

---

# Stand der Forschung

---

Der Hauptaugenmerk zur Übertragung natürlicher menschlicher Bewegungsdaten auf einen humanoiden Roboter, liegt auf dessen Funktions-, bzw. Bewegungsskelett und auf der Qualität der transkribierten Bewegungsdatensätze. Der Prozess, von der Transkription der Bewegungsdaten und die anschließende Übertragung auf das Bewegungsskelett des Roboterprototyps, rückt hierbei in den Fokus der Betrachtung. Als Grundlage der Untersuchung dienen Forschungsergebnisse aus dem Bereich der virtuellen Avatare/Charaktere, vgl. Frey (1999). Menschliche Bewegungsdaten konnten über diesen Weg erstmalig auf einen virtuellen Charakter übertragen werden, ohne beim Betrachter das *Uncanny Valley Phänomen* auszulösen. Um zu verstehen welche Ursachen hierbei eine Rolle spielen, gilt es die menschliche Anatomie mit der Anatomie virtueller Charaktere und humanoider Roboterprototypen zu vergleichen. Darüber hinaus muss der Prozess der Transkription und Übertragung von Signalen der nonverbalen Kommunikation einer näheren Betrachtung zu unterzogen werden.

## 2.1 Akzeptanzprobleme von virtuellen Charakteren

Der Begriff des *Uncanny Valley* übertrug sich in den 1970er Jahren rasch aus der Robotik auf den Bereich der virtuellen Charaktere oder auch Avatare genannt. Avatare dienen in modernen Medien (z.B. der Filmbranche, aktuell auch in Anwendungen des Web 2.0) als künstliche Schauspieler sowie als Protagonisten, um die Aufmerksamkeit des Benutzers auf sich zu lenken. Diese virtuellen Charaktere hatten mit ähnlichen Problemen zu kämpfen, wie sie auch in der humanoiden Robotik auftraten. Sie waren statisch sehr schön anzusehen und erreichten hohe Sympathiewerte. Sobald sie sich

aber in Bewegung setzten, verfiel die gewonnene Sympathie, die ihnen der Betrachter entgegen brachte. Virtuelle Darsteller hatten nicht den durchschlagenden Effekt, den sich die Entwickler damals von ihnen erhofft hatten. Trotzdem wurde sogar laut spekuliert, ob sie nicht sogar in der Lage seien, reale Schauspieler in Film und Fernsehen abzulösen. Doch das erwies sich schon bald als Irrglaube. Ein Beispiel, an dem dies besonders deutlich geworden ist, ist die in den 1980er Jahren gestartete Serie „Max Headroom“. Die Zuschauer hatten vor dem virtuellen Reporter Edison Carter dermaßen große Aversionen, dass die Serie wegen sinkender Einschaltquoten nach zwei Staffeln vorzeitig abgesetzt wurde, vgl. King (1987). Damals wurde noch alle Hoffnung in den technologischen Fortschritt gesetzt, der dieses Problem durch mehr Rechenleistung und verbesserte Grafik lösen sollte. Diese Hoffnung wurde allerdings schnell wieder zunichte gemacht.

An Beispielen wie dem vom Bundesministerium für Bildung und Forschung (BMBF) geförderten *Virtual Human-Projekt*<sup>1</sup> ist zu sehen, dass selbst mit den besten Mitteln in der grafischen Darstellung kein Fortschritt zur Lösung des *Uncanny Valley*-Problems erreicht wurde. Die Avatare wirken teilweise sehr unnatürlich und verfehlten den Effekt, den Zuschauer für sich zu gewinnen. Die Arbeitsgruppe um S. Frey hat sich im Rahmen des vom BMBF<sup>2</sup> geförderten Projekts „Anthropomorphe Tutoren für Lernprozesse in mit Web 2.0-Technologie unterstützten Lehr-/Lernszenarien“ ebenfalls dem Problem des *Uncanny Valley Phänomens* gewidmet und wegweisende Erfolge bei der Lösung der Problematik erzielt. Ziel war es, über die Erfassung nonverbaler Komponenten des menschlichen Kommunikationsverhaltens, die Bewegungsrepräsentation virtueller Tutoren systematisch zu verbessern. Dabei entwickelte die Forschergruppe um Prof. S. Frey ein Kodierverfahren, das auf dem Ende der siebziger Jahre entwickelten *Berner System* basiert. Diese Transkriptionsmethode dient nun als theoretische Grundlage und wird im Rahmen des Projekts sukzessive verbessert.

S. Frey erkannte, dass diese Kodierungen, als Ergebnisse der Analyse des Bewegungsverhaltens virtueller Charaktere, besonders gut zur Untersuchung und Lösung des *Uncanny Valley Phänomens* geeignet sind. Hier werden zusätzliche auftretende Probleme der physikalisch, realen Welt in den Hintergrund und es kann im Modell der Fokus auf die eigentliche menschliche Bewegung gelegt werden. Bewegungsabläufe können am realen Prototypen durch das Getriebespiel, Masseträgheiten und Elastizitäten verfremdet werden, wie dies auch bei Mori's Experimenten mit humanoiden Robotern der Fall war. Durch eine vereinfachte Modellannahme und die Abstrahierung des Skelettmodells virtueller Avatare wurde der schnelle Fortschritt zur Lösung erst ermöglicht. Über den gefundenen Weg die Bewegungsdaten eines Menschen mit dem neuen Transkriptionswerkzeug ScanimFBX zu parametrisieren und die anschließende Übertragung der Kodierungen auf einen virtuellen Avatar, wurde ein Bewegungsmodell ohne physikalische Restriktionen generiert. Hierdurch lassen sich schon frühzeitig eine ganze Reihe an Problemen ausklammern, die unter Einbeziehung der physikalischen Probleme eines realen Roboterprototyps entstehen würden. Aufgrund dieser Vorgehensweise kann das

<sup>1</sup> BMBF-gefördertes Projekt, Laufzeit 2002 bis 2006, <http://www.virtual-human.org>.

<sup>2</sup> Bundesministerium für Bildung und Forschung.

Problem des *Uncanny Valley* für die virtuellen Avatare durch die Forschung von S. Frey als nahezu überwunden betrachtet werden, vgl. Frey (2013). Im Folgenden werden die Ergebnisse zurück auf den Bereich übertragen, in dem sie zum ersten Male sichtbar wurden: der humanoiden Robotik.

## 2.2 Das Berner System

Bereits im Jahr 1999 demonstrierte S. Frey, wie stark spezifische Aspekte der Human-kommunikation den Betrachter einer Person, sei es real oder in einer Abbildung, beeinflussen können. In seinem Buch „Die Macht des Bildes“, wurden grundlegenden Aspekte der visuellen Eindrucksbildung und die Wirkung optischer Eindrücke auf die Eigenschaftszuschreibungen beim Betrachter untersucht, vgl. Frey (1999). Ein hervorragendes Beispiel hierfür bietet die Darstellung in Abbildung 2.1.



Abbildung 2.1: Einfluss der Lateralflexion auf die Eigenschaftszuschreibung einer anderen Person, die „Madonna del Magnifico“, gemalt von Sandro Botticelli (Abbildung links), verwandelt sich vor unseren Augen von einer demütigen und bescheidenen Person in eine selbstbewusste Herrin, in dem der Kopf in eine aufrechte Position bewegt wird (Abbildung rechts), vgl. Frey (1999).

Wie zu sehen ist, entscheiden bereits winzige Nuancen der Körperhaltung über die erste, schwer zu revidierende Einschätzung einer Person. Bewegt ein Schauspieler seinen Kopf in einer bestimmten Art und Weise, löst eine Drehung um wenige Grad eine vollkommen unterschiedliche Reaktion bei einem menschlichen Betrachter aus. Eine Person, die vorher als sanft und empathisch beurteilt wurde, wirkt nun arrogant und abweisend, vgl. Abb. 2.1. Ein weiteres Beispiel für diese unterbewusste Eigenschaftszuschreibung ist in Abbildung 2.2 zu sehen. Durch die Veränderung der lateralen Kopfeigung der abgebildeten Mutter und des Kinds scheinen die beiden abgebildeten Personen nun plötzlich ängstlich vor dem Betrachter zurückzuweichen. Diese unbewusste Aussage wurde von verschiedenen Menschen spontan und vollkommen unabhängig von ethnischer und kultureller Herkunft getroffen, vgl. Frey (1999). Diese kurzen Beispiele zeigen deutlich, auf welche Weise das menschliche Auge über die visuelle Wahrnehmung die Beurteilung anderer Personen beeinflusst. Der Mensch hat sich im Laufe der Evolution zu

einem wahren Meister der visuellen Eindrucksbildung anderer Personen entwickelt. Je menschlicher ein Avatar/Roboter aussieht, desto stärker treten diese unterbewussten Aspekte der visuellen Eindrucksbildung zu Tage.



Abbildung 2.2: Portrait der Elisabeth-Louise Vigée Lebrun (linke Abbildung), die seitliche Neigung des Kopfs der Mutter zum Kind sowie des Kinds zum Betrachter wird verändert, wodurch die abgebildeten Personen ängstlich vor dem Betrachter zurückweichen zu scheinen (rechte Abbildung), s. Frey (1999).

Diese Erkenntnisse bildeten die Grundlage für die Entwicklung des Berner Systems, vgl. Frey (1999). Dieses Notationssystem bot erstmalig die Möglichkeit, Körpersprache transkriptiv zu erfassen und damit die nonverbale Komponente des menschlichen Kommunikationsverhaltens systematisch auszuwerten. „Zielvorgabe war dabei die Schaffung eines Kodiervorgangs, das einerseits so geringe Anforderungen an die Diskriminationsleistung des Kopierers stellt, dass es - ähnlich dem Alphabet - von jedermann ohne großen Trainingsaufwand erlernbar und zuverlässig handhabbar ist. Andererseits sollte es eine so hochauflösende, detailgenaue Beschreibung von Körperhaltungen ermöglichen, dass sich auf Grundlage der Positionskodierungen die ursprüngliche Bewegung/Pose wieder reproduzieren ließe - gerade so, wie sich auf Grundlage eines alphabetischen Transkripts Sprachäußerungen wieder reproduzieren lassen“, vgl. Frey (1999). Das Berner System besaß in seiner damaligen Form 104 Bewegungsausprägungen (Freiheitsgrade<sup>3</sup>), um beliebig komplexe Körperbewegungen nonsemantisch zu erfassen und sowohl hochdifferenziert als auch zuverlässig zu reproduzieren.

<sup>3</sup> Als Freiheitsgrade bezeichnet man die Bewegungsfreiheitsgrade eines Objektes in dessen räumlichen Dimensionen.

## 2.3 Untersuchung von Methoden zur Bewegungstranskription

### WELCHES VERFAHREN EIGNET SICH FÜR DIE PRÄZISE UND NATURGETREUE ERFASSUNG MENSCHLICHER BEWEGUNGEN?

Um eine präzise Erfassung und naturgetreue Abbildung menschlicher Bewegungsvorgänge zu ermöglichen, muss zunächst ein passendes Kodierverfahren gefunden werden, das diese Anforderungen erfüllen kann. Generisch erzeugte Algorithmen werden zum jetzigen Zeitpunkt der Entwicklung noch nicht berücksichtigt. Ziel ist es, eine Grundlage an transkribierten menschlichen Bewegungstrajektorien zu erschaffen, die sowohl auf die virtuellen Charaktere als auch auf einen humanoiden Roboterprototyp übertragen werden können. Je besser die Kodierqualität der menschlichen Bewegungsdatensätze ist, desto eher kann eine der wichtigsten Fehlerquellen bei der Ausführung von naturgetreuen Bewegungsabläufen ausgeschlossen werden. Hierdurch verschiebt sich letztendlich der Fokus auf das neu entwickelte Funktionsskelett, dass diese kodierten Bewegungsabläufe auch darstellen muss.

#### Keyframe Animation

Bei der Keyframe Animation (zu Deutsch Schlüsselbildanimation) fertigt ein 3D-Artist, Zeichner oder Grafiker eine Bildfolge von Bewegungsabläufen an, die nach dessen Auffassung den groben Ablauf einer Bewegung widerspiegeln, sogenannte *Keyframes*, vgl. Kempter (1999). Die fehlenden Bilder zwischen diesen *Keyframes* werden über Interpolationsverfahren per Computer ergänzt. Einer von mehreren Schwachpunkten dieses Verfahrens ist, dass die in der Regel nichtlineare, positiv wie negativ beschleunigte Bewegungen von Menschen mit diesem Verfahren nicht abbildbar sind. Die Bewegungsverläufe, die über diese Animationstechnik generiert werden, sehen für einen Betrachter aus diesem Grund vollkommen anders aus, als die am Menschen beobachtbaren Bewegungen.

#### Performance-Animation bzw. Motion Capturing

Bei diesem Verfahren zur Bewegungserfassung trägt eine Person, häufig Schauspieler, einen mit Bewegungssensoren ausgestatteten Anzug (z. B. XSENS®). Die auf diese Art erfassten Bewegungen, häufig nach den Anweisungen eines Regisseurs ausgeführt, werden zur Steuerung eines Computermodells verwendet. Außer dass diese Technik den Träger häufig in seiner Bewegungsfreiheit behindert, werden auch hier wieder die nicht-linear beschleunigten Bewegungen von Menschen nur unzureichend abgebildet. Auch das Motion Capturing mit optischen Systemen (z. B. VICON®)<sup>4</sup> ist mit schweren Mängeln behaftet. Die auf der Haut der Probanden befestigten Marker, deren räumliche Ortsveränderungen mit speziellen Kameras zur Steuerung des Computermodells erfasst

<sup>4</sup> Wurde 1984 in Oxford, Großbritannien gegründet und ist ein führender Anbieter von Motion Capturing-Systemen.

werden, sind häufig während der Bewegungen verdeckt, s. Aristidou u. a. (2008), Wu u. Boulanger (2011), oder rutschen zwischen Haut und Knochen des Trägers hin und her (Soft Tissue-Problem), vgl. Giovanni u. a. (2005), Benoit u. a. (2006), Kuo u. a. (2011). Diese Probleme erzeugen zahlreiche Bewegungs-Artefakte in der späteren Bewegungstranskription. Bei den markerlosen Systemen (z.B. MICROSOFT KINECT®) kommt es, neben der nur sehr geringen Anzahl an erfassten Körperteilen häufig zur Überlagerung und Verdeckung von Körperteilen, vgl. Smorkalov u. a. (2014), Dutta (2012) und Stoyanov u. a. (2011). Auch dieses Verfahren ist deshalb, ungeachtet der zeit- und kostenintensiven Nachbearbeitungen die damit verbunden ist, für die präzise und naturgetreue Erfassung sowie Nachbildung des sichtbaren Bewegungsverhaltens, nur sehr eingeschränkt geeignet.

### **Skript Animat ions-Methode**

Diese Methode zur Bewegungserfassung ist bisher eher unbekannt. Hierbei handelt es sich um ein 2D/3D-Animationsverfahren. Der Ausgangspunkt dieses Verfahrens bildet ein zweidimensionales Video. Das sichtbare Bewegungsverhalten von Personen, die in diesem Video zu sehen sind, wird auf eine 3D-Stickfigure, die entfernte Ähnlichkeit mit einem Menschen hat, von einem Operator auf diese Puppe übertragen. Jeder Teil dieser 3D-Puppe ist steuerbar. Der Operator positioniert nun jeden Teil dieser Puppe exakt so, wie er es Bild für Bild auf dem Video (25 fps) an einer Person beobachtet. Die mit dieser 3D-Puppe erzeugten Positionen werden als Zahlenwerte in ein programminternes, ausführbares Skript geschrieben, das nach Zeit (Zeilen) und räumlicher Position der einzelnen Puppenteile (Spalten) geordnet ist. Über diese Positions-Zeit-Notation (Skript), wird die selbe Puppe, mit der die Werte in dieses Skript geschrieben wurden, reanimiert, vgl. hierzu Kempter (1999), Frey (1999).

Vorteilhaft beim Skript Animation-Verfahren ist, dass alle Einzelteile des 3D-Computermodells (stick figure) präzise positionierbar sind, weil dessen Positionen Bild für Bild erfasst werden. Das ist gleichbedeutend mit der zeitlichen Auflösung der später reanimierten Bewegung, dadurch werden bei diesem Verfahren alle negativen wie positiven Beschleunigungen des Bewegungsverhaltens exakt erfasst und sind nachbildbar. Gleichzeitig sind diese auch analysierbar und modifizierbar, da in diesem Skript Zahlenwerte stehen, die genau die jeweiligen Beschleunigungen in Kombination mit der räumlichen Positionenbeschreibung wiedergeben.

Nachteilig bei diesem Verfahren wirkt sich aus, dass es mühsam und zeitaufwändig ist, diese auf dem Video beobachteten Körper-Positionen Bild für Bild von einem Operator nachstellen zu lassen. Sicherlich wäre es zeitsparender und von daher wünschenswert, wenn die auf dem Video beobachteten Positionen auf die Computermodelle automatisch übertragen und in das Skript geschrieben würden. Dies ist jedoch nicht möglich, da die Daten durch Algorithmen unnatürlich verfälscht würden. Ursprünglich gab es bei der Keyframe Animation das gleiche Problem. Wäre dieser Kritikpunkt zu der Zeit, als die Keyframe Animation eingeführt wurde, berücksichtigt und dieses Verfahren verworfen worden, wären beispielsweise die Disney-Studios und dessen Erfolge gar nicht erst entstanden. Filmproduzenten hatten bei Disney verstanden, dass es eher auf die Qualitäten

von solch herausragenden Zeichnern wie Carl Barks<sup>5</sup> ankommt und deswegen Aufwand und Mühen der Keyframe Animation erst einmal toleriert. Der Nachteil der Skriptanimation liegt darin, dass hier sozusagen noch kein Carl Barks aufgetaucht ist. Die Qualität der Werte, die in diesem Skript stehen, sind zweifellos von der Qualität des Schreibwerkzeugs und des zugrundeliegenden Funktionsskeletts abhängig. Die bisher verwendeten rudimentären 3D-Puppen der ersten Programmversionen nach Kempster (1999) besitzen diese Qualität noch nicht.

Es fehlt folglich ein qualitativ hochwertiges Computermodell, mit dessen Hilfe die räumlichen Positionen in das entsprechende Skript geschrieben werden können. Je besser dieses Modell die realen menschlichen anatomischen Verhältnisse abbildet und sich dessen Teile in der korrekten anatomischen Lage positionieren lassen, desto präziser und natürlicher gleicht die Animation der 3D-Puppe oder des virtuellen Avatars der natürlichen Bewegung eines Menschen. Darüber hinaus erreichen dadurch auch die im Skript hinterlegten Werte die Qualität, die für die Analyse und Modifikation des sichtbaren Verhaltens von Menschen notwendig sind. Es sind explizit jedoch nur solche menschengleich aussehenden und bewegbaren Modelle dazu in der Lage, nicht im *Uncanny Valley Phänomen* zu enden bzw. sich gegenüber diesem Kriterium zu bewähren. Ein zweiter gravierender Nachteil der Skriptanimation liegt darin, dass deren Funktions- und Steuerelemente der GUI (Graphical User Interface, zu Deutsch grafische Benutzeroberfläche) wenig benutzerfreundlich sind. Die Handhabung der Software bedarf einer recht langen Einarbeitungszeit, gestaltet sich sperrig und wirkt auf den Operator nicht selten frustrierend. Der dritte Nachteil ist, dass sich mit der dort verwendeten Figur keine Objekte manipulieren lassen und sich die Figur nicht in ein auf dem Video zu sehendes animiertes Setting übertragen lässt. Da jedoch die methodischen Vorteile der Keyframe- und Performance-Animation überwiegen, bauen die nachfolgend aufgezeigten Entwicklungen des hier vorgestellten neuen Kodierwerkzeugs auf der Methode der Skriptanimation auf. Um deren wichtigsten Nachteile auszuräumen, wurde als erstes die Konzeption und Herstellung eines Funktionsskeletts zur Integration in einen virtuellen Avatar verfolgt. Integrierte Funktionen helfen dem Kodierer, Animationen zeit optimal zu erzeugen.

## 2.4 Differenzierter Vergleich untersuchter Verfahren zur Bewegungskodierung

Im Rahmen der Voruntersuchungen in konkreten Versuchsszenarien wurde verifiziert, inwieweit sich die hier vorgestellten Verfahren zur Transkription menschlicher Bewegungsabläufe und zur anschließenden Übertragung auf das Funktionsskelett eines virtuellen Avatars, eignen.

---

<sup>5</sup> US-amerikanischer Comicautor und -zeichner sowie Cartoonist. Er galt als einer der bekanntesten Disneyzeichner.

### Transkription menschlicher Bewegungsabläufe über die Microsoft Kinect®

Bei der Wahl eines Sensors zur Detektion von Menschen und menschlichen Bewegungsabläufen im Kontext des *Uncanny Valley* und der Mensch-Maschine-Interaktion wurde in einem ersten Anlauf der kostengünstige Sensor Microsoft Kinect®, bekannt aus der Gaming-Industrie, einer näheren Betrachtung unterzogen. Konventionelle Sensoren, die beispielsweise nach dem Prinzip *Time of Flight* (ToF) arbeiten, kosten ein Vielfaches und bieten weit weniger Funktionalität. Ein Beispiel hierfür ist das *SR4000* ToF-Kamerasystem mit Anschaffungskosten von circa \$4295, vgl. Acroname (2014).

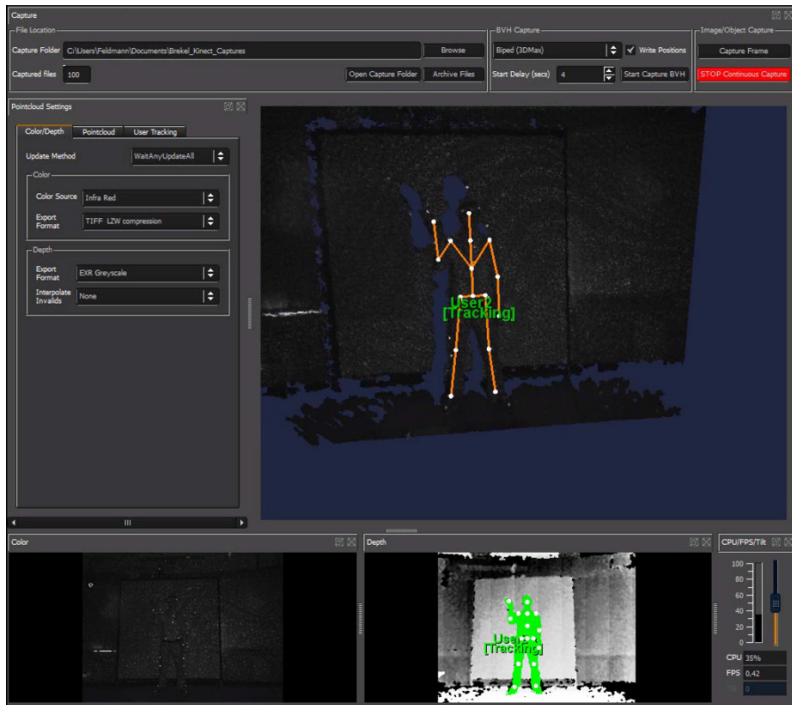


Abbildung 2.3: Motion-Capturing über den Time-of-Flight Sensor, Microsoft Kinect®.

Die Sensorikhardware der Kinect® besteht aus 4 separaten Sensoreinheiten: einem Infrarotlaser-Emitter, einer RGB-Kamera, einem Infrarot-Tiefensensor und einer Phalanx mit 4 zur Front gerichteter Mikrofone. Emissionen des eingesetzten Tiefensensors, der über ein strukturiertes Lasermuster Informationen der räumlichen Tiefe generiert, werden durch die Brechung der Laserstrahlen durch ein entsprechendes Beugungsgitter auf ein Minimum reduziert, vgl. hierzu Marcan (2015). Dies verhindert Schäden

am menschlichen Auge. Der Tiefendatenstrom wird über eine IR-Kamera erfasst, der das emittierte Licht des IR-Laser-Emitters mit einer Wellenlänge von  $\lambda = 830nm$  einfängt.

Abbildung 6.13.a veranschaulicht das Messprinzip des Infrarot Tiefensensors. Eine Laserquelle  $L$  emittiert hierbei ein strukturiertes Muster auf die Umgebung. Unter Kenntnis der Brennweite  $f$  kann über das Triangulationsverfahren, vgl. Khoshelham u. Elberink (2012), die resultierende räumliche Verschiebung bezüglich des Referenzpunkts auf dem Kamera-Chip berechnet werden. Über dieses Verfahren ist die räumliche Beschreibung eines Messpunkts über die  $[xyz]^T$ -Koordinaten komplett, vgl. Abb. 6.13 b. Das Messprinzip eignet sich ausschließlich für Objekte, die in der Lage sind Infrarotstrahlen zu reflektieren. Gläserne Objekte werden beispielsweise von der Infrarotstrahlung durchdrungen, weshalb keine exakte Positionsbestimmung erfolgen kann, vgl. Microsoft-Corporation (2015).

Die Auflösung des Punktrasters beträgt  $640 \times 480$  Bildpunkte bei einer Bildwiederholrate von 30 Frames pro Sekunde. Der Datenstrom besteht aus insgesamt 16 Bit pro Pixel (bpp) und enthält über 13 höherwertige Bits an Informationen zur Messung der Distanz in  $mm$  in einem Wertebereich von 0 bis 8,192 Einheiten. Die 3 verbleibenden Bits transportieren einen User-Index, der die erkannten Bildfeatures einem spezifischen Benutzer zuordnet, vgl. Microsoft-Corporation (2015). Dies ist insbesondere nötig, sobald sich mehrere Personen im Interaktionsraum des Roboters aufhalten. Abbildung 2.3 zeigt das Benutzerinterface zur Erfassung menschlicher Bewegungsdaten über die Microsoft Kinect®. Es wurde im Versuch die kommerziell erhältliche Software Brekel Kinect verwendet, vgl. Brekel (2015). Insgesamt wird eine Grundgesamtheit von 16 Gelenken erfasst. Das generierte, virtuelle Funktionsskelett wird über die erfassten Daten des Infrarot-Tiefensensors gelegt, vgl. Abb. 2.3. Es sind 13 Versuchsdurchläufe mit unterschiedlichen Bewegungen durchgeführt worden. Insgesamt haben sich folgende Vor-, bzw. Nachteile bezüglich der Tauglichkeit zur Transkription menschlicher Bewegungsmuster ergeben:

### Vorteile

- Geringer Anschaffungspreis ( $\leq \$140$ )
- Einfache, benutzerfreundliche Handhabung
- Gute Dokumentation
- Frei verfügbare Entwicklungsbibliotheken
- Einfache Implementierung von höheren Funktionen (Sprach- Gestensteuerung)
- Nahezu Beleuchtungsunabhängig (Infrarot-Laser)
- Hohe Erfassungsgenauigkeit (Kombination von antrainierten neuronalen Netzen und integrierter Feature-Erkennung)
- Markerfrei (Es werden keine Marker am Körper des Probanden benötigt)
- Direkte Umwandlung in BHV-Dateien zur Übertragung auf das Biped Funktionsskelett von Autodesk®
- Große Anzahl frei verfügbarer Software-Tools
- Mobil (Leicht beweglich)

## Nachteile

- Grobe Positionsauflösung ( $640 \times 480$  Bildpunkte), erzeugt Positions- Artefakte von bis zu  $5\text{cm}$  pro erfasstem Gelenk
- Unter Umständen geringe Bildwiederholrate ( $\leq 30\text{fps}$ )
- Hohe CPU-Auslastung je nach Erkennungswunsch
- Nur frontale Erfassung möglich (Körperteile können verdeckt werden)
- Geringe Anzahl an Gelenkfreiheitsgraden erfassbar (max. 22 Gelenkfreiheitsgrade)
- Rotation und Drehung von Körperteilen werden nicht erfasst

Insgesamt kann gesagt werden, dass die Microsoft Kinect® generell eine Reihe an Vorteilen besitzt. Diese sind hauptsächlich in der Sprach- und Gestensteuerung sowie in der hohen Zuverlässigkeit der Erfassungsalgorithmen zu finden. Aufgrund des Anwendungsgebietes im Marktsegment für Endbenutzer, bietet der Sensor eine sehr robuste Benutzererfassung. Grobe Bewegungsabläufe können leicht und intuitiv analysiert und aufgezeichnet werden. Erfasste Bewegungen können direkt auf das Funktionsskelett des virtuellen Avatars übertragen werden. Jedoch reicht dieser Sensor aufgrund häufig auftretender Bewegungsartefakte (Standortabhängig bis zu  $5\text{cm}$  Abweichung je erfaster Gelenkposition) nicht zur genauen Transkription menschlicher Bewegungsabläufe. Rotation und Drehung von Körperteilen können nicht erfasst werden.

## Transkription menschlicher Bewegungsabläufe über ein Motion-Capturing System

Das Unternehmen Vicon ist ein führender Entwickler von Motion-Capturing Systemen und besitzt seit nunmehr 30 Jahren Erfahrung auf dem Gebiet der Erfassung menschlicher Bewegungsabläufe, vgl. Vicon (2015). Ein System bestehend aus 8 Infrarotkameras steht am Lehrstuhl Mechanik und Robotik der Universität Duisburg-Essen zur Verfügung. Das System konnte im Rahmen von Probandenstudien für eine Reihe von Versuchsdurchläufen genutzt werden.

Zur Erfassung menschlicher Bewegungen muss der Proband vorher mit reflektierenden Markern versehen werden. Diese sind:

- Sphärische Marker mit  $3\text{-}4\text{mm}$  Durchmesser zur Erfassung einzelner Gesichts- oder Handbewegungen
- Sphärische Marker mit  $9\text{-}14\text{mm}$  Durchmesser zur Erfassung des kompletten Körpers
- $12\text{-}20\text{mm}$  Marker zur Bewegungserfassung mit Kontakt
- Sphärische Marker von  $25, 25$  und  $50\text{mm}$  Durchmesser zur Erfassung von Bewegungen in großen Volumen

Die Marker sollten entweder unmittelbar auf der Haut oder auf einem eng anliegendem Bodysuits befestigt werden. Die Marker stehen etwa  $30\text{mm}$  vom Körper ab und können hierdurch die Bewegungsfreiheit einschränken. Die Infrarotkameras werden an der Decke, bzw. an einem Aluminium Gestell, sphärisch um den Probanden platziert und

ermöglichen im Gegensatz zur Microsoft Kinect eine räumliche Rundumerfassung samt Erfassung der Orientierung der Extremitäten. Insgesamt müssen für eine komplette Aufnahme der Bewegungen des menschlichen Körpers 42 Marker an vorher fest definierten Stellen befestigt werden, vgl. Abb. 2.4. Weicht die Befestigungsposition von der Vorgabe ab, kann dies zu Messungenauigkeiten führen. Die Übertragung der gewonnen Bewegungsdaten geschieht über Autodesk Motion Builder<sup>®</sup> auf das Biped Skelett.

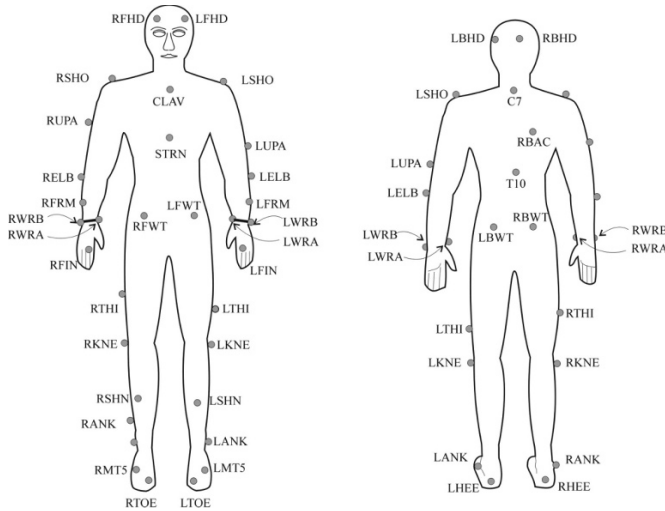


Abbildung 2.4: Befestigungspositionen der Reflektionsmarker beim Vicon<sup>®</sup> Motion-Capturing System, vgl. Vicon (2015).

### Vorteile

- Erfassung von räumlicher Orientierung der Extremitäten (8 Infrarotkameras werden um den Probanden platziert)
- Höhere Anzahl von Freiheitsgraden im Vergleich zur Microsoft Kinect<sup>®</sup>
- Kameraauflösung bis zu 16 Mega Pixel (MP), Bildwiederholrate bis zu 180 Frames pro Sekunde (fps)
- Übertragung der Bewegungsmuster auf ein Biped-Funktionsskelett über eine Drittsoftware (Autodesk Motion Builder<sup>®</sup>)
- Höhere Positioniergenauigkeit über redundante Erfassung eines Markers mit mehreren Kamerasystemen

### Nachteile

- Hoher Aufwand zur Vermarkierung (bis zu 40 Minuten)
- Marker schränken die Bewegungsfreiheit ein

- Soft-Tissue Artifacts können auftreten (Bewegung zwischen Haut- oder Kleidung und Marker während der Bewegung)
- Hohe Anschaffungskosten (ab ca. \$ 50.000)
- System ist unbeweglich, hoher Aufwand zur Platzierung des Systems außerhalb des Labors
- Es werden oftmals mehrere Aufnahmen derselben Bewegung benötigt, da vom System nicht immer alle Marker korrekt erfasst werden

Abbildung 2.5 zeigt einen Ausschnitt der Bewegungserfassung des Motion- Capturing Systems. Insgesamt konnten die Ergebnisse zur Erfassung dynamischer menschlicher Bewegungen überzeugen. Es konnten jedoch auch hier keine Bewegungen einzelner Finger erfasst werden. Würde sich der Proband auf einen Stuhl setzen so liefert dieses System falsche Datensätze, da Marker verdeckt werden. Anlehnen an eine Lehne ist nicht möglich, da sich Marker am Rücken befinden, hierdurch werden Bewegungen verfälscht. Werden beispielsweise Daten zur Pflege von Patienten in einem Krankenhaus erfasst, können bei einer engen Interaktion des Probanden mit einem Patienten nicht alle Marker erfasst werden. Es entstehen ebenfalls fehlerhafte Bewegungsdatensätze.

Abbildung 2.6 zeigt die Übertragung der Bewegungsdaten auf einen virtuellen Avatar. Die Bewegungen gestalten sich äußerst flüssig und sehen sehr natürlich aus. Es gibt so gut wie keine Artefakte in der Positioniergenauigkeit wie diese bei der Microsoft Kinect auftreten. Dennoch müssen von derselben Bewegung oft mehrere Aufnahmen durchgeführt werden. Dies ist nicht praxistauglich. Werden beispielsweise spontane Aufnahmen einer Person in seinem Arbeitsumfeld gewünscht, erweist sich die Aufnahmemethode als unpraktisch und geradezu untauglich. Die Kalibrierung des Kamerasystems dauert lange, die Vermarkierung der Probanden ebenfalls. Für weitere Untersuchungen können frei verfügbare Datensätze, der *Carnegie Mellon University* genutzt werden. Es wurden hierbei 12 Vicon infrarot MX-40 Kameras (Auflösung 4 MP) verwendet. Die Bildwiederholrate beträgt 120 fps. Es stehen 2605 Bewegungsaufnahmen in 6 Kategorien (Interaktion von Menschen, Interaktion von Menschen mit der Umgebung, Fortbewegung, physische Aktivität und Sport, diverse Situationen sowie Szenarien und Testbewegungen) zur Verfügung. Diese Datensätze bieten eine Reihe kodierter Bewegungsansätze und sind bestens geeignet zur weiteren Untersuchung von Bewegungsabläufen im Kontext des *Uncanny Valley*. Diese können bei Bedarf über das Vicon-System der Universität Duisburg-Essen ergänzt werden.

## 2.5 Anatomische Grundlagen des menschlichen Bewegungsapparats

Die Erkenntnisse aus den Untersuchungen zur Bewegungstranskription auf Grundlage des Berner Systems werden nun dazu genutzt, menschliche Bewegungen auf die humanoide Robotik zu übertragen. Dafür ist es zunächst notwendig, die Unterschiede zwischen menschlichen und maschinellen Bewegungsvorgängen zu verstehen.

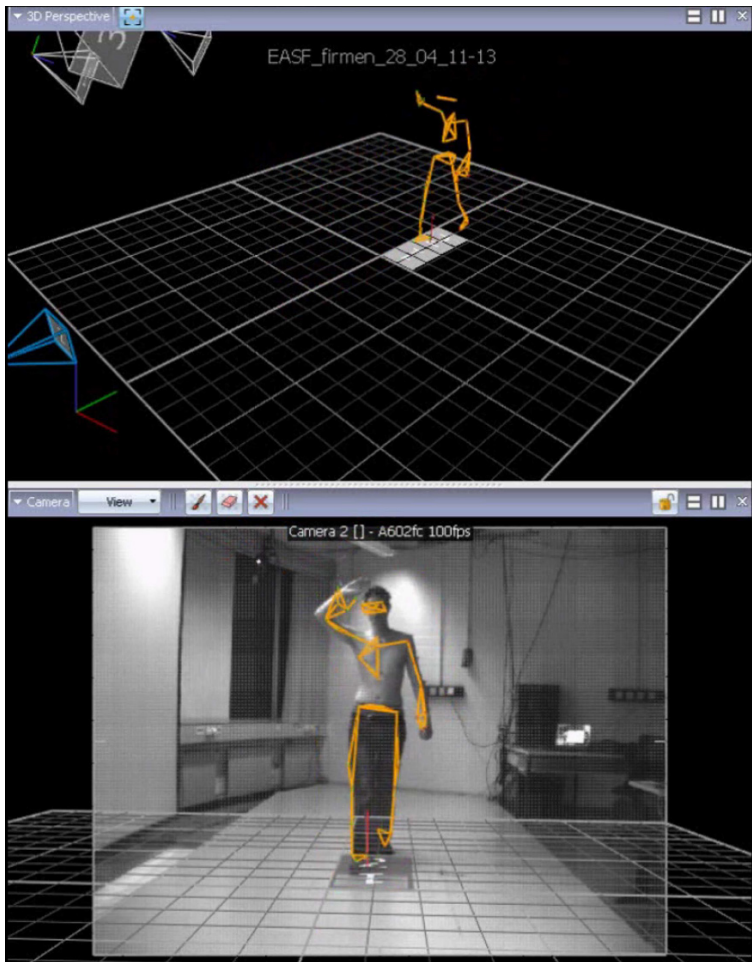


Abbildung 2.5: Motion-Capturing über das markerbasierte System von Vicon®.

Hierzu werden stellvertretend für eine Vielzahl anderer Gelenke des menschlichen Körpers, zwei typische Vertreter einer genaueren Betrachtung unterzogen. Das Ziel dabei besteht darin, Erkenntnisse über die spezifischen Bewegungsvorgänge der Gelenke des menschlichen Körpers zu gewinnen. Die zugrunde liegenden Untersuchungen konzentrieren sich zum einen auf das menschliche Schultergelenk (lat. *Articulatio humeri*, auch *Glenohumeralis*) und zum anderen auf das menschliche Hüftgelenk (lat. *Articulatio*

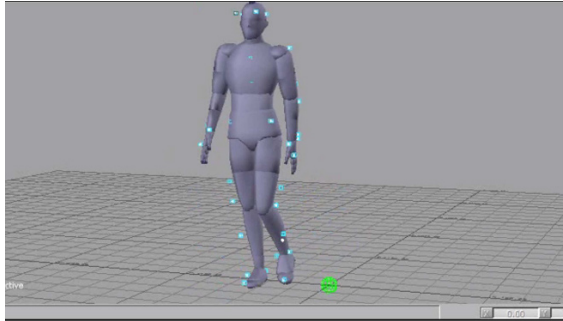


Abbildung 2.6: Übertragung der Vicon Bewegungsdaten auf einen virtuellen Avatar in Autodesk Motion Builder®.

coxae). Beide Gelenke werden bezüglich ihrer typischen Freiheitsgrade und Bewegungsmöglichkeiten untersucht. Vorab ist bereits festzustellen, dass sich die Bewegungsmöglichkeiten keineswegs so trivial (Punkt-Rotationen) darstellen, wie sie heutzutage von Ingenieuren in vielen Beispielen humanoider Roboterprototypen technisch umgesetzt worden sind. Oftmals kann ein maschinelles Gelenk die geforderten menschlichen Bewegungsabläufe aufgrund physikalischer und technischer Restriktionen gar nicht erst darstellen. Dies führt letztendlich dazu, dass ausgeführte Bewegungsvorgänge optisch stark von den menschlichen abweichen, auch wenn sie noch so gut programmiert oder mit technischen Hilfsmitteln erfasst wurden.

### 2.5.1 Beispiel 1: Die menschliche Schulter

Bei der menschlichen Schulter (lat. *Articulatio humeri* oder auch *Glenohumeralis*) handelt es sich um ein typisches Kugelgelenk mit 3 Freiheitsgraden. Der funktionelle Aufbau, der letztendlich die große Bewegungsfreiheit des menschlichen Arms ermöglicht, resultiert im wesentlichen aus dem Zusammenspiel von *Scapula* (Schulterblatt), *Clavicula* (Schlüsselbein) und *Humerus* (Oberarm), einschließlich der dazwischen befindlichen Gelenktypen.

Die menschliche Schulter besteht insgesamt aus drei Haupteinzelgelenken und zwei Scheingelenken, vgl. Aumueeller u. a. (2007). In Anlehnung zu Abbildung 2.7 sind zu nennen:

- Sternoclaviculargelenk: Gelenk zwischen *Clavicula* (Schlüsselbein) und Sternum (Brustbein) (a)
- Acromioclaviculargelenk: Gelenk zwischen *Clavicula* (Schlüsselbein) und *Scapula* (Schulterblatt) (b)

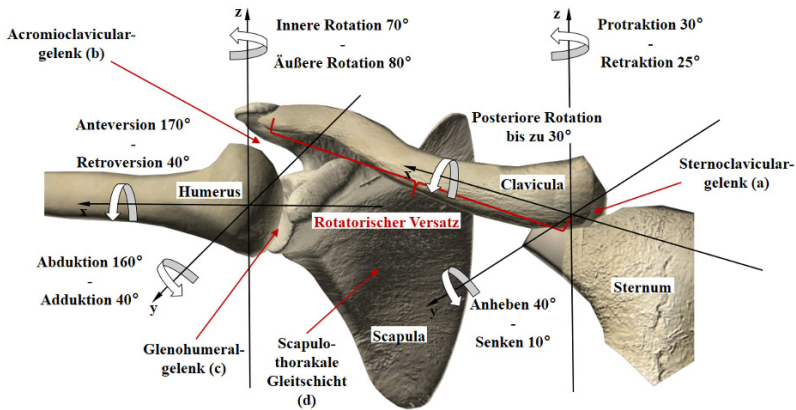


Abbildung 2.7: Erweitertes kinematisches Modell der menschlichen Schulter.

- Glenohumergelenk: „eigentliches Schultergelenk“ zwischen Humerus-Kopf (Oberarmkopf) und der Gelenkfläche des Schulterblattes dem Glenoid (c)
- Scapulothorakale Gleitschicht: Gleiten der Scapula (Schulterblatt) auf der Rückseite des Thorax (Brustkorb) (d)
- Humerus (Oberarm): angeschlossen an die Schulter, vgl. Abb. 2.7

Durch die Einbeziehung von Clavicula in den Bewegungsapparat, bietet sich dem menschlichen Oberarm ein bedeutend größerer Bewegungsspielraum, als dies ohne der Fall wäre. Was aber letztendlich Nachteile in der statischen Belastbarkeit zur Folge hat, vgl. Aumüller u. a. (2007). In Abbildung 2.7 werden zusätzlich zu den einzelnen Rotationsachsen und Freiheitsgeraden, typische Bewegungswinkel dargestellt. Die menschliche Schulter ist einer der komplexesten Bewegungsapparate, die der menschliche Körper zu bieten hat. Gerade das Zusammenspiel zwischen Clavicula und Scapula erlaubt es, den Arbeitsraum des Endeffektors, also der menschlichen Hand, entscheidend zu vergrößern. Hierdurch bewegt sich die menschliche Schulter zusätzlich zur eigentlichen Rotation im Glenohumergelenk auf einer konzentrischen Bahn um das Sternoclaviculargelenk. Dies führt letztendlich zu einer wesentlich höheren Flexibilität des Gesamtgelenks.

### 2.5.2 Beispiel 2: Das menschliche Hüftgelenk

In Abbildung 2.8 ist ein weiteres typisches Beispiel für die erweiterten Bewegungsmöglichkeiten in der menschlichen Anatomie zu sehen, das Hüftgelenk (lat. Articulatio coxae). Oftmals ist es der einfachen Bewegungskinetik oder der Art der zur Verfügung stehenden mechanischen Gelenkverbindungen geschuldet, dass bei rein funktionel-

dem Layout eines humanoiden Roboters kein Augenmerk auf die wirkliche Bewegungsausführung in der menschlichen Anatomie gelegt wird. Das menschliche Hüftgelenk besteht, wie in Abbildung 2.8 dargestellt, aus Hüftpfanne (Acetabulum) und Hüftkopf (Caput femoris). Die Pfanne wird von Anteilen des Os ilium, des Os pubis und des Os ischii gebildet, die über eine Y-förmige Fuge im Bereich des Acetabulum verbunden sind. Der obere Rand der Pfanne wird durch einen knorpeligen Rand, den Limbus acetabuli, verstärkt, vgl. AumueLLer u. a. (2007). Der Hüftkopf ist eine in etwa kugelförmige Extremität des Femurs (Oberschenkelknochen), der in die Pfanne drückt und damit eine Verbindung zwischen Bein und Rumpf ermöglicht. Das Hüftgelenk ist somit ein Nussgelenk, das eine Unterform des Kugelgelenks (Articulatio spherioidea) darstellt, vgl. AumueLLer u. a. (2007).

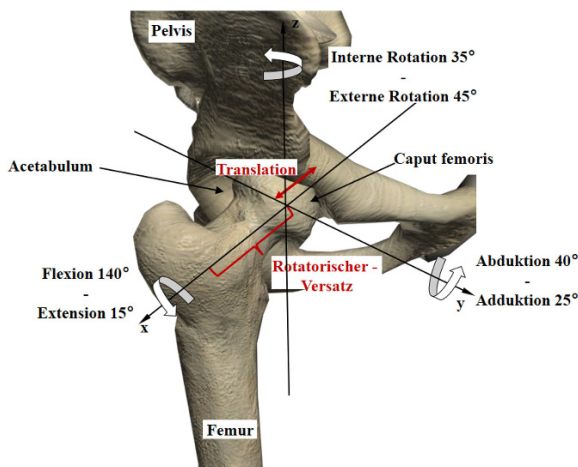


Abbildung 2.8: Erweitertes kinematisches Modell des menschlichen Hüftgelenks.

Ein konventionelles Kugelgelenk besitzt die Fähigkeit um drei orthogonale Achsen zu rotieren (XYZ-Achse). Der Femur (Oberschenkelknochen) besitzt zusätzlich die Fähigkeit, sich relativ zur Hüfte vor und zurück zu bewegen (flexion/extension). Weiterhin ist er in der Lage, sich medial und lateral (adduction/abduction) zu bewegen und intern/extern zu rotieren. Dies führt dazu, dass auch durch dieses Gelenk Femur eine konzentrische Bahn um das eigentliche Hüftgelenk vollzieht. Ein Abtasten der Hüfte zeigt, dass der Hüftkopf zusätzlich noch eine translatorische Bewegung ausführen kann. Er bewegt sich hierbei zur Hüftpfanne hin bzw. weg. Diese laterale Verschiebung ist eine wichtige Eigenschaft des Hüftgelenks, und wurde im Kontext der humanoiden Robotik bisher nur unzureichend untersucht. Der Betrag dieser Verschiebung kann als Laxität des Gelenks bezeichnet werden. Die maximale Ausprägung des Versatzes ist nicht, wie einst gedacht, abhängig von der Länge des Leistenbands. Vielmehr hängt

sie vom relativen Volumen der Gelenkflüssigkeit innerhalb des Gelenks ab, die mit der Gelenkkapsel interagiert, vgl. D'Amico u. a. (2011).

## 2.6 Der menschliche Körper als Vorbild in der humanoiden Robotik

Wird der menschliche Bewegungsapparat zum Vorbild genommen, ergeben sich ganz automatisch Anforderungen an die Bewegungsmöglichkeiten jeder einzelnen Gelenkverbindung eines technischen Systems, das diese imitieren möchte. Die beiden Beispiele menschliche Schulter und Hüftgelenk zeigen deutlich, dass die rein funktionale Umsetzung in mechanische Gelenke humanoider Roboter schnell zu deutlichen Problemen führen kann. Wird beispielsweise die konzentrische Bewegungsmöglichkeit des Femur vernachlässigt, können die Beine nur schwerlich übereinander geschlagen werden. Weiter verdeutlicht wird dies am menschlichen Schultergelenk. Vernachlässigt man die Bewegung des Glenohumeralgelenks, kann mit dem Arm die gegenüberliegende Schulter nicht erreicht werden. Wird die Bewegung von Clavicula vernachlässigt, ist ein einfaches *Achselzucken* nicht mehr möglich. Diese fehlenden Funktionalitäten führen dazu, dass allein strukturell gesehen schon starke Bewegungsdefizite, verglichen mit menschlichen Bewegungen, auftreten. Wichtige Signale der nonverbalen Kommunikation können daher nicht an den Empfänger einer Botschaft übermittelt werden.

Große Defizite entstehen also schon während des Entwicklungsprozesses eines humanoiden Roboters. Es bedarf demnach eines grundlegenden Paradigmenwechsels in der Herangehensweise bei der Entwicklung humanoider Roboterprototypen. Aspekte der Bewegungsrepräsentation müssen frühzeitig in den Vordergrund gerückt werden, um die Bewegungsabläufe eines humanoiden Roboters menschenähnlicher zu gestalten. Dieser Ansatz erfordert im Vorfeld eine ausführliche Untersuchung von Aspekten der nonverbalen Kommunikation. Interaktionsszenarien zwischen Mensch und humanoiden Roboter erfordern es, Aspekte der nonverbalen Kommunikation ausführlich zu untersuchen. ScanimFBX ermöglicht dies ohne die Erfordernis eines physischen Prototypen. Es kann damit das Phänomen des *Uncanny Valley* anhand eines virtuellen Prototypen untersucht werden und Erkenntnisse frühzeitig in den Entwicklungsprozess einfließen. Ziel ist es nun, genau diese Erkenntnisse zu erfassen und diese systematisch in einen Entwicklungsansatz nach Gesichtspunkten der Humankommunikation einfließen zu lassen. Dafür werden im ersten Schritt die charakteristischen Merkmale des humanoiden Roboterskelettes simulativ verifiziert, bevor es zum Aufbau des eigentlichen Prototypen und einer Übertragung menschlicher Bewegungen kommt.

Über einen geeigneten technischen Ansatz wird erforscht, wie die Defizite konventioneller humanoider Roboter, bezüglich der Humankommunikation, überwunden werden können. Hierzu werden folgende charakteristischen Merkmale ins Auge gefasst:

- Die dynamischen Eigenschaften (Bewegungsgeschwindigkeiten, Beschleunigungen, ...) des menschlichen Vorbilds sollen unter Berücksichtigung der technischen Restriktionen eingehalten werden.
- Bewegungsausprägungen und Freiheitsgrade eines Menschen sollen über eine geeignete Aktuatorik am humanoiden Roboter dargestellt werden.
- Die Reproduktionsgenauigkeit von kodierten Bewegungstrajektorien soll unter Berücksichtigung der Sensorik, Getriebeispiel und Bauteilelastizitäten dem Menschen nachempfunden werden.

## 2.7 Bewegungsdynamik der Gelenke des menschlichen Bewegungsapparats

Um das Ziel zu erreichen, die Dynamik menschlicher Bewegungen auf einen Roboterprototyp zu übertragen, wird nun exemplarisch analysiert, welche Anforderungen an Geschwindigkeiten und Beschleunigung an die Aktuatorik eines humanoiden Roboters gestellt werden. Eine Untersuchung des Ellbogengelenks aus dem Jahr 1989 dient als Grundlage, um eine genauere Einschätzung über die Anforderungen an die Kennwerte der dynamischen Bewegung des menschlichen Körpers zu erlangen, vgl. Nagasaki (1989). Um eine möglichst hohe Bewegungstreue bei der Übertragung menschlicher Bewegungstrajektorien zu gewährleisten, sind genaue Aussagen über die technischen Anforderungen an den späteren Prototypen notwendig.

Tabelle 2.1: Analyse bei Beugung des menschlichen Unterarms bei langsamen Winkelgeschwindigkeiten und Beschleunigungen (Mittelwert und Standardabweichung), Nagasaki (1989). Ziel- und Ist-Position in Grad, Zeit ( $T_{MU}$ ), Spitzengeschwindigkeit ( $V_{MU-max}$ ) und Winkelbeschleunigung ( $A_{MU}$ ).

| <i>Zielposition</i><br>(Grad) | <i>Ist-Position</i><br>(Grad) | $T_{MU}$<br>(ms) | $V_{MU-max}$<br>(Grad/s) | $A_{MU}$<br>(Grad/s <sup>2</sup> ) |
|-------------------------------|-------------------------------|------------------|--------------------------|------------------------------------|
| 20                            | 19,9                          | 874,2            | 43,6                     | 28,7                               |
|                               | 0,9                           | 149,0            | 9,4                      | 11,1                               |
| 30                            | 30,1                          | 978,7            | 57,7                     | 35,1                               |
|                               | 0,7                           | 173,6            | 14,0                     | 12,2                               |
| 40                            | 39,8                          | 1067,5           | 67,8                     | 37,7                               |
|                               | 0,8                           | 160,4            | 12,9                     | 9,4                                |
| 50                            | 49,8                          | 1142,8           | 81,9                     | 41,6                               |
|                               | 0,9                           | 169,8            | 15,8                     | 12,2                               |
| 60                            | 60,0                          | 1222,3           | 93,4                     | 40,9                               |
|                               | 0,1                           | 90,5             | 12,4                     | 4,7                                |

In den Tabellen 2.1, 2.2 und 2.3 werden die resultierenden Untersuchungsergebnisse zur Beugung des menschlichen Ellbogens dargestellt. Die erreichbaren Winkelgeschwindigkeiten des Arms der Probanden beim Beugen und Strecken, wurden unter Laborbedingungen erfasst. Im Experiment wurde der rechte Arm eines sitzenden Probanden

in ein schwenkbares Gestell gelegt, an dessen Ende ein vertikaler Griff befestigt ist. Die Versuchsperson sollte jeweils eins von 5 Zielen durch Beugung des rechten Vorderarms erreichen. Die Positionen der Ziele lagen bei 20, 30, 40, 50 und 60 Grad. Es wurden jeweils 10 einzelne Versuchsdurchläufe zu jedem untersuchten Zielwinkel bei 3 unterschiedlichen Geschwindigkeiten vollzogen.

Tabelle 2.2: Analyse bei Beugung des menschlichen Unterarms bei mittleren Winkelgeschwindigkeiten und Winkelbeschleunigungen (Mittelwert und Standardabweichung), Nagasaki (1989). Ziel- und Ist-Position in Grad, Zeit ( $T_{MU}$ ), Spitzengeschwindigkeit ( $V_{MU-max}$ ) und Winkelbeschleunigung ( $A_{MU}$ ).

| <b>Zielposition</b><br>(Grad) | <b>Ist-Position</b><br>(Grad) | $T_{MU}$<br>(ms) | $V_{MU-max}$<br>(Grad/s) | $A_{MU}$<br>(Grad/s <sup>2</sup> ) |
|-------------------------------|-------------------------------|------------------|--------------------------|------------------------------------|
| 20                            | 20,4                          | 567,1            | 74,1                     | 72,0                               |
|                               | 1,2                           | 122,9            | 16,4                     | 25,9                               |
| 30                            | 30,4                          | 624,4            | 99,5                     | 87,8                               |
|                               | 1,0                           | 122,5            | 24,8                     | 33,4                               |
| 40                            | 40,5                          | 635,2            | 128,4                    | 111,2                              |
|                               | 1,2                           | 119,2            | 26,1                     | 39,5                               |
| 50                            | 50,4                          | 688,3            | 143,6                    | 115,6                              |
|                               | 1,2                           | 115,3            | 25,4                     | 30,8                               |
| 60                            | 60,4                          | 699,6            | 172,5                    | 135,2                              |
|                               | 1,3                           | 124,6            | 32,2                     | 45,0                               |

Tabelle 2.3: Analyse bei Beugung des menschlichen Unterarms bei sehr schnellen Winkelgeschwindigkeiten und Winkelbeschleunigungen (Mittelwert und Standardabweichung), Nagasaki (1989). Ziel- und Ist-Position in Grad, Zeit ( $T_{MU}$ ), Spitzengeschwindigkeit ( $V_{MU-max}$ ) und Winkelbeschleunigung ( $A_{MU}$ ).

| <b>Zielposition</b><br>(Grad) | <b>Ist-Position</b><br>(Grad) | $T_{MU}$<br>(ms) | $V_{MU-max}$<br>(Grad/s) | $A_{MU}$<br>(Grad/s <sup>2</sup> ) |
|-------------------------------|-------------------------------|------------------|--------------------------|------------------------------------|
| 20                            | 22,5                          | 184,8            | 253,6                    | 720,8                              |
|                               | 2,1                           | 33,0             | 57,8                     | 228,6                              |
| 30                            | 32,0                          | 201,1            | 343,4                    | 914,4                              |
|                               | 2,4                           | 43,6             | 82,1                     | 413,4                              |
| 40                            | 42,2                          | 222,4            | 416,0                    | 985,5                              |
|                               | 2,4                           | 49,5             | 85,3                     | 434,3                              |
| 50                            | 52,7                          | 236,0            | 476,7                    | 1098,4                             |
|                               | 2,6                           | 52,7             | 94,6                     | 475,6                              |
| 60                            | 62,0                          | 247,2            | 540,3                    | 1151,1                             |
|                               | 2,3                           | 51,1             | 95,0                     | 480,7                              |

Die in Tabelle 2.2 aufgeführten maximalen Winkelgeschwindigkeiten treten bei schneller, ballistischer Bewegung auf. Bei der Bewegung auf die 60° Position liegt die erzielte Winkelgeschwindigkeit bei  $V_{MU-max} = 540,3 \frac{\text{Grad}}{\text{s}}$ . Dieser Maximalwert wird im folgenden als Referenzwert für Prämissen der menschlicher Bewegungsdynamik herangezogen.

An diesem maximal erreichten Wert, wird die Bewegungsgeschwindigkeit von potentiellen Antriebskonzepten des Roboters bemessen. Insbesondere für die spätere Auswahl von Aktuatoren des Roboterprototyps spielt die genaue Kenntnis über Eigenschaften menschlicher Gelenke eine überaus wichtige Rolle. Die Untersuchungsergebnisse dienen daher dazu eine Vorauswahl und Einordnung während der Systemauslegung zu treffen.

## 2.8 Betrachtung konventionell aufgebauter humanoider Roboterprototypen

Anhand zweier Beispiele autonomer humanoider Roboter des *Korea Institute of Science and Technology (KIST)* wird im Folgenden gezeigt, worin genau die Defizite liegen, die es verhindern, dass konventionell konstruierte Roboterprototypen Signale der nonverbalen Kommunikation übertragen. Blickt man auf die Erkenntnisse aus Untersuchungen mit dem *Berner System*, so formulierte S. Frey schon frühzeitig, dass es einer Mindestanzahl von 104 Bewegungsausprägungen bedarf, um die Komponenten der nonverbalen Kommunikation zu reproduzieren. Diese Erkenntnis wird derzeit im Entwicklungsprozess existierender Roboterprototypen schlicht und einfach vernachlässigt. Dies führt dazu, dass Signale der nonverbalen Kommunikation aufgrund dieser fehlenden Freiheitsgrade nicht übertragen werden können.

### Humanoider Roboter MAHRU-1:

Beispielhaft für eine Vielzahl weiterer Roboterprototypen wird der humanoide Roboter MAHRU-1, vgl. Park u. a. (2004), einer näheren Betrachtung unterzogen. Dieser ist ein hervorragendes Beispiel, um Defizite bezüglich der Fähigkeit zur Humankommunikation aufzuzeigen. Diese Defizite entstehen maßgeblich durch fehlende Freiheitsgrade und anatomisch inkorrekt positionierte Manipulatorgelenke, vgl. hierzu Abbildung 2.9. Das primäre Entwicklungsziel war es sicherlich nicht, menschliches Verhalten akkurat zu imitieren, dennoch können einige typische Defizite in der Struktur des Funktionsskeletts des Prototypen aufgezeigt werden.

MAHRU-1 besitzt insgesamt 35 Freiheitsgrade. Als Recheneinheit dient eine 3,2 GHz CPU. Die Motortreiber werden über den CAN<sup>6</sup>-Bus angesteuert, vgl. You u. a. (2008). Das Gesamtgewicht beträgt 67kg und die Bauhöhe 150cm. Die Systemfreiheitsgrade teilen sich auf wie in Tabelle 2.4 dargestellt. Im Bereich des Schultergelenks, vgl. Abb. 2.9 A, ist ein Gelenk mit lediglich 2 DoF (Rotation und Schwenk) zu sehen, das außen am Torso des Roboters angebracht ist. Es besitzt weder den nötigen dritten Freiheitsgrad, noch die erforderlichen Bewegungswinkel oder die übergeordnete Bewegungs-

<sup>6</sup> CAN steht für Controller Area Network und ist ein serielles Bussystem, welches zur Familie der Feldbusse gehört. CAN ist mit ISO 11898 international standardisiert und besteht je Kommunikationspaket aus einer physischen Schicht (Layer 1) und einer Datensicherungsschicht (Layer 2) im ISO/OSI Referenzmodell.

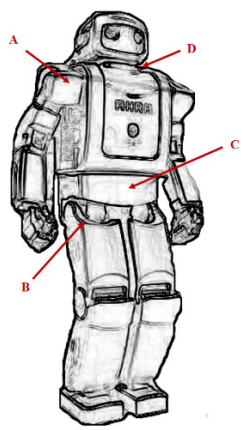


Abbildung 2.9: Defizite in der Bewegungsausführung demonstriert am humanoiden Roboter MAHRU-1, A: unzureichende Freiheitsgrade in der Schulter, B: unzureichende Bewegungsmöglichkeiten im unteren Bauch/Hüftbereich, C: fehlende Hüftknochen, D: Fehlende Bewegungsmöglichkeiten im Nackenbereich, in Anlehnung an You u. a. (2008).

Tabelle 2.4: Freiheitsgrade (DoF) des humanoiden Roboters MAHRU-1, vgl. You u. a. (2008).

| <i><b>Körperteile (einzeln)</b></i> | <i><b>Freiheitsgrade</b></i> |
|-------------------------------------|------------------------------|
| je Arm                              | 6 DoF x 2                    |
| je Hand                             | 4 DoF x 2                    |
| je Bein                             | 6 DoF x 2                    |
| Nacken                              | 2 DoF                        |
| Becken                              | 1 DoF                        |
| <b>Insgesamt:</b>                   | <b>35 DoF</b>                |

möglichkeit der Clavicula um die realen Bewegungsabläufe der menschlichen Schulter ausführen zu können, vgl. hierzu Abbildung 2.9. Fehlende Bewegungsmöglichkeiten im Nackenbereich, in der Hüfte und im Bauchbereich, lassen typische menschliche Bewegungen nicht zu. Daher können Signale der nonverbalen Kommunikation von diesem Prototypen schlichtweg nicht übermittelt werden. Das Problem der funktionalen Bewegungssteuerung wurde demgegenüber hervorragend gelöst. Der Prototyp ist in der Lage, autonom zu laufen und einfache Tätigkeiten mit seinen Armen auszuführen. Geht es jedoch darum, komplexe menschliche Bewegungsabläufe zu imitieren bzw. das Bewegungsverhalten eines Menschen zu übertragen, scheitert dieser Prototyp vollkommen.

Sicherlich lag der Sinn und Zweck darin einen möglichst autonomen Funktionsprototyp zu generieren. Rein funktionelle Aspekte und Kriterien wie eigenständiges Laufen, Greifen und die Manipulation und Interaktion mit der Umgebung standen im Vordergrund des Entwicklungsprozesses. Doch kann gerade an diesem Beispiel sehr gut gezeigt werden, welche Defizite bei gängigen humanoiden Roboterprototypen derzeit vorherrschen.

### **KIST Humanoid Robot-2 (KHR-2):**

Ein weiterer vom *Korea Institute of Science and Technology (KIST)* entwickelter Prototyp ist der humanoide Roboter KHR-2. Dieser ist mit 120cm und 54kg Gesamtgewicht etwas kleiner und leichter als MAHRU-1, vgl. Abb. 2.10.



Abbildung 2.10: Humanoider Roboter KHR-2, vgl. Kim u. a. (2005).

Die Aktuatoren werden ebenfalls über den CAN-Bus angesteuert. KHR-2 ist über Wireless LAN mit einem HOST-PC verbunden. Er verbraucht 5-6A bei 12V, bzw. 6–10A bei 24V und kann 50 Minuten netzunabhängig agieren. Angetrieben wird er über mehrere DC/Servomotoren. Als Sensoren kommen eine CCD-Kamera, Kraftsensoren in den Gelenken und Lagesensoren zum Einsatz, die ebenfalls über den CAN-Bus angesteuert werden. Die Gesamtzahl der Freiheitsgrade beträgt 41 und ist damit höher als die von MAHRU-1, vgl. Tabelle 2.5. Doch auch hier wurde der Aspekt der Ausführung menschlicher Bewegungen und damit die Möglichkeit zum Aussenden von Signalen der nonverbalen Kommunikation nicht berücksichtigt. Gerade im Schulter-, Torso- und

Hüftbereich liegen starke Defizite vor, die es verhindern naturgetreue menschliche Bewegungen auszuführen.

Tabelle 2.5: Freiheitsgrade (DoF) beim humanoiden Roboter KHR-2.

| <i><b>Körperteile (einzeln)</b></i> | <i><b>Freiheitsgrade</b></i> |
|-------------------------------------|------------------------------|
| Auge                                | 2 DoF x 2                    |
| Schulter                            | 3 DoF x 2                    |
| Ellbogen                            | 1 DoF x 2                    |
| Handgelenk                          | 2 DoF x 2                    |
| Finger                              | 1 DoF x 10                   |
| Nacken                              | 2 DoF                        |
| Bein                                | 6 DoF x 2                    |
| Torso                               | 1 DoF                        |
| <b>Insgesamt:</b>                   | <b>41 DoF</b>                |

Aufgrund der strukturellen Defizite werden diese Prototypen immer starke Unterschiede zu den tatsächlichen Bewegungen eines Menschen aufweisen. Dies mag auch der Grund sein, weswegen die Entwickler dieser Prototypen kein allzu menschliches Aussehen anstreben. Defizitäre Bewegungsabläufe werden nach Frey (1999) einem weniger menschlich aussehenden Roboter oder Avatar leichter verziehen, als einem stark menschlich aussehenden.

Wird nun ein sehr menschenähnlicher Roboter mit den selben Bewegungsmustern wie MAHRU-1 akuiert, treten schnell jene von Mori beschriebenen Probleme des *Uncanny Valley* bzw. des *Vally of Earines* wie es von K. F. MacDorman genannt wurde, auf, vgl. MacDorman (2006). Eine Ursache hängt unter anderem damit zusammen, dass der Mensch im Laufe der Evolution ein Experte in der Analyse und Auswahl anderer Menschen geworden ist. So ist er beispielsweise unterbewusst in der Lage, den am besten passenden Partner zur Fortpflanzung auszuwählen. Können nun diese von S. Frey untersuchten Signale der nonverbalen Kommunikation nicht vermittelt werden, so läuft ein evolutionär geprägtes „Verfahren“ beim Betrachter eines sehr menschenähnlichen humanoiden Roboters ab. Positive Sympathiewerte können demzufolge nach menschlichen Maßstäben nicht erreicht werden. Signale wie die Lateralflexion des Kopfs können beim Roboterprototyp MAHRU-1 aufgrund der defizitären Mechanik schlichtweg nicht ausgeführt werden. Dies wirkt sich auf genau jene elementaren Aspekte der nonverbalen Kommunikation negativ aus, so dass eine hohe Akzeptanz eines solchen Roboters durch den Menschen, nur schwer zu erreichen ist.

### Gemenoid Actroid-F

Um die Defizite des *Uncanny Valley* zu überwinden, wurden schon frühzeitig spezielle Roboterprototypen entwickelt, die menschliches Verhalten und Aussehen in Perfektion imitieren sollten. Die Bezeichnung dieser RoboterGattung lautet *Gemenoid*. Diese

homanoiden Klone sind in der Regel einer real existierenden Person nachempfunden. Ein typischer Vertreter dieser Gattung ist Actroid-F, entwickelt von Hiroshi Ishiguro<sup>7</sup>, vgl. Abb. 1.5, vgl. Yoshikawa u. a. (2011). Dieser besitzt im Oberkörper insgesamt 12 Freiheitsgrade und interagiert sitzend mit seiner Umgebung. Die Aktuatoren werden pneumatisch über einen Kompressor betrieben, der über einen 10m langen Anschluss mit dem Roboter verbunden ist. Die Entkopplung der Aktuatoren dient einerseits der Vermeidung von Geräuschemissionen, andererseits der Installation vieler Aktuatoren auf engstem Raum, wie dies im Gesichtsbereich eines Menschen von Nöten ist. Alleine 7 Freiheitsgrade entfallen bei Actroid-F auf die Bewegung der Gesichtsmuskulatur, vgl. Abb. 1.5. Hierdurch soll der Roboter in der Lage sein, Emotionen widerzuspiegeln und beispielsweise als Nachrichtensprecher, in der Telemedizin oder als Empfangsdame zu fungieren.

Tabelle 2.6: Systemspezifikationen von Actroid-F, vgl. Yoshikawa u. a. (2011).

|             |                                |
|-------------|--------------------------------|
| Abmessungen | W:600, D:800, H:140cm          |
| Gewicht     | Körper: 30kg, Kompressor 50kg  |
| Aktuatoren  | pneumatisch                    |
| DoF         | Augenbrauen (hoch, runter)     |
|             | Augenbrauen (zusammenkneifen)  |
|             | Augenlider (öffnen, schließen) |
|             | Augäpfel (pan)                 |
|             | Augäpfel (tilt)                |
|             | Mund (öffnen,schließen)        |
|             | Mundwinkel (heben,senken)      |
|             | Kopf (wenden)                  |
|             | linker Nackenbereich (beugen)  |
|             | rechter Nackenbereich (beugen) |
|             | Brust                          |
|             | Bauch (vorbeugen)              |

Tabelle 2.6 stellt die genauen Systemspezifikationen von Actroid-F dar. Trotz der Anstrengungen über die Entkopplung der Aktuatoren, der präzisen Modellierung eines menschlichen Gesichts über ein Silikonmodell und der Implementierung einer realistischen Sprachausgabe, empfanden in einer Probandenstudie einige Teilnehmer den Roboter in ihrer Umgebung als unangenehm und bevorzugten lieber dessen Abwesenheit, vgl. Yoshikawa u. a. (2011). Teilweise wurde der Roboter auch als mechanisch charakterisiert, obwohl sein Äußeres dem eines Menschen zum Verwechseln ähnlich sieht. Dies lässt den Schluss zu, dass allein die Gesichtsmuskulatur nicht ausreichend ist, Signale der Humankommunikation ausreichend genau zu übertragen. Die Darstellungen von Emotionen und Handlungen muss über mehr als nur die Gesichtsmimik gesche-

<sup>7</sup> Direktor des Intelligent Robotics Laboratory am Department of Adaptive Machine Systems, der Universität Osaka, Japan.

hen. Vielmehr müssen Hände, Schultern und Arme zusätzliche Signale der nonverbalen Kommunikation aussenden, um entsprechend wahrgenommen zu werden. Pneumatische Aktuatoren lassen keinen großen Bewegungsspielraum zu. Dies lässt darauf schließen, dass menschliche Bewegungen hier nicht korrekt in ihrer gesamten Ausprägung und Dynamik dargestellt werden können. Bewegungen eines humanoiden Roboters können vom Menschen unterbewusst nur dann korrekt interpretiert werden, wenn ein menschlicher Bewegungsvorgang in seiner gesamten Dynamik komplett abgebildet werden kann. Kann dies aus technischen Gründen nicht geschehen, wird ein solcher Prototyp immer wieder als mechanisch und nicht sympathisch charakterisiert werden, was augenscheinlich bei Actroid-F immer noch der Fall ist.

Insgesamt kann über diverse Beispiele der einschlägigen Fachliteratur gezeigt werden, dass aktuelle Modelle humanoider Roboter und Gemenoiden eine Reihe an Nachteilen besitzen, die diese daran hindern, entsprechende Signale der nonverbalen Kommunikation an einen menschlichen Interaktionspartner auszusenden. Über zu kleine Bewegungswinkel und fehlende Freiheitsgrade, können jene für die nonverbale Kommunikation notwendige Bewegungen, nur unzureichend repräsentiert werden. Darüber hinaus besitzen die verwendeten Aktuatoren konventioneller Prototypen schlechte dynamische Eigenschaften. Die Aktuatorik verhindert es, dass die menschliche Bewegungsdynamik abgebildet werden kann. Außerdem können in konventionellen konstruktiven Ansätzen Dimensionen und Abmessungen des menschlichen Körpers auf Grund technischer und struktureller Restriktionen oftmals nicht eingehalten werden. Hierdurch wird das Aussehen soweit verfremdet, dass die Menschenähnlichkeit des Prototyps nicht mehr gegeben ist.

## 2.9 Neubetrachtung des Entwicklungsprozesses humanoider Roboterprototypen

Die dargestellten Prototypen KHR-2, MAHRU-1 und Actroid-F stellen stellvertretend für die Vielzahl existierender Roboter, typische Defizite im Bereich der Humankommunikation dar. Mit dem Ziel, spezifische Schwachstellen bezüglich der Repräsentation menschlicher Bewegungen zu beseitigen, wurden einige elementare Einflussfaktoren identifiziert, die für die korrekte Ausführung menschlicher Bewegungsabläufe in der humanoiden Robotik von entscheidender Bedeutung sind:

- Anzahl der verfügbaren Freiheitsgrade (DoF)
- Anordnung und Positionierung der Manipulatorgelenke
- Abmessungen und Masse der Effektoren
- Antriebsleistung der Aktuatoren
- Getriebe- und Gelenkspiel
- Trajektorienplanung und Steuerung
- Strukturelle Elastizität des Funktionsskeletts
- Sensorauflösung und Geschwindigkeit der Datenübertragung

- Generelle Struktur der Software-/Hardwarearchitektur

Einflussfaktoren wie diese, werden nun Teil eines übergeordneten Entwicklungsansatzes. Dieser soll es ermöglichen, auf Grundlage eines neuartigen Konzepts zur integrierten Entwicklung humanoider Roboter, die Anforderungen an den späteren Prototyp vom Systementwurf bis zur Systemintegration zu definieren. Der Ablauf des Entwicklungsprozesses ist in Abbildung 2.11 dargestellt. Im unteren Abschnitt der Konzeptgrafik befinden sich die jeweils angewandten Methoden und Entwicklungsansätze des kompletten Entwicklungsprozesses bis hin zum fertigen Prototyp. Im ersten Schritt wurden die grundsätzlichen Konzeptanforderungen definiert. Es wurde bereits aufgezeigt, dass Bewegungen wie die laterale Kopfflexion oder die Bewegungen der Schulter auf die Aspekte der Humankommunikation grundlegenden Einfluss haben. Darüber hinaus wurden in Abschnitt 2.7 die Geschwindigkeits- und Beschleunigungsanforderungen an die Aktuatoren zur Ausführung menschlicher Bewegungen festgelegt.

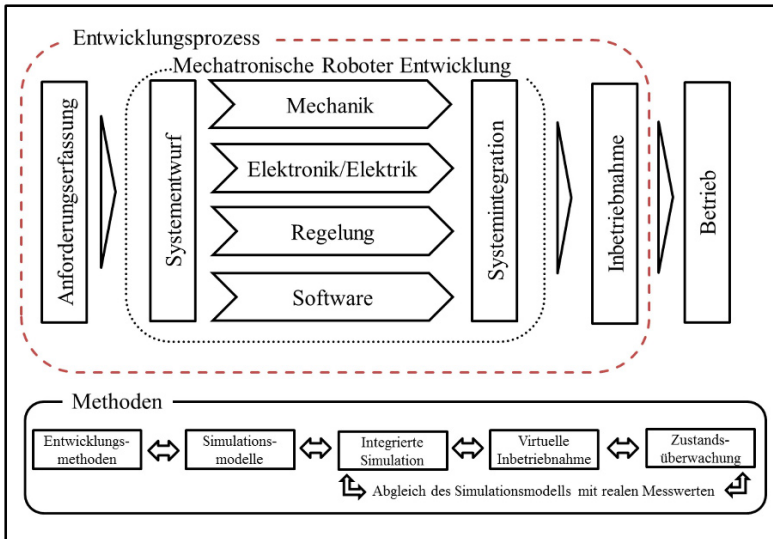


Abbildung 2.11: Mechatronischer Entwicklungsansatz in der humanoiden Robotik.

Nun gilt es, das Bewegungsskelett des humanoiden Roboters grundlegend neu zu gestalten. Zur Modellierung sind hierfür die Bewegungsskelette virtueller Avatare aus dem Bereich der 3D-Animation besonders gut geeignet. Über den Umweg eines abstrahierten Modells, können physikalische Kräfte, Aktuatoren und Platzprobleme außen vorgelassen werden.

Der Schritt der Anforderungsdefinition an das Funktionsskelett wird im folgenden über einen Reengineering-Ansatz zur Generierung der Skelettstruktur des humanoiden Ro-

boters erweitert. Mithilfe des hieraus entstandenen Modells wird anschließend ein Vergleich des menschlichen mit dem Arbeitsraum des humanoiden Roboters durchgeführt. Im Anschluss wird eine Verifizierung des dynamischen Bewegungsverhaltens simulativ durchgeführt und das resultierende Funktionsskelett am späteren Prototypen implementiert. Matlab Simulink<sup>®</sup> und Dymola<sup>®</sup> werden hierbei zur theoretischen Modellbildung und für den integrierten Simulationsprozess verwendet.

Die Inbetriebnahme geschieht in einem späteren Schritt am eigentlichen Funktionsprototypen. Über den Abgleich des Simulationsmodells mit realen Werten (Hardware-in-the-Loop) aus Untersuchungen des menschlichen Bewegungsapparats wird die fortlaufende Zustandsüberwachung des Steuerungsmodells sichergestellt. Die Schwierigkeit besteht hier in der Interdisziplinarität des Gesamtprojekts. Konzepte der Humankommunikation, der Biomechanik, der Softwareentwicklung, der Mechanik, der Elektrotechnik und der Regelungstheorie werden vereint und zu einem komplexen Gesamtsystem kombiniert. Die jeweiligen Stufen des Entwicklungsprozesses aus Abbildung 2.11 werden hierbei von den passenden Methoden von der Systementwicklung bis zur Zustandsüberwachung unterstützt.

Entwicklung des seilgetriebenen, humanoiden  
Roboterprototyps HUMECH  
Eine Verifizierung von Aspekten der nonverbalen  
Kommunikation  
Feldmann, S.  
2017, XXVIII, 246 S. 112 Abb., 56 Abb. in Farbe.,  
Softcover  
ISBN: 978-3-658-18789-7