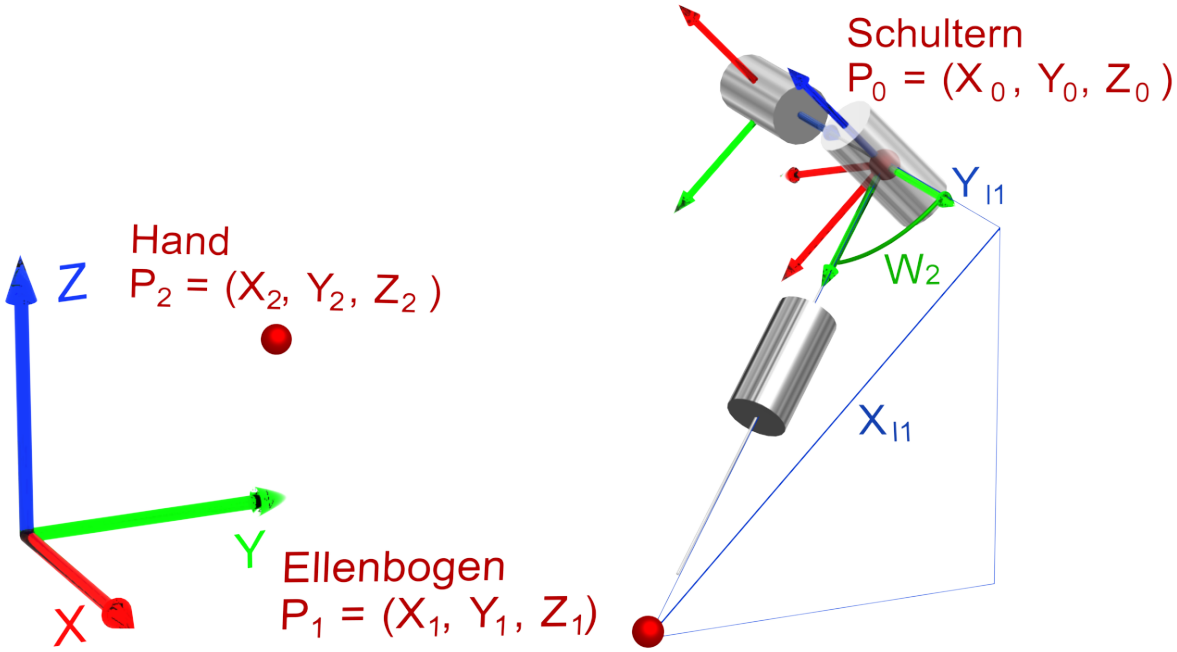

Abbildung 31 Iteratives Vorgehen bei der Berechnung der Gelenkstellungen am Beispiel des zweiten Schultergelenks.

Es sind verschiedene Positionen und Orientierungen nötig, um die einzelnen Gelenkwinkel auszurechnen. Zur Ermittlung der Winkel in den Endeffektoren wie Hand und Kopf kann wieder die Winkelextraktion im Anhang in der Sektion verwendet werden, allerdings muss nun die aktuelle Stellung des Skelettes mit berücksichtigt werden. Es muss die Differenztransformation des Handgelenkes, gegeben durch die Stellung des Skeletts vor dem Handgelenk, und die eigentlichen Orientierung der Hand, welche durch den Rigidbody auf dem Handrücken bestimmt ist, berechnet werden. Dazu wird die Orientierung vor dem Handgelenk invertiert und mit der Orientierung des Rigidbodys der Hand multipliziert.

$$M_{Differenz} = (M_{FormHandgelenk})^{-1} \times M_{Endausrichtung} \tag{10}$$

Aus dieser Differenzmatrix können mit Hilfe der Winkelextraktion im Anhang in der Sektion A die gewünschten Gelenkstellungen für die Hände und den Kopf berechnet werden. Im Fall des Ganzkörper-Motion-Capturing sind die Winkelberechnungen der Beine ziemlich ähnlich durchzuführen wie die der Arme. Für Motion-Capturings von mehreren Personen werden jeweils dem Skelett die entsprechenden Rigidbodys für die inverse Kinematik zugeordnet. Die eigens implementierte inverse Kinematik ist in der Lage, die Stellungen der Gelenke zu berechnen, wenn nicht alle Rigidbodys aufgezeichnet werden können. Dazu wird die vorherige

Position sehr nah bei der alten Position geschätzt und die Berechnung der in der kinematischen Kette folgenden Gelenke durchgeführt. Dadurch wird die kinematische Darstellung der Bewegung neben der Robustheit der Rigidbodys wegen der besseren Sichtbarkeit[49] erhöht

4.7 Zusammenfassung

In diesem Kapitel wurde vorgestellt, wie Motion-Capturing für mehrere Personen mittels Rigidbodys robust über einen längeren Zeitraum durchgeführt werden kann. Dazu müssen die verschiedenen Körperteile unterschieden werden, welches durch einen Eigenentwurf von Rigidbodys, die eine entsprechende Anzahl an Variationen zulassen, ermöglicht wird. Auf Basis der Körperteile mit Position und Ausrichtung können anschließend alle Gelenkwinkel nacheinander berechnet werden. Durch die verwendete inverse Kinematik ist es möglich, robust das Motion-Capturing mit minimalen Vor- und Nachbereitungszeiten durchzuführen [72].

[49] Fünf Kugeln können immer von mehreren Kameras besser gesehen werden.

5 Korpora

Diese Arbeit ist im Zusammenhang mit der Erstellung von verschiedenen Korpora entstanden und einer diesbezüglichen Unterstützung der Auswertung von verschiedenen Interaktionsverhaltensszenarien. Dabei wurde angestrebt, eine möglichst gute Abdeckung des Forschungsszenarios erhalten zu können. Dazu musste geklärt werden, welche Aspekte technisch realisierbar sind und welche mögliche Funktionalität in zukünftiger Software umsetzbar sein würde. Zu diesem Zweck wurden meistens mit unterschiedlichen Umfang Vorstudien durchgeführt. Das Tool PAMOCAT wurde diesbezüglich direkt auf die Korpora zugeschnitten oder nachher in der folgenden Zeit erweitert. Die Kernfunktionalität, menschliches Interaktionsverhalten aufzuzeichnen und auswerten zu können, ist bei allen Korpora ähnlich. Die technischen Gegebenheiten sind jeweils sehr ähnlich und wurden bereits im Kapitel 4 vorgestellt. Diese bezieht sich auf das Aufzeichnen von menschlicher Bewegung zusammen mit mehreren Video Kameras. Die Unterschiede liegen in der verwendeten Position und Anzahl der Kameras. Im Folgenden werden diese erstellten Korpora mit ihren gegebenenfalls technischen Abweichungen[50] vorgestellt.

5.1 Obersee

Der Korpus „Obersee“ wurde 2009 unter der Leitung von Karola Pitsch erstellt [72]. Der Gedanke dabei war, die Probanden miteinander interagieren zu lassen und genauestens zu analysieren, wie der Sprecherwechsel in Zusammenhang mit verschiedenen Interaktionselementen zusammenhängt. Dabei wurden einer Gruppe mit drei Probanden jeweils verschiedene Rollen zugeteilt, die sie in einer Verhandlungsrunde vertreten sollten. Diese Rollen waren ein Vertreter der Stadt, ein Umweltschützer und ein Investor. Die Grundlage der Verhandlung war das Planen eines Freizeitgebiets um einen See herum in der Nähe eines Vogelschutzgebietes. Dazu wurden den Personen verschiedene Figuren zur Hand gegeben, die verschiedene Teilgebiete des Erholungsgebiets repräsentieren. Diese konnten auf eine Karte positioniert werden, um eine mögliche neue Lage des Teilgebiets zu symbolisieren. Die Hauptanforderung bezüglich dieser Arbeit war, im Nachhinein zu ermitteln, wie Motion-Capturing überhaupt genutzt werden kann, um die Analyse bei der Interaktionsforschung zu unterstützen. Dabei spielte der zeitliche Nutzen in Bezug zum zusätzlichen Arbeitsaufwand, der durch das Motion-Capturing anfällt, eine wichtige Rolle. Nebenanforderungen waren es, möglichst viele verschiedene automatische Annotationen zu erstellen.

In der folgenden **Abbildung 32** ist diese Verhandlungssituation aus einem Versuchsablauf dargestellt. Insgesamt wurden 15 Versuchsdurchläufe aufgezeichnet, jeweils mit einer Länge von 30 bis 35 Minuten. Davon sind 9 Durchläufe mit drei Probanden und 6 mit zwei Proban

[50] Abweichung vom Grundsetup verwendet beim „Obersee“ Korpus.

Abbildung 32 Der „Obersee“ Korpus von 2009 (K. Pitsch, 2010) mit der ersten Version von Rigidbodys, die noch zu groß waren, um die nötige Variabilität zu erreichen.

den durchgeführt worden. Aufgezeichnet wurde die Interaktion der Probanden mit 4 HD Kameras und einem Mikrophon neben dem Motion-Capture-System, bestehend aus einem Vicon MX mit 10 x T10 Kameras. Die Anforderung an die Motion-Capture-Aufnahme war, Aufzeichnungen zu erstellen, die robust über einen längeren Zeitraum von mehr als 30 Minuten durchgeführt werden konnten. Eine weitere Anforderung war, dass die Motion-Capture- Daten mit absehbarem Zeitaufwand zur Analyse nutzbar sind, ohne ein Vielfaches der aufgenommenen Zeit mit der Korrektur der Motion-Capture-Aufnahmen aufwenden zu müssen. Hierzu wurde der Gedanke entwickelt, das Motion-Capturing mit den sogenannten Rigidbodys mit einer großen Variation durchzuführen und auszuprobieren, wie diese am besten an den Probanden angebracht werden konnten.

5.2 Kunsthalle

Der „Kunsthallen“ Korpus wurde 2010 ebenfalls unter der Leitung von Karola Pitsch erstellt [4]. Bei diesem wurde die Interaktion eines Roboters (Nao der Firma Aldebaran) mit Menschen analysiert, speziell wie Menschen auf Maschinen in „realen“ Lebenssituationen reagieren. Eine Fragestellung dazu war, wie die Museumsbesucher auf den Roboter reagieren würden, und damit verbunden, wie stark die Museumsbesucher dem Roboter Aufmerksamkeit schenkten (um sich Informationen zu den Bildern geben zu lassen) oder ob sie selber die Beschreibungen der Bilder durchlesen würden.

Dazu wurden die freiwilligen Probanden mit Rigidbodys ausgestattet, diesmal allerdings pro Person nur ein einziger. Damit waren das Experiment und die damit verbundene Vorbereitung nicht zu zeitintensiv, und die Probanden konnten sich spontan entscheiden, an der Studie teilzunehmen oder nicht. Zudem war nicht die Bewegung des Skeletts, sondern der Aufmerksamkeitsfokus in Relation zu den Gemälden und des Roboters von Interesse. Im Versuch sel-

ber konnten sich die Probanden in Ruhe die Gemälde anschauen und frei durch einen kleinen Teil der Ausstellung gehen. Im Falle, dass sie sich dem Roboter näherten, reagierte dieser unterschiedlich, je nachdem, in welcher Entfernung die Probanden waren. Dabei wurden zwei unterschiedliche Radien verwendet. Bei der ersten Entfernung wurde versucht, das Interesse für den Roboter zu wecken, und bei der zweiten Entfernung, die Probanden über die Gemälde zu informieren.

Dabei lag die Anforderung darin, die genaue Position der Personen mit ihrer Kopforientierung über das gesamte Zeitintervall aufzuzeichnen, in dem sie im Teilbereich der Kunsthalle waren. Das Interesse lag dabei über den Zeitverlauf hinweg auch in den Trajektorien mit ihrer jeweiligen Kopforientierung. Zusätzlich sollten die Rigidbodys besser handhabbar sein und nicht abstoßend aussehen, da es eine freiwillige spontane Studie war. Es wurde mit 3 HD Kameras gearbeitet zusätzlich zum aufzeichnenden Motion-Capture-System, das aus einem Vicon MX mit 10 x T10 Kameras bestand. Zusätzlich gibt es noch Ton- und Bildaufnahmen aus dem Roboter selber. Es wurde tageweise über eine Woche hinweg aufgezeichnet (Dienstag bis Sontag). Insgesamt wurden so 50 Aufnahmen mit einer durchschnittlichen Laufzeit von 45 min aufgezeichnet. Wie in der **Abbildung 33** ersichtlich, wurden hier erstmals die überarbeiteten 3D Rigidbodys verwendet, die deutlich kleiner als der erste Prototyp sind. Der Roboter Nao hat dabei den Input des Motion-Capture-Systems genutzt, um mit den Probanden zu interagieren. Dazu hat dieser die Entfernung aus dem Motion-Capture-System zwischen sich selber und den Probanden genutzt, um mit ihnen zu agieren. Bei dieser Studie war die Bewegung in Relation zur Aufnahmeumgebung, den Gemälden und dem Roboter von besonderem Interesse.

Abbildung 33 Kunsthallen Korpus, bei dem mit 3 Kameras gearbeitet wurde

5.3 Sagaland

Der Korpus Sagaland ist unter der Leitung von Kirsten Bergmann und Stefan Kopp im Jahre 2013 entstanden, basierend auf einer Studie aus dem Jahre 2008 mit Motion-Capturing [5]. Für diesen Korpus wurde im Rahmen dieser Arbeit eine Vorstudie durchgeführt, um zu testen, ob der geplante Ablauf die gewünschten Ergebnisse liefern würde. Diese wurde ein halbes Jahr vor der eigentlichen Studie im Jahre 2012 durchgeführt. Bei dem Korpus Sagaland ist weniger das „Turntaking“ (Sprechwechsel) von Interesse, sondern hauptsächlich die Verbindung von Sprache und körperlichen Gesten. Dabei war das Hauptinteresse, wie im Detail verschiedene interaktive Gesten im Bezug zum Ausgesprochenen benutzt wurden. Damit die Probanden über den gleichen Inhalt sprechen, wurden bei ihnen unter kontrollierten Bedingungen die gleichen Erinnerungen erzeugt. Dadurch, dass sie gleiche Dinge aus der Erinnerung erzählen, kann verglichen werden, wie Menschen im Allgemeinen mittels Gesten die gesprochene Sprache untermalen. Die gleichen Erinnerungen werden in den Probanden durch ein gleiches Erlebnis erzeugt. Dazu nehmen die Probanden an einer virtuellen Busfahrt mit insgesamt fünf Haltestellen teil. Damit die Probanden sich möglichst gut die Details der Umgebung einprägen können, dürfen sie den Zeitpunkt der Weiterfahrt selber bestimmen. Die Aufgabe der Versuchspersonen ist es, sich die Strecke bzw. einzelne Orte mit ihren verschiedenen Merkmal einzuprägen, siehe dazu **Abbildung 34**.

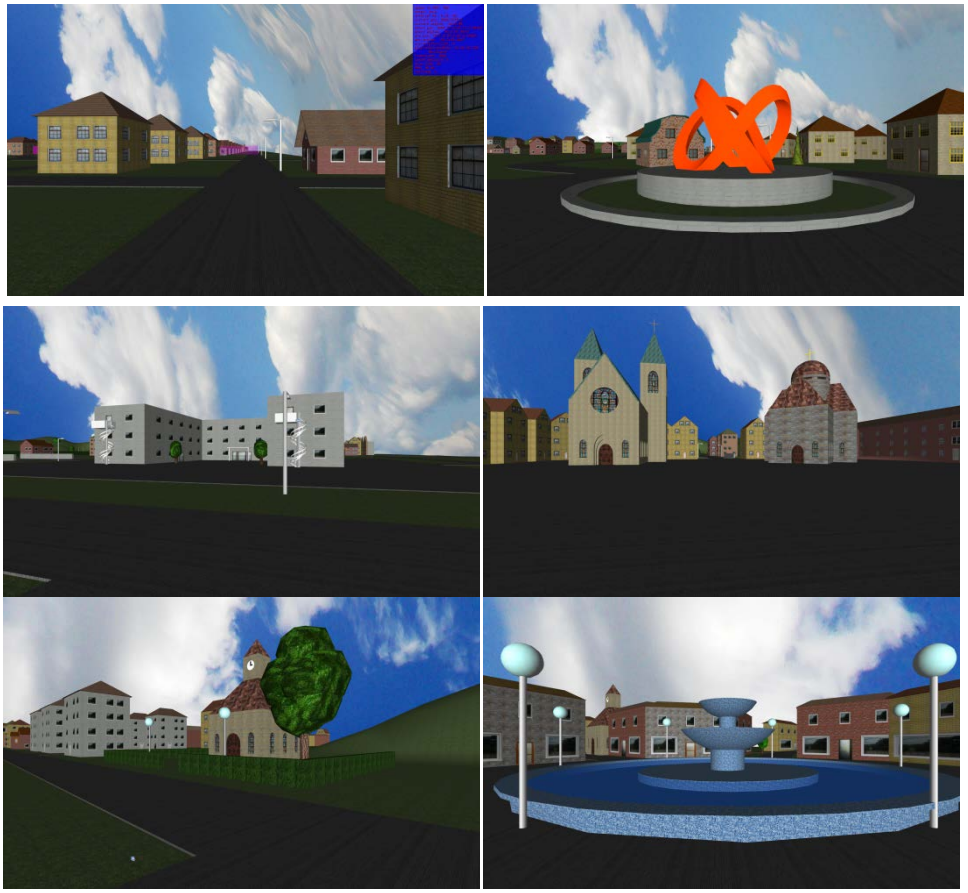

Abbildung 34 Sagalands Startposition und fünf Schauplätze, die sich die Probanden einprägen sollen.

Um die Teilnehmer zu motivieren, einander den Weg zu beschreiben, fahren zwei Personen eine leicht unterschiedliche Strecke und sollen diese einer dritten Person beschreiben. Die dritte Person soll sich beide Fahrten einprägen, damit sie diese voneinander unterscheiden kann. Anschließend soll sie eine eigene Fahrt durchführen und entscheiden, welche Fahrt sie erlebt hat. In der **Abbildung 35** sind beide Routen mit rosafarbenen Linien aus der Vogelperspektive eingezeichnet. Damit die Probanden immer noch über die gleichen Dinge intensiv reden, führen beide Routen an den gleichen Bushaltestellen vorbei. Allerdings sind bei manchen kleine Veränderungen vorgenommen worden, wie zum Beispiel das Vertauschen der Plätze zweier Kirchen an einer Haltestelle. Die Durchführung der Studie war insgesamt deutlich aufwendiger als die Durchführung der „Obersee“ Studie. Zusätzlicher Aufwand wurde durch die Tatsache nötig, dass den Probanden die jeweiligen einzelnen Busfahrten gezeigt

Abbildung 35 Sagaland, unterschiedliche Wege der Probanden

werden mussten. Zudem mussten die anderen Probanden einzeln vorbereitet und die Durchführung des Experiments erklärt werden. Außerdem durften die Probanden nicht vorher miteinander reden, damit sie nicht im Vorfeld bereits Informationen über das Experiment erfahren bzw. austauschen und diese Beschreibung daher nicht mit aufgezeichnet werden kann.

Bei dem Ablauf der Studie wurden allen Probanden zuerst die Rigidbodys angelegt. Zwei von drei Probanden durften die Busfahrt sehen. Anschließend wurden alle Probanden zusammengesetzt, um dem dritten Probanden diese Busfahrten zu beschreiben (siehe **Abbildung 36**). Zum Schluss durfte der dritte Proband die Busfahrten sehen und musste herausfinden, welche Busfahrt er selber im Vergleich zu den beiden anderen Probanden durchgeführt hatte.

Die Vorstudie hatte insgesamt 10 Aufzeichnungsrunden und die eigentliche Studie hatte 25 Durchläufe. Die Interaktionen wurden mit einem Vicon MX mit 14 x T10 Kameras und 4 HD Kameras aufgezeichnet.

Abbildung 36 Sagaland Vorstudie Ansicht durch Kontrollkamera

5.4 Fazit

In diesem Kapitel wurden die Korpora, die im Zusammenhang mit dieser Arbeit entstanden sind, vorgestellt. Diese verschiedenen Korpora wurden mit jeweils unterschiedlichen Forschungszielen erstellt. Dabei wurde jeweils auf den Erstellungshintergrund, das durchgeführte Szenario, die Anforderungen und die technischen Besonderheiten eingegangen. Die hier vorgestellten Korpora werden im späteren Verlauf dieser Arbeit aufgegriffen, um verschiedene Funktionalitäten, die für einzelne Korpora entwickelt wurden, vorzustellen.

6 Automatische Annotation und Analyse Möglichkeiten

Ein weiterer zentraler Aspekt dieser Arbeit ist es, durch die Motion-Capture-Daten den Analyseprozess in der Verhaltensforschung sinnvoll zu ergänzen. In der Interaktionsforschung müssen viele Daten mit möglichst geringem Zeitaufwand ausgewertet werden. Oft sind viele Stunden an Videodaten aufgezeichnet worden, die anschließende Analyse dauert oft ein Vielfaches (manchmal bis zum Hundertfachen) der Aufnahmezeit [13]. Die Motion-Capture-Daten bieten eine viel höhere Genauigkeit und Robustheit bei der Erfassung der Orientierungen einzelner Körperteile im Gegensatz zu aus Videodaten extrahierten Bewegungsdaten, bei denen nur eine wahrscheinliche Pose geschätzt wird. Zum Beispiel ist die Bestimmung des Fokus[51] einer Person aus einer Videodatei eher eine Schätzung und kann nicht automatisch durchgeführt werden. Zudem ist die Ermittlung, ob eine Person etwas anfokussiert, schwierig, wenn dieses nicht im gleichen Video enthalten ist. Im Gegensatz dazu sind bei Motion-Capture-Daten solche Analysen automatisierbar. Allerdings stellt sich auch die Frage, ob es Grenzen der automatischen Analysen basierend auf Motion Capture-Daten gibt und wo diese liegen. Das manuelle Annotieren kann in vielen Bereichen gut unterstützt, aber auch in manchen Bereichen ganz ersetzt werden. Im Folgenden wird auf verschiedene dieser Analyseaspekte, bezogen auf einzelne Personen und Gruppen, eingegangen. Dabei geht es darum, Zeitpunkte zu finden, in denen diese verschiedenen Phänomene auftreten. Damit verbunden ist das Auffinden von verschiedenen Interaktionsbestandteilen (Phänomena), die zusammen an einem Zeitpunkt vorkommen. Dadurch wird das Auffinden von komplexeren Verhaltensweisen als Kombination von Phänomenen ermöglicht, und eine detaillierte Analyse dieser Zeitpunkte kann z. B. auch anhand der Videoaufnahmen durchgeführt werden. Diese Funktionen sind unter anderem im Annotationstool namens PAMOCAT – „**P**re **A**nnotation **Mo**tion **Ca**pture **T**ool“ integriert. Dieses Tool bietet eine Benutzerschnittstelle, die für jede interessierte Person leicht zu bedienen ist.

6.1 Einzelpersonen-Phänomene

Im Folgenden wird auf die Einzelpersonen-Phänomene im Detail eingegangen. Dabei ist zu betonen, dass diese Entwicklung und die Auswahl der Phänomene eng mit dem Forschungshintergrund und der Gestaltung des Experiments zusammenhängen.

6.1.1 Die Zerlegung der Bewegung in Aktivitäten von einzelnen Freiheitsgraden

Eine dieser automatischen Annotationen bestimmt, zu welchen Zeitpunkten bestimmte Freiheitsgerade aktiv waren. Die Bewegung der einzelnen Personen wird nicht global in der Posi-

[51] Ausrichtung des Kopfes.

tionsebene der Hände oder auf Gelenkebene durchgeführt, sondern auf dem Level der einzelnen Freiheitsgrade. Dadurch können verschiedene Verhaltensweisen, basierend auf aktiven DOFs, gefunden werden. Dazu wird die Bewegung in sogenannte Key-Intervalle zerlegt. Dieses ermöglicht das schnelle Finden der Zeitpunkte, zu denen sich z. B. der Kopf seitlich bewegt hatte. Durch den Zusammenhang der einzelnen Key-Intervalle zu den DOFs lässt sich nach der Art der Bewegung suchen. Eine allgemeine Bewegung des Kopfes kann durch die Zuordnung der Aktivität im Gelenk als eine Geste klassifiziert werden. Je nach aktivem DOF kann eine Bejahung oder Verneinung interpretiert werden. Der Unterschied liegt in dem DOF, der aktiv ist; bei einer Verneinung durch (horizontale) Kopfrichtungsänderung ist ein anderer DOF aktiv als bei einer vertikalen Bewegung zur Bejahung.

Ein weiteres Beispiel dazu ist die Interpretation der Handbewegung. Je nach Verhaltensszenario kann das seitliche Bewegen der Hand „Hallo“ heißen, oder bei der Aktivität des anderen Freiheitsgrades, mit dem die Hand aufgerichtet werden kann, als „komm her“. Da bei der Verhaltensanalyse meistens unter kontrollierten Bedingungen gearbeitet wird, können verschiedenste Verhaltensweisen automatisch auf diese Weise ermittelt werden. Das heißt, dass die Bedeutung von Bewegungen bezüglich eines vorgegebenen und eingeschränkten Kontexts in Bezug gebracht wird. Manchmal ist aber auch das Zusammenspiel von verschiedenen DOFs entscheidend für das Finden von bestimmten Gesten und Verhaltensweisen; zum Beispiel, wenn eine bestimmte Geste mit einem Arm durch eine einzelne Aktivität im Schultergelenk in Kombination mit einer einzelnen Aktivität im Handgelenk vorkommt. Es ist nicht sichergestellt, dass immer die gesuchte Verhaltensweise gefunden wird, aber es werden alle Zeitpunkte gefunden, bei der möglicherweise diese bestimmte Verhaltensweise auftritt. Dieses erspart bei der späteren Analyse viel Zeit.

Ein Key-Intervall wird als Winkeländerung über einen Zeitraum mit gegebenem Anfangs- und Endzeitpunkt definiert. Der Unterschied zur Key-Frame-Animation ist, dass eine unterschiedliche Verwendungsebene vorliegt, bei dem einen ist eine ganze Animation (Key-Frame-Animation) gemeint, bei dem anderen nur ein Bestandteil, welcher mit vielen anderen zusammen eine Animation ergibt. Die Idee dahinter ist, dass man ähnliche Informationen zusammenfassen kann. Am Beispiel der Bewegung eines einzelnen DOFs des Ellenbogengelenks, welches den Unterarm zum Oberarm bewegen kann, kann genau diese Bewegung zusammengefasst werden, wenn sie bei mehreren einzelnen Zeitpunkten ähnlich ist. Das heißt, dass die einzelnen Zwischenschritte, bei denen die Bewegungsänderungen ähnlich sind[52], zu einem Key-Intervall mit einem Startzeitpunkt, einem Endzeitpunkt, einem Anfangs- und einem Endwinkel zusammengefasst werden [74]. In der **Abbildung 37** wird die Bewegung von vier einzelnen Zeitpunkten zu einer Zeitspanne mit Anfangs- und Endwinkel zusammengefasst. Je nachdem, wie klein der Parameter für die Winkelähnlichkeit ausgewählt wird, kann eine Bewegung stärker oder schwächer komprimiert werden. Da für die Bewegungsanalyse, aber auch für die Wiedererkennung ähnlicher Bewegungssequenzen die Änderung der Bewe-

[52] Diese Ähnlichkeit kann z. B. in der Geschwindigkeit definiert werden.

gungsrichtung ein entscheidendes Merkmal ist, wird eine reale Armbewegung, wie gerade betrachtet, in zwei einzelne Bewegungen zerlegt. Dies ist einmal eine Beschleunigungsphase,

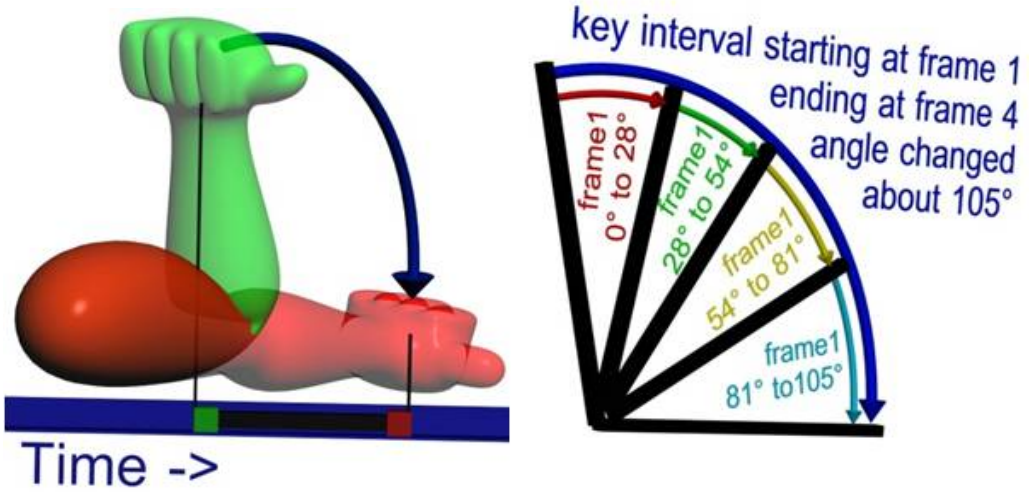

Abbildung 37 Ein Beispiel für ein Key-Intervall bezogen auf den Freiheitsgrad eines Ellenbogengelenks (a) Bewegung des Unterarmes um ein Ellenbogengelenk (b) Darstellung der einzelnen Bewegungsänderungen in verschiedenen Zeitpunkten und als zusammengefasste Zeitspanne („Bild Deutsch übersetzen“).

die bis zum Maximum der Geschwindigkeit geht, gefolgt von einer Abbremsphase, bei der die Hand zum Stillstand kommt. Dieser Zusammenhang von Winkel, Geschwindigkeit, Beschleunigung mit einem Key-Intervall ist in der **Abbildung 38** dargestellt. Nur wenn die Winkelähnlichkeit sehr groß ausgewählt ist, wird die Bewegung wie in der **Abbildung 38** in zwei Teile zerlegt, andernfalls werden Untersegmente mit ähnlichen Geschwindigkeiten

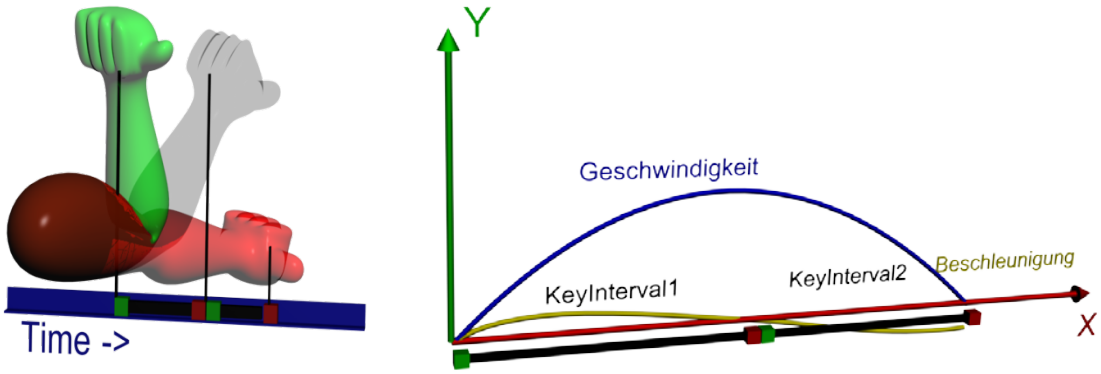

Abbildung 38 Die Beziehung der Bewegung in 3 D in Verbindung mit der lokalen Winkeländerung entlang eines DOFs, und Beschleunigung mit den jeweiligen dazugehörigen Key-Intervall Interpretationen

erzeugt, bei denen z. B. vier Key-Intervalle erzeugt wurden. Dabei wird ein Maximalwert angegeben, bis zu welchem Geschwindigkeitsunterschied die Geschwindigkeiten in einem Key-Intervall zusammengefasst werden dürfen. Mittels dieses Wertes wird beschrieben, welche Schwankungen in der Geschwindigkeit als gleichwertig angesehen werden und zusam-

mengefasst werden können. Die Länge der erzeugten Key-Intervalle hängt vom Geschwindigkeitsprofil ab, angefangen bei einer Geschwindigkeit gleich Null bis zu einem Maximum oder Minimum. Mathematisch gesehen erfordert das die Bestimmung eines Extremums, welches durch die Ableitung der entsprechenden Ortsfunktion f und das Ausrechnen der Nullstellen bestimmt wird.

$$f'(t_0) = g(t_0) = 0 \quad \textbf{(11)}$$

$$f''(t_0) = a(t_0) \neq 0 \quad \textbf{(12)}$$

Die erste Ableitung der Ortsfunktion beschreibt die Geschwindigkeit. Wenn die Geschwindigkeit gleich Null ist, liegt ein lokales Extremum vor, falls die zweite Ableitung von Null verschieden ist. Durch das Vorzeichen der zweiten Ableitung in dieser Nullstelle ist ermittelbar, ob es ein Minimum oder Maximum ist. Die Ableitung der Geschwindigkeit beschreibt die Beschleunigung. Daher hängen Geschwindigkeit und Beschleunigung wie folgt zusammen: Wenn die Geschwindigkeit maximal ist, wird keine Beschleunigung ausgeübt. Siehe dazu **Abbildung 38**, bei der die Winkeländerung in globalen Koordinaten auf eine Änderung in lokalen Koordinaten bezüglich eines DOFs im Bezug zur Geschwindigkeit und Beschleunigung gebracht wird. Dazu wird auch das dazugehörige Key-Intervall in Relation gebracht. Die eigentliche Analyse wird im lokalen Koordinatensystem bezüglich jedes einzelnen Gelenks durchgeführt.

$$T_0^{Ellenbogengelenk} = A_0 \times A_1 \times \ldots \times A_{Ellenbogengelenk} \quad \textbf{(13)}$$

Dazu werden alle Gelenke, die in der Hierarchie vor dem aktuell betrachteten Gelenk liegen, aufaddiert, um die exakte Position und Ausrichtung des aktuellen Gelenkes zu ermitteln. Um die Geschwindigkeit entlang eines DOFs auszurechnen, werden die einzelnen Änderungen entlang des DOFs über die Zeit durch Subtraktion ermittelt. Die Beschleunigung kann durch die zeitliche Geschwindigkeitsänderung ermittelt werden. Ein Key-Intervall beginnt, wenn die Beschleunigung anfängt, von Null verschieden zu sein, und endet, wenn sie wieder Null wird.

$$t_{anfang} = f''(t) \neq 0 \quad \textbf{(14)}$$

$$t_{ende} = f''(t) = 0 \quad \textbf{(15)}$$

Dieses so definierte Zeitintervall kann für jedes einzelne DOF weiter unterteilt werden, indem eine Winkelabweichung definiert wird, bei der die Geschwindigkeiten bis zu einer definierten Größe zusammengefasst werden. Dadurch wird die Beschleunigungsphase durch die Kurve mit mehreren linearen Phasen mit unterschiedlicher Steigung näherungsweise beschrieben.

$$f''(t) > Schwellwert \quad \textbf{(16)}$$

Wird ein gewisser Grenzwert der Beschleunigung überschritten, wird das Intervall an diesem Zeitpunkt geteilt, und ab diesem Zeitpunkt wird die Beschleunigung neu betrachtet. Bei der

Ermittlung der Key-Intervalle zur Bewegung wird eine Matrix verwendet, die als Elemente wiederum Vektoren besitzt.

$$\begin{pmatrix} w \\ v \\ a \\ c \end{pmatrix}_t \begin{pmatrix} w \\ v \\ a \\ c \end{pmatrix}_{t+1} \cdots \begin{pmatrix} w \\ v \\ a \\ c \end{pmatrix}_{t+n} \tag{17}$$

Diese Matrix beinhaltet auf der vertikalen Achse alle Freiheitsgrade[53] mit bis zu 41 elementaren Gelenken[54]. Auf der zweiten horizontalen Achse ist die Zeit enthalten, mit bis zu 200 Hz an Daten bei einer schnellen Aufnahme. In einem Element der Matrix sind insgesamt vier Werte enthalten, und zwar der Winkel, die Geschwindigkeit, die Beschleunigung und eine Zahl, die einen Interpretationscode enthält (siehe dazu **Abbildung 39**).

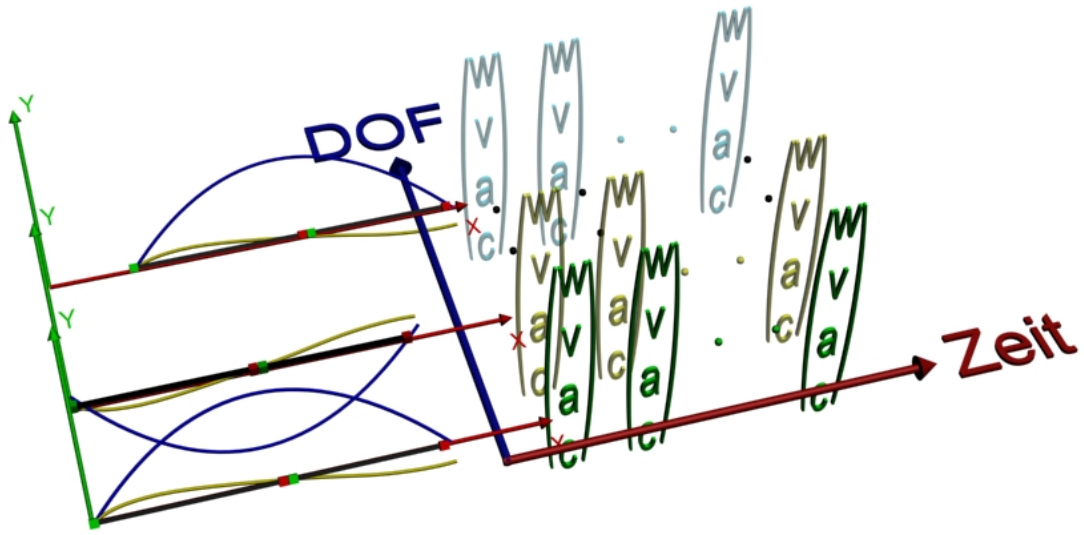

Abbildung 39 Value Over Time Matrix

Dieser automatisch berechnete Code besagt, ob das Intervall mit Aktivität begonnen hat, ob es anhält, ob ein mögliches Ende gefunden wurde oder ob das Intervall mit einem finalen Ende geschlossen werden kann. In diesem Fall wird rückwärts in den vorherigen Aufnahmezeitpunkten nach dem Anfang mit dem interpretierten Codewert für „Begin“ gesucht, und mit der Position des Anfangs- und des Endwinkels ein Key-Intervall erzeugt.

6.1.2 Automatische-Pose-Annotation

Um verschiedene Verhaltensweisen finden zu können, ist es ein Hilfsmittel, relevante Posen zu detektieren. Dieses kann z. B. eine Pose sein, bei der auf etwas gezeigt wird, oder wenn zwei Hände nach vorne gehalten werden, um mit den Händen etwas zu beschreiben. Eine

[53] Freiheitsgrad kann hier auch als elementares Gelenk aufgefasst werden.

[54] Je nachdem, welches Skelett zugrunde liegt.

Geste lässt sich meistens durch eine spezifische und signifikante Pose erkennen, wenn diese in einem Kontext mit eingeschränkten Themen und Interaktionsmöglichkeiten entstand. Eine Bedingung, um Posen zu suchen, ist, dass die gesuchten Gesten markante Unterschiede haben müssen, um diese auseinanderhalten zu können. Beim „Obersee“ Korpus sind solche Gesten, die rechte oder linke Hand zum Kopf zu führen, sich nach vorne zu lehnen, die Arme zu verschränken und das Zeigen mit dem rechten oder linken Arm auf eine Stelle auf dem Tisch. Beim „Sagaland“ Korpus sind relevante Posen das Zeigen, eine symbolische Haltegeste, bei der beide Hände nach vorne gestreckt sind und beide Arme weit auseinander gehalten werden.

Ein komplexerer Bewegungsablauf kann durch mehrere signifikante Posen als Sequenz erkannt werden. Alternative Überlegungen gingen in die Richtung, Bewegung anhand der Key-Intervalle zu ermitteln. Doch für den Einsatz in der Verhaltensforschung ist die Art und Weise, wie sich Leute in Gesprächen ausdrücken, zu unterschiedlich. Es gibt eine große Variation der Bewegungsabläufe bei der Beschreibung gleicher Dinge, sodass durch das Finden einzelner entscheidender Posen anstatt einer Bewegungssequenz ein höherer Gewinn für die Forschung erreicht wird, um Verhaltensweisen zu untersuchen. Dazu kann die Pose selbst durch die Stellung der Gelenke des Skeletts ausgewählt werden. Zu jedem Gelenk kann ein Gelenkwinkelbereich und ein Gewichtungsfaktor angegeben werden. Der Gelenkwinkelbereich gibt einen Bereich an, in welchen Stellungen sich das Gelenk für die jeweilige Pose befinden darf (siehe **Abbildung 40**).

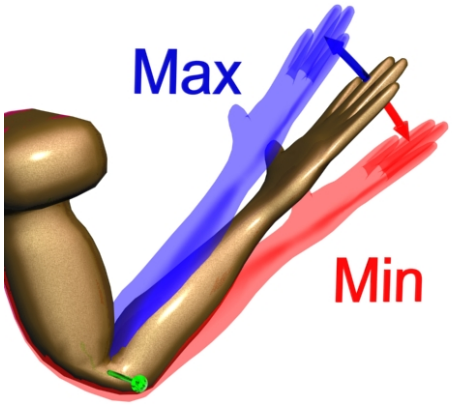

Abbildung 40 Eine Armpose mit dem zulässigen Winkelbereich bzw. Gültigkeitsbereich

Durch einen zugehörigen Gewichtungsfaktor kann definiert werden, welche Gelenke wichtig für eine Pose sind. Zum Beispiel ist bei einer Zeigegeste nur die Gelenkstellung des zeigenden Arms wichtig, die anderen Gelenke können sich in irgendeiner Stellung befinden. Um Übereinstimmung zu detektieren, werden zwei Merkmale berechnet, nämlich die Anzahl der Gelenke, die im definierten Wertebereich liegen, und die Gradabweichung aller Gelenke zur idealen Pose, wie sie definiert wurde, jeweils unter Berücksichtigung des zugehörigen Gewichtungsfaktors. Diese Berechnungen werden für alle ausgewählten Posen durchgeführt, für jeden Zeitpunkt der Aufnahme. Mathematisch ist die Berechnung eine Aufsummierung von

Winkeln, die im definierten Wertebereich liegen (siehe Formel 18) und die Aufsummierung der Winkeldifferenz (siehe Formel 19) der jeweiligen Gelenke zu der Winkelstellung der jeweiligen Pose [75].

$$similar(x) = \frac{\sum_{k=1}^{n} sim(k)}{n} \times 100$$

$$mit\ \ sim(k) = \begin{cases} 1, \min(k) < j(k) < \max(k) \\ 0, \min(k) > j(k) > \max(k) \end{cases} \quad \textbf{(18)}$$

Dabei ist k der Laufindex für die Gelenke. Die Funktion j(k) gibt den aktuellen Wert für das Gelenk an, min(k) den minimal und max(k) den maximal definierten Wert. $differenzialSimilar(x) = \sum_{k=1}^{n} diffSim(k) \times 100$

$$mit\ \ diffSim(k) = 1 - \frac{abs(\frac{\max(k)-\min(k)}{2} - j(k))}{\max(k)-\min(k)} \quad \textbf{(19)}$$

Der Gelenkbereich und die Gewichtung der Gelenke kann definiert oder durch Beispiele gelernt werden. Ist bei einem Gelenk die Gelenkstellung sehr nahe am Rande des Gelenkbereichs (min oder max), wird dieser Rand erweitert (siehe Formel 20 und 21).

$$expRangMin = \begin{cases} \min(k) \times \frac{\frac{\min(k)}{0.9}}{j(k)}, \min(k) < j(k) < \frac{\min(k)}{0.9}) \\ \min(k), \frac{\min(k)}{0.9} \leq j(k) < \max(k) \end{cases} \quad \textbf{(20)}$$

$$expRangMax(k) = \begin{cases} \min(k) \times \frac{j(k)}{\frac{\max(k)}{0.9}}, \max(k) > j(k) > \frac{\max(k)}{0.9}) \\ \max(k), \frac{\max(k)}{0.9} \geq j(k) > \min(k) \end{cases} \quad \textbf{(21)}$$

Liegt der Winkel eines Gelenkes außerhalb des Gelenkbereichs, ist dieses weniger wichtig für die Pose insgesamt.

$$importance(k) = \begin{cases} \frac{oldimportance(k) \times \frac{\min(k)}{j(k)}}{2}, j(k) < \min(k)) \\ oldimportance(k), \min(k) < j(k) < \max(k) \\ \frac{oldimportance(k) \times \frac{j(k)}{\max(k)}}{2}, \ j(k) > \max(k) \end{cases} \quad \textbf{(22)}$$

Für jede der ausgewählten Posen, die zu erkennen sind, wird die Gleichheit berechnet, wie ähnlich die aktuelle Pose ist.

$$totalSimilar(x) = \frac{similar(x) + differenzialSimilar(x)}{2} \quad \textbf{(23)}$$

Die Pose, die der aktuellen Pose am meisten ähnelt und über einem definierbaren Grenzwert (z. B. 85 %) liegt, wird entsprechend klassifiziert.

6.1.3 Ruheposition und Aktivitätsfindung von Händen

Das Finden von Zeitpunkten, bei denen die Probanden miteinander gestisch in Interaktion sind, ist eine große Hilfe bei der Annotation und für die spätere Analyse. Dieses entspricht dem Finden von Zeitpunkten, bei denen die Hände in Bewegung sind. Eine ähnliche Information ist indirekt aus der Bewegungszerlegung in Key-Intervalle ersichtlich, aber diese Funktion ist auf die Bewegungen aus lokaler Sichtweise der einzelnen Gelenke bezogen. Um die aktiven Bewegungsphasen der Hände zu detektieren, wird die Bewegung aus globaler Sicht analysiert. Eine Hand kann in einer Position bleiben, während andere Gelenke der kinematischen Kette sich bewegen. Ein Beispiel dafür aus dem „Obersee“ Korpus ist, wenn ein Proband sich aufrecht hinsetzt und sich dazu an den Armlehnen abstützt. Bei dieser Bewegung bleiben die Hände in einer globalen Sichtweise an der gleichen Stelle. Daher wird hierzu die Bewegung der Körperteile im globalen Raum betrachtet und nicht die lokalen Aktivitäten (wie es vorher der Fall war). Um dies zu berechnen, wird die Geschwindigkeit der Hände überwacht, und sobald eine definierte Geschwindigkeit überschritten ist, wird dies als Aktivität interpretiert. Die Geschwindigkeit wird als die innerhalb eines Zeitintervalls zurückgelegte Distanz definiert:

$$Geschwindigkeit(i) = \frac{p_{i-1} - p_i}{t_{i-1} - t_i} \quad \textbf{(24)}$$

Dabei steht p_{i-1} für die Position zum Zeitpunkt t_{i-1}und p_i für die Position zu einem späteren Zeitpunkt t_i. Alternativ hierzu kann auch die Berechnung von Bewegungssegmenten anhand von Richtungen genutzt werden. Ein Bewegungssegment ist eine Reihe von gleich klassifizierten Richtungsvektoren. Dabei werden nur solche als aktive Handbewegungsphasen angesehen, wenn eine minimale Distanz zurückgelegt wurde. Der Vorteil ist Stabilität gegenüber kleinen Bewegungsschwankungen, die um einen Punkt herum erfolgen (mehr hierzu folgt im nächsten Teilkapitel). Allgemein gibt es bei Zeigegesten eine Hold-Phase[55] (Haltephase), bei denen die Hände nicht bewegt werden; diese sind aber ein Teil der Geste und sollen daher auch als aktive Phasen klassifiziert werden. Um dies zu erreichen, wird eine einstellbare Zeitspanne verwendet, um zu bestimmen, was zu einer Haltephase gehören könnte und was zu einer Ruhephase gehört. Allgemein sind die Ruhephasen um ein Vielfaches größer als die Haltephasen. Die Position der Ruhe kann auch ermittelt und visualisiert werden. Eine solche Position, in der die Hand ruht, wird auch als „homeposition“ bezeichnet [76]. Die Zeitspanne, in der die Hand in einer Ruheposition ist, bezieht sich auf das genaue Gegenstück der Zeitspanne, in der Aktivität detektiert wurde. Zur Berechnung, wann die Hände sich in einer

[55] Eine Hold-Phase beschreibt eine Teilphase bei einer Geste, bei der die Hand sich für einen kurzen Augenblick nicht bewegt, um zum Beispiel auf etwas zu zeigen. Siehe hierzu in Sektion 2.4.2 „Bestandteile von Gesten“ für Details.

„homeposition“ befinden, kann die Funktionalität Handaktivitäten negativiert verwendet werden. Also wird nach den Zeitpunkten gesucht, bei denen keine Hand aktiv ist. Jede Hand kann verschiedene Ruhepositionen haben. Um alle Ruhepositionen zu berechnen, werden erst einmal alle Positionen ermittelt, bei denen sich eine Hand nicht bewegt. Da es teilweise Ruhepositionen gibt, die sehr nah beieinander liegen, wurde die Position im „Obersee“ Korpus durch Annotation bezüglich der Ruheposition der annotierenden Person zusammengefasst. Dazu wurde geprüft, wann der annotierenden Person auffällt, dass eine andere Ruheposition eingenommen wurde. Diese Distanz, die beieinanderliegende Ruhepositionen zusammenführt, ist einstellbar. Sie ist relativ groß, ca. 10 cm, da sich die annotierenden Personen nicht die genaue Position, sondern eher Anhaltspunkte merken. Daher werden im zweiten Schritt alle Ruhepositionen, die unterhalb dieser Distanz auseinander liegen, zusammengefasst. Dabei wird der Mittelwert dieser Positionen, die in diesem Umkreis beieinander liegen, gewählt.

6.1.4 Bewegungsrichtungen relativ zum Körper

Bei der Untersuchung von Sprache begleitenden Gesten ist es wichtig, die Handbewegung relativ zum Körper zu kennen. Anhand dieser lassen sich verschiedene Gesten identifizieren. Damit können Bewegungssequenzen mit Richtungen analysiert werden, die Aufschluss auf verschiedene Verhaltensmuster geben. Bei der Analyse des Zusammenhangs von Sprache und Gestik ist es hilfreich, verschiedene Wörter mit Bewegungsrichtungen zu kennen. Speziell bei der Analyse von Gesten in Relation zu Wegbeschreibungen ist es aufschlussreich, die Worte mit der Bewegungsrichtung in Bezug zu setzen. Damit kann die aktive gestische Beschreibung mittels Handbewegungsrichtungen verschiedener Personen genau miteinander verglichen werden.

Zur Berechnung der Bewegungsrichtung wird die Orientierung des Rückens verwendet, um die Positionen der Hände bezüglich des Körpers zu ermitteln. Dazu wird die homogene Matrix mit der Position und Orientierung des Rückens invertiert. Anschließend können die globalen Koordinaten in lokale Koordinaten bezüglich des Rückens durch eine Multiplikation ausgerechnet werden. Wird eine Position mit einer Matrix multipliziert, wird eine Transformation (Rotation und Translation) auf diese Position ausgeübt. Durch die Invertierung der Transformation werden globale Koordinaten in lokale umgerechnet.

$$M^{-1} * P_{Global} = P_{Lokal} \quad \textbf{(25)}$$

Die Bewegungsrichtung lässt sich aus zwei aufeinanderfolgenden Frames durch Subtraktion einer Position von einer anderen Position ausrechnen[56]. Je nachdem, welche Komponente des Differenzvektors (X, Y, oder Z) den größten absoluten Wert hat, ist dieses die entsprechende Bewegungsrichtung[57], die den verschiedenen Richtungsbezeichnungen entspricht (siehe dazu

[56] Die Berechnungen finden in einem diskreten Raummodell statt.

[57] Die Transformation des lokalen Koordinatensystems wird zu jedem Frame neu durchgeführt, um auch Änderungen der Ausrichtung der Probanden zu ihren Händen mit zu berücksichtigen.

Abbildung 41). Bei dem aktuellen Koordinatensystem, welches aus dem Rigidbody vom Rücken verwendet wird, ist nur die Yaw-Achse (Y-Achse in der Abbildung) von Interesse. Zum Annotieren der Bewegungsrichtungen der Hände gilt das Interesse nur der Orientierung des Probanden, nicht aber, wie schräg oder wie gebeugt sie sich positioniert[58].

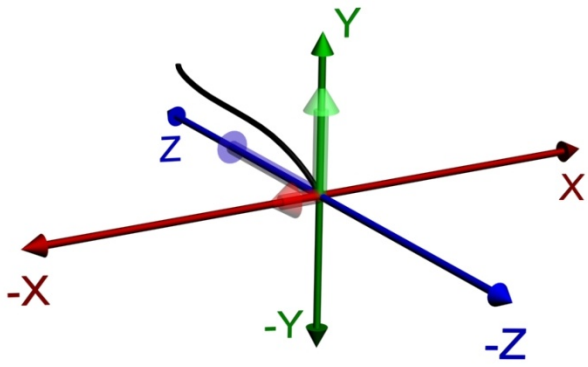

Abbildung 41 Bewegungsrichtungserkennung, bei der die größte Bewegung entlang der Z-Achse und entlang der Y-Achse aufgetreten ist

Nur die Bewegung aus der Sicht des Körpers ist nun von Interesse, nicht der Bezug zu den anderen Achsen. Ist der Wert der Komponente x am größten, liegt eine Bewegungsrichtung wie „rechts“ oder „links“ vor, entsprechend y „hoch“ oder „runter„ und z „vor“ oder „zurück“. Das Vorzeichen dieses Wertes gibt die genaue Richtung an, ob links oder rechts im Falle der X-Achse. Darüber hinaus können noch detailliertere Einschränkungen gemacht werden. Wenn zwei der drei Werte ähnlich groß sind oder eine der Komponenten nur mindestens 50 % kleiner ist, kann eine Klassifizierung als schräg, z. B. nach links-oben, vorgenommen werden. Bei der **Abbildung 41** ist die größte Bewegungskomponente entlang der Z-Achse und entlang der Y-Achse ist sie am zweitgrößten (größer als 50 % von der Z-Achse); daher würde die Bewegungsrichtung als nach „hinten“ und „hoch“ klassifiziert werden.

6.1.5 Segmentierung der Bewegungsrichtungen

Damit die Daten der Bewegungsrichtungen, z. B. von Händen, übersichtlich dargestellt werden können, ist es praktisch, die Bewegungsrichtung nicht für jeden Frame einzeln zu klassifizieren, sondern Intervalle mit gleicher Bewegungsrichtung zu detektieren. Um eine Bewegungfolge enthalten in mehreren Frames zu segmentieren, werden alle Frames nacheinanderdurchgegangen und es werden jeweils zwei aufeinander folgende Differenzvektoren der Bewegung betrachtet. Diese müssen nicht zwei direkt aufeinander folgenden Frames entsprechen. Durch die Benutzung der Definitionen des Skalarproduktes zweier Vektoren kann der Winkel zwischen diesen berechnet werden.

[58] Die Erwartungshaltung bei der Annotation verlangt hier die globalen Koordinatenachsen mit den modifizierten Yaw-Winkeln und der Positionierung im Raum.

$$\alpha = \text{ArcCos}\left(\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}\right) \quad \text{mit } \vec{a} \text{ und } \vec{b} \text{ Vektoren} \qquad \textbf{(26)}$$

Die Berechnung des Winkels zwischen diesen beiden Vektoren lässt sich als ein Hilfsmittel für die Unterteilung oder Segmentierung verwenden. Der Schwellwert oder Grenzwinkel kann frei definiert werden, je nachdem, wie genau die Unterteilung berechnet werden soll. Ein weiterer Parameter besagt, wie weit die Positionen in der zeitlichen Abfolge auseinander liegen, entsprechend den Frames, zu denen die jeweiligen Differenzvektoren gebildet werden. Dadurch wird der Differenzwinkel zwischen den Vektoren bei einer nicht linearen Bewegung größer, und es können globalere Änderungen robuster detektiert werden. Werden unterschiedlich weit auseinanderliegende Frames verwendet, muss die Differenzwinkelgrenze entsprechend angepasst werden. In der **Abbildung 42** wird eine Sequenz von Differenzvektoren entlang einer Trajektorie einer Handbewegung abgebildet, bei der ein Winkel zwischen zwei aufeinanderfolgenden Frames dargestellt wird. An dieser Stelle ist der Winkel besonders groß, und die Folge ist eine Unterteilung der Trajektorie bei der Berechnung der Segmente. Finden zu einem Zeitpunkt keine Bewegungen statt, stellt dieses auch eine Grenze eines Segments dar.

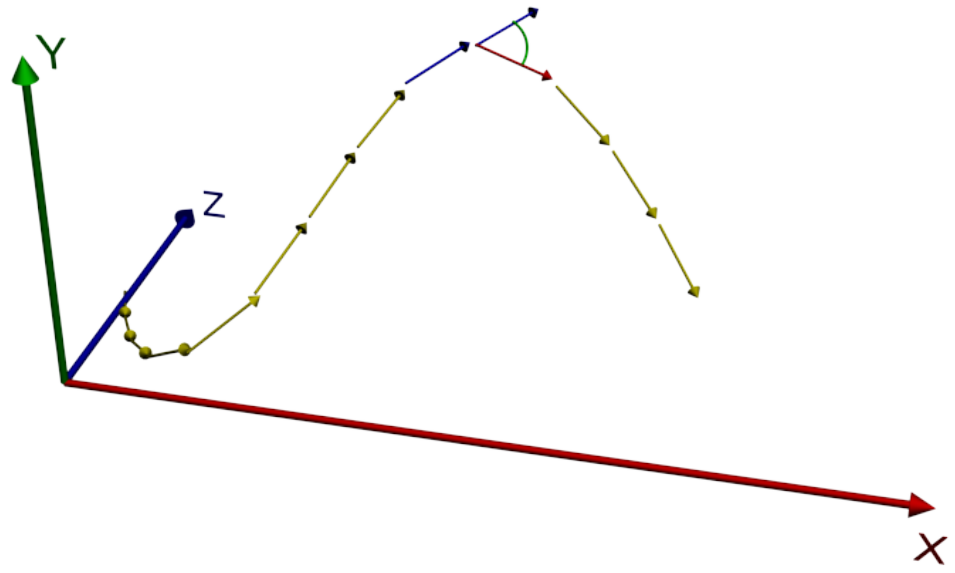

Abbildung 42 Trajektorie mit Differenzvektoren über mehrere Frames hinweg, für die zu einem Zeitpunkt ein Differenzwinkel berechnet wird.

6.1.6 Phasen der Bewegungssegmentierung und Erkennung

Kita [44] hat eine Unterteilung der Bewegung in Phasen (siehe Abschnitt 2.4.2) als Annotationsschema beschrieben, um diese einheitlich bezeichnen und vergleichen zu können. Auf Basis dieser Vorarbeit[59] wird im Folgenden beschrieben, wie diese Unterteilung in Bewegungsphasen automatisch annotiert werden soll. Dazu werden verschiedene Bestandteile benötigt. Der erste Bestandteil, der dazu nötig wird, ist die Ermittlung der aktiven Bewegungs-

[59] Genauer beschrieben in Sektion 2.4.2.

phasen der Hände. Der zweite Bestandteil ist die Ermittlung einer gegebenen auftretenden kurzen Haltephase. Der dritte Bestandteil ist die Berechnung der Ruhepositionen, um einen unvollständigen Rückzug der Hände detektieren zu können, wie er in Abschnitt 6.1.3 beschrieben wurde. Der vierte Bestandteil ist die Segmentierung der relativ zum Körper durchgeführten Bewegung, basierend auf der Bewegungsrichtung in Sektion 6.1.4 und der Segmentierung, die in Sektion 6.1.5 beschrieben wurde. Damit kann die Bewegungsrichtung mit ihrer Position und Orientierung zu den jeweiligen Segmenten bestimmt werden. Nach [44] muss zusätzlich die verwendete Kraft in den einzelnen Segmenten der Bewegung berechnet werden, damit diese Kräfte, die in den jeweiligen Segmenten aufgebracht wurden, zueinander ins Verhältnis gesetzt werden können. Dieses wird im Folgenden beschrieben. Damit lassen sich die bedeutungsvollen Segmente von den übrigen Segmenten wie zum Beispiel der Vorbereitungs- oder der Rückzugsphase unterscheiden.

6.1.6.1 Berechnung der Kraft in einzelnen Phasen

Die benötigte Funktionalität der Messung der Kraft in den einzelnen Bewegungsphasen wird durch das physikalische Gesetz F=m*a beschrieben. Allgemein ist nicht die eigentliche Kraft von Interesse bezogen auf die einzelnen Phasen, sondern die aufgebrachte Beschleunigung relativ zu den anderen Phasen[60]. Daher wird die jeweilige absolute Beschleunigung von einem Zeitpunkt bis zum nächsten Zeitpunkt innerhalb eines Bewegungsrichtungssegments aufsummiert.

$$aufgewendeteBeschleunigung(x) = \sum_{k=start}^{end} Beschleunigung(k)$$

$$Beschleunigung(i) = abs(\frac{g_{i-1} - g_i}{t_{i-1} - t_i}) \quad \textbf{(27)}$$

Dabei steht g für die Geschwindigkeit, t für einen Zeitpunkt, Start für die Nummer des Frames, bei dem das Bewegungssegment beginnt und entsprechend endet. Um die Kraft bei der Rückzugsphase in einer Ruhephase angemessen zu gewichten, wenn die Bewegung von oben nach unten durch die Gravitationsbeschleunigung beeinträchtigt werden kann, muss die entsprechende Beschleunigung der Bewegungsrichtung der Hand nach unten reduziert und bei einer Aufwärtsbewegung vergrößert werden[61], da sonst die Unterscheidung der Bewegungsphasen bei kleineren Bewegungen und bei der Zurücksetzung in die Ruheposition verfälscht würde.

[60] Das Interesse gilt dem aufgebrachten Unterschied der Kraft in verschiedenen Bewegungsphasen. Dabei ist die beschleunigte Masse konstant, es sei denn, ein Gegenstand würde gegriffen, was jedoch bei den betrachteten Szenarien nicht der Fall sein wird.

[61] Es ist nicht genau klar, wieviel Einfluss die Gravitation bei der Auf- und Abwärtsbewegung im Detail hat, da der Arm bei der Abwärtsbewegung abgebremst und nicht einfach fallen gelassen wird. Hierzu hat Kita [43] keine Angaben gemacht, natürlich gelten die physikalischen Gesetze der Gravitation.

Allerdings haben sich die Menschen mit der Zeit an die Gravitationskraft gewöhnt und können diese für sich nutzen. Daher ist die Beeinträchtigung durch die Gravitation bezogen auf die Bewegung sehr gering. Aus diesem Grund wird hier nur eine Annäherung benutzt, sodass die Auswirkungen der Gravitationsbeschleunigung nur zu einem Teil zur Gesamtbeschleunigung aufsummiert werden. Die Richtung der Gravitation wird entlang der y-Achse angesetzt und mit der Bewegungsrichtung der Hand in Relation gebracht, sodass die relative Beschleunigung durch die Länge des resultierenden Differenzvektors der Bewegungsrichtung der Hand und der Gravitation beschrieben werden kann. Wenn g die Gravitationsbeschleunigung bezeichnet und b_x, b_y, b_z die Komponenten der Beschleunigung in x-, y- und z-Richtung, so resultiert die Gesamtbeschleunigung

$$a = \sqrt{(g + b_y)^2 + {b_x}^2 + {b_z}^2} \quad \textbf{(28)}$$

Diese trifft für alle Zeitpunkte der Bewegung zu, daher lässt sich die gesamte absolute Beschleunigung in einer Bewegungsphase als die Summe der einzelnen Beschleunigungen ausrechnen.

$$\sum_{i=Anfang}^{end} a_i = \sum_{i=Anfang}^{end} \sqrt{(g + b_{i_y})^2 + {b_{i_x}}^2 + {b_{i_z}}^2} \quad \textbf{(29)}$$

6.1.6.2 Berechnung der verschiedenen Phasen

Nun sind alle Bestandteile zusammen beschrieben, die nötig sind, um die fünf verschiedenen Bewegungsphasen automatisch detektieren zu können. Nach Kita [44] sind diese verschiedenen Phasen (siehe Sektion 2.4.2):

- Bewegungszug (engl. **stroke**)
- Halten (engl. **hold**)
- Vorbereitung (engl. **preparation**)
- Rückzug (engl. **retraction**)
- unvollständiger Rückzug (engl. **partial retraction**)

Um diese verschiedenen Phasen zu unterscheiden, werden verschiedene Schritte nacheinander abgearbeitet.

1. Einzelne Phasen ermitteln

(a) Aktivitätsermittlung

Die Unterteilung in einzelne Phasen fängt mit der Ermittlung der Zeitpunkte an, wann die Hände überhaupt aktiv sind. Dabei können auch verschiedene kleine Phasen mit in diesen

Phasen enthalten sein, bei denen sich die Hand nicht bewegt, z. B. wie beim Zeigen. Damit sind der Anfang und das Ende einer Geste bekannt, gegebenenfalls auch der Anfang und das Ende einer Haltephase dazwischen.

(b) Segmentierung

Um die einzelnen Phasen voneinander zu segmentieren (trennen), wird der Winkel zwischen aufeinanderfolgenden Zeitpunkten in der Bewegung bestimmt, und wenn dieser einen Grenzwinkel überschreitet, wird eine Unterteilung vorgenommen. Dabei kann der Winkel je nach gewünschter Feinheit justiert werden. Ein weiteres Merkmal zur Segmentierung ist, wann die Geschwindigkeit zu Null wird. Treten eine Richtungsänderung und eine Geschwindigkeitsunterbrechung an einer Stelle auf, wird dort eine Segmentgrenze erstellt. Tritt nur eine Richtungsänderung ohne eine Unterbrechung der Geschwindigkeit auf, wird das Segment als „Multisegmentphase“ bezeichnet.

2. Phasen-Kategorisierung

(a) Stroke-Phase

Eine Stroke-Phase beinhaltet mehr Kraft als die umliegenden Phasen, dabei wird die Gravitation zum Teil mit berücksichtigt.

(b) Hold-Phase

Ist eine Phase einer aktiven Phase zugeordnet, in der keine Bewegung stattfindet, spricht man von einer Hold-Phase, ausgehend von der Definition der aktiven Phase.

(c) Preparation-Phase

Eine Vorbereitungsphase (engl. **preparation**) beginnt, nachdem sich die Hand in einer Ruheposition befindet, die keine Stroke-Phase ist oder zwischen zwei Stroke-Phasen liegt.

(d) Retraktion-Phase

Eine Nachbereitungsphase ist eine Phase, bevor die Hand in die Ruheposition geht.

(e) Partial-Retraction-Phase

Durch die Ermittlung der Ruhepositionen kann unterschieden werden, ob vielleicht nur ein schneller Neustart einer neuen Bewegungsgeste stattgefunden hat. Die Phase des teilweisen Rückzuges in die Ruheposition wird einmal durch die Bewegungsrichtung und durch die Distanz zu einer möglichen Ruheposition berechnet. Dabei muss die Bewegungsrichtung auf eine Ruheposition zugehen und darf diese nicht erreichen.

6.2 Gruppeninteraktionsphänomene

Phänomene, bei denen mehrere Personen beteiligt sind, werden als Gruppeninteraktionsphänomene bezeichnet. Ein Beispiel hierfür ist die Situation, in der sich zwei Personen zueinander orientieren. Im Folgenden wird beschrieben, welche Interaktionsphänomene bei den er-

stellten Korpora von Interesse waren und zur automatischen Annotation zur Verfügung stehen.

6.2.1 Orientierungsfokus

Das erste Phänomen ist die Situation, in der sich eine Person auf eine andere Person (oder ein Objekt) orientiert. Dabei wird unterschieden, ob dies mit den Augen oder mit dem Kopf geschieht. Mittels Eyetracking[62] könnte zwar die genaue Blickrichtung untersucht werden, dieses ist aber nicht in den Motion-Capture-Daten enthalten. Allerdings ist das Fokussieren mit dem Kopf eine interaktive Geste und wird schneller von einer anderen Person bemerkt. Diesbezügliche Studien mittels Motion-Capture- und Eyetracking-Systemen zeigen, dass das Fokussieren mit dem Kopf bei der Interaktion eine größere (interaktive) Rolle spielt als reine Augenbewegungen [77] [78]. Dieses Anschauen nur mittels der Augen wird nicht so schnell bemerkt, im Gegensatz zum Fokussieren durch Ausrichtung des Kopfes. Um Eyetracking nutzen zu können, würde spezielle Hardware benötigt, die einmal sehr teuer ist und zusätzlich störenden Einfluss auf das Geschehen in der Interaktion hat. Dadurch wiederum kann die automatische Erkennung von Emotionen in Gesichtern verhindert werden, welches zu einem späteren Zeitpunkt benötigt werden kann. Um zu ermitteln, wann eine Person möglicherweise eine andere Person oder ein Objekt anschaut, wird die Orientierung des Kopfes verwendet. Dabei ist es von Person zu Person und Situation unterschiedlich, wie stark dieses passiert.

Um zu ermitteln, wann eine Person einer andere in den Fokus nimmt, wird ein Strahl verwendet, der von der Position zwischen den Augen in Richtung der Z-Achse des Kopfes ausgeht. Der Versatz der Position zwischen den Augen und dem Kopf-Rigidbody muss im Vorfeld manuell für jede Person individuell ausgemessen werden. Damit kann die bestmögliche Genauigkeit bei der Detektion der Fokussierung von Personen erzielt werden. Der Strahl kann dann verwendet werden, um mathematisch zu prüfen, ob dieser mit einer virtuellen Geometrie kollidiert. Diese Geometrie kann von einer anderen Person sein, aber auch von einem Gegenstand, welcher im Aufnahmebereich bei der Motion-Capture-Aufzeichnung war. Um auszugleichen, dass die Augen sich relativ zum Kopf zusätzlich bewegen können, wird eine virtuelle Sphäre um die Objekte von Interesse (Köpfe der Personen) gelegt. Der Radius dieser Sphäre kann angepasst werden. Der Effekt ist, dass ermittelt werden kann, wann sich eine Person auch nur etwas auf eine andere orientiert. Die umschließende Sphäre muss deutlich größer sein als das eigentliche Objekt, um die Augenbewegung zu relativieren. Der Radius hängt von der Entfernung der einzelnen Personen voneinander und dem Sichtfeld des Auges, das ausgeglichen werden soll, ab. Um alle möglichen Zeitpunkte zu finden, wann eine Person anfokussiert wird, kann der Radius angepasst werden, wodurch die Detektion sensibler und auch fehleranfälliger wird. Dadurch kann die Größe einer kollisionsgeometrischen Kugel verändert werden. Eine größere Kugel wird von einem entsprechenden Sichtstrahl früher getroffen, wodurch die Detektion sensibler wird. Durch Untersuchungen ist ermittelt worden, dass ein

[62] Erfassung der Augen, um die genauen Blickrichtungen ermitteln zu können.

durchschnittliches Sichtfeld eines Menschen horizontal 190° und vertikal 150° umfasst [79] [80]. Durch dieses große Sichtfeld der Augen kann es vorkommen, dass ein Zeitpunkt ermittelt wird, bei dem nicht wirklich das Objekt angeschaut, sondern nur in die Nähe geblickt wurde. Es ist allerdings eine enorme Hilfe, nur einen möglichen Zeitpunkt zu kennen, wann ein Gegenstand, der von Interesse ist, betrachtet wurde. Im Detail kann der Zeitpunkt dann synchron mit dem Video angeschaut werden, aus dem dann eindeutig ersichtlich wird, ob ein Objekt bzw. eine Person betrachtet wurde oder nicht. Dabei wird in einem Dialog genauestens angezeigt, zu welchem Objekt hin sich die Person orientiert. In der 3D-Visualisierung der Motion-Capture-Daten wird das anvisierte Objekt durch eine umschließende dreidimensionale Box hervorgehoben, dargestellt durch Linien in einer individuellen Farbe[63], die jeder aufgenommenen Person zugewiesen wird [81].

6.2.2 Aufeinander orientieren

Das Phänomen des Aufeinanderorientierens basiert auf der Funktionalität des Orientierungsfokus. Es wird zusätzlich geprüft, ob es zwei Personen gibt, die sich jeweils auf den anderen orientieren. Dies ist speziell für die Analyse der Interaktion von Triaden interessant, da damit alleine auf der Basis der Motion-Capture-Daten ermittelt werden kann, wie die Rollenverteilungen im Gespräch sind. Die beiden anderen Probanden orientieren sich dem Sprecher zu. Der Sprecher orientiert sich dem Hauptgesprächspartner zu. Die dritte Person ist nur Zuhörer. In der **Abbildung 43** sind um die Köpfe herum einzelne Kollisionssphären visualisiert. Der jeweilige Kollisionsstrahl (versetzt zum Kopf-Rigidbody, ausgehend von den Augen) und das aktuell anorientierte Objekt werden durch eine Box in der Farbe der schauenden Person visualisiert. Der grüne und blaue Proband fokussieren den roten Probanden an, daher ist ersichtlich, dass der rot gefärbte Proband spricht. Da der blaue Proband von niemandem anfokussiert wird, kann geschlussfolgert werden, dass dieser gerade nur am Rande zuhört.

[63] Die Farbe ist dieselbe, die jedem Rigidbody-Set zugewiesen ist, um die einzelnen Sets voneinander zu trennen.

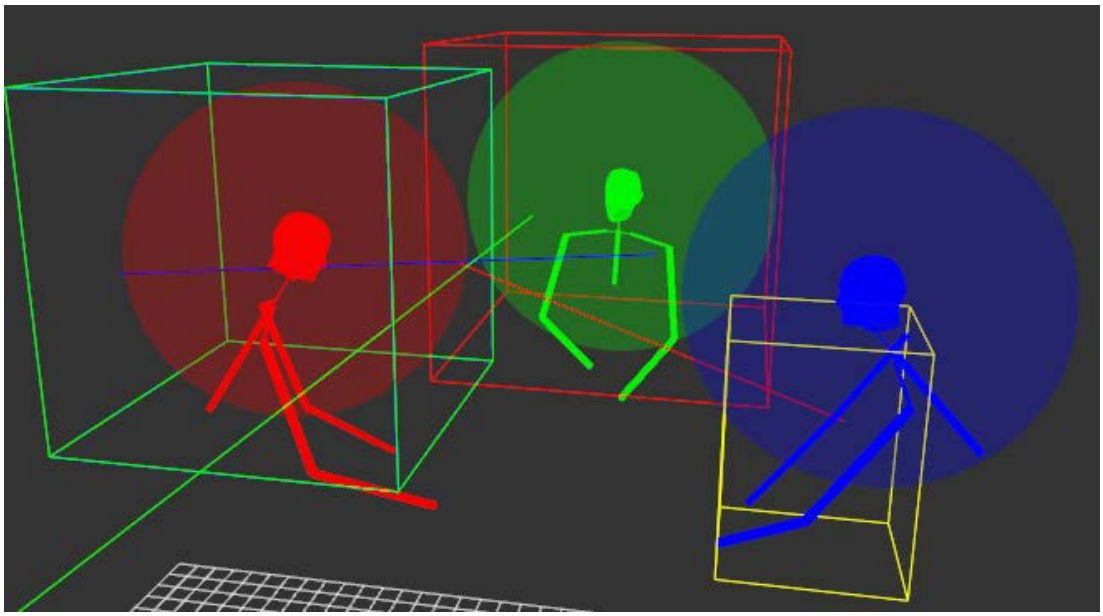

Abbildung 43 Ansicht der Detektion des Phänomens „sich zueinander Orientieren“ in einer Triade, bei der der grüne und der rote Proband sich gegenseitig ansehen und der blaue Proband dabei zuhört

6.2.3 Eindringen in den Personal-Space von anderen

Der sogenannte „personal space“ ist der Bereich um jede Person herum, der zu einem „gehört“, bzw. der Bereich, in dem man sich gestört fühlt, wenn eine fremde Person in diesen eindringt [82]. Dieser Bereich variiert je nach Person und Empfinden. Gemäß einer allgemeinen Schätzung wird dieser Bereich hier durch eine Armlänge zu jeder Person definiert. Diese Länge kann aus den Gelenkpositionen der Motion-Capture-Daten ausgerechnet werden und muss nicht manuell abgemessen werden. Da diese variiert, kann diese manuell nachgestellt werden. Konkret handelt es sich um die Annäherung einer fremden Hand zu einer anderen Person. Durch das Annähern einer Person zu einer anderen wird normalerweise der Aufmerksamkeitsfokus geändert, und es kann je nach Situation zu einem Sprecherwechsel kommen. Mathematisch wird der Abstand aller Hände (außer denen, die zu der jeweiligen Person gehören) zu dem Mittelpunkt des Torsos der anderen Personen auf die 2D x- und y-Ebene ohne Höhe projiziert und berechnet[64]. Dadurch wird eine Person als Zylinder repräsentiert, da Personen auch das Gefühl haben, ihnen wird zu nahe getreten, wenn sich etwas ihren Beinen oder Füßen nähert.

$$d_{x,y} = \sqrt{(x_p - x_H)^2 + (y_p - y_H)^2} \tag{30}$$

[64] Einige Analysen von Christian Schnier haben im Rahmen der Analysen zum Obersee Korpus gezeigt, dass Personen sich auch angesprochen fühlen, wenn sich ihnen eine Hand nähert.

Die Berechnung der Distanz d einer Hand zu einem Torso einer anderen Person wird somit in zweidimensionalen Räumen durchgeführt, da die Höhe als irrelevant für die Distanz zur Person gesehen wird[65].

6.3 Fehlerannotation

Leider ist das Motion-Capturing mittels Rigidbodys nicht ganz fehlerfrei. Bei den ersten Aufnahmen kam es vor, dass einzelne Rigidbodys nicht erkannt wurden oder dass die Orientierung springt. Um trotzdem mit diesen Daten arbeiten zu können, müssen diese Zeitpunkte des Springens markiert werden. Dazu wurde eine Detektion von Zeitspannen, in denen Rigidbodys nicht vorhanden waren, mit in die Annotation integriert. Zusätzlich wird auch annotiert, wann die Orientierung eines Rigidbodys flippt bzw. springt. Zur Ermittlung, wann ein Rigidbody verloren gegangen ist, werden die aktuellen Frames mit den vorherigen verglichen. Zur Berechnung der zeitlichen Abwesenheit wird der Zeitpunkt, an dem ein Rigidbody verloren gegangen ist, bis zum Wiedererscheinen gespeichert. Um Rotationsflips zu ermitteln, werden die Koordinatenachsen der jeweiligen Orientierungen mit den darauffolgenden verglichen. Ist der Winkel zwischen einer einzelnen Achse von einem bis zum nächsten Zeitpunkt größer als 90°, wurde ein Rotationsflip gefunden. Die Orientierungsänderung eines Körperteils kann sich nicht auf normalen Weg innerhalb 10 m/sec um 90° ändern. Da diese Rotationsflips vom Trackingsystem selber kommen und nicht um einen definierten oder berechenbaren Winkel erfolgen, wird nur der Bereich gekennzeichnet, bei dem dieser Flip stattfand. Die Resultate, wie häufig solche Fehler in den Korpora vorkommen, werden in Kapitel 9.1 vorgestellt. Im Folgenden wird eine Orientierungsmatrix mit ihren einzelnen Bestandteilen der Achsen des Koordinatensystems einzeln analysiert. Dabei entsprechen n, s, a den verschiedenen Achsen dieses Koordinatensystems.

$$M_i = (n_i \quad s_i \quad a_i) = \begin{pmatrix} x_{n_i} & x_{s_i} & x_{a_i} \\ y_{n_i} & y_{s_i} & y_{a_i} \\ z_{n_i} & z_{s_i} & z_{a_i} \end{pmatrix} \tag{31}$$

$$\text{ArcCos}\left(\frac{\vec{n_i}\vec{n_{i+1}}}{|\vec{n_i}||\vec{n_{i+1}}|}\right) > 90° \qquad \text{gleich für } s_i \text{ und } a_i \tag{32}$$

6.4 Zusätzliche Analyse Features

Neben den Annotationsfeatures, aufgeteilt in Einzel- und Gruppenpersonen-Phänomene, gibt es weitere Funktionen, die den Analyseprozess unterstützen. Diese werden im Folgenden beschrieben. Dazu zählen verschiedene Arten der Visualisierung als Ergänzung und Hervorhebung der Motion-Capture-Daten und die der GUI zugrundeliegende Funktionalität zur Inter-

[65] Es wird als besser angesehen, die Näherung des Personal Space als Zylinder anzusehen und nicht als Kreis, ausgehend vom Rücken oder Körperzentrum, da somit die eigentliche Körperform besser abgedeckt werden kann.

aktion mit dem Benutzer. Dazu zählen Funktionalitäten wie das Synchronisieren aller Visualisierungen von Daten wie die Motion-Capturing- und entsprechend Video-, Plot- und Annotationsdaten.

6.4.1 Multiple-Personen-Motion-Capture-View

Neben der automatischen Detektion von verschiedenen Phänomenen, die auf den Motion-Capture-Daten basieren, ist die Visualisierung von diesen und die Darstellung von mehreren Personen eine wichtige Funktionalität, um Gruppenverhaltensweisen aus allen möglichen Blickwinkeln zu analysieren. Es dient aber auch der Überprüfung vor der eigentlichen Aufnahme, ob alle Rigidbodys an den richtigen Personen und Körperteilen angebracht wurden. Es müssen die Positionen der jeweils acht Rigidbodys am Körper von drei Personen geprüft werden. Mit der Visualisierung der Motion-Capture-Daten ist es möglich, die Bewegungen der einzelnen Probanden in Relation zueinander und anderen Objekten zu analysieren. Dies ist die Grundlage für das Analysieren von Interaktionsverhalten in Gruppen. Um bei mehreren Personen im dreidimensionalen Raum den Überblick bewahren und eine detailliertere Analyse durchführen zu können, ist es hilfreich, die Darstellung in 3D-Stereo wie in der Realität zu betrachten. Dabei ist für jedes Auge ein eigenes speziell für dieses gerendertes Bild verfügbar, sodass es möglich ist, zu unterscheiden, welche Objekte in der Darstellung näher und welche weiter entfernt liegen. Damit ist genauestens zu sehen, welches Objekt oder welche Bewegung einer Person vor einer anderen Person oder einem anderen Objekt liegt. Die Steuerung bei der Betrachtung der virtuellen Aufnahme kann durch eine Wiimote[66] als Fernbedienung genutzt werden, mit der die virtuelle Umgebung mit den Motion-Capture-Daten durchlaufen werden kann. Mit dem Drücken von verschiedenen Richtungsknöpfen kann vorwärts und rückwärts gegangen werden, andere Knöpfe ermöglichen das Umdrehen und noch andere das Umschauen. Die Abspielzeit kann so manipuliert werden, dass ein Zeitpunkt festgehalten und aus verschiedenen Positionen analysiert werden kann. Es sind auch Beschleunigung und Verlangsamung der Abspielzeit der Motion-Capture-Aufnahme steuerbar. Darüber hinaus können diese Ansichten als Video gespeichert werden, um Einzelheiten präsentieren oder nachträglich mit anderen Leuten diskutieren zu können.

6.4.2 Virtuelle Aufnahmeumgebung

In ihrer reinen Darstellungsform sehen die Motion-Capture-Daten aus, als wären sie aus dem Kontext gerissen worden. Einerseits ist das gut, um die reine Bewegung genauestens analysieren zu können. Andererseits, da die Aktionen und Interaktionen der Probanden ohne die Umgebung, mit der sie interagieren, visualisiert werden, weiß man nicht genau, was die Probanden während der Interaktion genau machen[67].

---

[66] Es handelt sich um ein Bluetooth basiertes Eingabegerät mit Beschleunigungssensoren, das für eine Spielekonsole der Firma Nintendo entwickelt wurde.

[67] Besonderen Einfluss hat dieses, wenn ein Objekt wie ein Tisch ein zentraler Interaktionspunkt ist.

Es ist schwierig, die Bewegung in Relation zur realen Aufnahmeumgebung zu setzen. Zum Beispiel, wenn eine Gruppe um einen Tisch herum sitzt, ist es nützlich, Bewegungen in Relation zum Tisch zu sehen. In manchen Szenarios kann auch gerade diese Relation aus Bewegung und Umgebung das einzig Wichtige sein, an dem man interessiert ist. In dem „Kunsthallen“ Korpus wurde die Bewegung von Köpfen mit ihrer Orientierung aufgezeichnet. Das Experiment wurde in einer lokalen Kunsthalle aufgenommen, bei dem sich die Besucher verschiedene Gemälde angeguckt haben und dazu von einem kleinen Roboter namens Nao[68] über die Gemälde informiert wurden. Das Ziel der Studie war es, das genaue Interaktionsverhalten zwischen bis zu fünf Menschen mit dem Roboter und den Gemälden zu untersuchen [4]. Daher ist es wichtig, die Interaktion der Probanden im Bezug auf die Umgebung zu visualisieren und zu annotieren. Speziell die Orientierung der Köpfe ist von Interesse und ob diese auf eines der Gemälde oder den Roboter Nao ausgerichtet waren. Um dabei die Analyse zu erleichtern, wurde die Umgebung in exakter Relation zu der Bewegung virtuell nachgebildet. Dadurch kann nicht nur die Bewegung der Köpfe und separat eine Videoaufnahme zur Analyse verwendet werden, sondern auch die Bewegung in direkter Relation zu der Aufnahmeumgebung. Daraus ergeben sich neue Annotationsmöglichkeiten, z. B. kann automatisch mit annotiert werden, wann eine Person ein bestimmtes Objekt wie ein Gemälde, einen Tisch oder den Roboter anfokussiert.

Abbildung 44 Bewegung eines Kopfes mit einer virtuellen Rekonstruktion der Aufnahmeumgebung

Dazu wurden die wichtigen Gegenstände der Aufnahmeumgebung nachmodelliert, welches mit einem Zeitaufwand von ca. 1 Manntag durchgeführt werden konnte, da die genauen Abmessungen ermittelt und auf das Modell mit Texturen übertragen werden mussten. Auch der

[68] Nao ist ein kleiner humanoider Roboter, siehe dazu die **Abbildung 44** rechts unten.

Roboter Nao wurde als Modell mit Kopfbewegungen in die Aufnahme integriert; diese wurden mittels Motion-Capturing erfasst. Das Modellieren geschah in dem CAD Programm 3D Studie Max 2010, welches einen Export nach dem Standarddatenformat VRML besitzt. In der **Abbildung 44** ist ein Rigidbody mit einem zugehörigen Kopf gelb dargestellt und auf ein Gemälde blickend zu sehen. Der Roboter namens Nao steht in der rechten Ecke und sieht den Besucher an; dabei reagiert dieser unterschiedlich in den Bereichen, die durch die Distanzen zum Roboter auf dem Boden eingezeichnet sind. Zunächst versucht der Roboter, die Besucher zu interessieren, und dann - im näheren Bereich - zu informieren. Das Einbinden von 3D Models erfolgt optional über das Projektfile.

6.4.3 Visualisierung von Trajektorien

Eine weitere Funktionalität der Motion-Capture-Ansicht ist das Darstellen von Trajektorien. Dabei kann hier zwischen verschiedenen Visualisierungen gewählt werden. Die einfachste ist das Darstellen von Linien (ein Pixel Breite) und eine Darstellungsform, bei der nur jedes zweite Element visualisiert wird. Dadurch entsteht eine gestrichelte Linie, die es ermöglicht, leicht die Geschwindigkeit in den einzelnen Bereichen der Trajektorie zu erkennen. Dieses kann für jeden Rigidbody angezeigt werden, je nachdem, wie es gewünscht ist. Es kann auch nur eine Teilstrecke von einem bestimmten Frame an und bis zu einem anderen bestimmten Frame eingezeichnet werden. Das Ganze bietet eine Unterstützung bei der Analyse von komplexeren Bewegungen, da man so eine genaue Ansicht erhält, wann ein Körperteil sich wo und in welcher zeitlichen Abfolge aufgehalten hat. Es kann aber auch verwendet werden, um die Bewegungsmuster von Personen im Raum zu analysieren dem, wie es z. B. bei Kunsthallenexperiment der Fall ist. Mit Hilfe der Trajektorien ist es auch leicht, Bewegungsrichtungsänderungen zu annotieren, da man genau sieht, wann das Maximum bei der Bewegung in eine Richtung erreicht wird.

Abbildung 45 PAMOCAT im "Kunsthallenmodus" mit Trajektorien von drei Probanden

6.4.4 Multiple-synchroner Video-Player

Leider sind nicht alle Einzelheiten der Interaktion aus den Motion-Capture-Daten ersichtlich, da die Gesichtsmimik und die Finger nicht mit Motion-Capturing aufgezeichnet[69] werden. Allerdings können verschiedene Merkmale (wie Gelenkaktivitäten oder die Orientierung auf etwas wenden) in den Bewegungen des Körpers eine gute Identifikation für verschiedene

[69] Die Finger- und Gesichtsbewegungen könnten theoretisch auch mit aufgezeichnet werden, dabei wäre es aber nicht möglich, mit Rigidbodys zu arbeiten.

Verhaltensweisen sein. Damit können Zeitpunkte automatisch gefunden werden, bei denen verschiedene Verhaltensmerkmale auftreten. Diese können dann im Detail zusammen mit den verschiedenen Videos aus unterschiedlichen Sichten analysiert werden. Es ist möglich, an die Zeitpunkte, die durch Körperbewegung identifizierbar sind, zu springen und die Feinheiten gegebenenfalls in Videoaufnahmen zu analysieren. Dazu werden alle Videoaufnahmen zu anderen Videos und auch zu den Motion-Capture-Daten synchron gehalten. Zusätzlich kann die Zeit durch das Verschieben eines Sliders[70] frei gesteuert werden, um so frei die Video- und Motion-Capture-Daten analysieren zu können. Der Vorteil hierbei ist, dass man sich einen Zeitpunkt im Detail aus allen Kameraansichten[71] und den zugehörigen frei wählbaren Motion-Capture-Daten ansehen kann, um die genaue Interaktion der verschiedenen Personen miteinander analysieren zu können.

6.4.5 Plot von Winkel, Geschwindigkeit, Beschleunigung und Key-Intervalle der einzelnen Gelenke in einer Übersicht

Um Bewegungen im Detail zu analysieren, werden die Winkel, Geschwindigkeit und Beschleunigung der einzelnen Gelenke eingezeichnet. Dazu können einzelne Personen und einzelne Gelenke ausgewählt werden. Damit ist ein genaues Analysieren möglich, wann welcher Freiheitsgrad sich mit welcher Stärke verändert hat und wie einzelne DOFs zusammenhängen. Dies ist wiederum die Grundlage, um verschiedene Suchmuster (siehe in Sektion 6.5 für Details) zu definieren, um nach Zeitpunkten zu suchen, bei denen diese verschiedenen DOFs aktiv sind, und um verschiedene Verhaltensmuster in einem großen Korpus zu finden und zu prüfen. Zum Beispiel können alle möglichen Zeitpunkte, bei denen mit einer Hand gewunken wird, über einen aktiven Freiheitsgrad im Handgelenk gesucht werden. Aber auch eine Bejahung kann durch den entsprechenden einzelnen DOF im Gelenk am Kopf gefunden werden. Die Plots der Winkel liegen eng mit den im Key-Intervall berechneten zusammen und werden synchron angezeigt. Die Übersicht in **Abbildung 46** zeigt alle Gelenke und den genauen Zeitpunkt an, an dem alle DOFs der Gelenke aktiv sind. Die Key-Intervalle des selektierten Gelenks werden durch eine blaue transparente Linie hervorgehoben (siehe **Abbildung 46**). Der Winkel wird im unteren Bereich rot eingezeichnet, grün ist der rekonstruierte Winkel und blau der Winkel, nachdem er geglättet wurde (über Mittelung einer einstellbaren Anzahl). Durch die Erzeugung der Key-Intervalle gehen Informationen in dieser Darstellung der Bewegung verloren, der rekonstruierte Winkel zeigt die Auswirkungen diesbezüglich an, wodurch eine Anpassung der Parameter vorgenommen werden kann. Die Mittelung oder Glättung ist nötig, da die Daten leicht verrauscht ankommen. Dieses Verrauschen der Daten liegt daran, dass die ViconSDK zwar vorsieht, die aufgenommenen Zeitpunkte jedes einzelnen Frames mit zu verschicken, dies aber in der Praxis nicht tut. Daher muss nachträglich ein

[70] Ein Slider heißt zu Deutsch Schieber und ist ein GUI-Element, mit dem Einstellungen leicht durch das Verschieben verändert werden können.

[71] Es können beliebig viele Videos synchron zueinander gehalten werden, limitierende Faktoren sind die CPU und der Arbeitsspeicher des Rechners.

Zeitstempel erzeugt werden. Dieses hat zur Folge, dass Netzwerkverzögerungen mit in den Zeitstempel in kleinem Maße integriert werden. Bei der Betrachtung des reinen Winkels ist das noch kein Problem, aber bei der Betrachtung von kleineren Werten kommt der Zeitstempeleinfluss größer zur Geltung. Daher müssen die Daten für die Interpretation bezüglich der Key-Intervalle vorverarbeitet werden, damit eine exakte Trennung in Intervalle mit Bewegungen entlang der jeweiligen einzelnen DOFs bis zu den Extrema möglich ist. In der **Abbildung 46** wird die Geschwindigkeit orange und die Beschleunigung durch eine pinkfarbende Linie eingezeichnet. Damit die Geschwindigkeit und Beschleunigung sichtbar sind, sind sie in der Winkelanzeige skaliert dargestellt. Die Winkelanzeige und die Key-Intervallanzeige sind zeitlich unterschiedlich skaliert.

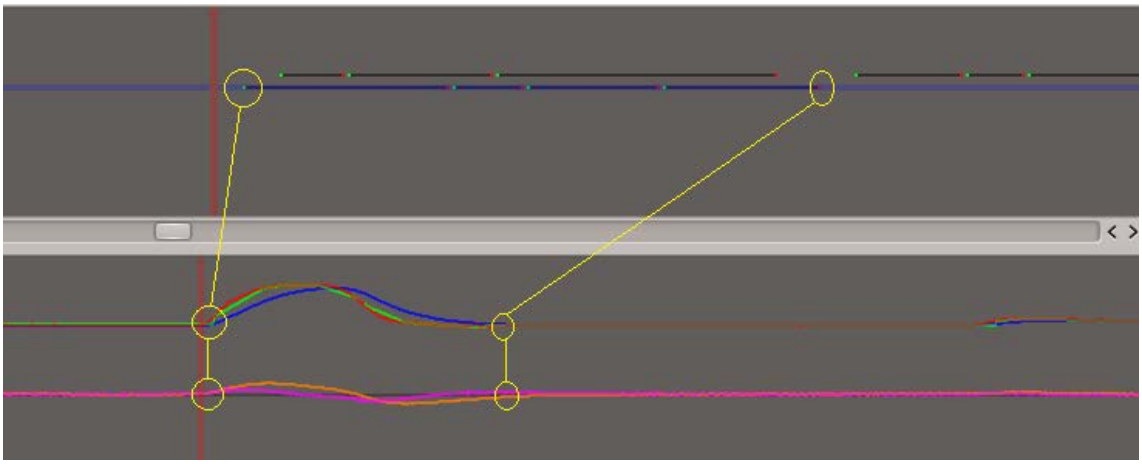

Abbildung 46 Key-Intervall Übersicht und Plot von Winkel, Geschwindigkeit[72] und Beschleunigung, dabei sind die Key-Intervall-Darstellung und der Plot zeitlich im Verhältnis 1:3 skaliert und in der Darstellung wurde mit gelber Farbe nachträglich die Key-Intervall Übersicht mit dem Winkelplot in Relation gebracht.

6.4.6 Zusammenführen von Annotationen

Die Funktionalität des Zusammenführens von Annotationen kann genutzt werden, um z. B. manuelle Annotationen zu vereinigen, aus denen eine Annotation mit höherer Qualität entsteht. Dabei können beliebige (auch externe) Annotationen Tier-weise zusammengeführt werden. Intern entspricht die Zusammenführung beider Tiers der Suche mittels des „und“ Operators auf beiden Tiers, das heißt, nur Annotationselemente, die auf beiden Tiers existieren, werden in das neue Tier übernommen. Eine weitere Verwendung der Zusammenführung ist, die Abhängigkeiten von anderen oder Gemeinsamkeiten einzelner Phänomene zu ermitteln. Eine einfache Abhängigkeit kann zum Beispiel zwischen verschiedenen DOFs bestehen. Bei der Geste des Winkens ist die gemeinsame Aktivität in den Gelenken der Hand und der Schulter zu finden. Vereinigt man diese Suche entlang beider Tiers miteinander (zu einem neuen

[72] Um Geschwindigkeit und Beschleunigung mit darstellen zu können, sind diese skaliert, damit Änderungen wahrnehmbar werden.

Tier), kann ein Korpus viel schneller auf diese Verhaltensweise hin durchsucht werden, da weniger Elemente insgesamt existieren.

6.4.7 Vergleichen

Um einzelne Phänomene auf Abhängigkeiten zu analysieren, ist es wichtig zu wissen, wie genau sie übereinstimmen. Das Vergleichen passiert in drei Schritten. Zum einen wird ein exakter Vergleich Frame für Frame durchgeführt. Da die verschiedenen Annotationen (bzw. einzelne Tier) nicht genau gleich sind[73], ist es wichtig zu wissen, ob während eines Annotationselements aus der einen Annotationen eine Aktivität auch in der anderen Annotationen vorhanden ist. Daher wird für jedes einzelne Element geprüft, ob während der Zeit auch Aktivität im anderen vorkommt. Außerdem ist es wichtig, wie genau diese aktiven Elemente miteinander übereinstimmen; dazu wird geprüft, ob während der gesamten aktiven Zeit auch ein Element im anderen Tier aktiv war. Dieses wird für beide, jeweils ausgehend von beiden Tiers, durchgeführt und mit der kompletten Übereinstimmung zu gleichen Teilen gewichtet. Das Resultat ist eine Angabe in Prozent.

6.5 Konstellationensuche

Komplexere Gesten, bei denen gleich mehrere DOFs und auch Gelenke beteiligt sind, können durch die Kombination von Aktivitäten der verschiedenen DOFs gefunden werden. Zum Beispiel können alle Zeitpunkte, bei denen ein Schlag vorgekommen sein könnte, durch eine Aktivität im Schultergelenk und im Ellenbogengelenk gefunden werden. Alle bis jetzt beschriebenen Phänomene können für die Untersuchung verschiedener Verhaltensweisen in Kombination miteinander von Interesse sein. Dazu hat jedes Phänomen ein eigenes Tier.

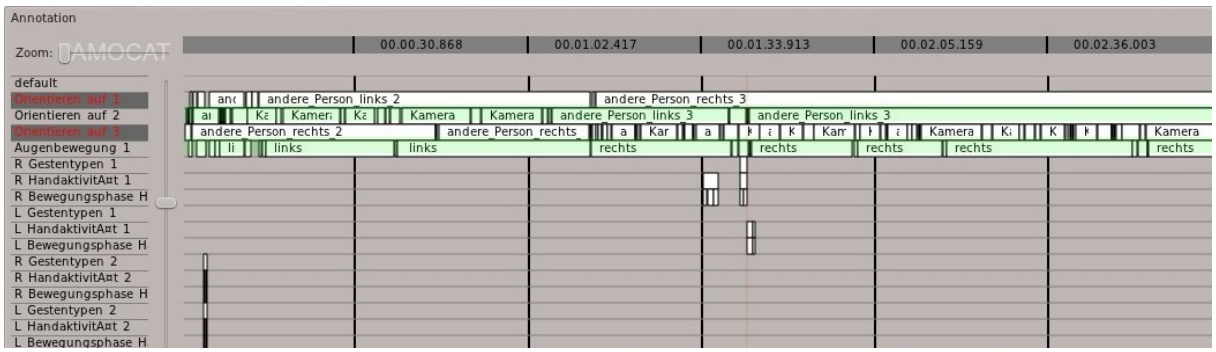

Abbildung 47 Tiers bezogen auf Phänomene, bei denen verschiedene Phänomene zur Suche ausgewählt werden können mit einem Knopf zur Änderung des logischen „Oder“ Operators zwischen den Pfeilen für die Vorwärts- und Rückwärts-Suche

Um Konstellationen von verschiedenen Phänomenen zu finden, kann ein logischer Operator ausgewählt werden und dann kann anschließend nach den speziellen Phänomenen in Kombination mit anderen Phänomenen gesuchen werden (siehe **Abbildung 47**). Dazu können auch

[73] Z. B. wenn mehrere Personen das gleiche Annotieren.

dynamische und statische Phänomene miteinander kombiniert werden; zum Beispiel bei der Suche nach einer Zeigegeste, bei der eine Pose in Kombination mit einer aktiven Hand gefunden werden muss. Im Normalfall muss nur die Pose mit einer bestimmten Armhaltung gefunden werden; allerdings, wenn die Versuchspersonen um einen Tisch sitzen, kann es vorkommen, dass einer dieser Probanden die Arme und Hände auf den Tisch legt und dies der gesuchten Pose zu ähnlich ist. Dies kann durch die Kombination des Phänomens der speziellen Pose und Handaktivität unterschieden werden. Es wird jeweils zum Anfangszeitpunkt der Intervalle gesprungen, damit diese dann abgespielt werden können. Ein solches Suchmuster beschreibt dabei die Suche nach Aktivität entlang von verschiedenen Tiers. Das Suchmuster umfasst dabei ein oder mehrere Tiers und einen logischen Operator wie „and“ oder „or“, der für alle gleich gilt. Beim Operator „or“ wird zu der Aktivität entlang der Tiers gesprungen und bei dem Operator „and“ nur zu den Zeitpunkten, bei denen Aktivitäten entlang aller selektierten Tiers vorkommen.

6.6 Zusammenfassung

In diesem Kapitel wurden verschiedene Funktionalitäten vorgestellt, die Aspekte von automatischem Annotieren beinhalten, wie „Einzelpersonen-Phänomene“ und „Gruppen-Interaktions-Phänomene“. Darüber hinaus wurden auch weitere funktionale Features vorgestellt, die das spätere Analysieren der Daten unterstützen und auch neue Möglichkeiten bieten, diese zu analysieren. Dazu zählt, dass verschiedene Kombinationen von Phänomenen ausgewählt werden können. Die entsprechenden Zeitpunkte können mit verschiedenen Datenaufbereitungen wie synchronen multiplen Videos, Winkel-, Geschwindigkeits- und Beschleunigungsplots analysiert werden. Und dazu zählt auch die Funktionalität, die es ermöglicht, die Bewegung in Relation zu Objekten im Aufnahmebereich von allen Seiten mit Trajektorien und Kopforientierung zu betrachten. Zudem kann die gesamte Bewegung in elementare Bewegungen zerlegt werden, wenn die Bewegung von einem einzelnen Gelenk von Interesse ist, um z. B. herauszufinden, wann sich der Kopf entsprechend einer Verneinung bewegt oder wann sich die Hand seitlich bewegt hat. Diese Bewegungsbestandteile werden jeweils als einzelne DOF in Tier mit entsprechender Aktivität annotiert. Mit diesen elementaren Bewegungen (dynamischen Bewegungsbestandteile) kann in Kombination (z.B. mit statischen Posen) nach verschiedenen Verhaltensweisen gesucht werden, bei denen typische Aktivitäten bei bestimmten DOFs herrschen. Dabei sind die integrierten automatischen Annotationen als Phänomene in **Tabelle 10** zusammengeführt.

Phänomen	Beschreibung
Key-Intervalle	Annotiert die Bewegung des ausgewählten Probanden entsprechend der Aktivität in allen Gelenken.
Posen	Annotiert die Körperstellung der entsprechenden ausgesuchten Posen, wenn eine einstellbare Übereinstimmung auftritt.
Handaktivität	Annotiert den Umstand, wenn die Hände eine einstellbare Geschwindigkeit überschreiten, und schließt eine definierbare Ruhephase (Zeigegesten) mit ein.
Ruheposen	Annotiert alle Ruhepositionen der Hände und clustert diese entsprechend einer einstellbaren Entfernung.
Segmentierte Trajektorien	Annotiert segmentierte Bewegungen entsprechend eines einstellbaren Winkels oder entlang von Weltkoordinaten.
Bewegungsrichtungen	Annotiert Bewegungsrichtungen bezüglich Weltkoordinaten, aber auch bewegliche Koordinaten wie die eines Rückens.
Orientierungsfokus	Annotiert, wann und welches virtuelle Objekt vom Kopf eines Probanden anfokussiert wurde.
Zueinander Orientieren	Annotiert, wann zwei Personen aufeinander ausgerichtet sind.
Personal Space	Annotiert, wann eine Hand einer Person einer anderen näher kommt.
Fehlerannotation	Detektiert und annotiert, wann Rigidbodys fehlen und Rotationsflips auftraten.
Zusammenführen	Führt zwei Tiers zusammen, um z. B.eine höhere Qualität von manuellen Annotationen zu erzeugen.

Tabelle 10 Aktuelle automatische Annotationen von PAMOCAT

7 Implementierung

In diesem Kapitel wird näher beschrieben, wie die Softwarearchitektur von PAMOCAT ist. Der Schwerpunkt liegt auf der statischen Beschreibung der Komponenten mit ihren Abhängigkeiten untereinander, um einen leichten Einstieg zur Erweiterung der Software zu ermöglichen. Dazu wird ein Gesamtüberblick der Softwarearchitektur gegeben, der anschließend durch die Beschreibung der einzelnen Komponenten im Detail vertieft wird. Eine detailliertere Beschreibung der grundlegendsten Klassen und deren Beziehung zu anderen Klassen mit Abhängigkeiten wird im Anhang B gegeben. Dazu wird die Entwicklungsumgebung mit Abhängigkeiten zu anderen Softwarekomponenten vorgestellt.

7.1 Softwareumgebung

Die Software wurde anfangs unter Suse Linux entwickelt; die wesentliche Entwicklung geschah später unter verschiedenen Ubuntuversionen. Zum Zeitpunkt der Veröffentlichung dieser Arbeit wird die Version 11.04 von Ubuntu verwendet. Bei der Entwicklung wurde darauf geachtet, möglichst plattformunabhängig zu sein. Daher sind die verwendeten Bibliotheken, auf denen PAMOCAT basiert, für Linux, Windows und MacOS erhältlich. Dadurch und dass Teile auch unter Windows entwickelt wurden, können andere Betriebssysteme[74] leichter unterstützt werden. Um nur die Annotationsfunktionen zu nutzen, kann ein einfacher PC (Dual Core 2,5 Ghz mit 3 GB RAM) verwendet werden. Um aktiv zu analysieren, ist ein PC mit mittlerer bis hoher Leistung (Quad Core 3 GHz und 6 GB RAM) vorteilhaft. Die Motion-Capture-Daten müssen von der Festplatte in den RAM gelesen werden, um in diesem schnell durch verschiedene Zeitpunkte navigieren und so Analysen durchzuführen zu können. Eine schnelle SSD Festplatte verkürzt dabei die Ladezeiten deutlich, und ein großer Arbeitsspeicher ermöglicht es, schnell sehr große Aufnahmen zu verarbeiten. Die Motion-Capture-Daten werden in den Arbeitsspeicher geladen, daher muss dieser ausreichend groß sein (eine Aufzeichnung mit 3 Personen mit einer Rate von 200 Hz und 30 min benötigt ca. 4GB RAM zusätzlich zum Betriebssystem). Dabei kann für langsamere PCs auch eine weniger hohe zeitliche Auflösung geladen werden, wie z. B. nur 25 Hz, wodurch der Speicherbedarf auf 500MB RAM sinkt. Die CPU sollte mehrere Kerne haben; allerdings ist hier noch Potential zur Optimierung einzelner Berechnungen durch eine stärkere Verteilung auf mehrere Kerne vorhanden. Der Hauptvorteil mehrerer Kerne ist aktuell bei der Darstellung der Analysen in den verschiedenen Modalitäten zu finden. Dabei arbeiten die Sensoren, das Motion-Capturing-View, jedes einzelne Video und die GUI in jeweils einem einzelnen Thread.

[74] Als nächster Schritt wird die Portierung von PAMOCAT nach Windows angesehen, um die Software möglichst vielen Anwendern bereitzustellen.

7.2 Abhängigkeiten

Die Software ist in C++ geschrieben und basiert auf verschiedenen Bibliotheken, die im Folgenden aufgeführt sind. Die größte Abhängigkeit besteht zu OpenSG, einer Bibliothek zum verteilten Rendering, was bei dieser Anwendung nicht Hauptmerkmal ist, aber von der historischen Entwicklung der Software stammt. Benutzt wird OpenSG zur Visualisierung von Bewegung und zur Darstellung von Bewegung in Relation zu relevanten Interaktionsobjekten. Außerdem wird OpenSG verwendet, um durch die aufgezeichneten Motion-Capture-Daten auch in Stereo 3D zu navigieren. OpenSG wurde vom Fraunhofer Institut in Deutschland zum verteilten Rendering in großen virtuellen Reality Anlagen wie einer CAVE[75] entwickelt. Die zweite Bibliothek ist QT4, welche die Basis der graphischen Benutzerschnittstelle, kurz GUI[76], darstellt. Die restlichen Abhängigkeiten bestehen zu kleineren Bibliotheken. Zum Komprimieren und Entpacken von Dateien wie den Motion-Capture-Daten wird die Bibliothek Quazip benutzt. Um den Benutzerinput als Input für die virtuelle Navigation durch die Aufnahmeumgebung mit den Motion-Capture-Daten entgegenzunehmen, werden die Bibliotheken CWiid und Wiimote verwendet, welche über Bluetooth Verbindung zu einer Nintendo Wiimote herstellen und den Benutzerinput verarbeiten. Für die Möglichkeit der Darstellung von Videodaten mit Unterstützung der meisten Codecs wird die Bibliothek Phonon verwendet. Um Daten strukturiert zu speichern, wird die Bibliothek XML2 verwendet. Die Motion-Capture-Daten werden durch die Bibliothek ViconSDK der Firma Vicon in Empfang genommen, um die Informationen zu erhalten, wo und wie die verschiedenen Körperteile orientiert sind. Die strukturellen Zusammenhänge dieser externen Komponenten und die Zusammenhänge der Komponenten mit ihren Abhängigkeiten sind in der **Abbildung 48** dargestellt. In dieser Abbildung sind die externen Komponenten, die verwendet werden, weiß eingefärbt und die Eigenentwicklung von PAMOCAT ist hellblau hervorgehoben. Im Folgenden werden diese selbst entwickelten Komponenten mit ihren Abhängigkeiten im Detail betrachtet.

7.3 Die ToolKit-Bibliothek

Die Bibliothek ToolKit[77] ist eine Sammlung verschiedener allgemeiner Klassen und Funktionen, die eine Vielzahl von Einsatzgebieten unabhängig von Motion-Capturing haben. Intern gibt es eine Aufteilung durch verschiedene externe Abhängigkeiten. Die Entwicklung des ToolKits wurde für OpenSG mit Hilfsklassen und Funktionalität angefangen, die seinen größten Teil ausmachen. Später wurde dieses mit der Teilkomponente zur Unterstützung von GUI-Elementen auf Basis QT4 und Phonon erweitert: eine Teilkomponente für verschiedene Dateiformate und Operationen, eine für Sensoren, und eine Basiskomponente mit Standards fürs Programmieren und Algorithmen für verschiedene mathematische Operationen.

---

[75] Eine Cave besteht aus mindestens 3 Wänden, auf die ein stereoskopisches Bild projiziert wird.

[76] GUI - Graphical User Interface

[77] Eine detailliertere Beschreibung ist im Anhang Kapitel B zu finden.

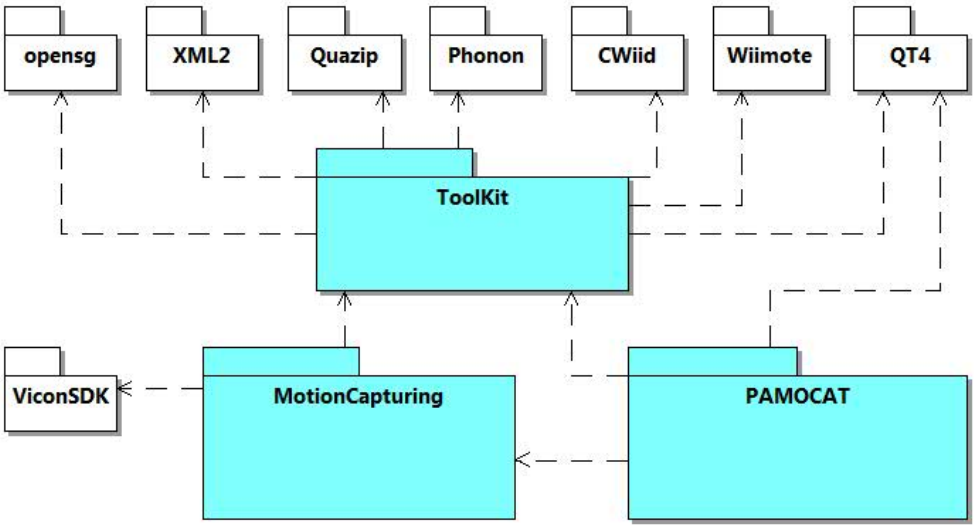

Abbildung 48 PAMOCAT Softwareabhängigkeiten

7.4 Die Motion-Capture-Bibliothek

Die Motion-Capture-Bibliothek bietet eine Gesamtfunktionalität für die Verarbeitung des Motion-Capturings und die automatischen Annotationen. Dazu ist sie wiederum aus verschiedenen Teilkomponenten aufgebaut, die unabhängig arbeiten und leicht in verschiedene Applikationen integriert werden können. Der Aufbau dieser Teilkomponenten wird im Folgenden vorgestellt. Anschließend werden die einzelnen Teilkomponenten nacheinander mit ihrer Funktionalität und Struktur beschrieben. Die **Abbildung 49** gibt eine Übersicht über die einzelnen Komponenten und deren Zusammenhänge. Angefangen wird mit den gespeicherten Daten, gefolgt von der darauf arbeitenden Kinematik. Um die Kinematik durchführen zu können, müssen die Bewegungsdaten und die Benutzerdaten geladen werden. Diese können dann verwendet werden, um automatische Annotationen zu speichern. Auf der Basis dieser Daten können Phänomene detektiert werde. Im Folgenden wird beschrieben, wie die jeweilig beteiligten Klassen zusammenhängen.

7.4.1 Datenstrukturen

Unter Datenstrukturen werden alle zu speichernden Daten aufgeführt. Dieses sind Rohdaten, Motion-Capture-Daten, benutzerspezifische Daten und die Annotationsdaten. In verschiedenen Komponenten werden die spezifischen Daten verarbeitet. Diese werden im Folgenden kurz vorgestellt. Sie bedienen dazu Interfaces (Schnittstellen), die von anderen Komponenten benötigt werden. Diese sind das Motion-Capture-Datenstruktur-Interface, das Benutzerdaten-Interface und das Annotationsdaten-Interface.

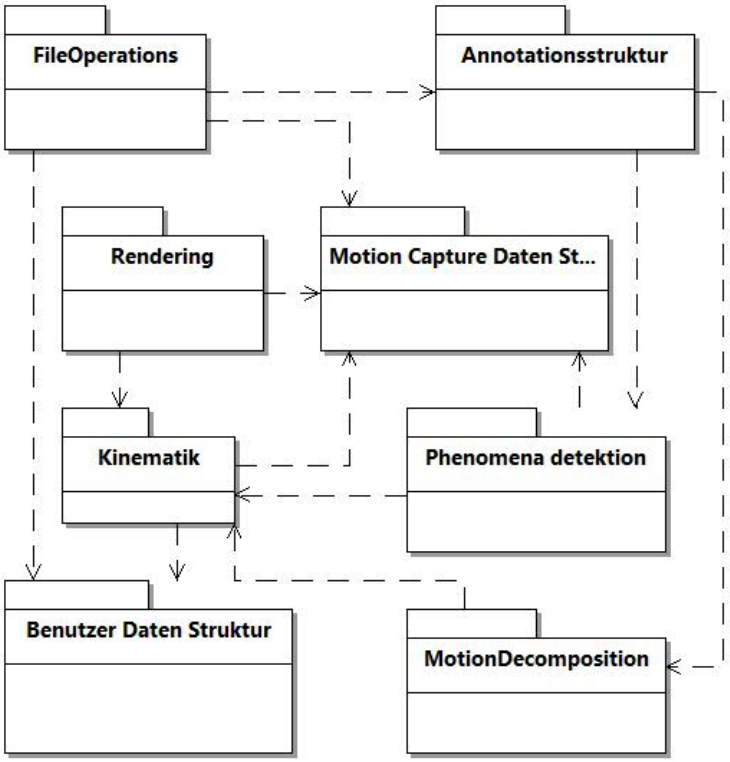

Abbildung 49 Übersicht über die Komponenten der Bibliothek Motion-Capturing

7.4.1.1 Motion-Capture Datenstruktur

Das Motion-Capture-Interface stellt die Bewegungsdaten der verschiedenen Personen bereit. Diese Bewegungsdaten beinhalten die Position und Orientierung einzelner Körperteile der verschiedenen Probanden. Jeder dieser Körperteile wird als ein Marker definiert mit einer ID und einer Position. Diese Markerdaten sind in jedem Frame gespeichert, in dem sie zu einem bestimmten Zeitpunkt existieren. Jeder Marker besitzt eine Referenz auf eine Klasse „MarkerProperties“, welche die Eigenschaften für jeden Marker verwaltet. Die Daten für „Rigidbody“ erweitern die vorhandenen Eigenschaften der „MarkerData“ um eine Orientierung. Ein Bestandteil der Funktionalität der Klasse „MarkerProperties“ ist es, Eigenschaften bezüglich der Marker bzw. Rigidbodys individuell zu verwalten. Diese sind Eigenschaften verschiedener Sichtbarkeiten, z. B. von Trajektorien, oder einer Beschriftung, aber auch die zugehörigen Offsets des Rigidbodys zum Mittelpunkt und der individuelle Offset zum Gelenk des Probanden. Eine vollständige Auflistung ist in der folgenden **Tabelle 11** aufgeführt. Der Offset der Ursprünge der Rigidbodys, der im ersten Marker des Rigidbodys liegt, zum eigentlichen Mittelpunkt des Rigidbodys ist in der „KalibrierungsDaten“ Klasse gespeichert. Der individuelle Offset der Probanden zu den einzelnen Gelenken wird auch von der Klasse „KalibrierungsDaten“ verwaltet. Zusätzlich können hier auch einzelne Korrekturen der Orientierungsoffsets mit eingebracht werden, falls die Rigidbodys nicht in der richtigen Position an dem Probanden platziert wurden. Jeder Marker und Rigidbody ist einem bestimmten Frame mit entsprechender Aufnahmezeit zugeordnet. Ist ein Rigidbody verloren gegangen, wird dieser nicht mehr im entsprechenden Frame gespeichert. Die gesamten Frames sind

Name (Attribute)	Beschreibung
Sichtbarkeit	Sichtbarkeit des gesamten Rigidbodys.
Beschriftung	Individuelle Sichtbarkeit des Bezeichners, der den Namen und die ID beinhaltet.
Trajektorien	Individuelle Sichtbarkeit der verschiedenen Trajektoriendarstellungen.
Koordinatenkreuz	Individuelle Sichtbarkeit der Orientierung durch mehrere orthogonal zueinander liegende Pfeile.
Zusatzgeometrie	Individuelle Sichtbarkeit von Zusatzgeometrie wie zum Beispiel ein Kopf.
Offset (Individuell)	Individueller personenabhängiger Offset des Rigidbodymittelpunkts vom eigentlichen Gelenk.
Offset (Center)	Rigidbody Offset der gelieferten Position in globalen Koordinaten zum Rigidbodymittelpunkt.

Tabelle 11 Eigenschaften der Klasse Markerproperties

entsprechend ihres Aufnahmezeitpunktes relativ zum Beginn der Aufnahme in einem Vektor[78] gespeichert. Dieses ist keine Liste, sondern ein Vektor, da meistens nicht auf das nächste Objekt, sondern auf ein bestimmtes Objekt an einer bestimmten Stelle zugegriffen wird (z. B. durch einen Slider[79]). Dieser Teilzusammenhang ist in der **Abbildung 50** dargestellt. Dieser Zugriff auf die einzelnen Frameelemente wird durch die Klasse „FrameDataVektor“ dem Interface bereitgestellt. Diese Klasse kontrolliert auch den zeitgesteuerten Zugriff auf die aktuellen Frames ausgehend vom Abspielzeitpunkt der Aufnahme und entscheidet, wann welcher

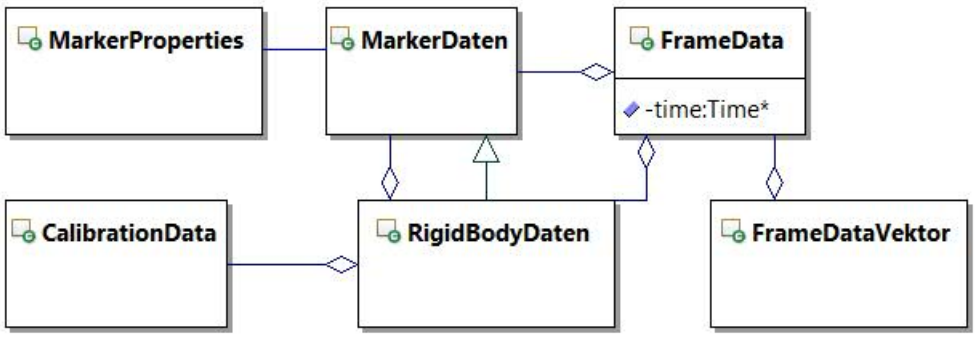

Abbildung 50 Klassendiagramm der Motion-Capture-Datenstruktur

[78] Ein Vektor kann als eine Liste mit dynamischen Größen von Elementen beschrieben werden.

[79] Ein Slider ist ein Schieberegler, dem ein minimaler und maximaler Wert zugewiesen ist und bei dem durch Verschiebung jeder Zwischenwert ausgewählt werden kann.

Frame aktiv ist und welcher ausgelassen wird. Die Abspielgeschwindigkeit kann in dieser Klasse beeinflusst werden, zum schnellen oder langsamen Abspielen.

7.4.1.2 Benutzerdatenstruktur

Alle Subkomponenten und Datenstrukturen sind im Core zusammengeführt. Der Zugriff auf diese Benutzerdaten wird über das entsprechende Interface bereitgestellt. Die Objekte der Klasse „MarkerProperties“ werden von einer Klasse „MarkerPropManager“ verwaltet, wodurch eine zentrale Schnittstelle bereitgestellt wird, die „MarkerProperties“ zu verändern. Die Klasse „SettingManager“ verwaltet die globalen Optionen wie z. B. „alle Trajektorien aus/an“, aber auch das Detektieren von Phänomenen wie z. B. „orientieren auf“ oder „Pose“, die damit für die aktuelle Darstellung berechnet werden. Benutzerspezifische Daten werden von der Klasse „UserManager“ verwaltet. Mit deren Hilfe kann bestimmt werden, welche Rigidbodys zu welchem Skelett gehören und wie groß die Offsets von den Rigidbodys zu den Gelenken sind. Dieser Teilzusammenhang ist in **Abbildung 51** dargestellt.

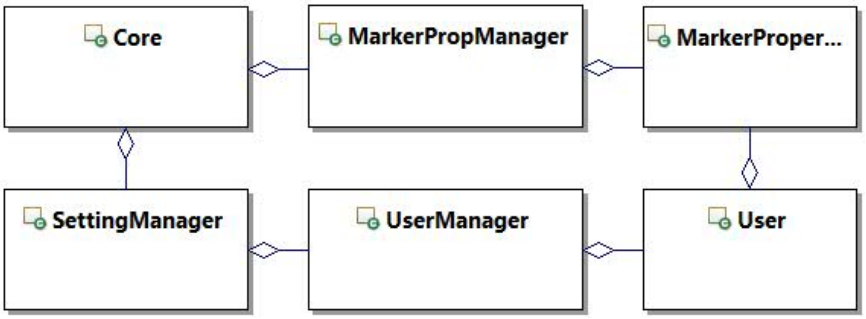

Abbildung 51 Klassendiagramm der Benutzerdaten

7.4.1.3 Annotationsdatenstrukturen

Die Annotationsdatenstruktur ist entwickelt worden, um das Fileformat „eaf“ zu integrieren, in dem das Annotationstool „ELAN“ seine Daten speichert. Das dazugehörige Interface ermöglicht, die Annotationsdaten für eine Visualisierung zu laden, zu manipulieren, hinzuzufügen und zu speichern. Dazu können Daten importiert und exportiert werden. Um z. B. manuell erstellte Annotationsdaten zusammenzuführen, können auch mehrere Daten untereinander gehängt werden. Ein einzelnes Annotat[80] besitzt einen Anfangs- und Endzeitpunkt, einen Annotationstext und eine Farbe. Diese verschiedenen Zeitpunkte werden global verwaltet, damit diese Informationen nicht mehrfach gespeichert werden. Das heißt, dass immer nur eine Referenz auf die eigentliche Zeit verwendet wird und ein Zeitpunkt nur einmal definiert ist. Die Verwaltung und auch die Garbagecollection[81] dieser Annotationszeitpunkte wird durch die Klasse „TimeReferenceManager“ durchgeführt. Ein Annotat ist immer genau einer „AnnotationsLinie“ zugeordnet. Die Annotationslinie bzw. Tier stellt sinngemäß eine Annotationska-

[80] Eine einzelne Beschreibung von etwas.

[81] Das Löschen von Zeitpunkten, auf die nicht mehr zurückgegriffen wird.

tegorie dar. Dieser Zusammenhang der Klassen ist in der **Abbildung 52** als Klassendiagramm dargestellt.

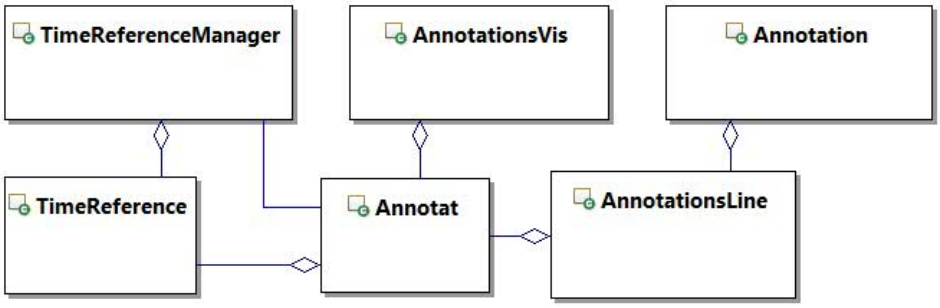

Abbildung 52 Klassendiagramm der Annotationsdatenstruktur

7.4.2 Kinematik

Die Kinematik ermöglicht es, einzelne Posen, aber auch komplexe Bewegungen des Skeletts mathematisch zu beschreiben. Wie auch in der Realität, besteht das Skelett aus einzelnen Knochen mit Gelenken (engl. „joint“). Die Klasse „Skeleton“ erbt von der Klasse „Joint“ die Fähigkeiten, weitere Gelenke einer Hierarchie unter sich zu verwalten und zu beeinflussen. Jedes Skelett besitzt benutzerspezifische Daten, die durch die Klasse „User“ dem Skelett zur Verfügung stehen. Dazu zählen die ID bzw. Namen der Rigidbodys, die einer Person und den verschiedenen Körperteilen zugeordnet sind, aber auch die Abstände der Gelenke von den einzelnen Rigidbodymittelpunkten. Die wichtigsten Informationen sind die jeweils aktuellen Positionen und Ausrichtungen der einzelnen Körperteile, um die gesamte Pose des Skelets zu berechnen. Diese Informationen können über die beiden Interfaces Motion-Capture und Benutzerdaten verwendet werden. Bei der Skelettinitialisierung werden die Längen der einzelnen Knochen berechnet, dazu wird geprüft, wann die erwarteten Körperteile alle zusammen vorhanden sind[82]. Siehe dazu die **Abbildung 53**, in der dieser Zusammenhang beschriebenen wird.

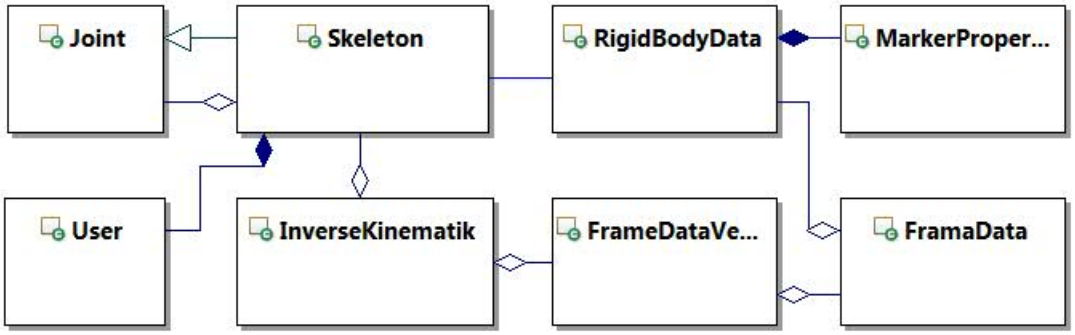

Abbildung 53 Diagramm der Klassen, die an der Kinematik beteiligt sind.

[82] Das heißt, wann alle Rigidbodys von den Sensoren erfasst werden.

7.4.3 File-Format

Im Rahmen des Motion-Capturings, der darauf basierenden Annotationen und des anschließenden Exportierens muss eine Reihe verschiedener Daten gelesen und gespeichert werden. Alle folgenden Datenformate basieren auf XML. Alle beteiligten Datenformate sind durch eine eigene Klasse beschrieben, deren Zusammenhang mit anderen Klassen wird in der **Abbildung 54** dargestellt. Die rot eingefärbten Klassen „FrameDataVector“ und „Annotation“ wurden schon zuvor detaillierter modelliert und sind zur Verdeutlichung der Relationen hier aufgeführt. Die rot und blau eingefärbten Klassen werden benötigt, um die Motion-Capture-Daten in Skelettbewegungen umrechnen zu können. In der Klasse „CMCFile“ werden die Skelettinformationen der Personen als Header[83] gespeichert, gefolgt von den Informationen, wann und wo welches Körperteil im Raum erfasst wurde. Um die Größe zu verringern, werden die Daten mit der Teilkomponente des „ToolKit“ und „File“ durch „FileZip“ Klasse komprimiert und beim Laden dekomprimiert. Die Klasse „Config“ beinhaltet allgemeine Optionen und Informationen, die z. B. zum Verbinden des PC mit einem Vicon PC[84] benötigt werden. Die Konfiguration beinhaltet Informationen wie Name, IP-Adresse, Verbringungsart und grafische Einstellungen entsprechend den Wünschen des Benutzers und der Leistung des Rechners.

Die Annotationen werden intern in der Elan Datenstruktur verwaltet. Das zentrale Datenformat wird „PAMFile“ genannt; es stellt eine Projektverwaltung der zugehörigen Filenamen dar, innerhalb derer alle beteiligten Daten gespeichert werden. Dazu zählen der Name der Motion-Capture-Daten, der Name der Videoaufnahmen und gegebenenfalls Namen von Annotationen im Format von ELAN. Die Kalibrationsdaten und die Markereigenschaften[85] sind nicht projektspezifisch und sind als Einstellungen des Setups gespeichert. Die Funktionalität der Klasse „ANVILExporters“ exportiert die Motion-Capture-Daten einer Person als BVH[86] File Format in das ANVIL-File-Format[87]. Die Klasse „Skelett“ bietet verschiedene Export unterstützende Funktionen, welche Gelenkwinkel in den verschiedenen Skelettmodellen (Anordnung der Gelenke) umrechnet. Damit lassen sich die Bewegungen von spezieller Bedeutung auch von virtuellen Agenten oder Robotern darstellen, um verschiedene Verhaltensweisen zu zeigen. Die gesamten Fileformate werden über das Interface „FileManager“ bereitgestellt und verwaltet, um die verschiedenen Daten zu laden oder zu speichern.

---

[83] Anfang einer Datei mit Initialisierungsdaten.

[84] Der Vicon-PC ist ein Windowsrechner, der mit den Hardware Geräten des Vicon-Nexus verbunden ist und die Steuer-Software enthält.

[85] Die Marker- und Rigidbody-Eigenschaften sind durch die Klasse „MarkerpropManager“ definiert und werden von dieser verwaltet.

[86] Die Datenstruktur, in der die BVH - Bio Vision Hierarchie Fileformat gespeichert wird, wurde von Holger Dierker implementiert (Aufbau der internen Struktur).

[87] Leider bietet ANVIL kaum automatische Annotationen und nur die Möglichkeit, eine einzelne Person mittels Motion-Capturing zu analysieren; daher wurden hier keine weiteren Entwicklungen durchgeführt.

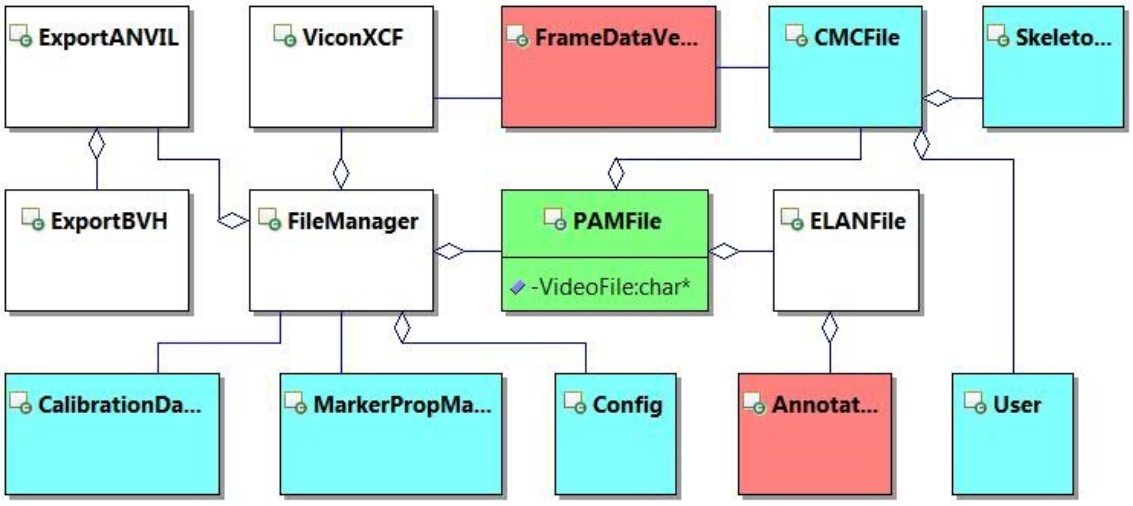

Abbildung 54 Klassendiagramm der Fileformate

7.4.4 Visualisierung von bewegungsrelevanten Inhalten

Um den Vorgang der Annotation zu unterstützen, werden Skelettbewegungen und weitere für die Annotation relevante Informationen hervorgehoben. Dieses sind Motion-Capture ergänzende Visualisierungen und Interaktionsphänomen-Visualisierungen. Neben der Skelettdarstellung gehört auch die Darstellung der Rigidbodys mit aktueller Position, Orientierung, Identifikation und der Hervorhebung, wann dieser verloren gegangen ist, dazu. Dazu werden die Standard Funktionalitäten des Szenengraphen OpenSG durch weitere Features ergänzt. Diese Ergänzungen beinhalten eine Schnittstelle zum permanenten Verschieben, zur Aktualisierung der Orientierung nur von Einzelteilen, damit das Koordinatensystem entsprechend ausgerichtet wird[88], außerdem Ergänzungen, um den Rigidbody einzufärben und die Information festzuhalten, zu welchem Zeitpunkt dieser zuletzt gerendert wurde. Diese Eigenschaften verbindet die Klasse „BasisVis“, mit der zusätzlich festgestellt werden kann, wann ein Objekt verloren wurde. Um hervorzuheben, dass ein Rigidbody verloren gegangen ist, wird dieser für die gesamte Länge durch eine Einfärbung entsprechend markiert. Die Klasse „MarkerVis“ erbt diese Eigenschaften und ergänzt die Möglichkeit, eine sichtbare Geometrie zu erzeugen und diese sichtbar und unsichtbar zu machen unter der Verwendung der „HideableNode“ des „ToolKits“. Die Klasse „RigidBodyVis“ erbt die gleichen Eigenschaften, hat allerdings eine eigene Implementierung der Darstellung und der Update-Funktion, die zusätzlich die Orientierung durch ein Koordinatensystem aktualisiert. Davon getrennt wird die Position, mit der die Beschriftung durch „Billboards“ aktualisiert wird, welche diese immer in Richtung der Kamera ausrichten. Diese beiden Visualisierungstypen werden durch die Klasse „MarkerVisualManager“ verwaltet. Dabei wird die Updatefunktion der aktuell vorhandenen Daten aufgerufen, außerdem wird die aktuelle Framenummer zu jedem dieser Datenelemente gespeichert. Die Klasse „MarkerVisualManager“ prüft, ob Objekte versteckt oder wieder gezeigt werden müssen. Die Klasse „Link“ stellt die Verbindung zwischen zwei Rigidbodys dar, um

[88] Bei einem zu großen Knoten im Szenengraph wird OpenSG langsam, daher sollte es vermieden werden, zusätzliche Knoten einzufügen.

schnell sehen zu können, wie Rigidbodys zueinander im Verhältnis stehen und gegebenenfalls vertauscht am Körper angebracht wurden. Dazu wird ein Link durch eine einfach gerade Linie im Dreidimensionalen dargestellt, von einer Rigidbodyposition zu einer anderen (z. B. ein Link vom Ellenbogen zur Hand oder von der Schulter zum Ellenbogen). Die gesamte Verwaltung der einzelnen Links wird von der Klasse „LinkManager“ übernommen. Die Klasse „Path“ stellt eine Trajektorie dar. Die einzelnen Trajektorien, welche jeder einzelne Rigidbody erzeugen kann, werden durch die Klasse „PathManager“ verwaltet, der diese gegebenenfalls ausblendet. Die Klasse „OfflineRendering“ kann verwendet werden, um jede beliebige Bewegungssequenz in ein Video mit sehr hoher zeitlicher Genauigkeit zu erstellen. Die Motion-Capture-Daten werden mit einer zeitlichen Auflösung von bis zu 200 Hz aufgezeichnet. Ein normaler Film wird mit 25 Hz dargestellt, welches das Auge ungefähr wahrnehmen kann. Dieses kann zum Beispiel auch nützlich sein, um die Bewegung aus einem Blickwinkel zu sehen, aus dem nicht gefilmt wurde. Diese Zusammenhänge sind in der **Abbildung 55** mit Relationen zu anderen Klassen dargestellt.

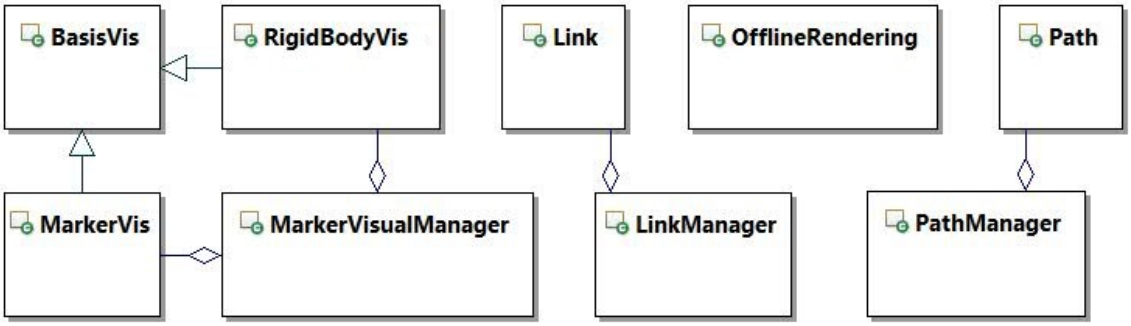

Abbildung 55 Klassendiagramm der Komponenten zur 3D-Visualisierung

7.4.5 Bewegungszerlegung in Aktivitäten einzelner Freiheitsgrade

Da nun die strukturellen Zusammenhänge der beteiligten Klassen zur Berechnung der einzelnen Skelettposen aus den Motion-Capture-Daten bekannt sind, können die strukturellen Zusammenhänge der Klassen bezüglich der automatischen Annotationen betrachtet werden. Dabei wird mit der Zerlegung der Bewegung in elementare Bestandteile (Gelenkaktivitäten) angefangen. Die Zerlegung der Bewegung in elementare Aktivität der einzelnen Freiheitsgrade wird durch mehrere Klassen durchgeführt. Dabei werden die Bewegungen der Probanden in eine Art Key-Frame-Animation umgewandelt, welche durch die Klasse „KeyMotion“ repräsentiert wird. Zu jedem der „KeyMotion“ Objekte existiert immer eine Anfangspose der Klasse „Posture“, zu der die Bewegungsänderungen durch die „KeyIntervalle“ definiert sind. Daher hat ein „KeyMotion“ Objekt ein oder mehrere Objekte der Klasse „TimeFrame“, bei dem die verschiedenen „KeyIntervalle“ ihre Aktivität beginnen. Ein „KeyInterval“ ist immer genau einem DOF zugeordnet. Ein „KeyInterval“ beinhaltet neben den Zeitangaben auch Informationen über eine Winkeländerung. Diese Klassen, die an der Keyframedarstellungsform beteiligt sind, haben eine gelbe Einfärbung in der **Abbildung 56**.

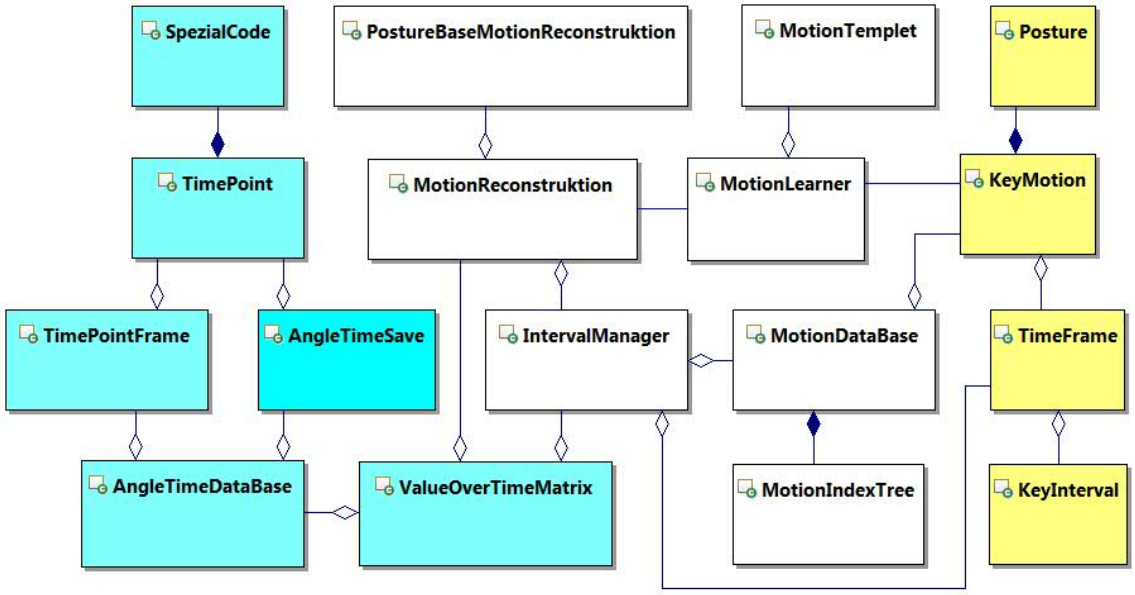

Abbildung 56 Klassendiagramm der Bewegungszerlegungsklassen.

Allerdings muss die Bewegung erst in dieses KeyFrameformat überführt werden. Die zentrale Klasse dabei ist die „ValueOverTimeMatrix“ Klasse mit ihren blau eingefärbten Abhängigkeiten. In dieser Klasse sind die gesamten Daten bezüglich der Frames und der Gelenke gespeichert. Intern ist sie wie eine dreidimensionale Matrix aufgebaut, bei der entlang einer Achse alle Freiheitsgrade der elementaren Gelenke (DOFs) aufgeführt sind. Entlang der zweiten Achse ist die zeitliche Abfolge der eingehenden Daten zu finden, auf der Daten mit einer Datenrate von bis zu 200 Hz ankommen können. Entlang der dritten Achse sind verschiedene Berechnungen mit einer Auswertung des Winkels bezüglich der Zeit zu finden. Dieses sind insgesamt vier Werte und zwar der Winkel, die Geschwindigkeit, die Beschleunigung und ein Interpretationscode. Die Klasse „SpezialCode“ ist eine Interpretation der aktuellen Information, sie kann verschiedene Zustände einnehmen. Diese Zustände können „Nichts“, „Anfang“, „Zwischendrin“, „Ende“, „finales Ende“ annehmen. Mittels dieses Codes wird bestimmt, ob ein Intervall mit Aktivität begonnen hat, anhält, ein mögliches Ende gefunden wurde oder das Intervall mit einem finalen Ende wirklich erstellt werden kann. In dem Fall, dass es geschlossen werden kann, wird rückwärts nach dem Anfang gesucht und mit den Zeitpunkten, der Position des Winkels am Anfang und am Ende, ein Key-Intervall erzeugt. Diese Interpretation basiert auf der Berechnung des Winkels, der Geschwindigkeit und der Beschleunigung in den jeweiligen Gelenken. In einem Objekt der Klasse „TimePoint“ wird die Analyse und Interpretation durchgeführt. Wenn die Geschwindigkeit ihr Maximum[89] erreicht oder es zum Stillstand kommt, kann die Interpretation als final angesehen werden, und ein Key-Intervall kann über den „IntervalManager“ an diesem Zeitpunkt (durch die Klasse „TimeFrame“ definiert) erzeugt werden. Die Klasse „MotionReconstruktion“ macht die Key-Animation wieder sichtbar und zeigt die komprimierte Bewegung in Relation zur real aufgenommenen Bewegung. Dieser Zusammenhang wird in der **Abbildung 56** dargestellt.

[89] Beschleunigung ist dann gleich Null.

7.4.6 Phänomena-Finden

In diesem Abschnitt wird auf weitere Interaktions-Phänomen-Analysen und die daran beteiligten Klassen eingegangen. Alle einzelnen Detektoren von Phänomenen werden von einem Manager verwaltet; dieser steuert, wann welche Detektion aktiv ist und bei den online Detektionen aktualisiert werden muss. Der Pose-Detektor “PostureDetektor“ arbeitet mit einer Klasse „Posture“ zusammen, in der die Gelenkstellung gespeichert wird. Dieser spiegelt die aktuelle Skelettpose wider. Die speziell ausgesuchten Posen, die gefunden werden sollen, sind durch die Klasse „PostureMask“ definiert. In dieser ist zu jedem DOF ein minimaler und maximaler Wert als Begrenzung des Gelenks (DOF) definiert, in dem sich das Gelenk befinden darf, falls eine dieser Posen detektiert werden soll. Außerdem ist zu jedem DOF ein Wert der Wichtigkeit des DOF zur insgesamt zu erkennenden Pose darstellt. Zum Beispiel spielt die Position des linken Armes meist keine Rolle, wenn mit dem rechten Arm auf ein Objekt gezeigt wird. Bei der Berechnung auf Übereinstimmung der aktuellen Pose mit einer zu detektierenden Posen-Schablone „PostureMask“ wird geprüft, ob die entsprechenden Winkel der wichtigen oder relevanten Gelenke innerhalb der Grenzen liegen.

Das Phänomen „PersonalSpaceIntrusion“ bezieht sich darauf, zu erkennen, wann eine Person in den persönlichen Bereich einer anderen Person eindringt. Die Klasse „HandAktivityDetection“ ermittelt, ob sich eine Hand bewegt. Die Klasse „TrajectoryAnalyser“ untersucht die Bewegungsrichtungen und unterteilt diese in Segmente, zu denen jeweils ein Bewegungsrichtungsvektor ausgerechnet wird. Die Detektion des Phänomens „Fokussiert auf“ findet in den Klassen „GazingAt“, „GazingAtManager“ und „CollisionSphere“ statt. Die Klasse „CollisionSphere“ repräsentiert eine Geometrie, die um die einzelnen Köpfe herum gelegt wird, um das Sichtfeld der anderen Probanden auszugleichen. Dieses kann durch einen Strahl, ausgehend vom Kopf eines anderen Probanden, geschnitten werden. Dazu wird diese kugelförmige

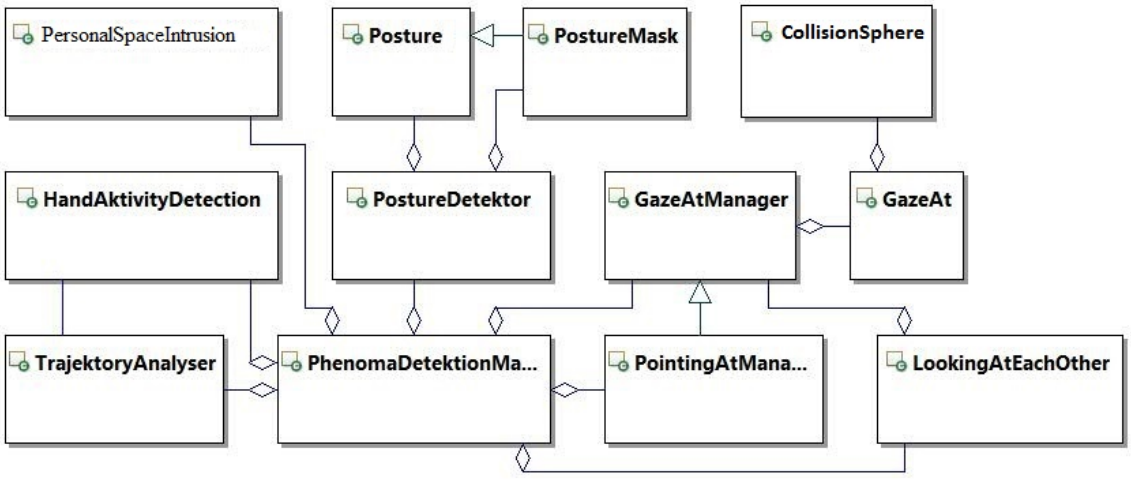

Abbildung 57 Klassendiagramm der Phänomenerkennungsklassen

Zusatzgeometrie, die entsprechend vorher abgemessener Distanz vom Rigidbody zum Mittelpunkt des Kopfes positioniert ist, mit der Bewegung mitgeführt. Die Klasse „GazingAt“ stellt einen Strahl dar, der von der Kopforientierung unter Berücksichtigung weiterer Abmessungen zu den Augen verläuft. Anschließend wird berechnet, ob ein Strahl mit einer Geometrie eines anderen Probanden kollidiert. Wenn eine andere Geometrie geschnitten wird, wird ermittelt, ob die Geometrie einen Namen besitzt. Das Phänomen des gegenseitigen Anfokussierens kann durch einen Vergleich der jeweiligen anfokussierten Namen detektiert werden. Diese Abhängigkeiten und Beziehungen der beteiligten Klassen sind in der **Abbildung 57** dargestellt.

7.4.7 Pluginstruktur

Um leicht Erweiterungen, die möglichst entkoppelt von PAMOCAT sind, zu ermöglichen, wird eine Pluginstruktur bereitgestellt (siehe **Abbildung 58**). Dieses ist eine Zusammenstellung der benötigten Dateninterfaces, um Erweiterungen zu erstellen. Dabei wird vor allem daran gedacht, mögliche Detektionen für Phänomene einzubauen. Um die Möglichkeit bieten zu können, flexible Erweiterungen zu testen, können alle Plugins über eine Konfigurationsdatei aktiviert und gegebenenfalls mit Optionen versehen werden. In der folgenden sind die beteiligten Klassen aufgeführt, die das Interface bilden und von dem die gesamte Funktionalität geerbt wird.

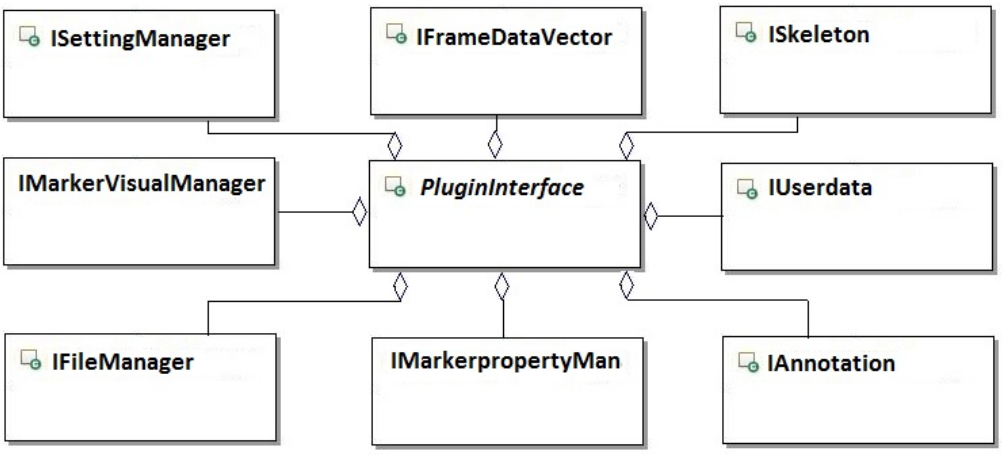

Abbildung 58 PluginInterface zur Erstellung von eigenen Plugins.

7.5 Die Anwendungsimplementierung PAMOCAT

Die Implementierung des Tools PAMOCAT ist losgelöst von der Implementierung der Bibliothek „Motion-Capture“ und des „ToolKits“. PAMOCAT bietet die GUI, um die Analysemöglichkeiten der Motion-Capture-Bibliothek einfach nutzen zu können, um einfache Annotationen durchzuführen und verschiedene Hypothesen basierend auf den Ergebnissen zu analysieren. Im Folgenden wird vorgestellt, wie die Benutzerschnittstelle aufgebaut ist.

7.5.1 Aufbau der GUI

PAMOCAT basiert auf dem Standard-Design der Klasse „QMainWindow“ der QT4 Bibliothek. Diese besitzt einen zentralen Hauptbereich in der Mitte, der hier die Motion-Capture-Visualisierung durch die Klasse „OSGWidget“ darstellt und die Möglichkeit bietet, mehrere dockbare GUI-Elemente darum zu platzieren. In der zentralen Klasse „MainWindow“ ist ein Timer-Objekt, mit dem je nach gewünschter Auslastung des PC eine Wiederholungsrate eingestellt werden kann[90]. Das Kürzel „W“ steht für „Widget“[91], „OV“ für „OverView“ und DW für „DockWidget“[92]. Die Klassen „MultiVideoPlayerDW“, „SkeletonOptionsDW“, „KeyIntervallOVDW“, „TimeShiftDW“, „DetektionResultViewDW“, „KeyIntervalOVDW“, „OptionDW“, „EditDW“, „GazingDW“ sind allesamt „DockingWidgets“ und um das zentrale „OSGWidget“ herum positioniert. Ihre jeweiligen Zusammenhänge und Abhängigkeiten zu anderen Klassen sind in dem Klassendiagramm **Abbildung 59** dargestellt. Die Klasse „KeyIntervalOVDW“ ist wiederum aus zwei einzelnen Komponenten zusammengebaut, einmal dem „KeyIntervalW“ zur Visualisierung der „KeyIntervalle“ und dem „AngleTrendView“, in dem Winkel, Geschwindigkeit und Beschleunigung angezeigt werden. Beide Klassen sind nicht im gleichen Maßstab skaliert, daher sind beide jeweils in verschiedene „ScrollArea“ oder „Verschiebungsbereiche“ eingebunden.

7.5.2 Globale Synchronisation aller Komponenten

Damit alle Komponenten beim Abspielen oder beim zeitlichen Scrollen des „TimeShiftSliders“ synchron zueinander laufen, wird die globale Zeit in einer Klasse „FrameDataVektor“ verwaltet. Diese Klasse entscheidet, welches der aktuelle „Frame“ ist, der von allen Komponenten dargestellt werden muss. Da die Video-Anzeige eine eigene Zeitabspielverwaltung besitzt, wird zur Sicherheit verglichen, wie die globale Zeit des „FrameDataVektors“ und der des Videoabspielers ist; laufen beide auseinander, wird hier die Zeit angepasst. Bei den manuellen Zeitänderungen wird die aktuelle Zeit in Millisekunden direkt in der „FrameDataVektor“ Klasse und dem Videospieler gesetzt.

---

[90] Über die GUI einstellbar.

[91] Widget ist ein Element einer Benutzeroberfläche, wie z. B. ein Texteingabefeld.

[92] Ein DockWidget ist ein GUI-Element, das an verschiedenen Stellen angehängt wird.

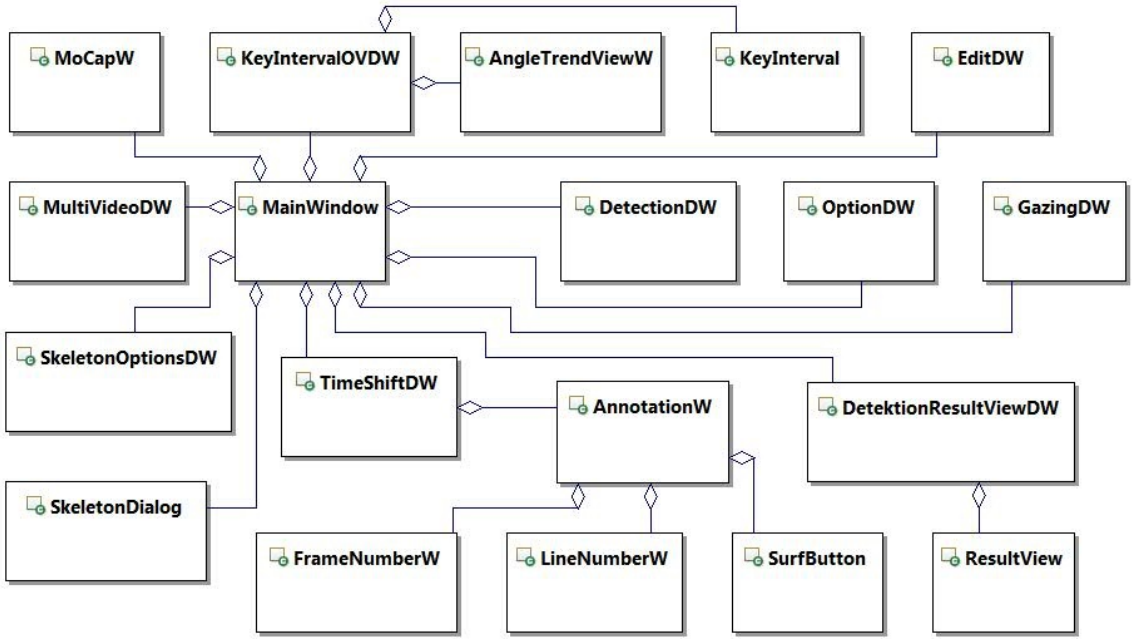

Abbildung 59 Klassendiagramm der Applikation PAMOCAT

7.6 Zusammenfassung

In diesem Kapitel wurde die Implementierung der Software PAMOCAT vorgestellt. Dazu wurden die beiden Bibliotheken, auf denen diese Software basiert, mit ihren jeweiligen Abhängigkeiten näher erläutert. Diese beiden Bibliotheken sind „ToolKit“ und „MotionCapture“. Durch die Entwicklung einzelner Komponenten mit Schnittstellen, das Verwenden von Designpattern[93] und eine klare Strukturierung ist die Software leicht erweiterbar. Dazu wurde außerdem eine Pluginschnittstelle definiert, mit der leicht Erweiterungen umgesetzt werden können. Um einen Einblick in die Software zu gewähren, wurden viele statische Aspekte der Komponenten aus der „MotionCapture“ Bibliothek vorgestellt. Ein dynamischer Aspekt des zeitlichen Ablaufs von PAMOCAT, wie die Anwendung in einem typischen Anwendungsfall arbeitet, ist im Anhang Kapitel B.6 zu finden.

[93] Vorlagen, wie verschiedene Strukturen und Funktionalitäten zu verwenden sind. Dazu zählen unter anderem Kompositum, Factory, Adapter, Fascade, Decorator und Observer.

8 PAMOCAT und seine Benutzung

In diesem Kapitel soll PAMOCAT vorgestellt werden. Dazu wird als Erstes die Benutzeroberfläche beschrieben und anschließend werden verschiedene Anwendungsfälle mit einer detaillierten Vorgehensweise vorgestellt. Es werden Beispiele gegeben, die einen leichten Einstieg in die Benutzung bieten sollen. Dazu werden verschiedene mögliche Anwendungsfälle durchgespielt. Um mit der GUI vertraut zu machen, ist diese in vielen Abbildungen verwendet worden, bei denen die entsprechenden Bedienelemente hervorgehoben wurden.

8.1 Die Benutzeroberfläche von PAMOCAT

Durch die verschiebbaren GUI-Elemente, die sogenannten „DockingWidgets“, kann die gesamte GUI entsprechend den Wünschen des Benutzers frei angeordnet werden. Dabei können auch mehrere Displays genutzt werden, um z. B. einzelne Elemente möglichst groß darstellen zu können, sodass bei komplexeren Annotationen mit vielen Tiers immer die Gesamtübersicht beibehalten werden kann. Bei dem „KeyIntervalOW“ werden neben dem aktiven DOF auch Plots von Winkel, Geschwindigkeit und Beschleunigung dargestellt. Dazu wird das ausgewählte Gelenk (DOF) durch eine blaue horizontale Linie hervorgehoben, zu dem der jeweilige Plot gezeigt wird. Durch eine „ComboBox“, die eine Auswahlliste darstellt, kann das ausgewählte Gelenk geändert werden (siehe **Abbildung 60**).

Im unteren linken Bereich ist der Navigationsmodus zu finden, bei dem Framenummern direkt eingegeben werden können, um zu diesen zu gelangen. Zudem kann hier auch die Suche nach Kombinationen von Tiers gesteuert werden. Im unteren mittleren bis rechten Bereich ist der Annotationsbereich. In der Mitte links sind die multiplen Videos zu finden. In der Mitte links sind verschiedene Eingabefenster positionierbar, um z. B. ein Skelett auszuwählen, verschiedene Visualisierungen zu aktivieren, aber auch, um Informationen anzuzeigen. Das Abspielen kann durch das „PlayToolbar“-Menü im oberen rechten Bereich ausgeführt werden.

8.2 Benutzerinteraktion mit PAMOCAT

Nachdem die Anordnung der GUI-Elemente bekannt ist, soll kurz eine typische Benutzung von PAMOCAT vorgestellt werden. Der Benutzer kann sich die Zeitpunkte berechnen lassen, wann die verschiedenen Phänomena eintreten. Diese verschiedenen Phänomene werden in Tiers oder Annotationskategorien im Annotationsbereich dargestellt. Darauf basierend kann eine Analyse durchgeführt werden. Dazu kann die Funktionalität genutzt werden, um nach den Zeitpunkten zu suchen, bei denen eine Kombination der ausgewählten Phänomene auftreten. Alternativ kann man die Liste der ausgewählten Phänomene durchgehen, die nicht

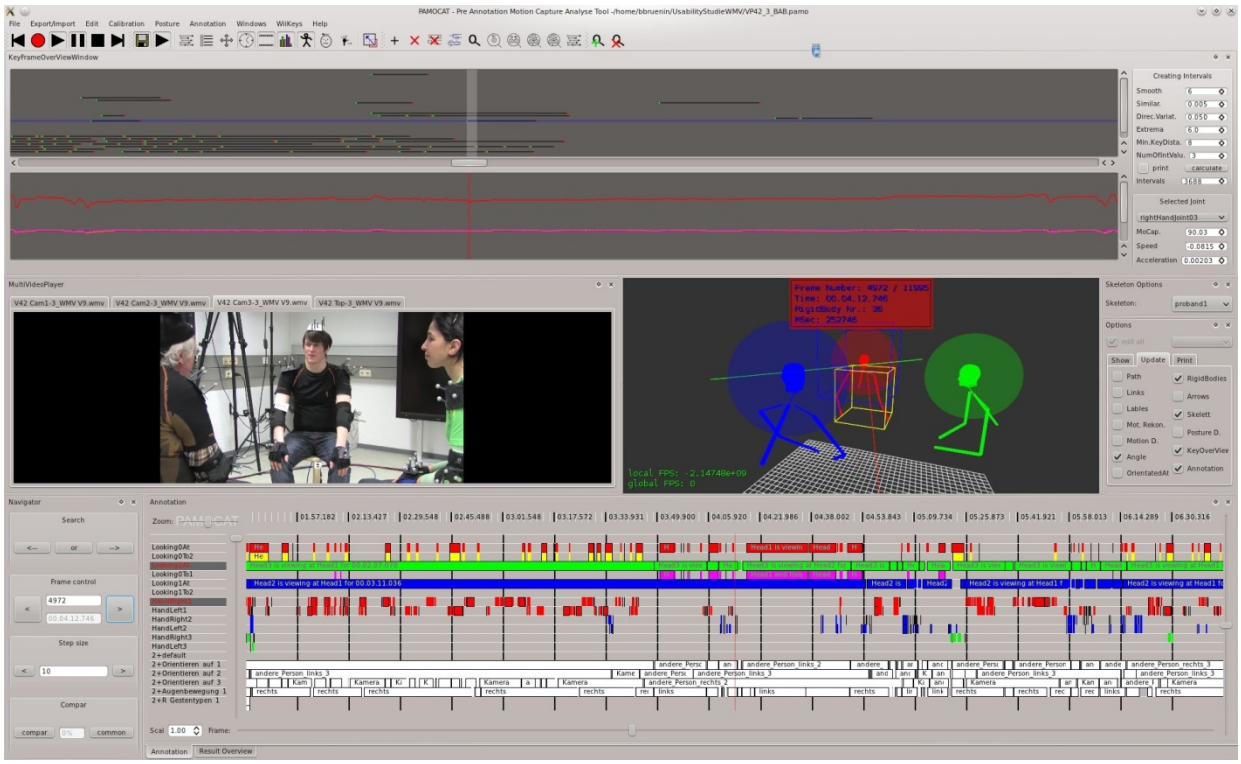

Abbildung 60 Die GUI von PAMOCAT mit seinen verschiedenen Dockingwidgets „KeyIntervallOverwiew“, „Plot“, „MultipleVideoPlayer“, „Annotation“, „TimeSlider“, „Edit“, „Options“ und „OSGWidget“

gleichzeitig aktiv sind. Dazu kann zwischen den logischen Operationen „und“ oder „oder“ für die Suche gewählt werden. Zur genauen Analyse dieser gefundenen Zeitspannen kann die Interaktion der Probanden in den synchron gehaltenen Videoaufnahmen betrachtet werden. Dabei spielt auch die Motion-Capture-View eine zentrale Rolle, in der die Bewegung aus allen möglichen Perspektiven mit einer fast uneingeschränkten Zoommöglichkeit betrachtet werden kann. Darüber hinaus werden verschiedene Phänomene optisch in der Motion-Capture-View hervorgehoben. Beispielsweise werden die Objekte, die von den Probanden anvisiert werden, mit einer dreidimensionalen Box in der Farbe des Skeletts umrandet und so dargestellt. Je nach Forschungskontext kann eine Analyse der räumlichen Positionsänderungen von Körperteilen mittels Trajektorien durchgeführt werden. Ein Plot des Winkels, der Geschwindigkeit und Beschleunigung einzelner oder aller Gelenke kann verwendet werden, um genauestens zu sehen, wie die zeitliche Abfolge einer Bewegung war. Um die beteiligten DOFs oder Gelenke zu bestimmen, die in einer Geste benutzt werden, kann die Key-Intervall-View genutzt werden. Dieses ermöglicht wiederum, ein Suchmuster nach Aktivitäten bei verschiedenen Gelenken zu finden. Es können auch eigene Annotationen den automatischen hinzugefügt oder davon entfernt werden. Darüber hinaus können weitere Annotationen anderer Tools mit geladen werden. Es wird auch eine Fehlerannotation durchgeführt, damit man schnell sehen kann, welche Interaktion durch Aufnahmefehler beeinträchtigt wurde.

8.2.1 Erstellen eines PAMOCAT-Project-Files

Ein PAMOCAT-Project-File beinhaltet die Informationen, welche Media-Daten zu den Aufnahmen gehören. Diese können Videos, Audio- oder Motion-Capture-Daten und 3D Modelle beinhalten. Bei den Multimedia-Datenformaten gibt es so gut wie keine Einschränkungen, da alle Formate, die das GStreamer-Framework benutzen, unterstützt werden. Zusätzlich kann die Projektdatei die Informationen bezüglich der Annotationen aufnehmen, welche im XML basierten *.eaf Format als ELAN-File gespeichert werden und dann in PAMOCAT mit visualisiert werden. Außerdem kann ein gegebenenfalls auftretender Zeitversatz zwischen Video/Audio-Daten und den Motion-Capture-Daten gespeichert werden[94]. Zusätzlich kann jede Video-Kamera auch direkt in der virtuellen Rekonstruktion der Aufnahmeumgebung eine Position zugewiesen bekommen.

```
<CMCFile File="MotionCapture.cmc"
ELANFile="V28_Annotationen.eaf" OffsetToVideo="-345" />
<Video File="Video28 Cam1.mp4" Position="1.64:1.89:0" />
<Video File="Video28 Cam2.mp4" Position="-1.34:1.34:0" />
<Video File="Video28 Cam3.mp4" Position="1.48:-1.58:0" />
<Video File="Video28 CamTop.mp4" Position="0:0:3.56" />
<Scene File="SagalandSetup.osb" Position="0:0:0" />
```

Tabelle 12 Inhalt eines PAMOCAT-Project-Files, in dem neben einem Motion-Capture-File auch eine ELAN-Annotation und vier Videos mit einem Zeitversatz von -345 Millisekunden definiert sind.

Um so ein PAMOCAT Projekt anzulegen, muss im Filemenü auf New geklickt werden. Daraufhin öffnet sich ein Dialog, bei dem die Parameter eingegeben werden können.

8.2.2 Synchronisation von Video- und Motion-Capture-Daten

Da die Funktionalität des gleichzeitigen Aufnehmens mit mehreren Kameras leider nicht zum Zeitpunkt der Erstellung der Korpora, die in Zusammenhang mit dieser Arbeit entstanden, existierte, müssen die Videodaten miteinander synchronisiert werden. Dazu müssen alle Videos mit dem gleichen Zeitversatz zueinander codiert werden[95]. Um einen guten Synchronisationspunkt zu haben, wird eine Filmklappe in den Videos zusammengeklappt, die in allen Videos sichtbar und hörbar sein sollte. Dieser Zeitpunkt oder ein davorliegender kann als

[94] Allgemein geht die Bestrebung dahin, auch Filmaufnahmen, die vor dem eigentlichen Start des Experiments aufgezeichnet wurden, mit aufzubewahren und keine Daten zu löschen; daher wird auf eine andere Datenquelle (Motion-Capture-Daten oder Video) gewartet.

[95] Zu diesem Zweck können einfache Video-Schnittprogramme verwendet werden.

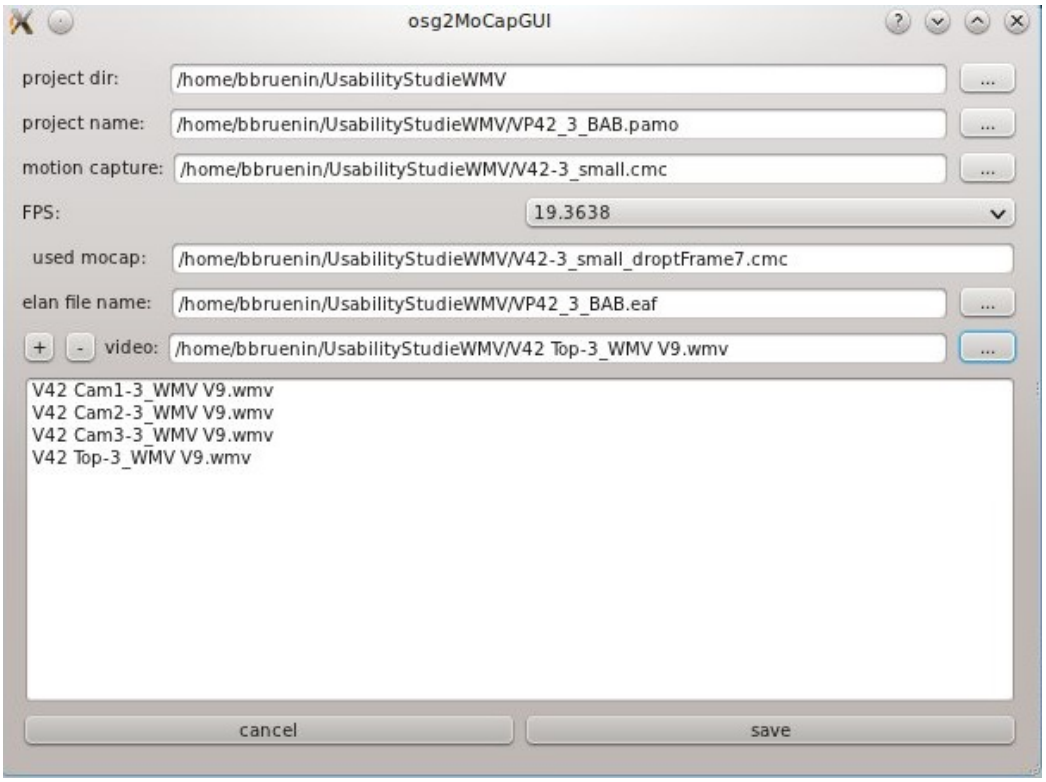

Abbildung 61 Projekt Dialog von PAMOCAT

Synchronisationszeitpunkt zwischen Video- und Motion-Capture-Daten genutzt werden. Diese exakte Zeitdifferenz muss im Millisekundenbereich dem Projekt mitgeteilt werden. Videoaufzeichnungen und 3D-Inhalte können einem Projekt in beliebiger Anzahl zugewiesen werden.

8.2.3 Virtuelle Aufnahmeumgebungen

Um ein Modell der Aufnahmeumgebung mit in einem Projekt anzeigen zu können, wird dieses unter einer XML-Node „Scene“ zum Projekt hinzugefügt werden. Die aktuell unterstützten 3D-Formate sind *.osb, ein internes OpenSG-Datenformat, welches sehr schnell geladen wird und nicht viel Speicherplatz verwendet, *.3ds, ein Datenformat, welches von 3D Studio Max benutzt wird, und das weit verbreitete *.vrml File-Format. Dabei bilden die 3D-Modelle die Realität im Maßstab 1 Meter zu 1 Meter ab.

8.2.4 Manuelles Annotieren in PAMOCAT

In PAMOCAT kann auch manuell annotiert werden. Um ein Tier zu erstellen, muss in der Toolbar Annotation der Button „add tier“ angeklickt werden. Anschließend kann der Name in einem Dialog eingegeben werden. Die Reihenfolge der Tiers kann beliebig beeinflusst werden. Dazu muss auf das zu verschiebende Tier geklickt werden, um es anschließend durch Lösen des Mausklicks an der gewünschten neuen Position zu platzieren. Durch das Klicken im Annotationsbereich entlang eines Tiers (in dem keine Annotation ist) wird der Startpunkt einer Annotation erzeugt, durch das Weiterbewegen des Mauszeigers zum Endpunkt und ein

Wiederloslassen des Klicks wird der Endpunkt markiert. Anschließend wird ein Dialog gestartet, bei dem manuell Start, End oder die Länge genauer justiert werden können. Bei der Bewegung des Mauszeigers zur Endzeitpunkt wird jeweils der aktuelle Frame dargestellt, um im Vorfeld eine schnelle Adjustierung des Endzeitpunktes zu ermöglichen. Bei dem Ändern einer Zeit wird im Hintergrund die aktuelle Zeit der gesamten Applikation mit Motion-Capture oder Videoview an diesen Zeitpunkt gesetzt. Zusätzlich kann die ausgewählte Zeitspanne von Start- bis Endzeitpunkt abgespielt werden. Die eigentliche Annotation kann in textueller Form eingegeben und es kann eine Farbe ausgewählt werden. Existiert in dem Bereich, auf den geklickt wird, bereits eine Annotation, wird ein Dialog zum Editieren dieser geöffnet. Siehe dazu die **Abbildung 62**.

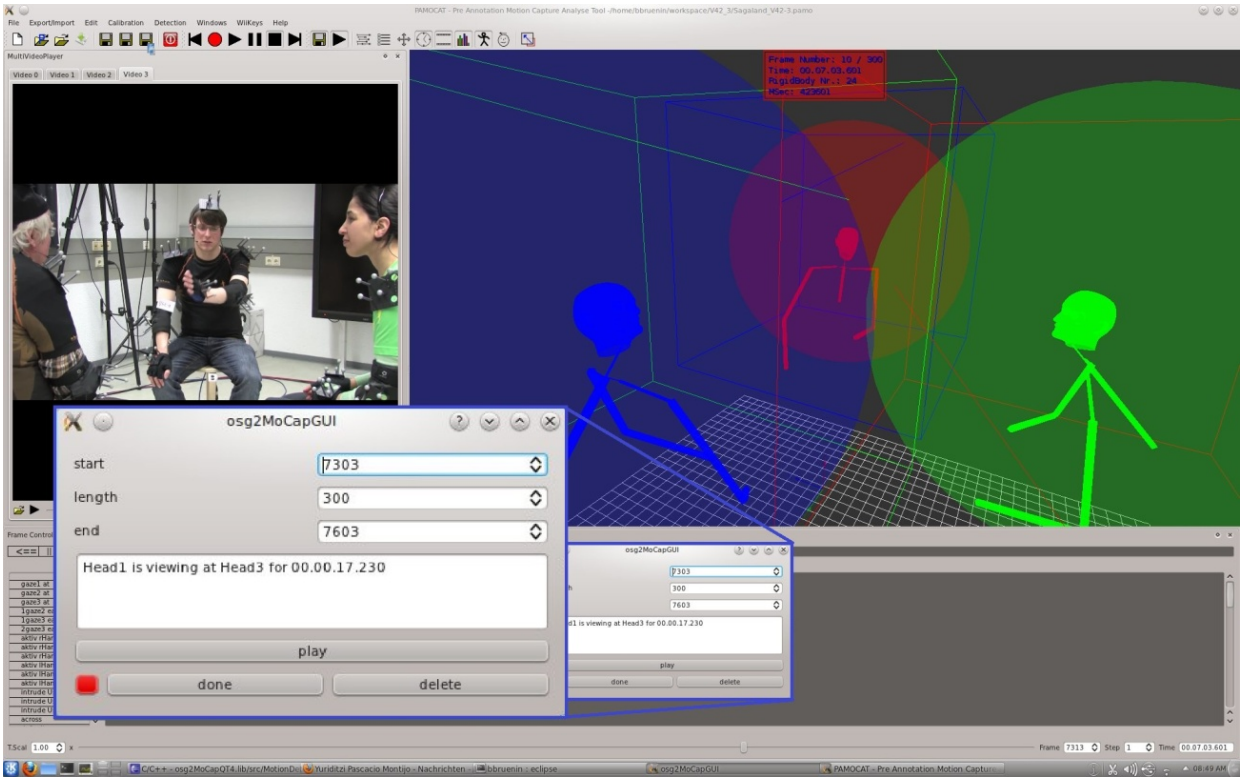

Abbildung 62 Annotationsdialog, der vergrößert wurde, mit Start, End, Längenänderungs- und Abspielmöglichkeit

8.2.5 Automatisches Annotieren

Die einzelnen automatischen Annotationen, die auf Motion-Capture-Aufnahmen basieren, können über das Detektion „MenüBar“ mit Default Parametern gestartet werden. Außerdem gibt es für verschiedene Phänomene einzelne „Dockingwindows“ mit speziellen GUI-

Elementen zur Anpassung der Parameter. Mit einem Rechts-Klick irgendwo auf der GUI, an der kein spezielles GUI-Element vorhanden ist, kann ein „Dockingwindow“-Verwaltungsmenü geöffnet werden oder auf der Toolbar (rot hervorgehoben in der **Abbildung 63**) im oberen GUI-Bereich können die „KeyFrameOverview-Dockingwindows“ (drittes Symbol in der Toolbar) und das allgemeine Detektionsdockingwindow sichtbar gemacht werden (vorletzten drei Symbole). In diesen kann z. B. der Detailgrad der Key-Intervall-

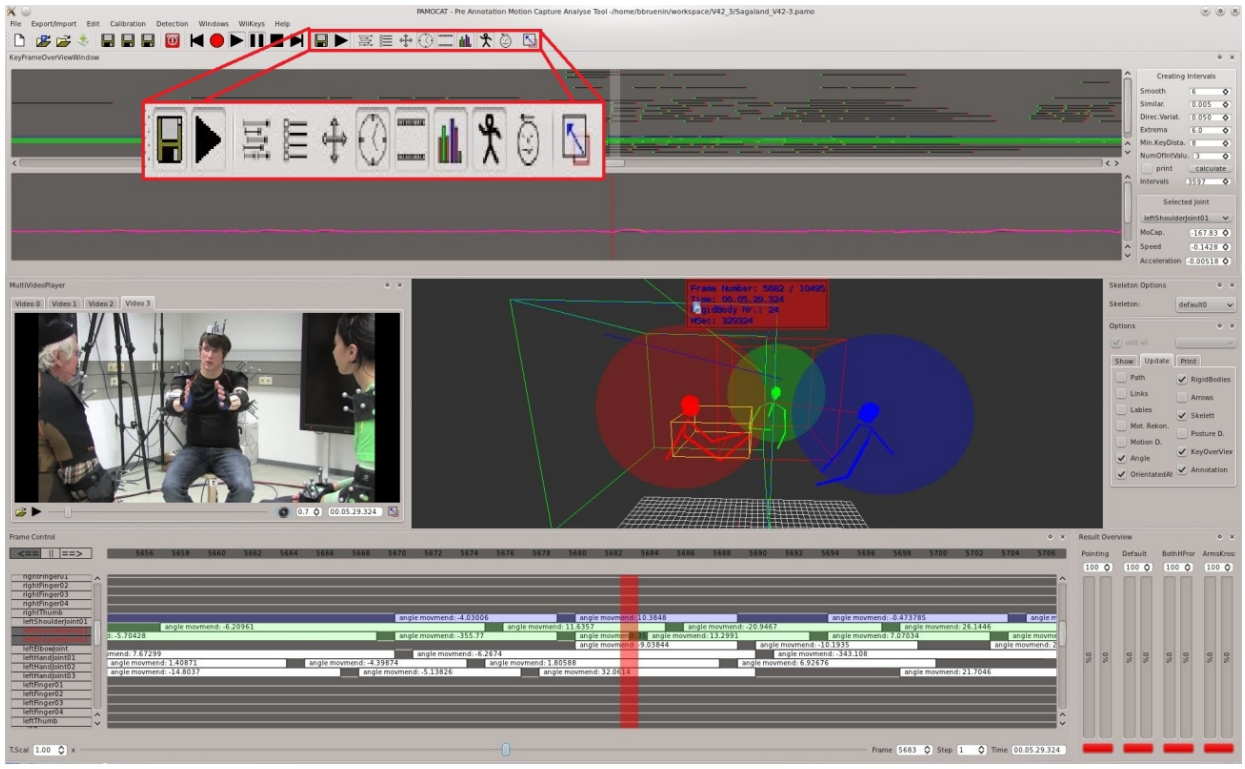

Abbildung 63 PAMOCAT mit aktivem KeyFrame-Detektions-DockingWindow und hervorgehobener Toolbar zum Verwalten der verschiedenen GUI-Dialoge

Detektion eingestellt werden. Ebenfalls kann die Größe der Kopf-Kollisions-Sphäre eingestellt werden, die für das Phänomen des Gegenseitig-aufeinander-Orientierens verwendet wird, oder die Genauigkeit der Geschwindigkeit bzw. ab welcher Länge ein gleichgerichtetes Bewegungssegment als Handaktivität klassifiziert werden soll. Außerdem können in einem weiteren Dockingwindow die zu detektierenden Posen hinzugefügt werden. Die Wichtigkeit und der Öffnungsbereich der Gelenke für eine Pose können manuell oder durch Lernen mit anderen Posen bestimmt werden. Diese verschiedenen Dockingwindows und Dialoge sind in der folgenden **Abbildung 64** aufgeführt.

8.2.6 Exportieren der Annotationen

Um Annotationen zu ELAN zu exportieren, muss im Dateien Menü der Punkt „Export-ELAN“ ausgewählt werden. Anschließend wird ein Filedialog gestartet, in dem die Position und der Name definiert werden können.

8.2.7 Benutzung der Kommandozeilenoptionen

Um verschiedene Vorgänge zu automatisieren, wie zum Beispiel Berechnungen, die Zeit kosten, können verschiedene Funktionen über eine Kommandozeile aufgerufen werden. Zu diesen Kommandozeilenoptionen zählen zum einen das automatische Annotieren und auch das

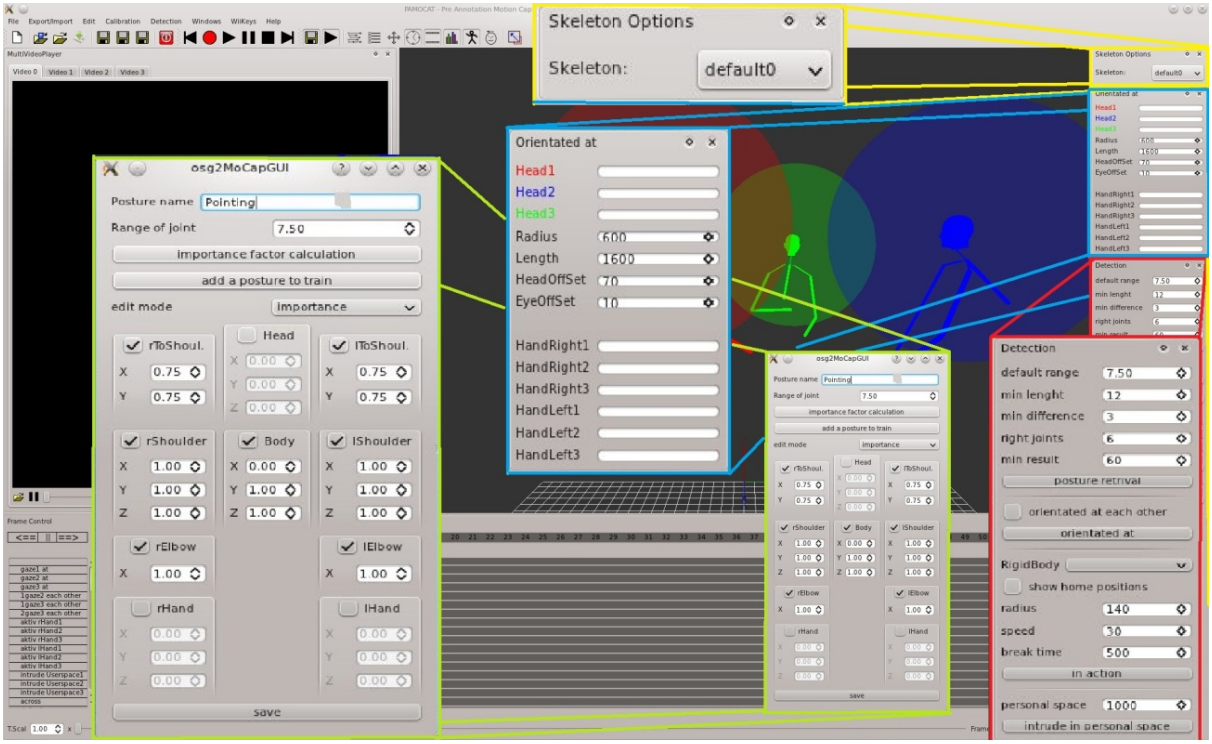

Abbildung 64 Detektions-Docking-Windows „Skelettselektion“ (gelb), „Fokussiert auf“ (blau), „Allgemein Detektion“ (rot) und der Posture-Detektion-Konfigurations-Dialog (grün)

Speichern verschiedener 3D Ansichten der Motion-Capture-Daten mit verschieden hervorgehobenen Phänomenen wie Skelettbewegungen, Trajektorien, Orientierungsfokus, Orientierung von einzelnen Körperteilen, Bewegung in Relation zu Aufnahmegegenständen usw. In der folgenden **Tabelle 13** sind alle aktuellen Optionen mit Parametern aufgeführt.

8.2.8 Programm Optionen

Die verschiedenen Optionen können im Optionsmenü ausgewählt werden. Dieses kann durch das vierte Symbol in der Toolbar (siehe **Abbildung 63**) geöffnet werden. Über diese Toolbar können auch das Abspiel-, Daten-Toolbar, Key-Intervall-, Optionen-, Edit-, Timeshift-(Zeitmanipulation), Video-, Posenerkennungs-, „Fokussiert auf“-Dockingwindow und der Fullscreenmodus aktiviert oder deaktiviert werden. Zu diesen Optionen zählen die Aktivierung von Trajektorien, verschiedene Rigidbody Darstellungsmöglichkeiten, aber auch die Einstellung des Sichtfeldes oder auch des Augenabstandes für den 3D-Stereo-Modus.

8.3 Zusammenfassung

In diesem Kapitel wurde die Benutzerschnittstelle von PAMOCAT zum einen beschrieben und zum anderen anhand von Beispielen vorgestellt. Dazu wurde zuerst der generelle Aufbau der GUI erläutert. Anschließend wurden verschiedene Beispiele vorgestellt, bei denen die Benutzung von PAMOCAT anhand von GUI-Bildern verdeutlicht wurde. Ergänzend wurde eine typische Benutzung von PAMOCAT im Forschungsalltag beschrieben.

Option	Parameter	Beschreibung
-help		Hilfe anzeigen für aktuelle Kommandozeilenoptionen
-pamocat	Filename	Öffnen eines PAMOCAT Projektfiles *.pamo
-open	Filename	Laden eines *.cmc Compressed-Motion-Capture Files
-load	Filename	Laden eines Motion-Capture-Files im ViconXCF Datenformat
-save	Filename	Speichern einer Motion-Capture-Aufnahme
-annotations	Filename	Speichern von Annotationen
-record		Motion-Capture-Aufnahme starten
-begin	Frame N r.	Starte PAMOCAT im Play Modus ab einem definierten Frame
-start	Frame Nr.	Definiere Start-Frame
-end	Frame Nr.	Definiere Ende-Frame
-from	X:Y:Z	3D-Positions/Vektor für die Sichtdefinition
-at	X:Y:Z	3D-Positions/Vektor für die Sichtdefinition
-up	X:Y:Z	3D-Positions/Vektor für die Sichtdefinition
-fullscreen		Maximierter Motion-Capture-View
-toVideo		Rendere Motion-Capture-View in gleichnamiges Video
-obersee		Aktiviere Obersee-Modus, bei dem entsprechende Sichten und Modelle geladen werden[96]
-kunsthalle		Aktiviere Kunsthalle-Modus, bei dem entsprechende Sichten und Modelle geladen werden
-stereo		Aktiviere Stereo-Rendering
-wiimote		Aktiviere die Wiimote-Steuerung für Demos
-walk		Aktiviere den Walk-Navigations-Modus
-links		Zeige Rigidbodys an, die durch Links in der Anordnung des Skelettes verbunden sind
-labels		Rigidbodys werden beschriftet angezeigt
-nolabels		Rigidbodys werden ohne Beschriftung angezeigt
-noLost Markers		Keine verloren gegangenen Rigidbodys pink färben
-geometrie		Rigidbodys werden anhand der Bezeichnung mit Zusatz-Geometrie geladen (z. B. Köpfe)
-xResult	Pixel	Fürs Rendern eine spezifische Pixelanzahl
-yResult	Pixel	Fürs Rendern eine spezifische Pixelanzahl
-min Visuali-		Minimale Visualisierung aktivieren

[96] Global verfügbarer Modus unabhängig vom Projekt.

sation		
-synchron VideoOff		Öffnet PAMOCAT in einem Modus, in dem unabhängig voneinander in dem Video und in dem Motion-Capture-View die Zeit geändert wird, um in beiden gleiche Zeitpunkte zu finden
-maximal Frames To-Load		Begrenzung der zu ladenden oder aufzunehmenden Frames
-loadEvery XFrameOnly		Um die Geschwindigkeit (bei Berechnungen oder auch zum Anzeigen) zu reduzieren, kann nur jeder x Frame geladen werden.
-create Annotations	Bezeichnung	Aktiviere annotation X (error, fokus, keyframes, hands, posture und personalspace)

Tabelle 13 Kommandozeilenoptionen des Tools PAMOCAT

9 Evaluation

In diesem Kapitel wird das Tool PAMOCAT mit den zugrundeliegenden Funktionen und Techniken evaluiert. Dazu wird als Erstes das Vorgehen zur Erstellung der Motion-Capture-Daten selber und der Einfluss durch diese Technik auf die Natürlichkeit der Bewegung evaluiert. Anschließend werden die verschiedenen Funktionen des automatischen Annotierens auf Korrektheit und Genauigkeit evaluiert. Damit soll gezeigt werden, dass das automatische Annotieren von Motion-Capture-Daten neben einer erheblichen Zeitersparnis eine höhere Genauigkeit durch eindeutige und immer gleiche Kriterien bietet. Darüber hinaus wird diskutiert, welche Phänomene sich automatisch annotieren lassen, um zu zeigen, wo die Grenzen des automatischen Annotierens basierend auf Motion-Capture-Daten liegen. Das generelle Vorgehen hierzu ist das Gegenüberstellen von manuell und automatisch annotierten Daten. Anschließend werden die Ergebnisse bezüglich der verschiedenen Phänomene einzeln diskutiert. Am Ende dieses Kapitel wird die Usability des Tools PAMOCAT untersucht.

9.1 Evaluierung des Motion-Capturings

Die Evaluierung des Motion-Capture-Verfahrens ist anhand der Videoaufzeichnungen und der automatischen Fehlerannotationen durchzuführen. Damit ist das Vorgehen in zwei aufeinander aufbauende Phasen unterteilt. In der ersten Phase wird eine automatische Annotation von verschiedenen Fehlern durchgeführt. Dazu werden die verschiedenen Motion-Capture-Daten aus den unterschiedlichen Korpora auf verlorene Rigidbodys und Rotationssprünge (Flips)[97] durchsucht. Dabei werden diese Ereignisse „wann ein Rigidbody verloren gegangen war“ „der Zeitraum, wie lange es verloren war“ und die Zeitpunkte von Rotationssprüngen als Annotation gespeichert. Diese automatische Annotation wird in der anschließenden Phase genutzt, um eine manuelle Analyse des Fehlers in PAMOCAT durchzuführen. Außerdem wird bei der manuellen Analyse die Korrektheit der Daten geprüft. Dazu zählt, ob alle Kameras vorhanden sind, ob überhaupt alle Rigidbodys vorhanden sind und ob deren Orientierung stimmt. Um im zeitlichen Rahmen dieser Analyse zu bleiben, wird die manuelle Analyse der automatischen Annotationen nur stichprobenartig durchgeführt. Diese Korrektheit beschreibt die Fehlzeit aller Rigidbodys während der gesamten Aufnahme im Verhältnis zur Aufnahmezeit und Rigidbodyanzahl.

$$\text{KorrektheitLeicht} = \left(1 - \frac{Gesamtfehlzeit}{Gesamtzeit * Rigidbodyanzahl}\right) * 100$$

$$\text{KorrektheitHard} = \left(1 - \frac{Gesamtfehlzeit}{Gesamtzeit}\right) * 100 \quad \textbf{(33)}$$

[97] Große Änderungen, die keinen natürlichen Ursprung haben können.

Zur Berechnung des „KorrektheitHard“ werden alle einzelnen Fehlzeiten jedes Rigidbodys aufsummiert und durch die Gesamtlaufzeit geteilt. Für die Berechnung der weichen Korrektheit wird berücksichtigt, dass jeder einzelne Rigidbody eine Fehlzeit hat. Da zu einem Zeitpunkt mehrere Rigidbodys fehlen könnten, wird dieses durch die Rigidbodyanzahl in der Formel mit berücksichtigt.

Name/ FPS	**Lost/Found/Total**	**Zeit/Fehlzeit**	**Gap(msec)/ Flip**	**NoError Length**	**Korrekt leicht/ Hard**
OS1/ 99.83	822/ 822/ 1466736	10.15.883/ 01.37.526	5646/ 475	00.28.900	99.34/84.16%
OS2/ 29.01	989/ 989/ 1398898	33.41.618/ 01.46.475	5646/ 584	22.38.452	99.78/94.73%
OS3/ 99.84	1757/ 1757/ 1895765	13.31.092/ 08.04.879	84593/ 3510	00.10.920	97.50/40.21%
OS4/ 99.99	519/ 519/ 1748283	12.10.627/ 00.56.779	3560/ 394	00.39.922	99.67/92.22%
OS5/ 99.99	1237/ 1237/ 1763222	12.21.625/ 02.58.490	56253/ 1964	00.18.711	98.99/75.93%
OS6/ 39.27	740/ 740/ 1037754	27.52.392/ 02.13.917	19791/ 963	16.53.537	99.49/91.99%
OS7/ 100.01	356/ 354/ 464617	03.16.768/ 01.19.852	8880/ 397	00.10.659	98.30/59.41%
OS8/ 34.83	845/ 845/ 944629	28.46.957/ 03.07.970	17921/ 1191	18.56.464	99.31/89.11%
OS9/ 99.85	1094/ 1094/ 1182239	12.28.421/ 02.26.123	17611/ 1454	00.15.341	98.77/80.47%
OS10/ 99.98	2937/ 2937/ 1384971	14.56.785/ 07.12.161	135375/ 2734	00.22.060	96.99/51.81%
OS11/ 99.94	2050/ 2049/ 3058337	33.33.157/ 26.58.956	4160/ 1614	16.48.409	94.97/19.58%
OS12/ 98.91	2178/ 2178/ 1667037	18.46.594/19.49.310	173106/ 2520	00.24.801	93.40/-5.56%
OS13/ 99.79	2631/ 2631/ 4533099	33.04.397/ 36.57.346	1189087/2177	19.49.087	95.34/-11.739%
OS14/ 99.93	1196/ 1195/ 2468950	27.30.695/ 04.22.366	26625/ 532	16.46.389	99.93/84.10%
OS15/ 99.96	937/ 937/ 2565069	27.04.352/ 05.38.403	89853/ 659	16.55.321	98.69/79.16%
Total / 79.5648	20290 / 20275/ 27579608	05.09.21.363 / 02.05.30.553	1189087/ 21168	22.38.452	98.91/80.43%

Tabelle 14 Automatische und manuelle Auswertung der Motion-Capture-Daten des Obersee Korpus

Das Analysieren der Zeitpunkte mit Fehlern kann in PAMOCAT gut durchgeführt werden, um die Videoaufzeichnungen der drei bis vier Kameras mit den Posen der Probanden in den Motion-Capture-Daten zu vergleichen. Zudem kann durch das Selektieren der verschiedenen Fehlerannotationen der Suchmodus genutzt werden, um mit Hilfe von PAMOCAT schnell die relevanten Fehlzeiten durchzusehen. Im Folgenden sind Zusammenfassungen der automatischen Fehlerannotationen der einzelnen Korpora bezüglich jeder einzelnen Aufnahme festgehalten. Die Fehlerannotationen und die Fehlerauflistung sind in Dateien mit der Endung „*.eaf“ und „*.mca“ im Korpus selber gespeichert. Dazu können die Tiers der automatisch erstellten Fehlerannotationen zur eigentlichen Annotation angehängt (als weitere Tiers) wer-

den. Die Ergebnisse vom Obersee Korpus und Sagaland Korpus sind in den **Tabelle 14** und **Tabelle 15** aufgeführt.

Name/ FPS	**Lost/Found/Total**	**Zeit/Fehlzeit**	**Gap/ Flip**	NoError-ror-Length	**Korrekt leicht/ Hard**
VP26/ 159.54	279/278/ 7221199	39.02.870/ 00.16.406	1682/ 94	15.35.197	99.96/ 99.37%
VP27/ 174.88	264/264/ 4573016	27.08.909/ 00.04.290	76/ 241	05.47.950	99.98/ 99.71%
VP28/ 152.38	936/ 936/ 6564296	35.57.696/ 00.55.922	3203/137	07.39.631	99.90/ 97.74%
VP29/ 163.20	256/ 256/ 8186848	42.29.883/ 00.06.441	827/ 4	10.42.019	99.98/ 99.74%
VP30/ 173.87	179/ 179/ 6104817	36.27.489/ 00.06.229	490/ 6	05.52.246	99.97/ 99.64%
VP31/ 154.30	152/ 152/ 5041590	27.51.871/ 00.06.071	978/ 7	05.50.423	99.98/ 99.63%
VP32/ 161.20	55/ 47/ 3066532	19.45.186/ 00.04.510	1246/1864	06.48.065	99.98/ 99.74%
VP33/ 176.35	44/ 44/ 4267592	25.11.864/ 00.01.739	347/ 55	07.41.961	99.99/ 99.85%
VP34/ 159.19	152/ 144/ 6789972	44.59.791/ 00.05.217	2152/ 2192	15.02.090	99.98/ 99.81%
VP35/ 170.38	133/ 120/ 5775450	35.27.824/ 00.03.237	199/ 16	08.44.391	99.99/ 99.88%
VP36/ 151.63	135/ 135/ 5851096	33.07.757/ 00.08.413	2121/ 34	06.13.857	99.97/ 99.55%
VP37/ 169.37	43/ 43/ 3931343	24.08.98/ 00.03.410	2031/ 5	07.46.256	99.98/ 99.70%
VP38/ 153.59	142/ 142/ 6385319	34.53.618/ 00.03.469	668/ 6	08.19.469	99.99/ 99.84%
VP39/ 165.72	172/172/ 5040493	31.42.650/ 00.08.251	3439/ 23	07.00.061	99.97/ 99.57%
VP40/ 162.97	510/502/13355702	1.11.11.25/00.17.151	1398/4020	04.24.702	99.97/ 99.63%
VP41/ 174.70	71/71/ 4737600	28.10.174/ 00.06.064	2077/ 56	09.39.227	99.98/ 99.74%
VP42/ 155.29	108/108/ 4696337	26.06.695/ 00.05.619	1244/12	07.25.094	99.98/ 99.63%
VP43/ 163.06	445/440/ 4196482	26.49.345/ 00.11.824	2926/ 875	06.55.995	99.94/ 99.15%
VP44/ 144.17	1040/1032/7892112	46.53.255/ 00.40.931	6571/1460	09.06.596	99.93/ 98.54%
VP45/ 177.28	134/134 / 5835490	34.19.491/ 00.05.061	922/ 0	11.47.974	99.98/ 99.76%
VP46/ 177.65	151/151/ 4533327	26.28.140/ 00.03.029	204/6	09.29.103	99.98/ 99.80%
VP47/ 176.32	202/201/ 5534983	32.41.156/ 00.04.174	330/ 1333	10.15.341	99.98/ 99.80%
VP48/ 164.96	1116/1115/6676550	34.09.506/ 00.41.582	6654/ 91	07.05.998	99.91/ 97.99%
VP49/ 158.54	258/250/ 8657799	46.04.332/ 00.46.162	21425/187	11.28.353	99.93/ 98.47%
VP50/ 159.26	949/949/ 4994913	27.14.928/ 00.33.583	3630/ 40	09.24.202	99.89/ 97.37%
Total/ 164,252	7928/7865/ 149910870	13.38.24.667/ 05.48.782	21425/ 12759	15.35.197	99.96/ 99.23%

Tabelle 15 Ergebnisse der automatischen und manuellen Auswertung der Motion-Capture-Daten vom Sagaland Korpus

Durch Vergleich dieser beiden Tabellen des ersten und letzten Korpus ist eine Steigerung der Qualität festzustellen. Da bei dem Kunsthallen Korpus das Motion-Capture-System über einen längeren Zeitraum lief und immer wieder Personen für einen kurzen Zeitraum in den Motion-Capture-Bereich hineingingen, kann das Fehlen von Rigidbodys nicht als Fehler wie bei den anderen Korpora gewertet werden. Außerdem wurden bei dem Kunsthallen Korpus keine Körper, sondern nur die Köpfe mittels Motion-Capturing aufgezeichnet. Deshalb fehlt dieser Korpus bei der Auswertung an dieser Stelle. Zu jeder Motion-Capture-Aufnahme aus den anderen Korpora werden die Anzahl an verlorenen Rigidbodys, die Anzahl der wieder gefundenen Rigidbodys, die Gap-Zeit der längsten Abwesenheit (in Millisekunden), die Anzahl an Rotationssprüngen und der Fehler dargestellt. Oft ist die gleiche Zahl bei verlorengegangenen und wiedergefundenen festzustellen, welches zeigt, dass alle Rigidbodys wiedergefunden wurden. Die Bezeichnung „NoErrorLength“ steht für den längsten Aufnahmezeitraum ohne Fehler. Dabei ist zu berücksichtigen, dass beim Sagaland Korpus die gesamten Motion-Capture-Daten in drei einzelnen Teilen aufgezeichnet wurden, wodurch die maximale Länge ohne Fehler auf etwa 1/3 reduziert wird. Ein weiterer informativer Aspekt ist, welche Rigidbodys der verschiedenen Körperteile wie oft verlorengegangen sind (siehe **Tabelle 16**).

Rigidbodyname	Lost Obersee (P1/ P2/ P3)	Flip Obersee (P1/ P2/ P3)	Lost Sagaland (P1/ P2/ P3)	Flip Sagaland (P1/ P2/ P3)
Linke Hand	1877/ 1391/ 489	1263/ 1187/ 574	214/76/1152	4/0/264
Rechte Hand	1278/ 5265/ 1171	1461/ 3204/ 1446	927/80/367	8/18/26
Linker Ellenbogen	1181/ 1628/ 42	1933/ 790/ 63	1612/656/142	99/123/192
Rechter Ellenbogen	522/ 1644/ 516	1271/ 2075/ 1427	181/1854/548	1/11620/382
Linke Schulter	639/ 308/ 1	1240/ 300/ 2	67/257/292	0/3/15
Rechte Schulter	424/ 1405/ 37	588/ 1939/ 37	52/21/50	0/0/0
Rücken	422/ 21/ 8	48/ 184/ 131	65/45/82	0/0/2
Kopf	9/ 0/ 12	5/ 0/ 0	26/5/16	0/0/2

Tabelle 16 Anzahl der verlorengegangenen Rigidbodys im Verhältnis zu den verschiedenen Körperteilen

Die Genauigkeit von Motion-Capturing bezüglich der aufgezeichneten Posen ist schwierig zu ermitteln. Zum einen müssen die Abmessungen der Probanden genau ermittelt werden, was schwierig durchzuführen ist, da die exakten Gelenkmittelpunkte nicht genau ermittelt werden können. Zum anderen aber auch, weil sich die Befestigung der Rigidbodys bei Bewegung verlagern kann. Geschätzt wird, dass die Rigidbodypositionen ca. 1 cm falsch liegen können.

Bei einer Armlänge von ca. 30 cm würde dadurch ein maximaler Fehler von ca. 2° möglich sein.

$$Maximalerposenfehler = atan(\frac{Fehlerhafteposition}{Armlänge}) = atan(1/30) = 2° \quad \textbf{(34)}$$

Bei den ersten Aufzeichnungen im Obersee Korpus war die Qualität nur befriedigend, da die Motion-Capture-Daten viele Ungenauigkeiten beinhalteten, sodass die eigentlichen Analysen nur unter der Berücksichtigung der Fehlerannotationen durchgeführt werden können. Zu den Fehlerquellen gehören neben der fehlerhaften Klassifikation des Viconsystems von den Rigidbodys auch Pannen bei der Vorbereitung der Probanden. Die fehlerhafte Vorbereitung beinhaltete, dass Rigidbodys vertauscht wurden, aber auch, dass die Klassifikation von einzelnen Rigidbodys bei einem Durchlauf ausgeschaltet war. Bei der falschen Klassifikation der Rigidbodys wurden einzelne Rigidbodys vom Viconsystem vertauscht, aber auch vereinzelt für einen kurzen Zeitraum nicht gefunden. Im ersten Obersee Korpus ist die Orientierung der einzelnen Rigidbodys nicht zuverlässig stabil. Dieses ist durch die höhere durchschnittliche Anzahl an Rotationsflips zu sehen (siehe **Tabelle 16**). Im Falle, dass Rigidbodys vertauscht sind, ist die betreffende Pose zu diesem Zeitpunkt unbrauchbar. Dieses kann aber manuell durch Angabe der betreffenden Rigidbodys korrigiert werden. Wenn ein Rigidbody nicht gefunden werden kann, ist die Pose möglicherweise auch nicht zu gebrauchen. Dieses ist oft der Fall, wenn die Rigidbodys nahe am Körper sind und sich nicht viel bewegen. Doch dann ist die letzte bekannte Position eine sehr gute Näherung und liefert so gute Resultate für die Berechnung der gesamten Pose. Die Rotationsflips haben bei den meisten Körperteilen kaum Auswirkungen, da dadurch nur die Position leicht verschoben wird; nur bei Körperteilen wie dem Kopf, bei dem die Orientierung von Interesse ist, entstehen dadurch Probleme. Der **Tabelle 16** ist aber zu entnehmen, dass diese Körperteile selten diese Art von Fehlern aufweisen. Durch eine Vertauschung von Rigidbodys wären fehlerhafte Körperstellungen möglich, wie es vereinzelt bei den ersten Aufnahmen im Obersee Korpus der Fall war. Allgemein deutet die **Tabelle 16** darauf hin, welche Entwürfe der Rigidbodys gegebenenfalls überarbeitet werden müssten, da sie vielleicht von der Anordnung der Marker dazu neigen, Rotationsflips zu produzieren, oder da manche Rigidbodys sehr viele und andere fast gar keine Rotationsflips aufweisen.

In diesen Tabellen werden die gesamten Durchläufe mit der Aufnahmezeit der verschiedenen Fehler gegenübergestellt. Dabei ist klar zu sehen, dass bei den ersten Korpora die Genauigkeit der Rigidbodys teilweise viel schlechter war. Dieses ist auf die Anzahl der Kameras und deren jeweilige Installation in Bezug zur gegebenen Fläche zurückzuführen. Die Verbesserung dieser Ergebnisse im neueren Korpus Sagaland ist auf drei Punkte zurückzuführen:

- Erfahrung mit dem Aufzeichnen mittels Rigidbodys und bestmögliche Schulung der Aufnahmehelfer
- Höhere Anzahl an Motion-Capture-Kameras (von 10 auf 14)

- Robustere Rigidbody-Detektion durch verändertes Design und verbesserte Softwareerkennung der Rigidbodys.

9.2 Evaluierung des Störfaktors der Rigidbodys

Da nun geklärt ist, wie stabil und genau das Motion-Capturing an sich arbeitet, wird jetzt evaluiert, ob das Motion-Capturing selber mittels Rigidbodys Einfluss auf die Natürlichkeit der Bewegung hat. Dazu wird untersucht, ob die Versuchspersonen sich durch diese Rigidbodys bei der Interaktion gestört gefühlt haben. Die Rigidbodys, welche in Abschnitt 4.2 vorgestellt wurden, sind relativ groß im Vergleich zu einzelnen Markern. Dadurch, dass die Rigidbodys automatisch den Körperteilen zugeordnet werden sollen, um mit den Motion-Capture-Daten ohne Nachbearbeitung arbeiten zu können, sind die Marker durch eine Variation als Rigidbody verwendet worden. Dieses könnte das Verhalten und die Bewegung bei der Interaktion zwischen den Probanden verfälschen. Diesbezüglich wurde zu jedem Korpus mit evaluiert, ob die Rigidbodys die Bewegung merklich verändern. Die Evaluation, die hier durchgeführt werden soll, bezieht sich auf die Daten des „Sagaland“ Korpus. Bei dem „Sagaland“ Korpus wurde diese Evaluierung in zwei verschiedenen Teilen durchgeführt, eine schriftliche Befragung der Probanden und eine Analyse des Verhaltes basierend auf den Videoaufnahmen und Motion-Capture-Daten. Diese einzelnen Evaluationsschritte werden im Folgenden beschrieben.

9.2.1 Schriftliche Evaluation

Um den Einfluss der Rigidbodys auf die Interaktion zu analysieren, wurden die Probanden nach den Versuchen gebeten, einen Fragebogen auszufüllen. Dieser anonymisierte Fragebogen beinhaltete Fragen bezüglich der Person (Geschlecht, Beruf, Alter und Sinnestype[98]), Teilnahmeerfahrung bei Studien, störende Elemente und zu dem Szenario selber. Die Fragen nach störenden Elementen bezogen sich auf die Rigidbodys, die am Körper getragen wurden, und auf die Kameras, mit denen die Interaktion gefilmt wurde. Die Evaluierung bezüglich der Rigidbodys wurde bei allen Korpora durchgeführt, um ausschließen zu können, dass die Bewegung durch die Verwendung der Rigidbodys beeinträchtigt oder verändert wird. Diese Evaluation wurde erneut durchgeführt, da sich die Rigidbodys im Bezug zur ersten Evaluation bezüglich Größe und Tiefe verändert haben[99]. In diesen Fragebögen konnten die Probanden das Tragen der Rigidbodys in 5 Unterstufen zwischen sehr störend bis gar nicht störend bewerten. Das Ergebnis ist in der folgenden Tabelle 17 aufgeführt.

[98] Es sollte unterscheidbar sein, wo die Stärken (auditiv, visuell oder kinästhetisch) einer Person beim Lernen liegen.

[99] Die großen Rigidbodys waren 16 cm statt 10 cm und die kleinen 10 cm statt 7 cm groß. Eine weitere Änderung ist, dass die Variationen zusätzlich eine weitere Dimension erfassen.

Position	Sehr viel	viel	etwas	kaum	Gar nicht
Kopf	0	0	6	5	19
Schultern	0	0	0	2	28
Rücken	0	1	2	2	25
Ellenbogen	0	3	8	6	13
Hände	0	2	9	4	15
Total	0	6	25	19	100

Tabelle 17 Evaluationsergebnis des störenden Einflusses von Rigidbodys an verschiedenen Körperteilen

Darüber hinaus haben 27 von 30 befragten Probanden des Sagaland Korpus 2012 ausgesagt, dass sie sich trotz Rigidbodys natürlich bewegen konnten. Die Ergebnisse decken sich mit denen aus der vorherigen Evaluation beim „Obersee“ Korpus [72]. Allgemein wurde wieder ausgesagt, dass die Rigidbodys an den Händen und Ellenbogen am ehesten als störend empfunden wurden. Eine Abweichung bei der Evaluation ist, dass der Kopf vereinzelt als etwas störend bewertet wurde. Dieses liegt daran, dass Haarreifen anstatt Hüten oder Caps verwendet wurden, um die Rigidbodys besser und stabiler am Kopf befestigen zu können. Eine Erklärung hierfür ist, dass diese hauptsächlich bei Personen ohne viele Haare als störend empfunden wurde. Außerdem wurden die Probanden gefragt, ob die Kameras sie bei der Interaktion beeinflusst haben. Dieses konnte wieder in 5 Unterstufen zwischen sehr störend bis gar nicht störend bewertet werden. Das Ergebnis ist in der Tabelle 18 aufgeführt und zeigt, dass die Rigidbodys als nicht störender empfunden wurden als die Tatsache, gefilmt zu werden.

Störfaktor	Sehr viel	viel	etwas	kaum	Gar nicht
Kamera	0	1	6	9	14

Tabelle 18 Störeinfluss der Kameras

Danach haben sich nur 14 der Probanden gar nicht durch die Kameras gestört gefühlt, aber 16 Probanden haben sich von kaum bis viel gestört gefühlt. Im Vergleich zu den Rigidbodys, bei denen sich ca. 33 % gestört gefühlt haben, sind das ca. 54 % der Probanden, die sich durch die Kameras gestört gefühlt haben.

9.2.2 Manuelle Evaluation

Der entscheidende Einfluss der Störfaktoren bei der Interaktion ist, wann genau sich die Probanden gestört gefühlt haben. Für die eigentliche Interaktionsanalyse ist nicht entscheidend, ob und wie stark die Probanden durch die Rigidbodys abgelenkt waren, sondern zu welchen Zeitpunkten (ob während einer Interaktion oder davor/danach). Wenn die Probanden während

der zu analysierenden Interaktion nicht abgelenkt sind, ist der Einfluss durch die Rigidbodys auf die Interaktion nicht relevant. Da im Sagaland Korpus nur die eigentliche Interaktion aufgezeichnet wurde, wird diese Analyse auf Basis der Vorstudie Sagaland2012 durchgeführt, bei der die Kameras während der gesamten Zeit liefen. Dabei wurden die Probanden auch während der einzelnen virtuellen Busfahrten gefilmt. Schon in der ersten manuellen Analyse der Videoaufnahmen [72] war ersichtlich, dass die meisten Probanden während des Wartens auf die Rigidbodys schauen und sich daran scheinbar stören. Dazu wurden in zwei Aufnahmen aus der Vorstudie Sagaland2012 genau die Zeitpunkte ermittelt, wann sich die Personen scheinbar durch die Rigidbodys abgelenkt gefühlt haben. Dabei war das Kriterium entscheidend, zu welchen Zeitpunkten die Probanden die eigenen Rigidbodys betrachtet hatten. Die gesamte Aufnahme wurde in Vorbereitungsphase, Aufgabenphase und Nachbereitungsphase unterteilt und genau festgehalten, in welchen Phasen die Probanden sich überhaupt an den Rigidbodys stören (diese anschauten). Bei der Aufgabenphase sind die Probanden sehr auf die Aufgabenstellung konzentriert und betrachten die Rigidbodys nicht, nur in der Vorbereitungsphase und Nachbereitungsphase passierte dieses. Diese Analyse, basierend auf den Videoaufnahmen und Motion-Capture-Annotationen, ergab, dass die Probanden während der Aufnahme vollkommen mit der Aufgabenstellung beschäftigt sind und sich nicht durch die Rigidbodys ablenken ließen. Dabei wurden zwei Aufnahmen des Korpus Sagaland, basierend auf dem Phänomen der Key-Intervalle mit aktiven Gelenken, genauestens betrachtet. Bei der Analyse wurde eine Suche mittels „or“ Operation bezüglich der Aktivität in den Gelenken des Kopfes und der Hände durchsucht und die gefundenen Zeitpunkte in den Videodaten angeschaut. Dabei wurde festgestellt, dass ca. 2 von 3 Probanden die Rigidbodys genau einmal anschauen, wenn aktuell keine Aufmerksamkeit gefordert wird. Dieses war der Fall, wenn ein Wechsel der Interaktionspartner stattfand.

9.2.3 Zusammenfassung der Ergebnisse in der Evaluation zur Ablenkung durch Rigidbodys bei der menschlichen Interaktion

Durch Vergleich der Fragebogenaussagen und der manuellen Analyse wurde ermittelt, dass die einzelnen Probanden sich in der Vorbereitungsphase teilweise durch die Rigidbodys gestört gefühlt haben, aber bei der Bewältigung der Hauptaufgabenstellung nicht. Das Tool PAMOCAT konnte mit seinen automatischen Annotationsfunktionen dazu genutzt werden, um schneller als mit anderen vergleichbaren Tools einen Korpus auf das gesuchte Phänomen zu prüfen. Dazu wurden die Zeitpunkte angeschaut, bei denen eine mögliche Ablenkung stattgefunden haben könnte, und eine manuelle Analyse hat quantitative mit qualitativen Methoden verbunden. Damit konnte die aus der ersten Evaluation [72] hervorgegangene These untermauert werden, dass sich die Probanden scheinbar nicht durch die Rigidbodys während der eigentlichen Interaktion gestört gefühlt haben.

9.3 Evaluierung der automatischen Annotationsfunktionen

In dieser Arbeit wurden verschiedene Korpora erstellt, zu denen das Tool PAMOCAT angepasst bzw. entwickelt wurde. Dazu zählen die Korpora „Obersee“ [72], „Kunsthalle“ [4] und „Sagaland“ [83], die in Kapitel 5 vorgestellt wurden. Zur Unterstützung der Verhaltensanalyse werden die aufgezeichneten Daten annotiert, damit verschiedene Interaktionsphänomene ausfindig gemacht werden können. Um die Richtigkeit der automatischen Annotationen unter Beweis zu stellen, werden die Daten von zwei verschiedenen Personen annotiert und mit den automatischen Annotationen verglichen. Dazu wird geprüft, inwieweit automatische Annotationen basierend auf Motion-Capture-Daten diesen Vorgang unterstützen können oder sogar besser sind. In der folgenden **Tabelle 19** sind die Phänomene aufgezählt, bei denen der Vergleich durchgeführt werden soll. Dazu sind die verschiedenen Phänomene, die annotiert werden sollen, mit allen möglichen Werten eingetragen. Im Folgenden werden die einzelnen Evaluationen bezüglich der Phänomene im Detail vorgestellt. Die Evaluation der manuellen und automatischen Daten wurde jeweils durch einen Vergleich der zeitlichen Tiers durchgeführt.

Phänomen	**Wert**
Orientieren auf	Kamera, andere Person, eigene Hand, zueinander orientieren
Handaktivität	hoch, runter, links, rechts, vorwärts, rückwärts, symmetrisch
Posen	Default (Ausgangsstellung), rechter Arm nach vorne, linker Arm nach vorne, beide Arme nach vorne, Arme weit auseinander, Arm zum Kopf
Aktivität in Gelenken	Hand seitlich, Hand aufrecht, Ellenbogenverdrehen, Ellenbogen zum Oberarm, Schulter seitlich, Schulter aufrecht, Schulterverdrehen, Kopf Orientierung, Kopf seitlich, Kopf aufrecht
Bewegungsphasen	Bewegungszug (engl. stroke), Haltung (engl. hold), Vorbereitung (engl. preparation), Rückzug (engl. retraction), unvollständiger Rückzug (engl. partial retraction)

Tabelle 19 Phänomene mit den möglichen spezifizierten Zuständen

In der folgenden **Abbildung 65** sind neben den automatischen die manuellen Annotationen angehängt, wodurch PAMOCAT zur Analyse mit der Suche nach Tiers in der Konstellation[100] bezogen auf das Phänomen verwendet kann, um diese zu vergleichen. Dabei sind die manuellen Annotationen weiß und weiter unter angeordnet. Der Vorteil dabei ist, dass die manuellen Annotationen, basierend auf den Videodaten, mit den automatischen von PAMOCAT direkt betrachtet werden können, und so gegebenenfalls die Gründe für die Abweichung mit größtmöglichem Input analysiert werden können. In PAMOCAT können die verschiedenen Annotationen geladen werden, und es kann die Reihenfolge der Tiers geändert werden. Damit kön-

[100] Konstellation aus manuell erhobenen und automatisch berechneten Annotationen mittels „Oder“ Suchfunktionalität.

nen die verschiedenen Annotations-Phänomene von beiden manuellen Annotationen direkt untereinander sortiert werden. Um die manuellen Annotationen automatisch zusammenzuführen, können die übereinstimmenden Zeitpunkte der Annotation in einem neuen Tier automatisch zusammengefasst werden. Es können anschließend zwei Tiers miteinander verglichen werden. Die Übereinstimmung wird berechnet, indem zum einen die komplette Aufnahmezeit daraufhin geprüft wird, ob die Annotationen der zwei jeweiligen Tiers beide gleichzeitig aktiv oder inaktiv sind. Darüber hinaus wird geprüft, ob jedes Annotat[101] des einen Tiers zu mindestens einem Zeitpunkt in der anderen Annotation aktiv war. Es wird aber auch die Übereinstimmung der jeweils aktiven Annotationen geprüft. Dieses wurde für jeden der zwei zu vergleichenden Tiers durchgeführt. Die gesamte Übereinstimmung ist das gleichgewichtete Verhältnis aller fünf Faktoren.

$$Gleichheit = \frac{gesamte\ Zeitlich\ Gleichheit}{5} + \frac{aktiver\ Elemente\ 1\ in\ 2}{5} + \frac{aktiver\ Elemente\ 2\ in\ 1}{5} + \frac{Elementgleichheit\ 1\ in\ 2}{5} + \frac{Elementgleichheit\ 2\ in\ 1}{5} \quad \textbf{(35)}$$

In den folgenden Kapitelteilen wird im Detail auf Übereinstimmung in der Genauigkeit der Annotationen eingegangen werden. Dabei heißt Übereinstimmung nicht gleich besser, es können auch Fehler von den Menschen und von der automatischen Annotationssoftware gemacht worden sein. Genau diese Tatsache wird bei den verschiedenen Phänomenen analysiert.

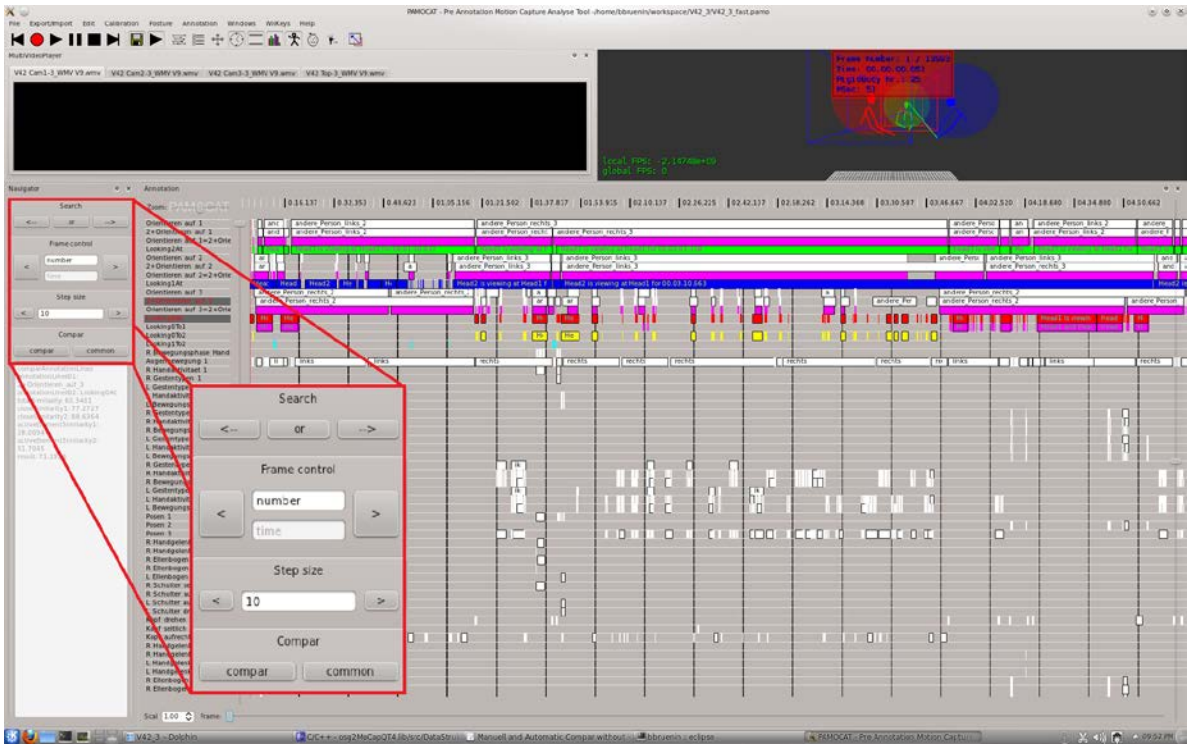

Abbildung 65 PAMOCAT mit automatischer und manuell erzeugter Annotation im Vergleich und einer ausgerechneten Übereinstimmung der beiden selektierten Tiers.

[101] Eine textuelle Beschreibung innerhalb eines Tiers.

Durch eine genauere Analyse der Aufnahmedaten und durch Gespräche mit den annotierenden Personen konnte mehr über die Qualität und auch die Hintergründe hierfür festgestellt werden. Die Unterschiede liegen meistens in Variationen der Anfangs- und Endzeitpunkte, aber auch in teilweise anders zugeordneten Beschreibungen. Die Unterschiede in der manuellen Annotation liegen in einer unterschiedlichen Herangehensweise und der Sorgfalt der einzelnen Personen. Dabei wird z. B. eine Hilfslinie anders gesetzt. Solch eine Hilfslinie kann z. B. im Gesicht zwischen den Augen definiert werden, um daraus eine Schätzung über den Sichtwinkel geben zu können. Dabei können aus dieser individuellen Herangehensweise Unterschiede entstehen. Die Unterschiede zu den automatischen Annotationen sind auch auf die Sichtwinkel und Position der filmenden Kameras zu den Personen zurückzuführen. Bei Objekten, die nahe um die Kamera herum aufgestellt sind, kann besser abgeschätzt werden, ob diese zu einem Zeitpunkt angeschaut werden[102]. Bei den annotierenden Personen kommen leider noch Ermüdungserscheinungen und auch Langeweile oder Unkonzentriertheit hinzu. Aus den einzelnen Frontalansichten kann nur geschätzt werden, wann sich zwei Leute gegenseitig ansehen, da meist nur eine Person sichtbar ist. Um dieses genauestens prüfen zu können, hätte man bei der Annotation zwischen den betreffenden Videos wechseln müssen. In der Praxis wurden bei der Suche nach den jeweiligen Phänomenen die Videos jedoch einzeln nach und nach abgearbeitet. Bei der Deckenkamera können zwar alle Personen in Relation zueinander gesehen werden, aber leider ist nicht ersichtlich, wohin genau die einzelnen Personen schauen.

9.3.1 Unterschiede der manuellen Annotationen zueinander

Um die manuellen Annotationen mit den automatisch von PAMOCAT erzeugten Annotationen bestmöglich vergleichen zu können, ist es erst nötig, eine zusammengeführte Annotation der beiden manuell erzeugten Annotationen zu erstellen. Dazu werden beide Annotationen in PAMOCAT geladen und automatisch zusammengeführt. Dieses geschieht durch eine Vereinigung beider Annotationen; dazu werden ähnliche Annotationen nur dann zusammengeführt, wenn sie in beiden Tiers existieren. Sind die annotierten Texte verschieden, werden beide Beschreibungen eingeführt. Anschließend können diese zusammengeführten Annotationen manuell überprüft werden. Bei den beiden manuellen Annotationen gab es Unterschiede, die in der folgenden **Tabelle 20** festgehalten wurden. Hierbei zeigt sich, dass schon die manuellen Annotationen teilweise eine erstaunlich hohe Abweichung zueinander aufweisen. Dabei muss bei dieser Berechnung berücksichtigt werden, dass hier auch die Zeitpunkte mit einfließen, bei denen nichts hervorgehoben wurde, wodurch der eigentliche Fehler relativ wenig ins Gewicht fällt. Zu diesen Fehlern kommt noch hinzu, dass die gleiche Situation vereinzelt unterschiedlich aufgefasst wird. Zum Beispiel wird „Arme nach vorne“ mit „Arme umschließen“ anders annotiert. Diesen Fall kann man allerdings als gleiche Annotation auffassen. Verwunderlich ist, dass einzelne Phänomena gar nicht annotiert wurden bzw. nur von einem der bei-

[102] Diese Annotationen wurden in ELAN durchgeführt und basieren nur auf den Videodaten.

den manuellen Annotatoren[103]. Eine Erklärung dafür ist Unaufmerksamkeit nach der Fortsetzung bei den Annotationen und dass nach einer Pause nicht zur richtigen Stelle zurück gefunden wurde, um weiter zu annotieren. In einem Einzelfall wurde auch links und rechts vertauscht. Eine weitere Erklärung ist, dass beim Vorgehen eine abweichende Definition verwendet wurde. Zum Beispiel

Phänomen	**Zeitliche**	**Aktive Elemente 1 / Übereinstimmung**	**Aktive Elemente 2/ Übereinstimmung**	**Resultat**
Orientieren auf 1	99.00 %	100 / 99.67%	97.40 / 99.28 %	99.07 %
Orientieren auf 2	98.24 %	98.78 / 98.74%	100 / 99.42 %	99.03 %
Orientieren auf 3	98.38 %	98.92 / 99.25%	100 / 99.06 %	99.12 %
R. Handaktivität 1	98.64 %	100 / 63.30 %	50 / 55.46 %	73.48 %
R. Handaktivität 2	98.70 %	81.81 / 75.49 %	78.57 / 71.58 %	81.23 %
R. Handaktivität 3	95.24 %	92.39 / 74.71 %	100 / 95.95 %	91. 66 %
L. Handaktivität 1	99.45 %	100 / 41.98 %	100 / 100 %	88.28 %
L. Handaktivität 2	98.80 %	90.24 / 89.53 %	89.74 / 75.09 %	88.68 %
L. Handaktivität 3	95.25 %	88.48 / 73.07 %	100 / 93.90%	90.14 %
Posen 1	98.44 %	85.71 / 63.81 %	66.66 / 56.55 %	74.23 %
Posen 2	98.59 %	96.77 / 90.23 %	90.00 / 73.13 %	89.74 %
Posen 3	87.40 %	96.47 / 89.07 %	62.82 / 56.50 %	78.45 %

Tabelle 20 Zusammenführung der manuellen Annotationen

hat die eine annotierende Person nur Posen bei aktiver Bewegung annotiert, die zweite ist aber von der eigentlichen Körperstellung ausgegangen und hat auch Posen annotiert, wenn diese der Ruhestellung sehr nah kamen. In der **Tabelle 20** sind Gelenkaktivitäten ausgelassen, da diese wegen des sehr hohen Zeitaufwandes nur von einer Person annotiert wurden.

9.3.2 Analyse der Unterschiede bei der Annotationsgenauigkeit bezogen auf das Phänomen „Orientiert auf“

Die zusammengeführten manuellen Annotationen des Phänomens „Orientiert auf“ werden nun mit den automatisch erzeugten Annotationen verglichen. Die Gleichheit der Annotationen ist bezüglich zeitlicher Übereinstimmung, aktiver Elemente und Gleichheit der aktiven Elemente von der einen zur anderen Annotation aufgeführt. Die Ergebnisse sind in der **Tabelle 21** aufgelistet. Bei der manuellen Überprüfung des Vergleiches der automatischen und manu-

[103] Natürlich wurden auch diese Zeitpunkte in der Original Software ELAN gegengeprüft, um auszuschließen, dass eine fehlerhafte Verarbeitung der Daten vorliegt.

ellen Annotationen zeigt sich, dass die Unterschiede darauf zurückzuführen sind, dass es für die annotierende Person schwierig ist, abzuschätzen, wann genau wohin geschaut wird, da nur die eine Person im Video sichtbar ist. **Abbildung 66** zeigt den Vergleich von automatischen und manuellen Annotationen; dabei sind die manuellen Annotationen weiß eingefärbt. Pink eingefärbt sind die automatisch zusammengeführten Annotationen und die restlichen (grün, blau und rot) eingefärbten Annotationen sind automatisch erzeugt.

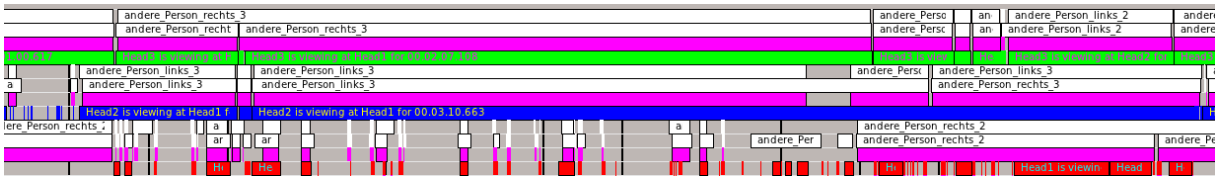

Abbildung 66 Vergleich von automatischen und manuell erzeugten Annotationen in PAMO-CAT

Zudem scheinen die Annotatoren die Augenbewegungen mit in die Analyse einzubeziehen, was eigentlich nicht der Aufgabenstellung entsprach, da nur die wirkliche Kopfausrichtung vom System erfasst werden kann.

Phänomen	Zeitliche	Aktive Elemente 1 / Übereinstimmung	Aktive Elemente 2/ Übereinstimmung	Resultat
Orientieren auf 1	93.99 %	100/ 99.58 %	79.66/ 94.20 %	93.48 %
Orientieren auf 2	82.87 %	97.56 / 98.38 %	76.19 / 83.32 %	87.66 %
Orientieren auf 3	67.33 %	78.18/ 41.10 %	60.60 / 69.14 %	63.27 %

Tabelle 21 Ergebnisse des Vergleichs der manuellen (1) und automatischen (2) Annotationen des Phänomens „Orientiert auf“

Bei der dritten Person sind die Ergebnisse schlechter, deutlich erkennbar aus der **Abbildung 66** und **Tabelle 21**. Um bessere Ergebnisse zu erzielen, könnten die Parameter angepasst werden.

9.3.3 Analyse der Unterschiede bei der Annotationsgenauigkeit bezogen auf das Phänomen „Handaktivität“

Die automatische Annotation des Phänomens „Handaktivität“ kann sehr präzise durchgeführt werden. Dazu muss allerdings das gewünschte Maß an Aktivität eingestellt werden. In manchen Fällen will man nicht jede Bewegung der Hände untersuchen, sondern nur gesprächsrelevante Gesten. Daher kann, je nach gewünschtem Resultat, die Geschwindigkeit, ab wann eine reale Geste interessant ist, eingestellt werden. Die automatische Erkennung detektiert auch Bewegungen, die von den Annotationen nicht erkannt wurden, falls gewünscht. Bei der Erkennung, wann Handaktivität herrscht, ist die automatische Annotation der manuellen Annotation überlegen. Zusätzlich wird die Richtung der Handbewegungen annotiert, bei denen

viele Unterschiede zwischen der automatischen und der manuellen Annotation erkennbar sind. Zum einen erfolgt eine natürliche Bewegung nicht nur entlang einer einzelnen mathematischen Achse, sondern entlang einer Kombination aus mehreren Achsen. Bei der Erkennung der realen Achsen bzw. welche Kombination der Achsen bei einer Bewegung auftritt, sind die manuellen Annotationen auch wieder unterlegen. Dies liegt einmal daran, dass aus einer frontalen Sicht nicht gut abgeschätzt werden kann, entlang welcher Achse eine Bewegung erfolgt; nur horizontale oder vertikale Bewegungsänderungen sind gut erkennbar. Die Bewegungen nach vorne oder nach hinten sind nur aus der Deckenkamera gut sichtbar, welche hierzu aber nicht verwendet wurde.

Phänomen	Zeitliche	Aktive Elemente 1 / Übereinstimmung	Aktive Elemente 2/ Übereinstimmung	Resultat
R. Handaktivität 1	98.38 %	100 / 82.35 %	35.89 / 35.62 %	70.45 %
L. Handaktivität 1	99.67 %	100 / 83.01 %	54.54 / 54.76 %	78.40 %
R. Handaktivität 2	94.89 %	100 / 85.14 %	26.92 / 24.46 %	66.28 %
L. Handaktivität 2	94.52 %	100 / 87.36 %	36.52 / 31.38 %	69.95 %
R. Handaktivität 3	92.27 %	100 / 94.00 %	65.05 /62.37 %	82.74 %
L. Handaktivität 3	92.79 %	100 / 94.31 %	67.74 / 60.99 %	83.17 %

Tabelle 22 Ergebnisse des Vergleichs von manuellen (1) und automatischen (2) Annotationen des Phänomens „Handaktivität“

9.3.4 Analyse der Unterschiede bei der Annotationsgenauigkeit bezogen auf das Phänomen „Posen“

Die manuellen Annotationen des Phänomens Posen sind von einer sehr schlechten Qualität. Viele automatisch gefundene Zeitpunkte sind nicht entsprechend annotiert. Daher wurde hier ein anderes Vorgehen ausgewählt, um die Qualität der automatischen Analysen sicherzustellen. Alle gefundenen Zeitpunkte wurden manuell nachgeprüft, ob die zugehörige Pose stimmt. Die Posen stimmen überein, lediglich bei der Anfangszeit und Endzeiten erkennt die automatische Annotation diese oft früher bzw. länger, als es der Mensch tun würde (durch Parameter änderbar). Dabei wurden alle Gelenke mit gleicher Gewichtigkeit verwendet, damit eindeutig unterschieden werden kann, dass beide Arme oder nur ein Arm nach vorne gegangen war.

9.3.5 Analyse der Unterschiede bei der Annotationsgenauigkeit bezogen auf das Phänomen „elementare Gelenkaktivität“

Die manuelle Annotation von elementaren Gelenkaktivitäten ist eine sehr zeitintensive Prozedur. Dabei hat sich gezeigt, dass die annotierenden Personen speziell geschult werden müssen, damit sie diese Aufgabe überhaupt durchführen können. Die Unterschiede sind allerdings erheblich, da sehr oft elementare Bewegungen nicht erkannt oder nicht den richtigen Gelen-

ken zugeordnet wurden. Es hat sich auch gezeigt, dass die annotierenden Personen wahrscheinlich unbewusst Bewegungen zusammenfassen, die zu verschiedenen Zeitpunkten anfangen und auch von verschiedenen elementaren Gelenken ausgeführt werden. Beispielsweise wird oft die Bewegung der Handgelenke als Gesamtbewegung erfasst und nicht im Detail ermittelt, welche einzelnen Freiheitsgrade der Hand aktiv waren. Die manuelle Annotation ist sehr ungenau und hat Fehler im Gegensatz zur automatischen Annotation. Dasselbe Vorgehen wie bei dem Phänomen „Pose“ wurde auch bei Gelenkaktivität angewendet, es wurden die automatischen Ergebnisse einzeln stichprobenartig auf Stimmigkeit geprüft, da die manuellen von sehr schlechter Qualität waren.

9.3.6 Analyse der Unterschiede bei der Annotationsgenauigkeit bezogen auf das Phänomen „Bewegungsphasen“

Die Unterteilung der Bewegung in ihre einzelnen Phasen funktioniert kaum, allerdings hat sich gezeigt, dass die erkannten Phasen, basierend auf den Bewegungsrichtungen, nicht den manuell annotierten Phasen entsprechen. Phasen, die keine Pause mit deutlicher Richtungsänderung beinhalten, werden nicht gefunden. In manchen Fällen sind Bewegungen nicht nach diesem klar definierten Schema aufgebaut, bzw. die Übergänge sind fließend; es wird keine Pause gemacht oder der Richtungswechsel ist nicht detektierbar (zu weich), wenn er sehr klein ist. Bei diesem Phänomen ist nur die manuelle Annotation für eine spätere Analyse zu gebrauchen.

9.3.7 Ergebnis des Vergleiches manueller und automatischer Annotation

Die Ergebnisse der Evaluierung sind in der folgenden **Tabelle 23** aufgeführt. Dabei werden die Zeit, die für die Erstellung gebraucht wurde, und die Genauigkeit der Annotationen gegenübergestellt. Zu der eigentlichen Zeit zur Berechnung der automatischen Phänomene (im Bereich von mehreren Minuten bis zu mehreren Stunden[104]) muss noch gegebenenfalls die Zeit zur Einstellung der Parameter mitberücksichtigt werden. Allerdings kann man sagen, dass die Annotationszeit von zwei Personen von etwa 32 h eingespart werden kann; zusätzlich müsste noch die Zeit für die manuelle Zusammenführung der Annotationen gerechnet werden (ca. 50 % mehr). Dabei sind z. B. je nach Szenario die Bewegungsgeschwindigkeit bzw. die zurückgelegte Strecke der Hände, der Radius der Sphäre um die Köpfe herum für das Phänomen „Orientiert auf“ und Einstellung der Posen mitzuberücksichtigen. Durch den Vergleich der beiden Annotationen ist ersichtlich, dass in den menschlichen Annotationen verschiedene Merkmale unterschiedlich annotiert wurden. Dies liegt einmal daran, dass selbst bei klar definierten Regeln immer noch viel Freiraum zur Interpretation der Daten vorhanden ist. Die annotierenden Personen waren zwei verschiedene Personen, die beide schon ausführliche

[104] Die tatsächliche Zeit ist abhängig von der Aufnahmelänge und den verwendeten Frameraten. Dabei kann eine 10-minütige Aufnahme mit 20 FPS in ca. 30 Minuten berechnet werden und eine mit 160 FPS in ca. 4 Stunden, abhängig von der Rechnerleistung.

Phänomen	**Manuelle Person 1/ 2**	**Übereinstimmung / Qualität**
Orientieren auf (Augen)	3.15 h/ 2.40 h (3.00 h)	92 % / beide gute Ergebnisse
Handaktivität	4.15 h/ 3.10 h	80 % / automatische Annotationen sind deutlich besser
Posen	1.46 h/ 1.25 h	Manuelle Annotationen sind sehr schlecht und schon die beiden manuellen Annotationen unterscheiden sich sehr.
Aktivität in Gelenke	8.54 h	Sehr schlechte Ergebnisse der manuellen Annotationen
Bewegungs-phasen	4.01.07 h / 2.55 h	An der Erkennung der Bewegungsphasen muss noch weiter gearbeitet werden, aktuell ist die automatische Erkennung noch nicht für den praktischen Einsatz zu gebrauchen.

Tabelle 23 Resultat des manuellen und des automatischen Annotierens

Erfahrungen mit dem Annotieren und dem Tool ELAN hatten. Wegen des immensen Zeitaufwandes wurden nicht alle Phänomene zweifach annotiert, da z. B. die Annotation von elementaren Aktivitäten sehr viel Zeit in Anspruch nimmt und nur extrem schwer durchzuführen ist. Es ist für die Menschen schwer zu interpretieren, einmal, wann in welchem Gelenk Aktivität vorkommt, und zusätzlich, die genaue Zuordnung der Aktivität zu den elementaren Gelenken durchzuführen. Ein weiterer Punkt scheint auch die Motivation zu sein. Auch wenn beide Personen sehr motiviert waren, so scheint Person 1 genauer gearbeitet zu haben als die Person 2. Dazu fallen auch die Unterschiede bei der Genauigkeit der Zeiterfassung auf, wobei die Vorgabe nur war, die Zeit zu protokollieren, und Sekunden eigentlich bei Annotationen weniger von Bedeutung sind. Die manuellen Annotationen sind teilweise stark unterschiedlich und haben auch Fehler. Da die Ergebnisse der manuellen Annotationen untereinander selber schon starke Abweichungen haben, wurde hier darauf verzichtet, eine ROC – „Receiver Operating Characteristic“[105]-Auswertung mit dem Einfluss von verschiedenen Parametern auf die Annotationsrichtigkeit durchzuführen.

9.4 Usability von PAMOCAT

PAMOCAT bietet neben den automatischen Annotationen auch die Möglichkeit, manuell zu annotieren. An dieser Stelle soll das manuelle Annotieren untersucht werden. Dazu wurde eine kleine Studie mit erfahrenen ELAN-Anwendern und unerfahrenen Probanden durchge-

[105] Bei der ROC-Auswertung wird eine Kurve aus den verschiedenen Eingabewerten und die resultierende Korrektheit dargestellt. Anhand dieser Kurve kann analysiert werden, ob eine Verbesserung durch Werte erreicht wird, ob die Werte richtig interpretiert werden oder ob gar kein Zusammenhang der Parameter und Ergebnisse existiert.

führt. Diese sollten das Tool PAMOCAT benutzen, um kleinere Annotationsaufgaben durchzuführen. Anschließend sollten die Erfahrungen durch einen Fragebogen festgehalten werden. Die Ergebnisse sind in der **Tabelle 24** zusammengefasst.

Aufgabenstellung	**Sehr Leicht**	**Leicht**	**Normal**	**Schwer**	**Sehr Schwer**
Projekterstellung	1	1	2	0	0
Speichern	2	0	1	0	0
Exportieren	0	1	1	0	0
Tiererstellung	2	0	2	0	0
Tierreihenfolge ändern	4	0	0	0	0
Annotationserstellung	1	1	2	0	0
Annotation ändern	1	1	1	1	0
Videowechsel	3	0	1	0	0
Automatisches Annotieren	0	0	2	0	0
Fensterverständlichkeit	0	0	2	0	0
Parameteranpassung	1	1	0	0	0
Anordnung der Visualisierungsfenster	0	1	1	0	0
Fensterverwaltung	0	1	1	0	0
Visualisierungsoptionen	1	1	0	0	0
Suchfunktionalität	0	0	2	0	0

Tabelle 24 Usability bezüglich des manuellen Annotierens in PAMOCAT

Die Aufgabenstellung gab den Annotatoren vor, die ersten drei Zeitpunkte zu finden, an denen sich zwei Probanden aufeinander zu orientierten. Dabei sollten das Erstellen eines PAMOCAT Projektes, das Erstellen von Tiers, das Erstellen von Annotationselementen und der Wechsel zwischen den verschiedenen Videos im Mittelpunkt stehen. Anschließend sollte die Analyse Funktionalität mittels der Suche nach Kombinationen von Tiers genutzt werden. Je nachdem, wie schnell die Probanden waren, konnten sie noch die Fensterverwaltung, die Visualisierungsoptionen, die automatischen Annotationen und das Exportieren nach ELAN testen. Die Probanden hatte dazu 30 Minuten Zeit erhalten, mit einem zusätzlichen Puffer von 15 Minuten. Zum Schluss wurden die Probanden gebeten, einen Fragebogen auszufüllen. Die Ergebnisse des Fragebogens sind in der folgenden Tabelle aufgeführt.

Zudem wurden die Probanden gebeten, eine Aussage zu machen, wie sich PAMOCAT im Vergleich zu ELAN bedienen lässt. Die Ergebnisse hieraus sind in der **Tabelle 25** zu finden. Sehr positiv zu vermerken ist, dass es viele der Probanden einfacher fanden, PAMOCAT zu bedienen als ELAN.

Aufgabenstellung	**Sehr Leicht**	**Leicht**	**Normal**	**Schwer**	**Sehr Schwer**
Manuelles Annotieren	0	1	1	0	0
Videoauswahl	1	1	0	0	0
Projekterstellung	1	0	1	0	0
Fensterverwaltung	0	1	1	0	0

Tabelle 25 Usability im Vergleich zu ELAN

Das Anpassen der Annotationen ist als Einziges etwas negativ aufgefallen, dazu meinten die Probanden aber, dass sie sich an ELAN schon sehr gewöhnt haben. Hier ist eine Änderung der Endzeit eines Annotationselements nur über ein Zahlenänderungselement innerhalb eines Dialoges und nicht durch mehrfaches Klicken im Annotationsbereich möglich. Diese Änderung wird in PAMOCAT integriert werden.

9.5 Zusammenfassung

In diesem Kapitel wurde erläutert, wie die Evaluation des Tools PAMOCAT durchgeführt wurde. Es wurde eine Analyse durchgeführt, wie gut das Motion-Capturing insgesamt arbeitet. Es wurde auch untersucht, wie die zum Motion-Capturing verwendeten Rigidbodys die Bewegungen der Probanden verändert haben könnten. Dabei ist der Nutzen bei der möglichen Bewegungsveränderung berücksichtigt worden. Damit würde durch das Motion-Capturing sonst zwar Zeit gewonnen, aber andererseits müsste mehr Zeit zum Nacharbeiten von Fehlern in den Motion-Capture-Daten aufgewendet werden. Darüber hinaus wurde die Genauigkeit der automatischen Annotation im Vergleich zur manuellen für verschiedene Phänomene im Detail geprüft. Dabei hat sich gezeigt, dass sehr viel Zeit gespart werden kann. Ein weiterer Punkt ist die gleichbleibende Qualität der Annotationen, wodurch eine einheitliche Qualität der Annotationsdaten gewährleistet wird, da gegebenenfalls Abweichungen einheitlich sind. Die manuell annotierten Daten sind teilweise schon bei der Annotation recht unterschiedlich und haben teilweise eine schlechte Qualität. Insgesamt ist zu sagen, dass bei den Phänomenen „Orientiert auf“, „Handaktivität“, „Zeigen auf“ und „Posen“ die automatischen Annotationen besser als die manuellen sind. Dabei ist die Zeitersparnis von 16 Stunden für eine Aufnahmezeit von 10 Minuten ein gewaltiger Fortschritt mittels der automatischen Berechnungen im Minutenbereich. Dadurch können detaillierte Annotationen auch für große und umfangreiche Korpora automatisch erstellt werden. Außerdem können die Berechnungen für einen großen

Korpus mit vielen einzelnen Aufnahmen durch die Verwendung von Scripts automatisch nacheinander abgearbeitet werden. Bei dem Phänomen „elementare Aktivität in Gelenken“, welches einen sehr großen Zeitaufwand bei der manuellen Annotation erfordert, konnte auch festgestellt werden, dass die Qualität der automatischen Annotationen viel besser ist als die der manuellen; genau wie bei dem Phänomen der „Posen“, bei dem die Körperstellungen doch teilweise sehr unterschiedlich annotiert wurden. Bei der Erkennung des Phänomens der „Bewegungsphasen“ ist hingegen die Qualität der manuellen Annotationen deutlich besser, bzw. hier muss die Erkennungseffektivität der automatischen Funktionen noch verbessert werden. Das Programm PAMOCAT ermöglicht eine Kombination aus klassischen qualitativen Videoannotationen und quantitativen automatischen Annotationen basierend auf Motion-Capture-Daten. Allerdings können nicht alle Phänomene auf Basis von Motion-Capture-Daten automatisch annotiert werden. In PAMOCAT wurden auch verschiedene Funktionalitäten eingebaut, um das Zusammenführen der Annotationen zu unterstützen bzw. Annotationen zu vergleichen. Zum Schluss wurde eine Evaluierung der Qualität des manuellen Annotierens mit PAMOCAT im Vergleich zu anderen Annotationstools vorgestellt. Allerdings ist PAMOCAT als „**P**re **A**nnotations Tool für **Mo**tion-**Ca**p**t**ure-Daten“ entwickelt worden und soll hauptsächlich zum automatischen Annotieren verwendet werden, auch wenn manuelle Annotationen sehr gut möglich sind.

10 Schlusswort

In dieser Arbeit wurde untersucht, wie Motion-Capture-Daten am besten nutzbar gemacht werden können, um menschliches Interaktionsverhalten zu analysieren. Dazu wurden sogenannte Rigidbodys mit einer eigenen inversen Kinematik verwendet, um ein stabiles Motion-Capturing über einen längeren Zeitraum zu erreichen, damit nicht übermäßig viel Zeit in die Vor- und Nachbearbeitung investiert werden muss. Damit wurde ein Weg gefunden, mit dem das Motion-Capturing mit einem angemessenen Zeitaufwand für die Verhaltensforschung genutzt werden kann. Diese Funktionalität, die es ermöglicht, multiple Personen gleichzeitig aufzuzeichnen, ist in die Software PAMOCAT eingeflossen, die von Verhaltensforschern genutzt werden kann, um Interaktionsanalysen durchzuführen. Mit diesen Resultaten können die Motion-Capture-Daten genutzt werden, um verschiedene elementare Phänomene automatisch zu finden, mit deren Hilfe dann komplexere Verhaltensweisen untersucht werden können. Dazu steht eine Reihe von elementaren Interaktionsphänomenen zur Verfügung, die individuell miteinander kombiniert werden können. Diese elementaren Phänomene sollen genutzt werden, um Hypothesen zu erstellen, und durch eine Suche nach Zeitpunkten, bei denen eine Kombination von Phänomenen auftritt, diese eventuell zu belegen. Dazu können diese Daten von allen Richtungen im Motion-Capture-View analysiert und mit den synchron gehaltenen multiplen Videos genauestens betrachtet werden. Mittels der Motion-Capture-Daten können Zeitpunkte gefunden werden, die in den Videos im Detail analysiert werden können. Zudem können verschiedene Tiers auf Gleichheit geprüft und zusammengefasst werden, um auf unterschiedlichen Phänomenen basierendes Verhalten zu untersuchen. Dabei soll das hier vorgestellte Tool nicht bereits existierende Tools ersetzen, sondern auf der Basis einer erweiterten Modalität zusätzliche (automatische) Analysen ermöglichen. Diese Ergebnisse können dann in altbewährten Annotationstools wie ELAN weiter analysiert werden, unter Ausnutzung der individuellen Funktionalitäten dieser Tools. In anderen aktuellen Annotationstools wie ELAN oder ANVIL müssen diese Phänomene mühsam stundenlang annotiert werden. PAMOCAT erspart hingegen viel Annotationszeit, die fast das 100-fache der Aufnahmezeit betragen kann. Die automatische Berechnung hingegen kann im Bereich von Minuten durchgeführt werden[106]. Zusätzlich können spezielle Suchfunktionen und Betrachtungen wie auf Motion-Capture basierende Visualisierungen die Analyse detaillierter machen. Ein weiterer Vorteil der automatischen Annotationen sind die immer gleichen Kriterien, die bei den annotierten Phänomenen genutzt werden. Dadurch wird der menschliche Fehlerfaktor bei den Annotationen ausgeschlossen und eine hohe Annotationsqualität erreicht. So können die Vorteile von quantitativen und qualitativen Methoden der Verhaltensforschung zusammen genutzt werden. In dieser Arbeit wurden aber auch die Grenzen des automatischen Annotierens auf-

[106] Für die im vorherigen Kapitel analysierte Video Sequenz von 10 Minuten wurden 31 Stunden benötigt, um die manuelle Annotation anzufertigen.

gezeigt, nicht alle Phänomene, wie z. B. die verschiedenen Gestenphasen, lassen sich mit einer hohen Qualität automatisch annotieren. Allgemein können mit der Möglichkeit, verschiedene Parameter anzupassen, auch die resultierenden Annotationen individuell sensibler gestaltet (bzw. die allgemeine Detektion von Phänomenen beeinflusst) werden. Im Folgenden werden nun abschließend weitere Möglichkeiten betrachtet, wie diese Software erweitert werden könnte, um an verschiedene Szenarien individuell angepasst zu werden. Dabei fällt der Schwerpunkt auf mögliche Erweiterungen bezüglich Softwarefunktionalitäten und auf die Anbindung verschiedener Hardware.

10.1 Mögliche Softwareerweiterungen

Es gibt viele Möglichkeiten, PAMOCAT zu erweitern, einige naheliegende Erweiterungen sollen an dieser Stelle kurz näher betrachtet werden.

1. **Optimierung der Benutzerschnittstelle**

   Eine mögliche Erweiterung wäre, die Flexibilität noch mehr den Benutzern anzupassen. Dabei wurde bei der Entwicklung darauf geachtet, dass die GUI flexibel und eigenständig konfigurierbar ist. Eine Verbesserung wäre, alle Videos einzeln in der GUI immer sichtbar platzieren zu können, damit diese bei verschiedenen Annotationen und Analyseaspekten individuell angepasst werden können. Zudem könnten die GUI durch Studien zur Benutzbarkeit allgemein optimiert werden.

2. **Kommunikation mit externen Annotationstools wie ELAN**

   Auch wenn PAMOCAT als eigenständiges Tool entwickelt wurde, wäre es hilfreich, die Möglichkeit zu haben, es an ELAN anzubinden. Damit könnte über die Netzwerk Schnittstelle von ELAN kommuniziert werden, um einmal die Motion-Capture-Daten in ELAN visualisierbar zu machen, aber auch, die automatischen Annotationen individuell aus ELAN heraus als Erweiterung steuerbar zu haben[107].

3. **Video basierte Detektionserkennung von Gesichtsmimik**

   Für Studien mit langen Aufnahmezeiten, wie sie in der auf Interaktionen basierenden Verhaltensforschung durchgeführt werden, ist Performance-Capturing[108] nicht einsetzbar. Zum einen ist der zeitliche Aufwand der Vorbereitung und der Nachbearbeitung zu groß. Andererseits wäre mit solch einer zusätzlichen technischen Ausrüstung auch die Ablenkung der Probanden zu stark. Daher müssen Verfahren aus der Bildverarbeitung verwendet werden, bei denen die Erkennung auf Bilddaten basiert. Dazu könnten externe Bibliotheken angebunden werden, die Gesichtsmimik erkennen und annotieren können. Mit die-

[107] Die automatischen Annotationen werden im ELAN-Datenformat gespeichert und können auch jetzt schon in ELAN verwendet werden.

[108] Motion-Capturing des ganzen Körpers mit zusätzlicher Bewegungserfassung des Gesichts.

ser Technik wäre auch die nachträgliche Ermittlung von Gesichtsmimik auf den bereits aufgezeichneten Korpussen möglich.

4. Spracherkennung

Die automatische Spracherkennung ist heutzutage noch schwierig und auch noch nicht sehr zuverlässig. Allerdings kann sie eine Erleichterung bei der Annotation im Allgemeinen sein und speziell bei dem in dieser Arbeit untersuchten Zusammenhang von Sprache und körperlicher Gestik. Dazu könnten auch Bibliotheken zur Übersetzung von geschriebener Sprache in andere Sprachen mit eingebunden werden.

5. Persönlichkeits-Erkennungsfunktion

Die Körpersprache kann viel über einen Menschen aussagen, durch die Detektierung von verschiedenen typischen Posen für verschiedene Verhaltensweisen wäre die Ermittlung verschiedener charakteristischer Merkmale einer Persönlichkeit möglich, wie zum Beispiel, ob es sich vielleicht um einen eher schüchternen Menschen oder einen sehr selbstbewussten Menschen handelt. Es könnte aber auch analysiert werden, ob es sich um einen links- oder rechtshändigen Menschen handelt. Oder auch, ob es wahrscheinlich ein eher aktiver oder passiver Mensch ist. Manche Posen und die Tatsache, ob es eher ein aktiver oder passiver Mensch ist, könnten Schüchternheit oder Selbstsicherheit vermuten lassen.

6. Einbindung einer SVM, um Bewegungen wiederzuerkennen

Auch wenn komplexere Bewegungen bei der Interaktion sehr unterschiedlich ausfallen und meist auf elementarerer Ebene Anhaltspunkte gesucht werden müssen, um diese zu vergleichen, kann es nützlich sein, Bewegungssequenzen wiederzufinden. Dazu würde sich die Anbindung einer Support-Vektor-Maschine zum Trainieren verschiedener typischer Bewegungen eignen, ähnlich wie es in der Arbeit [66] beschrieben ist.

7. Betriebssysteme

Linux ist ein Betriebssystem, mit dem nicht alle Benutzer arbeiten, da es in der Benutzung teilweise komplizierter ist, auch wenn es viele Möglichkeiten bietet, die mit anderen Betriebssystemen nicht möglich sind[109]. Daher ist ein wichtiger Schritt, PAMOCAT auf andere Betriebssysteme zu portieren. Zu diesen Betriebssystemen zählen Windows und MacOS in ihren verschiedenen Versionen.

10.2 Anbindung weiterer Hardware

Durch zusätzliche Hardware kann die Funktionalität in verschiedenen Bereichen erweitert und die Durchführung vereinfacht werden. Dabei gibt es eine Reihe von zusätzlicher Hardware, die während verschiedener Studienvorbereitungen als wünschenswert angesehen wurden.

1. Netzwerkkameras

[109] Zum Beispiel Einfluss auf den Kernel nehmen.

Mit Netzwerkkameras[110] kann die Videoaufnahme automatisch synchron und mit einem einzelnen Klick gestartet und ohne Weiteres direkt an einer zentralen Stelle gespeichert werden. Aktuell wurden die Videoaufnahmen von externen Kameras durchgeführt, die später entsprechend zusammengeführt bzw. kopiert werden mussten, um in zeitlichen Einklang gebracht zu werden. Die einzelnen Kameras haben immer etwas Versatz und ihre Aufnahmen müssen zurechtgeschnitten werden. Zudem erfordern manche Kameras, bedingt durch die Filegröße und das Komprimierungsverfahren, zusätzlichen Arbeitsaufwand, um die Videos in ein geeignetes Format zu bringen, in dem es gut analysiert werden kann. Durch ein System, das entweder auf verschiedene Rechner verteilt ist oder in einem sehr leistungsfähigen Rechner mit entsprechenden Festplattenschreibefähigkeiten untergebracht ist, könnten die Daten zwischengespeichert und dann zum Schluss an einem Ort als Projekt abgespeichert werden. Dabei wird das Aufzeichnen mit allen Kameras gleichzeitig begonnen, wodurch auch alle Aufnahmen zueinander synchron sind.

2. Eyetracker

Um exakt untersuchen zu können, worauf sich die Augen der Probanden richten, müssen diese mittels Eyetracking erfasst werden. Bei mehreren Versuchspersonen müssen diese Daten mit den Motion-Capture-Daten synchron kombiniert werden. Dazu müssen die verschiedenen Eyetracker jeweils über mehrere Rechner betrieben werden; die resultierenden Daten können über das Netzwerk an einen geeigneten Rechner geschickt werden.

3. Fingertracking

Speziell bei der Gestenforschung in Bezug zur Sprache, d. h. wenn Personen anderen Personen Sachverhalte beschreiben, werden die Finger mitbenutzt. Diese machen zusätzlich verschiedene spezielle Bewegungen, die für die genaue Analyse hilfreich sind, da somit zusätzliche Details erfassbar sind. Hierzu gibt es verschiedene Techniken, die verwendet werden könnten, um sie zu integrieren: Einmal Cybergloves, welche Handschuhe darstellen, die zusätzlich getragen werden müssten, oder auf aktiven infraroten Markern basierende Tracking-System-Erweiterungen wie das der Firma ART[111]. Basierend auf diesen Bewegungsdaten könnten wiederum Phänomene wie Fingerposen berechnet werden.

4. Microsoft-Kinect

Die Motion-Capture-Systeme sind in der Anschaffung teuer; eine vielversprechende Alternative ist die von Microsoft entwickelte Kinect 2, eine Verbesserung der ersten Version, die auch Orientierungen von Körperteilen, die Bewegung von Fingern und Gesichtsausdrücke erfassen kann. Die Vorteile sind, dass nicht einmal mit störenden Markern gearbeitet werden muss, die ablenken oder die Bewegung verändern. Zudem werden Tiefenbilder und Farbbilder automatisch durch eine entsprechende SDK OpenNI synchron erfasst. Ein großer Nachteil ist aktuell noch, dass mittels der aktuell verfügbaren Kinect-

[110] Aber auch anderen Schnittstellen, um eine synchrone Aufzeichnung starten zu können.

[111] ART ist die Abkürzung für „Advance Realtime Tracking“.

Sensoren nicht genau genug gearbeitet werden kann, sodass die Kopforientierung und die Orientierung der Hände nicht genau ermittelt werden kann. Zudem werden beim Aufzeichnen von mehreren Personen andere Personen durch diese verdeckt, und manche Posen sind aus manchen Blickwinkeln nicht oder nur ungenau zu ermitteln. Ein System, das aus mehreren Kinect 2 besteht und bei dem die Motion-Capture-Daten miteinander synchronisiert werden, wäre wünschenswert.

10.3 Fazit

Die Ergebnisse, die aus der Gegenüberstellung von manueller zu automatischer Annotation resultieren, lassen ein klares Ergebnis erkennen, nämlich dass mit dem Tool PAMOCAT viel Zeit beim Annotieren gespart werden kann und manche Annotationen viel genauer erstellt werden können. Dabei stellt PAMOCAT einen Schritt dar, Motion-Capturing in der Verhaltensforschung zu verwenden, bei der dieses mit normalem Arbeitsaufwand durchgeführt werden kann. Die Zeit für die Analyse liegt dabei weit unter der manuellen Annotationszeit. Zusätzlich wird hier ein klar definiertes Verfahren für das Annotieren (Coden) verwendet, bei dem der menschliche Fehlerfaktor keinen Einfluss mehr auf die resultierenden Annotationen hat. Zusätzlich müssen keine annotierenden Personen angelernt werden, was bei der Analyse von Verhaltensweisen wiederum Zeitersparnis darstellt. Insgesamt stellt PAMOCAT ein Annotationstool dar, das auf Motion-Capture-Daten basiert und in dem auch zusätzliche detailliertere Analysen durchgeführt werden können, um Hypothesen zu evaluieren und neue aufzustellen. Diese Ergebnisse können nachträglich mit anderen Tools wie ELAN weiter untersucht werden. Dabei kombiniert PAMOCAT quantitative mit qualitativen Methoden zur Analyse von Interaktionsverhaltensweisen. Die Usability Studie, in der PAMOCAT von Anwendern getested wurde, hat zudem gezeigt, dass es auch im Bereich einfacher Bedienbarkeit sogar Vorteile zu den existierenden Tools gibt neben der Funktionalität des Motion-Capturings.

A. Mathematische Grundlagen

In diesem Teil des Anhangs werden die mathematischen Grundlagen beschrieben, die für verschiedene Berechnungen speziell im Bereich der Vorwärtskinematik für die Gelenkwinkel erforderlich sind. Die Extraktion von Winkeln aus einer Rotationsmatrix wird benötigt, um die Orientierung z. B. einer Hand in einer mathematischen Darstellung auf einen Roboter zu übertragen. Je nachdem, wie der Roboter aufgebaut ist und sich mit seinen spezifischen DOFs von anderen Robotern unterscheidet, müssen entsprechend andere Winkelstellungen für die einzelnen Gelenke berechnet werden. Ist die Ausrichtung eines Gelenkes durch eine Matrix M bekannt und es müssen die Winkel in den verschiedenen Winkeldarstellungen ermittelt werden, so können diese extrahiert werden. Eine Rotationsmatrix M beschreibt in eindeutiger Weise eine Orientierung, da in ihren Spalten die Einheitsvektoren stehen, die ein Koordinatensystem aufspannen. Der Vektor in der ersten Spalte der Matrix steht für die x-Achse, der zweite für die y-Achse und der dritte für die z-Achse. Allerdings kann so eine Rotationsmatrix auf verschiedene Weise durch einzelne Rotationsmatrizen entsprechend verschiedener Konfigurationen entstehen. Dazu wird die Extraktion der einzelnen Winkel mittels zweier Arten von Konfigurationen betrachtet, die Euler-Winkel und die Roll-Pitch-Yaw Extraktion, die in Sektion 4.6 benötigt wurden [22].

A.1 Extraktion von Euler-Winkeln

Um die Winkel in der Euler-Darstellung zu berechnen, müssen diese allgemeinen Konfigurationen durch Multiplikation der elementaren Rotationsmatrizen ermittelt werden. Anschließend können die entsprechenden Euler-Winkel durch eine Analyse der Komponenten der Matrix ausgerechnet werden.

$$M_{Euler} = R_{z,\Phi} \times R_{y,\Theta} \times R_{z,\Psi} =$$

$$\begin{pmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos(\Theta) & 0 & -\sin(\Theta) \\ 0 & 1 & 0 \\ \sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \times \begin{pmatrix} \cos(\Psi) & -\sin(\Psi) & 0 \\ \sin(\Psi) & \cos(\Psi) & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} c(\Phi)\,c(\Theta)\,c(\Psi) - s(\Phi)\,s(\Psi) & -c(\Phi)\,c(\Psi) - c(\Phi)\,c(\Theta)\,s(\Psi) & -c(\Phi)\,s(\Theta) \\ s(\Phi)\,c(\Theta)\,c(\Psi) + c(\Phi)\,s(\Psi) & c(\Phi)\,c(\Psi) - s(\Phi)\,c(\Theta)\,s(\Psi) & -s(\Phi)\,s(\Theta) \\ -c(\Theta)\,c(\Psi) & s(\Theta)\,s(\Psi) & c(\Theta) \end{pmatrix} =$$

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} \quad \textbf{(36)}$$

Um die Lösung des Gleichungssystems zu finden, müssen zwei Fälle separat betrachten werden:

1. Fall: m_{13} und m_{23} ungleich Null

$\Phi = \text{atan2}(m_{13}; m_{23})$

$\tan(\Phi) = m_{32}/m_{13}\Theta = \arccos(m_{33})$

$\tan(\Psi) = -m_{32}/m_{31}$

2. Fall: m_{13} und m_{32} gleich Null

$\Phi = 0$ oder $\Phi = \pi$

$\Rightarrow \Phi + \varphi = \text{atan2}(m_{11}; m_{21})$

Der 1. Fall ist der am häufigsten eintretende Fall, bei dem die einzelnen Winkel eindeutig ausgerechnet werden können. Um den Winkel zu errechnen, wird die Umkehrfunktion des Cosinus bezogen auf das 33-Element der Matrix angewendet $\Theta = \arccos(m33)$. Die anderen beiden Winkel lassen sich mit Hilfe der Funktion atan2 ausrechnen. Der Tangens ist gegeben durch $\tan(x) = \sin(x) / \cos(x)$. Wenn die Elemente der Matrix m13 und m23 als Argument verwendet werden, kürzt sich der sin() heraus.

Der 2. Fall hingegen beschreibt die Situation, bei der die Rotation um die y-Achse mit dem Winkel gleich 0 oder 180 ist. Dieses kann man aus den Elementen m13 und m23 der Matrix schließen. Bei diesem Fall 2 sind die Achsen der ersten und letzten Rotation parallel oder antiparallel. Daher kann nur die Summe der beiden Winkel - + ausgerechnet werden. Es gibt somit eine unendliche Anzahl an Lösungen, wobei die Rotation der einen Achse um den Winkel - durch einen entsprechend entgegengesetzten Rotationsanteil des anderen Gelenks kompensiert werden kann. Hierbei spricht man von dem sogenannten „Gimbel-Lock“, siehe **Abbildung 67**; die Gimbel-Lock-Stellung entspricht in der Robotik einer singulären[112] Gelenkstellung eines Manipulators. Trotzdem beschreibt die Matrix eine eindeutige Lage des Koordinatensystems, nur gibt es unendlich viele Wege, diese Lage durch die Euler-Winkel zu erreichen.

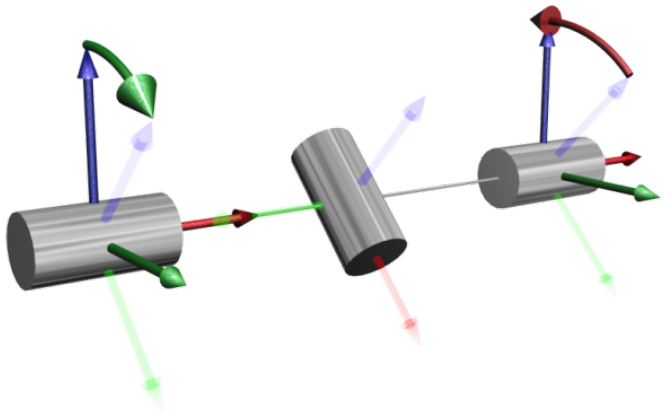

Abbildung 67 Gimbel-Lock, zwei Gelenke sind parallel, und es gibt eine unendliche Anzahl an möglichen Gelenkstellungen

[112] Roboterstellungen, die einer Singularität nahe sind, sind mathematisch schwierig zu handhaben, da z. B. unendliche Geschwindigkeiten aufgebracht werden müssen, um von einer Position in eine gewünschte andere zu kommen, was technisch nicht möglich ist.

A.2 Extraktion von Roll-Pitch-Yaw-Winkeln

Das gleiche Vorgehen ist bei der Extraktion der Roll-Pitch-Yaw-Winkel aus einer Rotationsmatrix zu verwenden. Dabei liegt der Unterschied in der Konfiguration, mit der die Rotationsmatrix zusammengesetzt ist. Zuerst wird die Berechnung der zusammengesetzten Rotationsmatrix durchgeführt, und anschließend werden die Elemente dieser Matrix analysiert [22]

$$R_{RollPitchYaw} = R_{z,\Phi} \times R_{y,\Theta} \times R_{x,\Psi} =$$

$$\begin{pmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos(\Theta) & 0 & -\sin(\Theta) \\ 0 & 1 & 0 \\ \sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Psi) & -\sin(\Psi) \\ 0 & \sin(\Psi)\,0 & \cos(\Psi) \end{pmatrix} =$$

$$\begin{pmatrix} c(\Phi)\,c(\Theta) & -s(\Phi)\,c(\Psi) - c(\Phi)\,s(\Theta)\,s(\Psi) & s(\Phi)\,c(\Psi) - c(\Phi)\,s(\Theta)\,s(\Psi) \\ s(\Phi)\,c(\Theta) & c(\Phi)\,c(\Psi) + s(\Phi)\,c(\Theta)\,s(\Psi) & -c(\Phi)\,s(\Psi) - s(\Phi)\,s(\Theta)\,c(\Psi) \\ s(\Theta) & c(\Theta)\,s(\Psi) & c(\Theta)\,c(\Psi) \end{pmatrix} =$$

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} \quad \textbf{(37)}$$

Die gesuchten Winkel können wie folgt aus einzelnen Elementen der Matrix berechnet werden:

$\Phi = \text{atan2}(m_{21}; m_{11})$ $\tan(\Phi) = m_{21}/m_{11}$

$\Theta = \arcsin(+m_{31})$

$\varphi = \text{atan2}(m_{32}; m_{33})$ $\tan(\Psi) = m_{32}/m_{33}$

B. Detaillierte Beschreibung der Implementierung des ToolKits und PAMOCAT

In diesem Teil des Anhangs geht es um eine genauere Beschreibung des Toolkits, welches einer Sammlung von Funktionalitäten entspricht. Um die Verständlichkeit, Erweiterbarkeit und Wartbarkeit zu gewährleisten, wurde versucht, allgemeine Funktionalitäten zu separieren, damit diese wieder verwendet werden können. Dabei zählt das Toolkit zu den allgemeinen Funktionalitäten, die keinen direkten Bezug zur Bewegungsanalyse haben. Anschließend wird kurz die Anwendung von PAMOCAT aus dynamischer Sicht beschrieben, um das allgemeine Laufzeitverhalten zu erläutern.

B.1 Die Basis Teilkomponenten des ToolKits

Die erste Teilkomponente heißt „Basis“. Die Namensgebung basiert auf der Abhängigkeit von C++ und den Standardklassen. Die Klasse „ToString“ ist programmiert worden, um alle möglichen Datentypen oder auch Klassen in einen String oder Zeichenkette umzuwandeln[113]. Die Klasse „Time“ verwaltet die Zeit. Intern ist die Zeit in Millisekunden seit 1970 gespeichert und bietet Funktionen zum Setzen, Addieren, Subtrahieren und zum Vergleichen von Zeitpunkten an. Außerdem werden die Millisekunden auch in andere Zeiteinheiten wie Jahr, Monat, Tag, Stunde und Minute umgerechnet. Die Klasse „BasicTypes“ beinhaltet eine Reihe von mathematischen Grundlagenklassen wie Punkt, Vektor, Color, Linie, Winkel, Dreieck und Punktwolke. Dazu sind alle Basisoperationen enthalten[114] und zusätzliche weitergehende Funktionen wie die Berechnung eines Winkels zwischen zwei Vektoren oder die Berechnung des Kreuzprodukts zweier Vektoren usw. Der Grund für die eigene Implementierung ist, dass möglichst unabhängig von anderen Bibliotheken gearbeitet werden sollte, um Abhängigkeiten von anderen Bibliotheken und Betriebssystemen zu minimieren und um möglichst einfache Funktionen auf embedded Systemen wie Robotern (z. B. Nao der Firma Aldebaran) portieren zu können. Die Klasse „MatrixPath“ erweitert die Klasse der „Matrix“ um die Möglichkeit, eine Kette von Matrizen mit definierten Zeitpunkten zu erzeugen. Diese wird benutzt, um auf einem Pfad verschiedene Orientierungsrichtungen speichern zu können. Damit ist eine interaktive Fahrt mit verschiedenen vorgegebenen Sichten möglich, zwischen denen die verschiedenen Matrizen zu jeweils einem individuellen Zeitpunkt interpoliert werden. Dieser gesamte Zusammenhang ist in der **Abbildung 68** dargestellt.

[113] Daher gibt es zu jeder Klasse eine Abhängigkeit, die es ermöglicht, verschiedene Informationen über diese Klasse als String abzufragen, z. B. zum Zwecke der Laufzeitdokumentation in einer Logdatei.

[114] Die Basisfunktionalität beinhaltet das Addieren, die Längeberechnung, die Flächenberechnung usw.

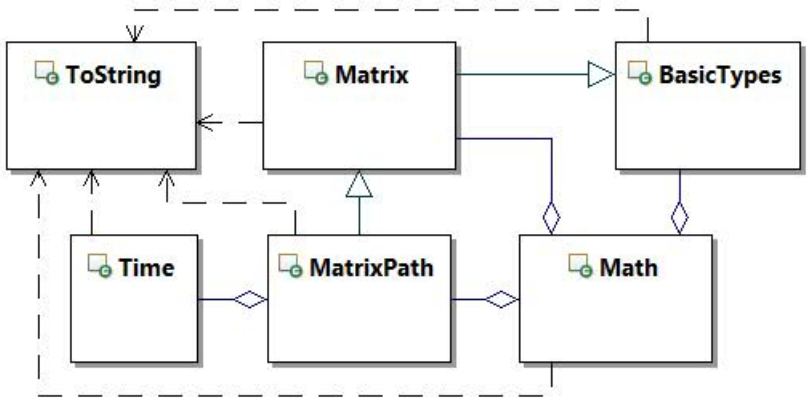

Abbildung 68 Klassendiagramm der ToolKit Basis Komponenten

B.2 Die Teilkomponente „OSG“ des ToolKits

Die zweite Teilkomponente des ToolKits heißt „OSGToolKit“ und beinhaltet viele nützliche Klassen mit Funktionen, die es ermöglichen, leichter mit OpenSG zu arbeiten. Durch diese kann schneller, einfacher und effektiver gearbeitet werden, und der Quellcode wird insgesamt deutlich kürzer, übersichtlicher, wartbarer und einfacher zu erweitern. Es gibt drei grundlegende Klassen, die am häufigsten in PAMOCAT benutzt werden, da sie elementare Funktionalitäten beinhalten. Diese sind „GeoEdit“, „MaterialLib“ und „NodeHelper“. Die Klasse „GeoEdit“ ermöglicht es, die Geometrie zu bearbeiten. Solche Bearbeitungsvorgänge können Translation, Rotation oder Skalierung sein, beinhalten aber auch verschiedene Funktionen, Geometrien zusammenzufügen, Geometrie zu normalisieren[115], ein Objekt in seinen Mittelpunkt zu verschieben und Funktionen um Normale auszurechnen (damit die Geometrie korrekt beleuchtet dargestellt werden kann). Die Klasse „MaterialLib“ stellt eine Sammlung von verschiedenen Materialien für Geometrie bereit und ermöglicht deren einfache Erzeugung. Sie stellt auch Shader[116] Objekte bereit, über die Vertex-, Geometrie- und Fragment-Shader als Material verwendet werden können. Die Klasse „NodeHelper“ besitzt Funktionen, um verschiedene Nodes mit Namen zu finden, das Hinzufügen von Namen an Nodes und das Hinzufügen von verschiedenen Transformationen.

Bei komplexeren grafischen Visualisierungen werden verschiedene Objekte oft mehrfach verwendet. Dabei resultiert dieses aus dem Datenformat (z. B. VRML), in dem es gespeichert wurde. Die gleiche Geometrie wird mehrfach für jede Referenz gespeichert. Leider sind die Importfunktionen von OpenSG noch nicht in der Lage, die mehrfach referenzierte Geometrie aus Modellierungstools wie „3D Studio Max“, „Maya“, „Softimage“ und „Cinema 4D“ zu erkennen. Das hat zur Folge, dass diese mehrfach im Speicher geladen werden, obwohl die Geometrie einfach mehrfach referenziert werden könnte. Daher brauchen große Szenarien

[115] Dabei ist gemeint, die Größe oder die längste x-, y-, z-Position eines Vertex der Geometrie auf 1 zu skalieren. Dieses ist eine praktische Funktionalität, die es ermöglicht, leicht Objekte aus verschiedenen Quellen bzw. Maßstäben in eine einheitliche Größe zu bringen.

[116] Ein Shader ist ein Programm, das Einfluss auf die Grafikpipeline nehmen kann und auf der Grafikkarte ausgeführt wird.

beim Laden und auch beim Darstellen recht lange. Um hier eine Optimierung zu erreichen, ist es wichtig, dass gleiche Objekte erkannt werden und in einen Szenengraph mit mehrfachen Referenzen überführt werden, in dem die Daten für die Geometrie und die Texturen nicht mehrfach gespeichert werden. Dazu können nicht einfach nur Knoten (engl. node) unter weitere Knoten gehängt werden, da ein Knoten immer nur unter genau einen anderen Knoten gehängt werden kann. Stattdessen müssen die Geometrie und Texturen mehrfach referenziert und in eine neue Graph-Struktur überführt werden, bei der die Objekte verschieden platziert sind. Genau diese Funktionalität bietet die „MultipleNodeClone“, welche eine komplexere Knotenstruktur mit verschieden positionierten Geometrieteilen optimiert. Dazu werden die mehrfach verwendeten Geometrien erkannt und zusätzlich durch das Übertragen verschiedener Transformationen auf die eigentliche Geometrie und das anschließende Zusammenführen optimiert. Dieser Vorgang beschleunigt die Darstellung in OpenSG, da die Anzahl der Knoten drastisch reduziert wird. Das ist ein Vorteil beim verteilten Rendering, wie es bei Caves verwendet wird, da deutlich weniger Knoten über das Netzwerk abgeglichen wird müssen. Weitere Klassen wie „Grid“, „Arror3D“, „SelectionBox“, „Nao“, „RigidBodyVis“, „Text3D“, „VideoTexture“ sind grafische Erweiterungen durch Zusatz-Geometrie-Primitive für OpenSG. Die Klasse „StereoProjektion“ ermöglicht es, auf einfache Weise 3D-Stereo-Rendering mit verschiedenen Setups wie „Polarisation“, „Infinitec“ oder „Shutter Brillen“ zu benutzen. Die Klasse „WindowImageRecording“ erlaubt es, die Interaktion mit der virtuellen Realität festzuhalten und daraus ein Video zu erstellen. Der Zusammenhang der einzelnen Klassen ist in **Abbildung 69** dargestellt.

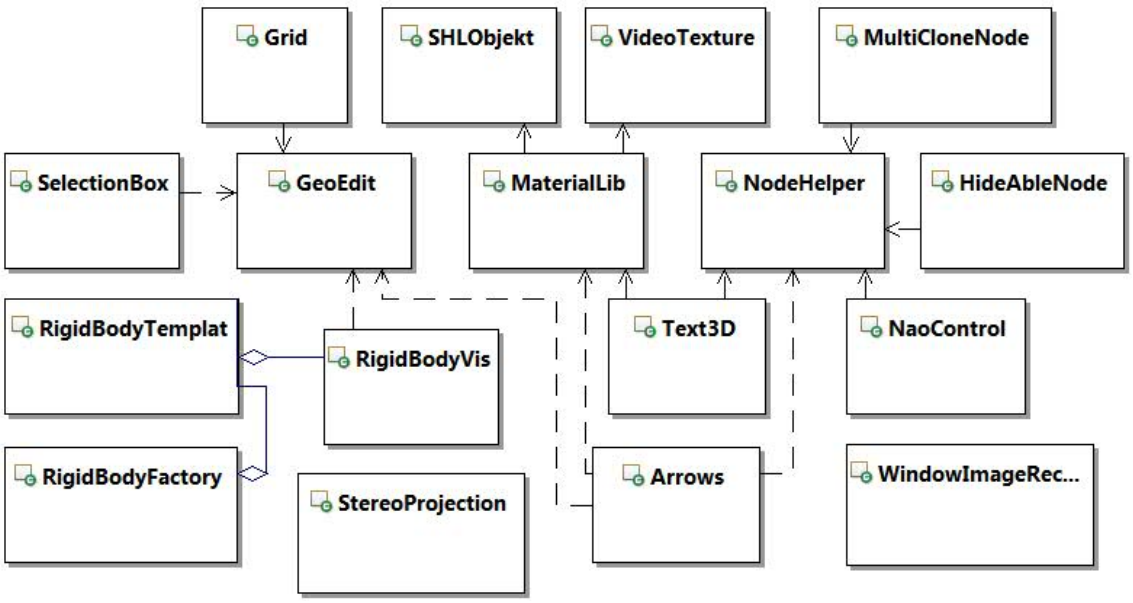

Abbildung 69 Klassendiagramm der ToolKit OSG Komponente (Ausschnitt)

B.3 Die Teilkomponente „File“ des ToolKits

Wie im früheren Verlauf deutlich wurde, wird eine Vielzahl an Daten gespeichert. Dieses sind Aufzeichnungsdaten und eine Menge an Konfigurationsdaten. Dazu wird hier eine Basisfunk-

tionalität bereitgestellt, die es ermöglicht, Dateien einfach und einheitlich zu verwalten. Dazu bietet die Klasse „File“ die Möglichkeiten zum Öffnen und dem Übergeben des Inhaltes in verschiedenen Formen, aber auch die Möglichkeit, Daten zu speichern, und nimmt dabei auf existierende Dateien Rücksicht und überschreibt diese nicht einfach. Die Klasse „LiveStream“ erlaubt es, Aufnahmedaten zu jedem Zeitpunkt direkt in eine Datei zu schreiben. Die Klasse „ZipFile“ ermöglicht es, Daten zu komprimieren, so dass z. B. die Motion-Capture-Daten auf mindestens 10 % der ursprünglichen Größe reduziert werden können. In der „LiveStream“ Klasse können Daten auch live komprimiert werden (siehe **Abbildung 70**).

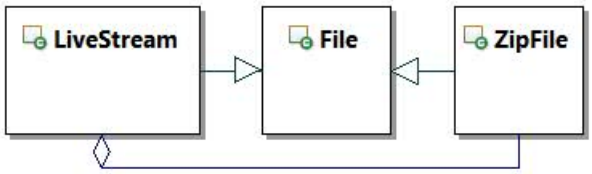

Abbildung 70 Klassendiagramm der ToolKit File Komponente

B.4 Die Teilkomponente „Input“ des ToolKits

Die „Input“ Komponente ermöglicht es, die Eingangsdaten von verschiedenen Inputgeräten einfach zu integrieren. Die Klasse „Wiimote“ erlaubt es, mit Bluetooth Geräten zu verbinden. Bei nicht bekannten Mac Adressen wird automatisch das erste erreichbare Bluetooth Gerät mit einer Wiimote Device ID zur Verbindung gewählt. Die Klasse „USBDevice“ bietet einheitlich die Möglichkeit, mit USB Geräten zu kommunizieren, auf denen die Klassen „Joystick“ zum Interagieren bassiert. Die Klasse „PowerSwitch“ ermöglicht das Koordinieren verschiedener externer elektronischer Geräte (zum z. B. synchronen Anschalten von Geräten), die Klasse „Temperatur“ um die Temperatur auszulesen. Durch die Klassen Temperatur und PowerSwitch können z. B. teure Geräte vor Überhitzung geschützt (siehe dazu **Abbildung 71**).

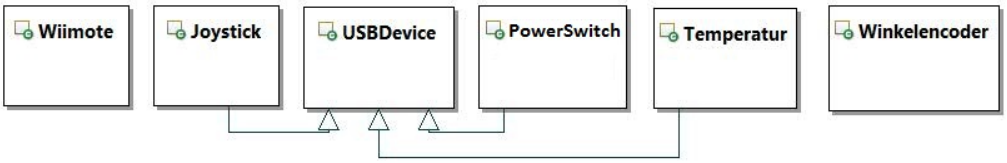

Abbildung 71 Klassendiagramm der ToolKit Input Komponente

B.5 Die Teilkomponente QT des ToolKits

Die Teilkomponente QTToolKit hat zwei Klassen, nämlich die Klassen „VideoPlayer“ und „MultiVideoPlayer“. Die Klasse VideoPlayer bietet einen Bereich, auf dem ein Video gezeigt werden kann, ermöglicht es, ein Video auszuwählen und durch grafische Elemente zu laden, abzuspielen, zu unterbrechen. Außerdem kann die Zeit durch einen Schieberegler manipuliert werden, der Ton lauter und leiser eingestellt und zwischen einem Vollbildmodus und dem Normalmodus gewechselt werden. Die Klasse „MultiVideoPlayer“ ermöglicht die gleichzeiti-

ge Verwaltung vieler Videos. Dabei sind diese verschiedenen multiplen Video-Aufnahmen der gleichen Interaktionssituation aus verschiedenen Richtungen darzustellen. Diese werden für eine detaillierte Analyse synchron gehalten, damit jede Bewegung aus verschiedenen Blickwinkeln analysiert werden kann. Dabei kann einfach zwischen den verschiedenen Videos bzw. Kameras gewechselt werden. Dieser Teilzusammenhang ist in der **Abbildung 72** dargestellt.

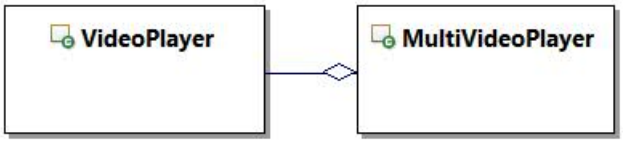

Abbildung 72 Klassendiagramm der ToolKit Komponente QT

B.6 Dynamischer sequentieller Programmablauf

Neben den statischen Eigenschaften der Software, die zeigen, wo welche Funktionalität angeordnet ist und wie Klassen, aber auch Komponenten miteinander arbeiten, ist auch die Kenntnis der dynamischen Aspekte der Software wichtig, um diese eventuell zu erweitern. Dabei steht das zeitliche Interagieren der verschiedenen Klassen miteinander im Vordergrund. An dieser Stelle wird ein sequentieller Ablauf bei der Zerlegung der Bewegung in elementare Aktivität einzelner Gelenke näher betrachtet.

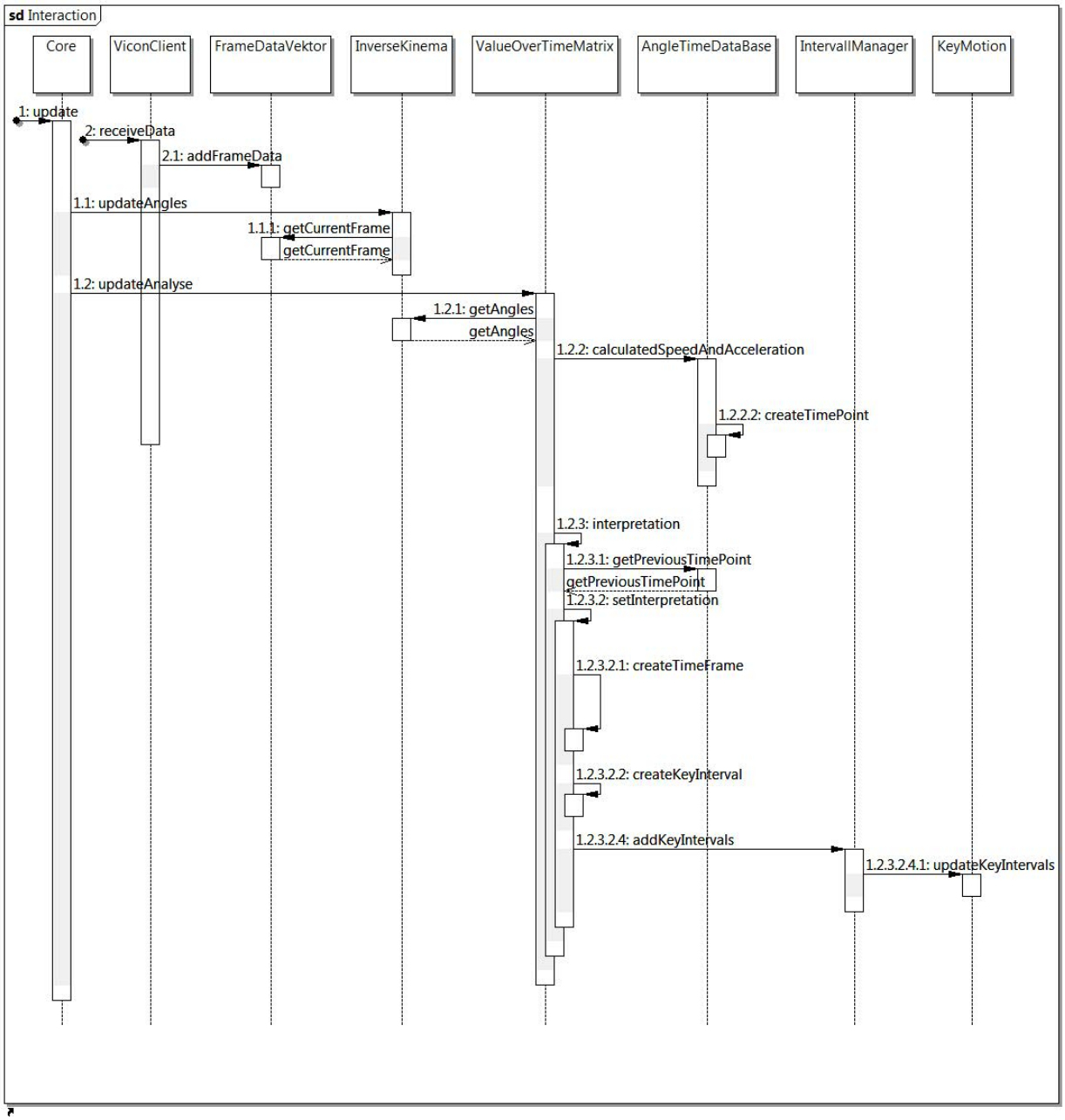

Abbildung 73 Vereinfachtes Sequence-Diagramm zur Online Zerlegung der Bewegung in eine Key-Frame-Animation

Der sequentielle Ablauf ist in **Abbildung 73** dargestellt und fängt durch einen externen Aufruf der „update“-Methode der „Core“ Klasse an. Die Klasse „ViconClient“ ist ein einzelner

Thread, der die Daten im Hintergrund vom Vicon Nexus über die ViconSDK in Empfang nimmt. Diese werden anschließend in die Motion-Capture-Datenstruktur überführt, sodass sie als Frames mit mehreren Rigidbodys oder Markern zur Verfügung stehen. Aus der Klasse „Core“ wird die inverse Kinematik aufgerufen, um die Skelettwinkel auszurechnen. Diese können von anderen Komponenten, zum Beispiel der Visualisierung des Skeletts, verwendet werden[117]. Falls die Option zur Zerlegung der Bewegung in eine Key-Frameanimation aktiviert ist, wird die „update“-Funktion der Klasse „ValueOverTimeMatrix“ aufgerufen. Diese verwendet die Winkelstellung aus vorherigen Zeitpunkten, um die aktuelle Geschwindigkeit und die Beschleunigung auszurechnen. Solange ein Gelenk weiter an Geschwindigkeit gewinnt, ist noch kein Ende eines Key-Intervalls erreicht. Erst wenn die Beschleunigung gleich null wird und negativ ist, kann zu dem Gelenk ein neues Key-Intervall erzeugt werden. Je nachdem, ob zu dem Zeitpunkt schon bei einem anderen Gelenk ein Key-Intervall beginnt, wird dieses dann zum existierenden „TimeFrame“ hinzugefügt, andernfalls wird ein neues Objekt der Klasse erzeugt. Diese Verwaltung der Key-Intervalle geschieht in der Klasse „IntervallManager“. Der Zweck dahinter ist die Erkennung von verschiedenen Bewegungsmustern anhand von aktiven Bewegungen in Gelenken, die zu einem späteren Zeitpunkt in eine Echtzeitinteraktion integriert werden können.

[117] Aber auch zur Posenerkennung.

11 Literaturverzeichnis

[1] A. Peräkylä, C. Antaki, S. Vehviläinen und I. Leudar, Conversation Analysis and Psychotherapy, Cambridge: University Press Cambridge, 2008.

[2] M. Salem, S. Kopp, I. Wachsmuth und F. Joublin, „Towards an Integrated Model of Speech and Gesture Production for multi modal robot behavior," in *Roman*, Viareggio, Italy, 2010.

[3] G. Wilcock, Jokinen und K., „Speech, gaze and gesturing- multimodal conversation interaction with nao robot," in *International Summer Workshop on Multimodal Interfaces*, Metz, 2012.

[4] K. Pitsch, S. Wrede, J.-C. Seele und L. Süssenbach, „Attitude of German Museum Visitors towards an Interactive Art Guide Robot," HRI2011, 2011.

[5] A. Lücking, K. Bergmann, F. Hahn, Kopp, S. und H. Rieser, „Data-based Analysis of Speech and Gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its Applications," in *Journal on Multimodal User Interfaces*, 2013.

[6] A. Kendon, Gesture Visible Actions as Utterance, Cambridge: University Press Cambridge, 2004.

[7] C. Müller, Redebegleitende Gesten: Kulturgeschichte-Theorie Sprachvergleich, Berlin: Arno Spitz Verlag, 1998.

[8] Duden Die deutsche Rechtschreibung, Berlin: Bibliographisches Institut, 2013.

[9] D. McNeill, Hand and Mind, Chicago: University of Chicago Press., 1992.

[10] P. Blache, R. Bertrand, B. Bigi, E. Bruno, E. Cela, R. Espesser, G. Ferré, M. Guardiola, D. Hirst, E.-P. Magro, J.-C. Martin, C. Meunier, M.-A. Morel, E. Murisasco, I. Nesterenko, P. Nocera, B. Pallaud und L.-V. B. J. S. Prévot, Multimodal Annotation of Conversational Data, Sweden: Proceedings of the Fourth Linguistic Annotation Workshop, 2010.

[11] J. Bressem, „A linguistic perspective on the notation of form features in gestures," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*, Boston, De Gruyter: Mouton, 2013.

[12] M. Kipp, Annotation Facilities for Reliable Analysis of Human Motion, Istanbul: LREC, 2012.

[13] E. Auer, A. Russel, H. Sloetjes, P. Wittenburgm, O. Schreer, S. Masnieri, D. Schneider und S. Tschöpel, ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors, Malta: LREC 2010, 2010.

[14] M. Kipp, „Multimedia Annotation, Querying and Analysis in ANVIL.," in *Multimedia Information Extraction*, MIT Press, 2010, p. Chapter 19.

[15] T. Schmidt and K. Wörner, EXMARaLDA CREATING, ANALYSING AND SHARING, International Pragmatic Association, 2009.

[16] J.-T. Milde und U. Gut, „The TASX-environment: an XML-based corpus database for time aligned language data," 2001.

[17] K. Rohlfing, D. Loehr, S. Duncan, A. Brown, A. Franklin, I. Kimbara, J.-T. Milde, F. Parril, T. Rose, T. Schmidt, H. Sloetjes, A. Thies und S. Wellinghoff, „Comparsion of multimodal annotation tools - workshop report," *Geschprächsforschung Online Zeitschrift zur verbalen Interaktion (ISSN 1617- 1837)*, pp. 99-123, 2006.

[18] A. Heloir, M. Neff und M. Kipp, Exploiting Motion Capture for virtual Human Animation - Data Collection and Annotation Visualisation, LREC Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality: ELDA, 2010.

[19] „Center of Excellence Cognitive Interaction Technology," 20 01 2014. [Online]. Available: http://www.cit-ec.de.

[20] J. Sobotta, Atlas der Anatomie des Menschen, München: Urban und Fischer, 2004.

[21] H. J. Wagner, Einf□uhrung in den aktiven und passiven Bewegungsapparat, Tübingen: Anatomisches Institut Universirät Tübingen, 2005.

[22] M. W. Spong, S. Hutchinson und M. Vidyasagar, Robot Modeling and Control, United States Naval Academy: Wiley, 2005.

[23] R. Möller, Robotik: Serielle Manipulatoren, Bielefeld: Universität Bielefeld, 2005.

[24] R. Parent, Computer Animations: Algorithms and Techniques., Amsterdam: Morgan Kaufmann, 2008.

[25] D. Jackel, S. Neunreither und F. Wagner, Methoden der Computeranimation, Berlin: Springer, 2006.

[26] I. Wachsmuth, Menschen, Tiere und Max: Natürliche Kommunikation und künstliche Intelligenz, Spektrum Akademischer Verlag, 2013.

[27] M. Giese, „Motion Capturing Vorlesung," Uni-Tuebingen, Tuebingen, 2005.

[28] E. Muybridge, „Finding Aid to Valley of the Yosemite, Sierra Nevada Mountains, and Mariposa Grove of Mammoth Trees," U.S. Copyright, San Francisco, 1872.

[29] B. Jung, H. B. Amor, G. Heumer und M. Weber, „From Motion Capture to Action Capture: A Review of Imitation Learning," ACM, Freiberg, 2006.

[30] „Motion Capture Artikel Wikipedia," 12 01 1014. [Online]. Available: http://de.wikipedia.org/wiki/Motion_Capture.

[31] „ar-tracking," 12 01 2014. [Online]. Available: http://www.ar-tracking.com/.

[32] „Phasespace," 12 01 2014. [Online]. Available: http://www.phasespace.com.

[33] „Ascension-Tech," 12 01 2014. [Online]. Available: http://www.ascension-tech.com/.

[34] „Xsens," 12 01 2014. [Online]. Available: http://www.xsens.com/.

[35] „MetaMotion," 12 01 2014. [Online]. Available: http://www.metamotion.com/.

[36] D. Vlasic, R. Adelsberg, G. Vannucci, J. Barnwell, M. Gross, W. Matusik und J. Popovic, „Practical Motion Capture in Everyday Surroundings," SIGGRAPH, Zürich, 2007.

[37] „OpenNI," 12 01 2014. [Online]. Available: http://www.openni.org/.

[38] Metamotion, „Metamotion," [Online]. Available: http://www.metamotion.com/gypsy/gypsy-motion-capture-system.htm. [Zugriff am 27 11 2013].

[39] VICON, „VICON," 27 11 2013. [Online]. Available: www.vicon.com. [Zugriff am 27 11 2013].

[40] H. Garfinkel, Studies in Ethnomethodology, Malden: Plackwell Publisher USA, 1984.

[41] E. Gülich, L. Mondada und I. Furchner, Konversationsanalyse: Eine Einführung am Beispiel des Französischen, Tübingen: Niemeyer, 2008.

[42] T. Schmidt, S. Duncan, O. Ehmer, J. Hoyt, M. Kipp, D. Loehr, M. Magnusson, T. Rose und H. Sloetje, „An exchange format for multimodal annotations," LREC, 2008.

[43] D. Spohr und P. Cimiano, „Information and Communication Technology," in *Studies on Subject-Specific Requirements for Open Access Infrastructure*, Bielefeld, 2011.

[44] S. Kita, I. van Gijn und H. van der Hulst, „Movment Phases in Signs and Co-Speech Gestures, and Their Transcription by Human Coders," Springer, Berlin Heidelberg, 1998.

[45] F. Kügler, „Einführung in die Grundlagen von Praat," 2007.

[46] H. Sloetjes und A. Somasundaram, ELAN development, keeping pace with communities

needs, Istanbul: LREC, 2012.

[47] T. Schmidt und W. Schütte, „FOLKER: An Annotation Tool For Efficient Transcription Of Natural, Multi-Party Interaction," LREC, Malta, 2010.

[48] „ANNEX - Annotated Explorer version 1.1," The language archiv, MPI for Psycholinguistics, Nijmegen, The Netherlands, 2012.

[49] „Ehmer, O.," http://www.oliverehmer.de/transformer/.

[50] T. Schmidt, K. Elenius und P. Trilsbeek, „Multimedia Corpora (Media encoding and annotation)," 2010.

[51] N. Bevan, J. Kirakowski und J. Maissel, „What is Usability?," Proceedings of the 4 th International Conference on HCI, Stuttgart, 1991.

[52] F. Meakins, „Computerized Language Analysis (CLAN)," 2007.

[53] K. Maeda, S. Brid, X. Ma und H. Lee, „The Annotation Graph Toolkit: Software Components for Building Linguistic Annotation Tools," 1999.

[54] P. Menke and P. Cimiano, MExiCo: A Library for Managing Multimodal Data Collections, Procedia Social and Behavioral Sciences, 2013.

[55] P. Menke, Multimodal data and multimodal corpora, Universität Bielefeld, 2013.

[56] C. Sminchisescu, „3D Human Motion Analysis in Monocular Video," in *Human Motion Computational Imaging and Vision*, Netherlands, Springer, 2008.

[57] V. Parameswaran und R. Chellappa, „View Invariants for Human Action Recognition," *International Journal of Computer Vision 66*, 2006.

[58] C.-F. Tasai und C. Hung, „Automaticaly Annotating Images with Keywords: A Review of Image Annotation Systems," *Recent Patents on Computer Science*, 2008.

[59] M. Chessa, F. Solari, S. P. Sabantini und G. M. Bisio, „Motion Interpretation Using Adjustable Linear Models," BMVC, 2008.

[60] W. Xu, M. Yang und K. Yu, „3D Convolutional Neural Networks for Human Action Recognition," in *ICML2010*, Haifa, Israel, 2010.

[61] O. Duchenne, I. Laptev, J. Sivic, F. Bach und J. Ponce, „Automatic Annotation of Human Actions in Video," in *IEEE*, 2009.

[62] S. Wu, „Indexing and Retrieval of Human Motion Data by Hierarchical Tree," *Proceedings of the 19th ACM Symposium on VirtualReality Software and Technology*, 2009.

[63] A. Bobick und J. Davis, „The recognition of human movement using temporal templates," PAMI, 2001.

[64] D. Tran und A. Sorokin, „Human Activity Recognition with metric learning," in *Computer Vision ECCV*, 2008.

[65] B. Schölkopf und A. Smola, „ Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond," MIT Press, Cambridge, 2002.

[66] D. Ramanan und D. A. Forsyth, „Automatic Annotation of Everyday Movemnts," 2003.

[67] M. Müller, A. Baak und H.-P. Seidel, „Efficient and Robust Annotation of Motion Capture Data," *Eurographics ACM SIGGRAPH*, 2009.

[68] M. Müller und T. Röder, Motion Templates for Automatic Classification and Retrieval, ACM Siggraph/ Eurographics Symposium on Computer Animation, 2006, pp. 137-146.

[69] O. Masaki, „Motion-Capture-Based Avatar Control Framework in Third-Person View Virtual Environments," ACM, Fukuoka Japan, 2006.

[70] V. B. Zordan, H. Van und C. Nicholas, „Mapping optical motion capture data to skeletal motion using a physical model," ACM, California Riverside, 2003.

[71] B.-A. Brüning, M. Latoschik und I. Wachsmuth, Interaktives Motion-Capturing zur Echtzeitanimation virtueller Agenten, Magdeburg: Virtuelle und Erweiterte Realität, 5. Workshop of the GI VR & AR special interest group, 2008.

[72] K. Pitsch, B.-A. Brüning, C. Schnier, H. Dierker und S. Wachsmuth, „Linking Conversation Analysis and Motion Capturing: How to robustly track multiple participants?," Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality., Malta, 2010.

[73] B.-A. Brüning, Entwicklung eines Motion-Capture Recorders f□ur einen virtuellen Agenten auf der Basis eines optischen Trackingsystems, Bielefeld, 2008.

[74] B. Brüning, C. Schnier, K. Pitsch und S. Wachsmuth, Automatic detection of motion sequences for motion analysis, Alicante, Spain: ICMI, 2011.

[75] B.-A. Brüning, C. Schnier, K. Pitsch und S. Wachsmuth, PAMOCAT: Automatic retrieval of specified postures, Istanbul: LREC, 2012.

[76] H. Sacks und E. A. Schegloff, „Home position," John Benjamins Publishing Company, Los Angeles, 2002.

[77] K. Jokinen, „Turn taking, Utterance Density, and Gaze Patterns," ICMI, Alicante, 2011.

[78] H. Furukawa, M. Nishida, K. Jokinem und S. Yamamoto, „A multimodal Corpus for

modeling turn management in multi-party conversations,“ in *Speech Database and Assessments*, Hsinchu, 2011.

[79] H. Schober, Das Sehen, Leipzig, 1970.

[80] G. Gerstbach, Augen und Sehen - der lange Weg zu digitalem Erkennen, Wien: Sternenbote Heft 11/99, 1999.

[81] B. Brüning, C. Schnier, K. Pitsch und S. Wasmuth, Integrating PAMOCAT in the research cycle. Linking Motion Capturing and Conversation Analysis, Santa Monica California: ICMI, 2012.

[82] R. Sommer, Studies in personal space, Bobbs-Merrill, 1967.

[83] A. Lücking, K. Bergmann, F. Hahn, S. Kopp und h. Rieser, „The Bielefelder Speech and Gesture Alignment Corpus (Saga),“ LREC, 2010.

[84] L. Johnson, „TX 77058 U.S.A. Man-System Integration Standards,“ NASA: U.S. National Aeronautics and Space Administration Space Center, 1994.

[85] D. McNeill, Hand and mind What Gestures reveal about Thought, Chicago: University of Chicago Press, 1992.

[86] „Wikipedia,“ 10 01 2013. [Online]. Available: http://de.wikipedia.org/wiki/Liste_von_Gesten. [Zugriff am 20 02 2013].

[87] M. Belke, „Gestik,“ Bielefeld, 2000.

[88] C. Andres, „Mündliche Wegauskünfte von Kindern und Jugendlichen im Spannungsfeld von Sprache, Interaktion, Kognition und Multimodalität,“ Weimar, 2009.

[89] C. L. Nehaniv, „Classifying Types of Gesture and Inferring Intent,“ AISB, 2005.

[90] S. Haykin, Neural Networks a comprehensive foundation, New Jersey: Prentice Hall, 1999.

[91] F. Argelaguet, C. Andujar und R. Trueba, „Overcoming Eye-Hand Visibility Mismatch in 3D Pointing Selection,“ in *VRST*, 2008.

[92] M. Fröhlich, Ein wissensbasiertes Rahmensystem zur merkmalsbasierten Gestenerkennung für multimediale Anwendungen, Bielefeld: Universität Bielefeld, 1999.

Eidesstattliche Versicherung

Name:	Brüning	Vorname:	Bernhard-Andreas
Matrikel-Nr.:	1659625	Studiengang:	Promotion ISY

Hiermit versichere ich, Bernhard-Andreas Brüning, an Eides statt, dass ich die vorliegende Dissertation mit dem Titel „PAMOCAT: Kombination von qualitativen und quantitativen Methoden zur automatischen Analyse von menschlichen Verhaltensweisen in der Kommunikation basierend auf Bewegungsdaten“ selbständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Auszug aus dem Strafgesetzbuch (StGB)

§ 156 StGB Falsche Versicherung an Eides Statt

Wer von einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

Ort, Datum

Unterschrift