



**INF**

Studiengang  
Medien- und  
Kommunikationsinformatik



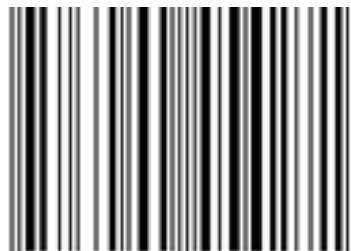
**Hochschule Reutlingen**  
Reutlingen University

Uwe Kloos, Natividad Martínez, Gabriela Tullius (Hrsg.)

# **Informatics Inside: Human-Centered Computing**

Informatik-Konferenz an der Hochschule Reutlingen  
30. April 2014

ISBN 978-3-00-045427-1



9 783000 454271 >

# Impressum

## **Anschrift:**

Hochschule Reutlingen  
Reutlingen University  
Fakultät Informatik  
Medien- und Kommunikationsinformatik  
Alteburgstraße 150  
D-72762 Reutlingen

Telefon: +49 7121 / 271-4002

Telefax: +49 7121 / 271-4042

E-Mail: [infoinside@reutlingen-university.de](mailto:infoinside@reutlingen-university.de)

Internet: <http://www.infoinside.reutlingen-university.de>

## **Organisationskomitee:**

Prof. Dr. Gabriela Tullius, Hochschule Reutlingen  
Prof. Dr. Natividad Martínez, Hochschule Reutlingen  
Prof. Dr. Uwe Kloos, Hochschule Reutlingen

André Antakli  
Thomas Bauer  
Olaya De la Rosa Avitia  
Matthias Gutekunst  
Viktoria Hoffmann  
Johannes Kartheininger  
René Mangold  
Stanislas Mauser  
Lars Schneider  
Arkadius Weister  
Anna Wellerdiek



**Hochschule Reutlingen**  
Reutlingen University

Copyright: © Hochschule Reutlingen, Reutlingen 2014

Herstellung und Verlag: Hochschule Reutlingen

ISBN 978-3-00-045427-1

# Inhaltsverzeichnis

## Gestenerkennung & Augumented Virtuality

---

### **Thomas Bauer**

*Anforderungsanalyse zur computergestützten Erkennung der Deutschen Gebärdensprache.....* 8

### **Matthias Gutekunst**

*Augmented Virtuality zur Steigerung der Immersion in virtuellen Umgebungen.....* 26

### **Stanislas Mauser**

*Analysis of Finger- and Palm-based interaction paradigms for Touch-Free Gesture-Based Control of Medical Devices with the Leap Motion Controller.....* 34

## Softwaretechnik

---

### **René Mangold**

*Selektion von Szenarien zur Optimierung von Simulationen im präventiven Krisenmanagement.....* 46

### **Arkadius Weister**

*Language Oriented Programming: Modulare domänenspezifische Sprachen.....* 54

## Entwicklung Mobiler Anwendungen

---

### **Olaya De la Rosa Avitia**

*Strategy to Test Mobile Apps.....* 70

### **Viktoria Hoffmann**

*Optimierung der Usability von digitalen Fahrtenbüchern durch automatisches Erfassen von fahrzeugspezifischen Daten.....* 80

### **Johannes Kartheininger**

*Vergleich der Single Sign On Verfahren SAML und OpenID Connect.....* 92

## Virtuelle Welten

---

### **André Antakli**

*Umgebungswahrnehmung von agentenbasierten simulierten Menschmodellen in virtuellen Welten im Kontext C3D.....* 100

# Anforderungsanalyse zur computergestützten Erkennung der Deutschen Gebärdensprache\*

Thomas Bauer  
Reutlingen University  
Thomas.Bauer@Student.  
Reutlingen-University.DE

## Abstract

Im Rahmen der wissenschaftlichen Vertiefung an der Hochschule Reutlingen befasst sich diese Arbeit mit der Untersuchung der Anforderungen und der Machbarkeit zur computergestützten Erkennung der Deutschen Gebärdensprache - im Folgenden DGS - und des deutschen Fingeralphabets. Die Erkenntnisse aus dieser Arbeit dienen als Grundlage zur Entwicklung eines Systems zur Übersetzung von Gebärden der DGS oder des Fingeralphabets in die deutsche Schriftsprache. Zunächst werden grundlegende Informationen zu Geschichte, Aufbau und Grammatik der DGS und des Fingeralphabets aufgeführt. Die Erkennung der Gebärden soll durch optische Bewegungssensoren erfolgen. Hierfür werden unterschiedliche Sensortypen betrachtet und verglichen. Im weiteren Verlauf erfolgt die Analyse der benutzerspezifischen und technischen Anforderungen. Erstere basieren auf der Befragung einer Fokusgruppe aus gehörlosen und hörenden Menschen aus dem Bereich der Gehörlosen-, Schwerhörigen- und Sprachbehindertpädagogik. Abgeleitet aus den Informationen der Anforderungsanalyse ergibt sich, bis zu einem gewissen Grad, die Machbarkeit aus technischer und benutzerspezifischer

Sicht. Abschließend erfolgen die Zusammenfassung der Anforderungen, welche an das zu entwickelnde System gestellt werden sowie eine Handlungsempfehlung für die Entwicklung eines Prototyps.

## Schlüsselwörter

Gestenerkennung, Deutsche Gebärdensprache, Fingeralphabet, Übersetzung, Dolmetscher, Grammatik, Anforderungsanalyse, Prototyp, Bewegungssensor, Asus Xtion, PrimeSense Carmine, Leap Motion

## CR-Kategorien

A.1 [Introductory and survey]; D.2.1 [Requirements/Specifications]: Elicitation methods (e.g., rapid prototyping, interviews, JAD); H.1.2 [User/Machine Systems]: Human information processing; I.4.8 [Scene Analysis]: Sensor fusion, Tracking

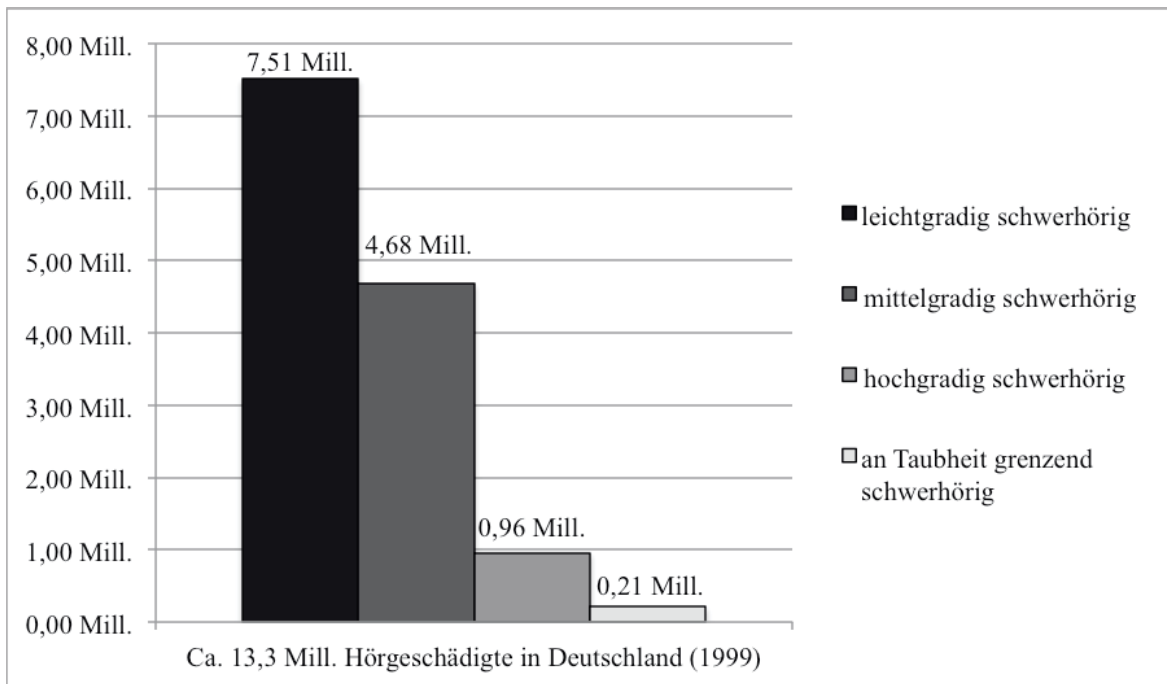
## 1 Thematische Einführung

Neben der Linguistik, welche sich bereits seit Jahrhunderten mit der exakten wissenschaftlichen Erforschung der menschlichen Sprache befasst, charakterisiert die Computerlinguistik (CL) seit Mitte des 20. Jahrhunderts das Fachgebiet der maschinellen Verarbeitung natürlicher Sprache. Sie orientiert sich dabei an den linguistischen Teilbereichen der Phonetik und Phonologie, der Morphologie, der Syntax und Semantik sowie der Pragmatik (vgl. [2], S. 1ff). Ein Anwendungsbereich der Computerlinguistik stellt die maschinelle Übersetzung (MÜ) oder auch Machine Translation (MT) dar.

\*

Betreuer Hochschule: Prof. Dr. rer. nat. Gabriela Tullius  
Hochschule Reutlingen  
Gabriela.Tullius@Reutlingen-  
University.de

Informatics Inside 2014  
Wissenschaftliche Vertiefungskonferenz 2014  
30. April 2014, Hochschule Reutlingen  
Copyright 2014 Thomas Bauer



**Abbildung 1: Verteilung der Hörschädigung nach Sohn**

*Quelle: vgl. [15], Abbildung eigen*

Die fortschreitende Globalisierung erfordert zunehmend zeitnahe Übersetzungen unterschiedlicher digitaler Dokumente wie z.B. von Handbüchern, Gebrauchsanweisungen oder Websites. Dabei behandelt die zugrunde liegende Übersetzung die Übertragung eines Textes aus einer Quellsprache in die jeweilige Zielsprache. Das Verfahren der maschinellen Übersetzung aus der Computerlinguistik soll hierbei unterstützend eingesetzt werden, indem beispielsweise Übersetzungsprozesse sowie Sprachein- und ausgaben durch eine Recheneinheit vollautomatisch analysiert und übernommen werden. Damit vereint die maschinelle Übersetzung Einflüsse aus der Linguistik, Mathematik, Informatik, Übersetzungswissenschaft und der Elektrotechnik (vgl. [2], S. 642ff).

Auch für Menschen mit Behinderung kann die maschinelle Übersetzung unterstützend eingesetzt werden. So besteht bereits die Möglichkeit der Ein- und Ausgabe am Computer für sehbehinderte bzw. blinde Menschen mittels Lautsprache. Die automatische Spracherkennung (Automatic Speech Recognition - ASR) liefert diesbezüglich bereits gute und brauchbare Ergebnisse, dennoch

steigt die Fehlerquote durch unterschiedliche Faktoren an. So erschweren bspw. Störgeräusche, ein umfangreiches Vokabular oder aber auch Konstruktionsabbrüche, Dialekte und Akzente die korrekte Spracherkennung (vgl. [9], S. 15). Braillezeilen und Brailletastaturen hingegen sind vergleichbar mit einem Monitor und einer Tastatur und somit nahezu fehlerlos. Die Fehleranfälligkeit wird hier durch den Faktor Mensch bestimmt. Die Verarbeitung der maschinellen Übersetzung bezieht sich in diesem Falle ausschließlich auf die Substitution der einzelnen Schriftzeichen eines Textes, da für jedes dieser Zeichen ein passendes Pendant in Brailleschrift existiert und grammatikalische Unterschiede, wie bei verschiedenen Sprachen, nicht existieren.

Ein weitaus komplexerer Anwendungsfall ergibt sich bei Menschen mit einer starken Hörschädigung sowie gehörlosen Menschen. Dabei steht vor allem die Kommunikationsunterstützung im Vordergrund. Der Deutsche Schwerhörigenbund e.V. geht derzeit von rund 80 000 gehörlosen und rund 16 Millionen schwerhörigen Menschen in Deutschland aus (vgl. [5]). Diese Zahlen beruhen auf einer Schätzung, der die wissenschaftliche

Untersuchung „Hörscreening“ von Dr. med. Wolfgang Sohn aus dem Jahre 1999 zugrunde liegt. Dabei wurden Daten von ca. 2000 Probanden aus Deutschland im Alter über 14 Jahren ermittelt und ausgewertet (vgl. [15]). Das Diagramm in Abbildung 1 zeigt die Verteilung nach Grad der Schwerhörigkeit zum Zeitpunkt der Datenerhebung.

Als leichtgradig schwerhörig wird ein mittlerer Hörverlust bis 35dB (Dezibel) bezeichnet, wobei als Faustregel davon ausgegangen werden kann, dass eine Abnahme des Schalldrucks um ca. 10 dB einer Lautstärkenhalbierung entspricht. Bei einem Hörverlust von 30dB entspräche dies etwa  $\frac{1}{8}$  des ursprünglichen Schalldrucks. Mittelgradig schwerhörig entspricht einem mittleren Hörverlust zwischen 40 – 65 dB, hochgradig bereits 70 – 85 dB. Ab einem mittleren Hörverlust von 90 dB und größer gilt eine an Taubheit grenzende Schwerhörigkeit (vgl. [15]). Vielen der Betroffenen bleibt für die Kommunikation, aufgrund ihrer Behinderung, ausschließlich der Schriftverkehr oder die Gebärdensprache. Im Gegensatz zur Brailleschrift stellen die Gebärdensprachen eigenständige Sprachen mit eigener Grammatik dar. Die Anforderungen an eine maschinelle Übersetzung der Gebärdensprache sind daher weitaus höher gesetzt und nicht so einfach zu realisieren, da zunächst die Erkennung der Gebärden und im weiteren Verlauf die grammatikalische Übersetzung erfolgen muss.

## ***1.1 Motivation und Stand der Technik***

Verschiedene Möglichkeiten wie z.B. Dolmetscher, Ferndolmetschdienste oder aber auch mobile Gebärdensprachlexika ermöglichen bzw. erleichtern die Kommunikation zwischen Gehörlosen und Hörenden. Dennoch ergeben sich daraus diverse Einschränkungen. So ist der Einsatz eines Dolmetschers oder Ferndolmetschdienstes immer von einer realen Person abhängig, welche die eigentliche Übersetzung durchführt. Des Weiteren ist die Nutzung solcher Dienste mit relativ hohen Kosten verbunden. Diese werden jedoch in

der Regel von einem Kostenträger übernommen und belasten somit nicht den Anwender selbst. Gebärdensprachlexika erlauben es dem Benutzer einzelne Begriffe und deren Gebärden nachzuschlagen, jedoch wird hier keine Übersetzung ganzer Sätze angeboten. Eine computergestützte Übersetzung könnte dabei eine höhere Verfügbarkeit für den Anwender und ggf. gleichzeitig eine Kostenreduktion für den Kostenträger bedeuten.

Es gibt einzelne Ansätze zur computergestützten Erkennung und Übersetzung von Gebärdensprachen. Sowohl Eingabetechniken als auch die Methoden zur Erfassung und Übersetzung der Gebärdensprache oder des Fingeralphabets variieren hier. In früheren Ansätzen fanden für die Eingabe meist Datenhandschuhe Verwendung. So beschreiben Jangquin et al. in [11] bereits 1998 einen Ansatz zur Erkennung der Chinese Sign Language (CSL) unter Verwendung eines Datenhandschuhs. Mittlerweile werden jedoch meist rein visuelle Ansätze verfolgt (vgl. [3], S. 1). Diese reichen von der einfachen Webkamera bis hin zu dreidimensionalen Tiefenkameras wie z.B. die Microsoft Kinect. So beschreiben Dos Santos Anjo et al. in [6] einen Ansatz zur Erkennung einiger statischer Gesten des Fingeralphabets der Brazilian Sign Language (Libras) mit Hilfe der Kinect.

Ebenfalls auf Basis der Microsoft Kinect Technologie entwickelten Forscher der Chinese Academy of Science und der Beijing Union University in Zusammenarbeit mit Microsoft Research Asia einen Prototyp zur Gebärdensprachübersetzung. Der Software-Prototyp VisualComm ermöglicht die Kommunikation zwischen gebärdensprachorientierten und lautsprachenorientierten Menschen sowie eine Übersetzung zwischen unterschiedlichen Gebärdensprachen (vgl. [4], S. 1ff). Den Entwicklern war es wichtig, eine Zweiwegekommunikation anzubieten. So werden einerseits Gebärden in Text oder Sprache übersetzt und andererseits Wörter und Sätze durch einen 3D Avatar in Gebärdensprache dargestellt. Zur Eingabe der

Gebärden wird der optische Tiefensensor Kinect verwendet. Mit einem Wortschatz von 370 chinesischen Gebärden und einer Erfassungsquote von ca. 95% beweist dieser Ansatz, dass die Erkennung gestenbasierter Sprachen mit bereits verfügbaren und kostengünstigen 2D und 3D-Sensoren möglich ist (vgl. [3], S. 1f).

Im Vergleich hierzu bietet das britische Unternehmen Technabling Ltd. bereits ein plattformunabhängiges Softwaresystem zur Echtzeitübersetzung gestenbasierter Sprachen. Der Einsatz des Portable Sign Language Translator (PSLT) soll auf den meisten unterschiedlichen Geräten wie z.B. Smartphones, Tablets und Laptops möglich sein. Voraussetzung ist eine einfache Kamera um die jeweiligen Gesten aufzuzeichnen. Aktuell beschreibt Technabling Ltd. die Möglichkeit einer Übersetzung in geschriebenen Text aus der British Sign Language (BSL), der Chinese Sign Language (CSL) und aus Makaton<sup>1</sup> (vgl. [20]).

Die Ansätze von Microsoft und Technabling Ltd. wurden gegen Ende 2013 veröffentlicht und befinden sich derzeit in unterschiedlichen Entwicklungsstufen. Dies lässt auf das noch junge Thema der Gebärdenspracherkennung und -übersetzung schließen, welches weiterhin viel Raum für Forschungsmöglichkeiten bietet.

## 1.2 Zielsetzung

Im Zuge einer Master-Thesis soll ein Prototyp zur computergestützten Übersetzung der Deutschen Gebärdensprache und/oder des deutschen Fingeralphabets entstehen. In einem ersten Schritt werden daher in dieser Arbeit sowohl erste benutzerspezifische als auch technische Anforderungen ermittelt und analysiert. Aus den daraus gewonnen Informationen lässt sich die Machbarkeit des Projekts ableiten. Hinsichtlich der benutzerspezifischen

---

<sup>1</sup>Makaton ist ein Förderansatz welcher ein Kernvokabular von Alltagsbegriffen in Form von Gebärden und Symbolen bereitstellt (vgl. [13]).

Anforderungen werden sowohl Gehörlose und Schwerhörige als auch Hörende in einem Interview befragt. Hier gilt es unter Anderem die Relevanz festzustellen, ob eine Übersetzung der Deutschen Gebärdensprache und/oder des deutschen Fingeralphabets erfolgen soll. Wie bereits in Abschnitt 1.1 aufgeführt, werden mittlerweile meist nur visuelle Eingabetechniken für die Erfassung der Gebärdensprache verwendet. Daher werden für die technischen Anforderungen die derzeit aktuellen optischen Bewegungssensoren Leap Motion, Asus Xtion Pro Live sowie die PrimeSense Carmine 1.09 verglichen und auf ihre Einsatztauglichkeit untersucht.

## 2 Gebärdensprache und Fingeralphabet

In den folgenden Abschnitten werden die Grundlagen und Besonderheiten der Deutschen Gebärdensprache und des deutschen Fingeralphabets aufgeführt.

### 2.1 Deutsche Gebärdensprache (DGS)

Woher Gebärdensprachen stammen und wann sie entstanden sind, lässt sich nicht genau erklären. Dies ist darauf zurückzuführen, dass in den vergangenen Jahrhunderten gehörlose, schwerhörige und sprachbehinderte Menschen meist unterdrückt und als Minderheit ausgeschlossen wurden, weshalb es auch nur wenig Zeugnisse dieser Zeitgeschichte gibt. Es wird jedoch angenommen, dass Gebärdensprachen in einfacher Form schon vor der Entstehung von Lautsprachen verwendet wurden (vgl. [10], S. 4). Durch die Entwicklung und Verbreitung der Lautsprachen wurden die Gebärden nur noch in abgeschwächter Form zur Unterstützung eingesetzt oder zum Teil ganz verdrängt. Erst Mitte des 16. Jahrhunderts erkannte der Mönch Pedro Ponce de Leon, dass die bis dahin als „geistarm“ geltenden gehörlosen und sprachbehinderten Menschen ebenfalls lernfähig sind. Er unterrichtete Jungen in Sprechen, Schreiben und Lesen. Bis ins späte 19. Jahrhundert entwickelten sich aus dieser Kenntnis die ersten Schulen für Gehörlose

in Frankreich, den USA und Deutschland. 1880 erfolgte der Umbruch beim Mailänder Kongress. Insgesamt 255 Lehrer aus Europa und Amerika – darunter nur zwei Gehörlose – entschieden über die Entwicklung des Gehörlosen-Unterrichts und verboten ihn. Erst 1988 wurde durch das Europäische Parlament der Entschluss zur Anerkennung nationaler Gebärdensprachen gefasst. Seit dem 1. Mai 2002 ist die Gebärdensprache in der Bundesrepublik Deutschland als eigenständige Sprache anerkannt (vgl. [14]).

Gebärdensprache ist nicht gleich Gebärdensprache. Oftmals wird davon ausgegangen, dass sich gebärdensprachorientierte Menschen auf der ganzen Welt miteinander verständigen können. Dem ist jedoch nicht so. Wie bei den Lautsprachen haben sich kulturell unterschiedliche Sprachen und Dialekte entwickelt. In Deutschland spricht man daher von der Deutschen Gebärdensprache (DGS). Sie ist wie alle anderen Gebärdensprachen eine rein visuelle Sprache und besitzt eine eigene Grammatik. Neben den Armen und Händen werden vor allem das Gesicht, dessen Mimik und Mundbild sowie der gesamte Oberkörper zur Zeichenproduktion eingesetzt. Im sogenannten Gebärdenraum, welcher sich im Bereich des Oberkörpers befindet, werden die Gebärden in dreidimensionaler Form dargestellt. In seltenen Fällen kann der Gebärdenraum auch verlassen werden wie z.B. bei einer Form der Gebärde für Hund, bei welcher mit der flachen Hand zweimal auf den Oberschenkel geklopft wird (vgl. [7], S.34f).

Auch in Bezug auf die Grammatik unterscheidet sich die DGS von der deutschen Lautsprache. In beiden Sprachen werden zur Bildung eines einfachen aktiven Satzes mindestens die Satzglieder Subjekt (S) und Prädikat/Verb (V) benötigt – oft auch mit der Erweiterung durch ein Objekt (O). Ist unklar, ob die Sprache die Satzfunktionen Subjekt und Objekt aufweist, wird stattdessen von Agens (A) und dem Gegenüber Patiens (P) gesprochen. So auch bei der DGS. Die Anord-



**Abbildung 2: Gebärdenraum**  
Quelle: [1], S. 23

nung der Satzglieder der beiden Sprachen ist meist unterschiedlich. Während die deutsche Lautsprache in der Regel einer Anordnung von AVP bzw. SVO folgt und somit das Prädikat als bindendes Glied zwischen Agens und Patiens dient, steht das Prädikat in der DGS meist am Ende eines Satzes (vgl. [7], S. 175ff).

**Lautsprache:**

Das Mädchen - putzt - das Auto

**DGS:**

Mädchen - Auto – putzen

Aber auch hier gibt es unterschiedliche Wortstellungsvarianten sodass, neben der grundlegenden Variante APV, auch die Varianten AVP oder PAV vorkommen können. Dies hängt von unterschiedlichen Faktoren ab (vgl. [7], S. 178).

Nicht nur die Syntax beeinflusst den grammatikalischen Unterschied der DGS zur deutschen Lautsprache. Beispielsweise existieren, wie im obigen Beispiel zu sehen, auch keine Artikel (der, die, das) oder Kopulae<sup>2</sup>. Der grammatikalische Aufbau setzt

<sup>2</sup>Kopula ist eine Verbform von „sein“, „bleiben“ und „werden“, welches Subjekt und Prä-



sich zudem aus unterschiedlichen Elementen zusammen. So spielen neben dem Produktionstempo und der Modalitätsunterschiede der Gebärden auch die Nutzung des Gebärdenraums sowie unterschiedliche Formen der Hände, des Oberkörpers oder der Mimik eine elementare Rolle in der Grammatik einer visuell-gestischen Sprache. Im Folgenden werden daher nur die einzelnen Bereiche gestreift, welche für die vorliegende Arbeit relevant sind. Eine ausführliche Beschreibung des grammatikalischen Unterschieds der DGS zur deutschen Lautsprache bieten Boyes, Braem und Penny in [1].

Modalitätsunterschiede beziehen sich primär auf die lineare Anordnung der gesprochenen Sprache im Vergleich zur räumlichen Anordnung der Gebärdensprache und sind eng mit dem Produktionstempo verknüpft. Ein Versuch von Bellugi und Fischer aus dem Jahr 1972 vergleicht dabei die Geschwindigkeiten der beiden Sprachen bezogen auf einen vorgegebenen Text im Vergleich zum Ausdruck eines eigens gefassten Gedanken bzw. Satzinhaltes. Die Ergebnisse, dargestellt in Tabelle 1, überraschten die Forscher. Während beim vorgegebenen Text das Produktionstempo der Lautsprache annähernd doppelt so schnell war wie bei der Gebärdensprache ist das Tempo um den Inhalt in eigenen Worten auszudrücken annähernd gleich (vgl. [1], S. 46f). Die Annahme der Forscher, dass Satzinhalte in Lautsprache schneller zu übermitteln sind als in Gebärdensprache wurde damit widerlegt.

**Tabelle 1: Vergleich Produktionstempo**

*Quelle: ([1], S. 47)*

*nach Bellugi und Fischer 1972*

Gebärdensprache	Lautsprache
2,37 Gebärden/Sek.	4,7 Wörter/Sek.
1,27 Sek./Satzinhalt	1,47 Sek./Satzinhalt

dikat zu einer Aussage verbindet. So ist bei dem Satz „Er ist krank“ das Wort „ist“ als Koppula anzusehen. In der DGS würde somit nur „Er krank“ bestehen (vgl. [1], S. 45).

Wie ist es dennoch möglich, dass bei einer zeitlich doppelt so langen Produktion der Gebärden, Satzinhalte in gleicher oder sogar schnellerer Zeit übermittelt werden können? Dies liegt am Aufbau und der Form der Sprache. Während die Lautsprache durch eine sukzessive Organisation und Verkettung linguistischer Formen einen grammatikalischen Aufbau in linearer Form verfolgt, können in der Gebärdensprache durch die räumliche Anordnung grammatikalische Formen gleichzeitig produziert werden. Auch die Wahrnehmung spielt dabei eine große Rolle. Das menschliche Gehör kann einzelne aufeinanderfolgende Elemente einfacher erfassen als eine Überlagerung der Elemente. Die visuelle Wahrnehmung hingegen erlaubt die simultane Erfassung mehrerer Elemente. So können bei der Gebärdensprache alle Komponenten (Handform, Bewegung, Mimik) gleichzeitig wahrgenommen und verarbeitet werden. Je nach Ausführung bzw. Abwandlung einer Gebärde können dadurch unterschiedliche Inhalte übermittelt werden, für die in der Lautsprache mehrere Wörter notwendig wären. Folgende Abbildung zeigt ein Beispiel der Gebärdensprache in welchem zum Einen dargestellt wird, dass in manchen Fällen nur eine Gebärde für die Übermittlung eines ganzen Satzes benötigt wird und zum Anderen, dass die geringe Abwandlung der gleichen Gebärde eine etwas andere Information übermittelt (vgl. [1], S. 47ff).



**Abbildung 3: Modifizierung von Gebärden**

*Quelle: vgl. [1], S. 51*

Der dreidimensionale Gebärdenraum stellt hier ein unabkömmliches Werkzeug zur grammatikalischen und semantischen Satzbildung dar. So können unter Anderem auch Ausgangsort und Ziel einer Handlung oder aber auch Zeitangaben durch die Position und die Bewegungsrichtung der Gebärde im Gebärdenraum übermittelt werden (vgl. [1], S. 74). Allgemein lässt sich also sagen, dass die Gebärdensprache ein Konstrukt einzelner Gebärden ist und aus Bewegungen und Formen der Hände, der Arme, des Oberkörpers und des Kopfes besteht. Zudem müssen Mimik und Mundbild sowie die eigene Syntax berücksichtigt werden. Durch den Einsatz dieser Elemente bzw. durch die Modifikation einer Gebärde durch diese Elemente lassen sich unterschiedliche Bedeutungen eines Satzes ausdrücken.

## 2.2 Deutsches Fingeralphabet

Das Fingeralphabet stellt keine eigenständige Sprache dar. Es ist ein Teilbereich der Gebärdensprache und ermöglicht die Darstellung von Buchstaben mittels einzelner Handformen. So ist jedem Buchstaben des Alphabets eine eigene Handform zugewiesen. Meist dient das Fingeralphabet nur unterstützend zur Beschreibung von Eigennamen oder unklaren Wörtern in der DGS (vgl. [1], S. 146f).



**Abbildung 4: Beispiel deutsches Fingeralphabet**

Quelle: [www.visuelles-denken.de/Schnupperkurs3.html](http://www.visuelles-denken.de/Schnupperkurs3.html)

Das Buchstabieren von Wörtern erfolgt in der Regel wie bei der Gebärdensprache innerhalb des Gebärdenraumes vor der Brust. Es ist möglich ganze Sätze zu buchstabieren. Dies ist jedoch über einen längeren Zeitraum so-

wohl für den Produzent als auch für den Zuschauer anstrengend, sodass durchschnittlich nur 56% der gezeigten Buchstaben klar verstanden werden können (vgl. [17] zitiert nach [1], S. 146). Wie die Gebärdensprache selbst, unterscheiden sich Fingeralphabete über Ländergrenzen hinweg oft sehr stark. Das deutsche Fingeralphabet wird beispielsweise einhändig buchstabiert während das britische Fingeralphabet durch beide Hände dargestellt wird (vgl. [1], S. 146f).

## 3 Erfassung durch optische Bewegungssensoren

Bewegungssensoren ermöglichen im Allgemeinen die Erfassung von kinematischen Vorgängen innerhalb des zu überwachenden Raumes. Dies können Lageänderungen, Geschwindigkeiten oder auch Beschleunigungen sein. Für jede dieser Größen existieren spezielle Sensoren (vgl. [18], S. 182). Im Bereich der Mensch-Maschine-Interaktion werden Bewegungssensoren meist zur Erfassung von Gestik und Mimik eingesetzt. Dies setzt voraus, dass Position und Lageveränderungen von Händen, Kopf, Augen, Augenbrauen oder Lippen erkannt werden können. Bei der Bewegungserfassung von Händen und Fingern werden derzeit überwiegend optische Bewegungssensoren oder Datenhandschuhe eingesetzt. Für die Erfassung der Gesten der DGS und des Fingeralphabets eignen sich Datenhandschuhe. Sie bieten im Vergleich zu den optischen Systemen eine exaktere Erfassung der Anatomie und Bewegung von Händen und Fingern. Dies wird über Dehnungssensoren entlang der Finger und Hände sowie durch Beschleunigungssensoren ermöglicht. Dadurch lassen sich die Freiheitsgrade von Hand- und Fingergelenken sowie deren Bewegungen genau abbilden. Optische Bewegungssensoren hingegen können bei Verdeckungen keine zuverlässige Erfassung, wie sie bspw. bei einer Faust vorkommen können, ohne weiteres gewährleisten (vgl. [21], S. 1). Datenhandschuhe sind jedoch im Vergleich zu optischen Eingabegeräten umständlicher und nicht leicht adaptierbar. Im folgenden Abschnitt werden daher zwei Arten von optischen Bewegungssen-

soren in ihrem Aufbau und der Funktionsweise beschrieben, welche für die Gebärdenspracherkennung eingesetzt werden könnten. In Abschnitt 4 werden diese Sensoren auf ihre technischen Anforderungen und deren Einsetzbarkeit überprüft.

### 3.1 *Asus Xtion Pro Live / PrimeSense Carmine*

Sowohl die Asus Xtion Pro Live als auch die PrimeSense Carmine gehören zu der Familie der RGB-D Sensoren oder auch Tiefenkameras. Sie bieten die Möglichkeit einer dreidimensionalen Erfassung von Räumen und Objekten und unterscheiden sich prinzipiell nur im Design voneinander. Vergleichbar ist die Asus Xtion Pro Live mit der PrimeSense Carmine 1.08. Beide Kameras besitzen dieselben Spezifikationen und sind für Entfernungen zwischen 0,8 – 3,5 m ausgelegt. Parallel hierzu bietet PrimeSense noch zusätzlich die Carmine 1.09 für kurze Distanzen zwischen 0,35 – 1,4 m an (vgl. [16], S. 4).



**Abbildung 5: PrimeSense Carmine 1.09**

Quelle: [www.faceshift.com/wordpress/wp-content/uploads/2013/01/PrimeSense\\_sensor.jpg](http://www.faceshift.com/wordpress/wp-content/uploads/2013/01/PrimeSense_sensor.jpg)

Die Erfassung des Raumes und dessen Objekte wird mittels zweier integrierter CMOS-Sensoren sowie einer Infrarotlichtquelle realisiert. Dabei sendet der IR-Emitter ein codiertes Lichttraster in den Raum. Das reflektierende Raster wird von einem der beiden CMOS-Sensoren aufgezeichnet. Aus den so gewonnenen Daten lässt sich ein Tiefenbild erzeugen. Der zusätzliche CMOS-Sensor ermöglicht die Aufnahme eines einfachen Kamerafarbbildes (vgl. [16], S. 2).



**Abbildung 6: Erzeugung eines Tiefenbildes**

Quelle: [16], S. 2

Auf Basis der Daten des Tiefenbildes kann dann, beispielsweise durch Software, eine Identifikation von Personen und deren Bewegungen erfolgen, eine Klassifikation von Objekten, eine Messung der Größe oder des Volumens sowie die Lage von Wänden und Boden ermittelt werden (vgl. [16], S. 2). Mit dem OpenNI (Open Natural Interaction) Framework der gleichnamigen Organisation wird eine Schnittstelle geboten, über welche mit der Hardware kommuniziert und Daten ausgelesen werden können. OpenNI ist ein quelloffenes Projekt und kann somit durch weitere Organisationen oder auch einzelne Entwickler beliebig erweitert und verändert werden (vgl. [8], S. 9).

### 3.2 *Leap Motion*

Im Juli 2013 brachte das Startup-Unternehmen Leap Motion seinen gleichnamigen Bewegungssensor auf den Markt. Dahinter verbirgt sich eine kleine Box, in der Größe eines Feuerzeugs.



**Abbildung 7: Leap Motion Controller**

Quelle: [www5.pcmag.com/media/images/393268-leap-motion-controller.jpg?thumb=y](http://www5.pcmag.com/media/images/393268-leap-motion-controller.jpg?thumb=y)

Platziert zwischen Display und Anwender erlaubt der Controller eine dreidimensionale Bewegungs- und Gestenerkennung der Hände und Finger innerhalb eines ca. 60 cm hohen und rund 150° großen Feldes über dem Controller. Leap Motion bewirbt sein Produkt mit der Erkennung aller zehn Finger und einer Genauigkeit von 0,001 mm (vgl. [12]). Technische Details werden Seitens Leap Motion nicht veröffentlicht. Nach Informationen der Firma SparkFun Electronics<sup>3</sup> sind nur einfache elektronische Komponenten verbaut. Lediglich zwei CMOS Bildsensoren mit einer optischen Linse sowie drei Infrarot-LEDs sind Hauptbestandteil des Controllers. Die genaue Funktionsweise ist dabei nicht bekannt. SparkFun Electronics vermutet, dass die beiden Kameras auf jeden Fall zur stereoskopischen Bildverarbeitung dienen. Wird einer der beiden Sensoren verdeckt, funktioniert der Controller nicht mehr. Des Weiteren ist nur noch die USB Anschlussstechnik verbaut. Dies lässt darauf schließen, dass die Verarbeitung der Daten überwiegend vom Rechner selbst und der mitgelieferten Software erledigt werden muss (vgl. [19]). Über die Airspace Oberfläche lassen sich Anwendungen aus unterschiedlichen Bereichen verwalten. Diese können über den Airspace Store, überwiegend kostenpflichtig, bezogen werden. Aktuell stehen ca. 170 Anwendungen zur Verfügung, wobei der größte Teil auf die Kategorie „Games“

<sup>3</sup>SparkFun Electronics ist ein Elektronikhändler welcher neben Herstellung und Verkauf von MicroController-Entwicklerboards, neue Gadgets wie die Leap Motion hinsichtlich technischer Bestandteile und Funktionen untersucht (vgl. [19]).

fällt. Über die Leap Motion API ist es Entwicklern möglich, eigene Anwendungen für den Bewegungssensor zu implementieren.

## 4 Anforderungsanalyse

Im folgenden Abschnitt werden sowohl die benutzerspezifischen als auch die technischen Anforderungen hinsichtlich des zu entwickelnden Systems zur Gebärdenspracherkennung aufgeführt. Die benutzerspezifischen Anforderungen basieren dabei auf den Aussagen von betroffenen bzw. fachnahen Personen aus dem Bereich der Gehörlosen-, Schwerhörigen- und Sprachbehindertenpädagogik. Hierzu wurden Interviews mit den entsprechenden Personen durchgeführt. Diese eignen sich vor allem für die Ermittlung der funktionalen und nicht-funktionalen Anforderungen aus Benutzersicht. Die funktionalen Anforderungen beschreiben dabei, welche Aufgabe bzw. welches erwartete Verhalten an das System gestellt werden soll. Nicht-funktionale Anforderungen beziehen sich unter anderem auf die Qualität, Performance oder Bedienbarkeit des Systems. Hierbei hilft die Befragung der betroffenen und fachnahen Personen, da diese, als zukünftige Anwender, ein qualifiziertes Urteil bezüglich der benutzerspezifischen Anforderungen abgeben können.

Die technischen Anforderungen beziehen sich auf Auflösung, Geschwindigkeit und Bewegungsfreiheit der optischen Sensoren sowie auf das Produktionstempo der Gebärden. Hierfür wurden von den gehörlosen Personen jeweils zwei Sätze in eine Kamera gebärdet bzw. buchstabiert. Durch diese Aufnahmen erfolgt eine Zeitmessung der Gebärden pro Sekunde. Aus den gewonnenen Informationen der Anforderungsanalyse lässt sich bis zu einem gewissen Maß die Machbarkeit eines Systems zur Gebärdenspracherkennung bzw. zur Erkennung des Fingeralphabets ableiten. Daraus ergibt sich auch die abschließende Entscheidung, welcher Ansatz weiterhin in der Master-Thesis verfolgt wird.

## 4.1 Benutzerspezifische Anforderungen

Für die Ermittlung der benutzerspezifischen Anforderungen wurden zwei Interviews durchgeführt. Als Gesprächsgrundlage wurde ein Szenario in Form einer Alltagssituation vorgegeben. Dabei sollten sich die Interviewpartner einen Arztbesuch vorstellen, bei welchem kein Dolmetscher vorhanden ist und die Kommunikation zwischen Gehörlosem und Arzt direkt stattfindet. Im ersten Interview fand die Befragung mit einer hörenden Person statt. Die Fokusgruppe des zweiten Interviews bestand aus insgesamt drei gehörlosen Personen im Alter zwischen 30 – 60 Jahren. Dieses Interview entwickelte sehr schnell eine Eigen- dynamik und wurde zur Gruppendiskussion mit teils unterschiedlichen Ansichten.

Zusammengefasst sind alle befragten Personen in der Lage mit Hilfe der DGS zu kommunizieren, wobei es starke Unterschiede gibt, seit wann, wie lange und intensiv die DGS gelernt und eingesetzt wurde. Auch unter den Gehörlosen ist es möglich, dass diese nicht DGS-orientiert aufgewachsen sind und die Sprache teilweise erst spät erlernt haben. Grundsätzlich lässt sich jedoch sagen, dass die DGS für Gehörlose als Muttersprache anzusehen ist.

In Bezug auf die Selbsteinschätzung über die Kenntnis und Erfahrung von technischen Hilfsmitteln wie E-Mail, SMS, Mobiltelefon und Computer gehen alle Interviewpartner von einem guten bis sehr guten Kenntnisstand aus. Der Unterschied in der Nutzung von technischen Hilfsmitteln zwischen den gehörlosen und hörenden Personen beschränkt sich lediglich auf die Nutzung des Telefons.

### 4.1.1 DGS oder Fingeralphabet

In Bezug auf die Frage, ob eine Übersetzung der DGS in die deutsche Schriftsprache oder die Übersetzung des Fingeralphabets sinnvoller wäre, sind alle Interviewpartner gleicher Meinung. Hierbei steht die DGS eindeutig im Vordergrund. Das Fingeralphabet ist nur als Hilfsmittel zur Erklärung einzelner Wör-

ter oder Eigennamen anzusehen. Meist werden schnell neue Eigengebärden erfunden. Zudem erfordert das Gebärden und Verstehen des Fingeralphabets mehr Konzentration und oftmals werden nur Bruchstücke verstanden. Ganze Sätze werden in der Regel nicht durch das Fingeralphabet gebärdet. Die DGS bietet hier eine weitaus einfachere und vor allem schnellere Möglichkeit der Kommunikation. Die Erkennung des Fingeralphabets wäre demnach nur eine zusätzliche Option.

### 4.1.2 Übersetzungsrichtung

Innerhalb der Fokusgruppe wird eine Übersetzungsrichtung von der DGS in die deutsche Schriftsprache bevorzugt. Die Interviewpartner beschreiben dies damit, dass DGS-orientierte Menschen in der Regel beide Sprachen lernen. Auch wenn bei der deutschen Schriftsprache Defizite existieren, ist die Verständlichkeit für geschriebene Texte relativ groß. Dem gegenüber wird das Verfassen eigener Texte entsprechend schwieriger empfunden. Verglichen wird dies mit einer anderen Fremdsprache wie z.B. Englisch. Meist wird auch hier mehr verstanden wie selbst artikuliert. Einige DGS-orientierte Menschen können zudem viele Informationen über das Mundbild gesprochener Sätze aufnehmen. Hörende Menschen hingegen verstehen kaum etwas von der Gebärdensprache, es sei denn sie wurde erlernt. Somit steigt die Relevanz einer Übersetzung aus der DGS in die deutsche Schriftsprache. Dennoch sollte auch der umgekehrte Weg nicht ganz außer Acht gelassen werden. Vor allem bei Grup- pengesprächen eines Gehörlosen mit mehreren hörenden Personen sinkt die Aufnahmefähigkeit der gehörlosen Person stark. Hier kam der Vorschlag seitens der Interviewpartner, die gesprochene Sprache über eine virtuell gebärdende Person in die DGS übersetzen zu lassen.

### 4.1.3 Grammatikalische Übersetzung

Über die grammatikalische Übersetzung der DGS in die deutsche Schriftsprache herrscht eine geteilte Meinung innerhalb der Fokus- gruppe. Zwei Interviewpartner sind der Mei-

nung, dass zunächst eine Übersetzung der Inhalte wichtig ist, da der Kontext dennoch verstanden werden kann, auch wenn der jeweilige Satz nicht in grammatikalisch richtigem Schriftdeutsch vorliegt. Viel wichtiger als eine grammatikalische Übersetzung wäre eine Erkennung der richtigen Gebärde auf Basis der gesamten Körpersprache. Dazu zählt auch die Beachtung der Mimik, des Mundbildes und der Körpersprache. So wäre es beispielsweise auch möglich, statt der DGS aus der LBG zu übersetzen. Die LBG beschreibt dabei lautsprachen begleitende Gebärden, welche sich an der deutschen Schriftgrammatik orientieren. Dies würde jedoch für DGS-orientierte Menschen bedeuten, dass sie entgegen ihrer Muttersprache die Sätze in deutscher Grammatik bilden müssten. Für die anderen Interviewpartner ist eine grammatikalische Übersetzung wichtig und sollte auf jeden Fall durchgeführt werden. Dennoch sehen alle Interviewpartner auch eine gewisse Schwierigkeit der grammatikalischen Übersetzung, da hier mehrere Faktoren eine Rolle spielen. Eine Übersetzung mit einfachen grammatikalischen Regeln, wie das bereits in Abschnitt 2.1 aufgeführte AVP  $\Rightarrow$  APV, sollten für den Anfang verfolgt und umgesetzt werden.

#### 4.1.4 Persönliche Erwartungen

Alle Interviewpartner können sich ein System zur Übersetzung der DGS vorstellen und würden dies auch effektiv einsetzen. Vor allem in Alltagssituationen wie beim Einkaufen, wo keine komplexen Themeninhalte gesprochen werden, sehen sie ein solches System als sinnvolle Unterstützung an. Bei komplexen Themenbereichen wie bei Ärzten oder Ämtern ziehen die gehörlosen Interviewpartner jedoch immer noch einen realen Dolmetscher vor. Dies liegt ein wenig in der Vorstellungskraft der Interviewpartner, da sich derzeit keiner ein voll funktionsfähiges und vor allem fehlerfreies System vorstellen kann.

Ein weiterer wichtiger Punkt bezieht sich auf die Größe und damit die Einsetzbarkeit des Systems. Es muss klein, handlich und

praktikabel sein. Als Beispiel wird hier das Handy bzw. Smartphone genannt. Die Übersetzung muss annähernd in Echtzeit erfolgen. Kleine Latenzzeiten werden akzeptiert, dürfen jedoch nicht störend sein. Neben der reinen Übersetzung aus der DGS in die deutsche Schriftsprache werden Erweiterungen wie z.B. eine Sprachein- und -ausgabe, eine Speichermöglichkeit oder auch Mustertexte vorgeschlagen. Auch sollte es möglich sein, neue Gebärden in das System übertragen zu können.

## 4.2 Technische Anforderungen

Bezüglich der technischen Anforderungen wurden die bereits aufgeführten Bewegungssensoren mit Hilfe ihrer Demoanwendungen auf ihre Einsatzmöglichkeit hinsichtlich der Gebärdensprach- und Fingererkennung getestet. Hier spielen die Auflösung und damit die Genauigkeit der Systeme sowie die Bewegungsfreiheit und die Bewegungsrichtung eine ausschlaggebende Rolle. Auch die Ermittlung des durchschnittlichen Produktionstempos einer Gebärde ist entscheidend, da hiermit eine Einschätzung gegeben werden kann, mit welcher Mindestgeschwindigkeit das System reagieren sollte.

### 4.2.1 Produktionstempo

Unter dem Begriff des Produktionstempos versteht man, wie bereits in Abschnitt 2.1 angesprochen, den zeitlichen Zusammenhang hinsichtlich der Erzeugung des sprachlichen Ausdrucks durch die Gebärden oder durch die Darstellung der Buchstaben des Fingeralphabets (vgl. [1], S.46f). Die Ermittlung der durchschnittlichen Zeit pro Gebärde bestimmt die Mindestgeschwindigkeit, mit welcher ein System zur Gebärdenspracherkennung agieren muss. Werden Gebärden erst später oder langsamer erfasst, entstehen Latenzzeiten, welche sich nachteilig auf die Benutzerfreundlichkeit auswirken. Dies bedeutet gleichzeitig, dass die Akzeptanz eines solchen Systems stark abnehmen würde. In Bezug auf den Versuch von Bellugi und Fischer (1972) wurden folgende Sätze von den Gehörlosen der Fokusgruppe in eine

Kamera gebärdet:

*„Hallo, mein Name ist V-o-r-n-a-m-e-N-a-c-h-n-a-m-e.“*

*„Ich arbeite bei der Paulinenpflege in Winnenden.“*

Im ersten Satz sollten dabei Vorname und Nachname im Fingeralphabet buchstabiert werden. Die Aufnahme erlaubt, bedingt durch das Aufnahmeformat, eine zeitliche Messung mit einer Genauigkeit von bis zu 25 Bildern pro Sekunde (fps). Dies entspricht 40 Millisekunden (ms). Die Auswertung ergab dabei die in Tabelle 2 aufgeführten Ergebnisse.

Um beide Sätze, inklusive dem Buchstabieren des Namens, zu gebärden, benötigten die Testpersonen im Schnitt 9,4 sek. Betrachtet man lediglich die Gebärden der DGS, so werden ca. 0,55 sek pro Gebärde und ca. 1,87 sek pro Satzinhalt benötigt. Die minimale Abweichung im Vergleich zum Messergebnis von Bellugi und Fischer ist auf die geringe Anzahl von Sätzen und gebärdenden Personen zurückzuführen. Dennoch ähnelt das Ergebnis der Studie von 1972 (vgl. [15]). Betrachtet man die durchschnittlich benötigte Zeit pro Gebärde im Fingeralphabet, fällt auf, dass diese mit ca. 0,37 sek deutlich schneller durchgeführt werden kann. Dies liegt schätzungsweise am kleineren Gebärdenbereich und der damit geringeren Bewegungsstrecke. Für ein System zur Gebärdenspracherkennung müsste demnach die Erfassung der jeweiligen Gebärde mindestens innerhalb von ca. 500 ms in der DGS und ca. 350 ms im Fingeralphabet erfolgreich durchgeführt und verarbeitet werden.

#### 4.2.2 Räumliche Auflösung und Erfassungsgeschwindigkeit

Mit der räumlichen Auflösung und der Erfassungsgeschwindigkeit lässt sich die Genauigkeit des Bewegungssensors beschreiben, mit welcher Objekte erkannt sowie deren

Bewegungen erfasst und verarbeitet werden können. Vor allem bei der Erkennung der Hände und einzelner Finger sollte eine möglichst hohe Auflösung gegeben sein. Tabelle 3 zeigt die Auflösungen der Bewegungssensoren sowie deren Erfassungsgeschwindigkeiten im Vergleich.

Die aufgeführten Werte sind Angaben der Hersteller und können nicht messbar überprüft werden. Demnach bietet die Leap Motion eine deutlich höhere Auflösung und Erfassungsgeschwindigkeit. Sie wurde primär für die Erkennung der Hände und Finger entwickelt, während die Asus Xtion als auch die Sensoren von PrimeSense primär zur Erkennung des gesamten Körpers konzipiert wurden. Für den Einsatz zur Erkennung des Fingeralphabets steht demnach die Leap Motion an oberster Stelle. Aber auch die Sensoren von Asus und PrimeSense ermöglichen, mittels der speziellen API des amerikanischen Unternehmens 3Gear Systems, die Erkennung beider Hände und aller zehn Finger. Die API basiert dabei auf einer Erweiterung des OpenNI Frameworks. Da es mit der Leap Motion nicht möglich ist ein Körpertracking zu realisieren, wird im Anschluss lediglich das Fingertracking betrachtet und getestet.

Ein messbarer Vergleich der Hardware ist mit den vorhandenen Mitteln nicht zu realisieren. Daher wird die Hardware auf Basis der dazugehörigen Demoanwendungen getestet und beurteilt. Als Testumgebung dient ein Mac Book Pro mit einem 2,4 GHz Intel Core i5 Prozessor, 8GB RAM und dem aktuellen Betriebssystem OS X in der Version 10.9.1. Hinsichtlich des Fingertrackings wird die Leap Motion mit der PrimeSense Carmine 1.09 und deren jeweils aktuellsten Softwareversionen verglichen.

Die Leap Motion wird standardmäßig mit dem sogenannten Visualizer ausgeliefert. Diese Demoanwendung erlaubt die Erfassung und grafische Darstellung der Hände und Finger. Im Einstellungsfeld der Leap Motion lässt sich während der Nutzung die aktuelle Daten-

**Tabelle 2: Messergebnis Produktionstempo**

Messwert	Gesamt Sätze bzw. Name		Durchschnitt pro Gebärde	
	fps	ms	fps	ms
Gebärden DGS	93,66	3746,66	13,73	549,33
Fingeralphabet	141,33	5653,33	9,15	365,86
Gesamt	234,99	9399,99	-	-

**Tabelle 3: Auflösung und Geschwindigkeit**

*Quelle: Vgl. [12] und [16]*

	Leap Motion	Asus Xtion/Carmine 1.08	Carmine 1.09
Räumliche Auflösung	0,001 mm	3,4 mm	0,9 mm
Tiefenauflösung	0,001 mm	12 mm	1 mm
Geschwindigkeit	>200 fps	max. 60 fps	max. 60 fps

übertragungsrate sowie die Verarbeitungszeit anzeigen. Durchschnittlich stellte sich während des Tests eine Datenübertragungsrate von ca. 90 fps und einer Verarbeitungszeit von ca. 8 ms ein. Vergleichbar hierzu wird das Fingertracking der Carmine 1.09 mit der Demoanwendung HandViewer aus der bereits erwähnten API von 3Gear Systems getestet.



**Abbildung 8: Darstellung Visualizer (oben), HandViewer (unten)**

Während des Tests stellte sich hier eine durchschnittliche Datenübertragungsrate von ca. 59 fps ein. Eine Verarbeitungszeit wird seitens der Anwendung nicht gegeben. Beide

Ergebnisse übertreffen dennoch die Anforderungen des in Abschnitt 3.2.1 ermittelten Produktionstempos deutlich und könnten somit zur reinen Erfassung von Gebärden der DGS und des Fingeralphabets eingesetzt werden. Zum Test selbst werden nacheinander verschiedene Handformen und Bewegungen durchgeführt und deren visueller Eindruck beurteilt.

Bezüglich der Geschwindigkeit ist ein deutlicher Unterschied, vor allem bei schnellen Bewegungen, zu verzeichnen. Hier bietet die Leap Motion mit ihrer höheren Datenübertragungsrate einen flüssigeren und annähernd realen Bewegungsablauf. Über den HandViewer wirken die Bewegungen etwas langsamer und stockend. Schnelle Bewegungen der Finger können teilweise nicht erfasst werden.

Bei unterschiedlichen Handformen fällt im Visualizer der Leap Motion auf, dass eng beieinander liegende Finger schnell zu einer Erfassungsschwierigkeit und damit zum Verschwinden der Finger führt. Dies wird gerade bei einer Faust oder einer Rotation der Hand, bei welcher sich die Finger überlagern, stark wahrgenommen. Im HandViewer in Verbindung mit der Carmine 1.09 werden



solche Überlagerungen teilweise kompensiert. Selbst eine zur Kamera orthogonal stehende Hand wird dargestellt und bei einer Bewegung einzelner Finger werden diese trotz Überlagerung erfasst. Allerdings sind auch hier Erfassungsfehler zu verzeichnen. Beispielsweise wird bei einer Faust, die den Daumen unter oder über den restlichen Fingern hat, kein Unterschied wahrgenommen. In Bezug auf die Bewegungsauflösung ist die Leap Motion deutlich filigraner. Kleinste Bewegungen des Zeigefingers im Millimeterbereich werden erfasst. Hier kann es bei der Carmine 1.09 vorkommen, dass solch geringe Bewegungen erst gar nicht erfasst werden können.

Hinsichtlich der Gebärden des Fingeralphabets, bei welchem die Finger oft eng beieinander liegen, verzeichnen beide Systeme deutliche Schwächen bei der Auflösung. Hier entstehen in der Regel stärkere Erfassungsfehler. Dennoch ist es denkbar, dass diese Fehler durch die Entwicklung bestimmter Algorithmen kompensiert werden könnten. Ein Beispiel hierfür bieten Zagel, Süßmuth und Bodendorf in [22]. In ihrer Arbeit „*Automatische Rekonstruktion eines 3D Körpermodells aus Kinect Sensordaten*“ wird ein dreidimensionales Körpermodell aus dem Tiefenbild des Kinect Sensors durch einen speziellen Algorithmus errechnet. Der verwendete Algorithmus gleicht dabei Artefakte wie z.B. Rauschen oder Quantisierungsartefakte, welche durch die limitierte Auflösung des Sensors entstehen, aus (vgl. [22], S. 554f). Eine Aussage darüber, ob und wie gut die Erfassung mit beiden Systemen erfolgen kann, ist zum aktuellen Zeitpunkt nicht möglich, da dies einen konkreten Test mit einer Software zur Erkennung der Gebärden voraussetzen würde. Dennoch ist der Einsatz beider Systeme hinsichtlich der räumlichen Auflösung und der Erfassungsgeschwindigkeit denkbar.

### 4.2.3 Bewegungsfreiheit und Bewegungsrichtung

Sowohl die Gebärden der DGS als auch die des Fingeralphabets werden innerhalb

des Gebärdenraums in Sprachrichtung zum Zuseher gebärdet. Dies erfordert eine frontale Erkennung der Gebärden. Die Asus Xtion sowie die PrimeSense Carmine sind in technischer Hinsicht für diesen Frontaleinsatz konzipiert. Leap Motion bietet hingegen nur eine Erkennung der Hände und Finger über dem Controller. Bezüglich der Erfassung von Gebärden der DGS kann die Leap Motion nicht eingesetzt werden, da der Erfassungsraum zu klein und damit kein Körpertracking möglich ist. Sowohl die Asus Xtion als auch die Sensoren von PrimeSense ermöglichen bereits, mittels OpenNI Framework, ein Körpertracking. Hier kann beispielsweise der gesamte Körper oder aber auch nur der Oberkörper erfasst werden. Durch eine Erweiterung des Frameworks könnte demnach die gewünschte Gebärdenspracherkennung realisiert werden.

Für die Erfassung des Fingeralphabets müsste die Leap Motion um 90° gedreht werden und sich die gebärdende Person sehr nahe am Controller befinden, da die Erfassungstiefe lediglich max. 60 cm beträgt. Ein Test mit dem Visualizer zeigt, dass dies nicht so einfach zu realisieren ist. Die reine Erfassung der Hände bei einem gedrehten Controller funktioniert, solange sich keine Fläche oder kein Objekt hinter den Händen befindet. Da das Fingeralphabet jedoch vor dem Körper in Brusthöhe gebärdet wird, kann hier keine Erfassung erfolgen. Beim Fingertracking in Verbindung mit der PrimeSense Carmine 1.09 erfolgt das Fingertracking von oben. Auch hier muss der Sensor zunächst gedreht werden. Die Frontalerfassung ist hier nur eingeschränkt möglich, wenn sich die Hände vor einem ebenen Hintergrund befinden. Vor dem Körper steigt die Fehlerrate jedoch stark an. Somit können beide Sensorarten nicht für eine Frontalerkennung des Fingeralphabets eingesetzt werden.

## 5 Zusammenfassung

Durch die mit der Fokusgruppe geführten Interviews wird deutlich, dass aktuell immer noch Probleme in der Kommunikation zwischen gehörlosen und hörenden Menschen

bestehen. Es gibt nur wenig effektive Hilfsmittel, sodass meist ein realer Dolmetscher zum Einsatz kommt. Die Idee einer computergestützten Erkennung und Übersetzung der Gebärdensprache findet Anklang bei der Fokusgruppe. Allerdings begrenzt sich die Vorstellung derzeit noch auf ein System, welches nur für einfache Gespräche verwendet wird. Für komplexe Gespräche, wie z.B. bei Ärzten oder Ämtern, wird weiterhin noch der reale Dolmetscher bevorzugt. Das System selbst sollte natürlich portabel, einfach zu nutzen und schnell sein. Geringe Latenzzeiten werden akzeptiert. Zunächst wird die Übersetzungsrichtung aus der DGS in die deutsche Schriftsprache gewünscht. Aber auch der umgekehrte Weg sollte nicht außer Acht gelassen werden. Weitere Optionen könnten eine Sprachein- und -ausgabe, eine Speicherfunktion sowie die Möglichkeit zur Erweiterung des Wortschatzes durch neue Gebärden sein.

In technischer Hinsicht zeichnet sich der Trend zu den Sensoren von Asus und PrimeSense ab. Trotz der geringeren räumlichen Auflösung sowie der langsameren Erfassungsgeschwindigkeit, lassen sich diese Sensoren vielseitiger einsetzen. Auch die Möglichkeit mit den Sensoren eine Erkennung der Mimik realisieren zu können, spricht stark für den Einsatz bei der computergestützten Übersetzung aus der DGS in die deutsche Schriftsprache. Alle ermittelten Anforderungen werden nochmals in Tabelle 4 aufgeführt.

## 6 Handlungsempfehlung

In Anbetracht der technischen und benutzerspezifischen Anforderungsanalyse wird für die Entwicklung eines Systems zur Erkennung und Übersetzung der DGS in die deutsche Schriftsprache folgende Handlungsempfehlung ausgesprochen. Im ersten Schritt sollte die Erfassung der Gebärden im Vordergrund stehen. In technischer Hinsicht werden dafür die Sensoren Asus Xtion oder PrimeSense Carmine empfohlen. Die Software muss daher zunächst auf einem stationären PC oder Laptop lauffähig sein. Dies begründet sich durch die vorgeschlagenen Bewegungssenso-

ren, welche für den mobilen Einsatz aufgrund ihrer Größe ungeeignet sind. PrimeSense entwickelt aktuell diesbezüglich den Bewegungssensor Capri 1.25, welcher für den mobilen Einsatz vorgesehen ist (vgl. [16], S. 3). Die Entwicklung einer mobilen Version könnte dann, nach erfolgreichem Testlauf des stationären Systems, erfolgen. Aus Entwicklersicht müssen, für die Erfassung der Gebärden, eigene Gesten programmiert werden, welche beim Erfassungsvorgang ggf. über eine Datenbank abgeglichen werden können. Ist die Erfassung mehrerer Gebärden möglich, kann im zweiten Schritt die grammatikalische Übersetzung implementiert werden. Zunächst sollten hier jedoch nur einfache grammatikalische Regeln in Betracht gezogen werden. Ziel für einen Prototyp ist die Möglichkeit zur Erkennung von ausgewählten Gebärden, aus welchen bspw. zwei unterschiedliche Sätze erzeugt und übersetzt werden können.

## 7 Danksagung

Mein Dank gilt zunächst meiner Lebensgefährtin Frau Manuela Schreiber, ohne die ich vermutlich nicht auf dieses interessante und vor allem wichtige Thema gestoßen wäre. Sie ermöglichte mir auch die einfache und problemlose Kommunikation und Organisation mit den Interviewpartnern der Paulinenpflege Winnenden e.V. Ebenfalls geht mein Dank an die Schulleiterin der Schule beim Jakobsweg Frau Beate Löffler sowie die stellvertretende Schulleiterin Frau Christiane Sättler-Adel, welche mir zunächst die Durchführung des Interviews an ihrer Schule ermöglichten. Auch möchte ich mich diesbezüglich für die Bereitstellung des Raumes und der Verpflegung bedanken sowie für den Einsatz der Dolmetscherin Frau Sonja Lewandowsky und ihrer Praktikantin Antje Voss. Ganz großer Dank gilt natürlich den Interviewpartnern Frau Gabi Braig, Herrn Christian Hermann und Herrn Martin Lindenberger von der Paulinenpflege Winnenden e.V. sowie Herrn Martin Müller von der Stiftung St. Franziskus Heiligenbrunn. Allen zusammen nochmals ein herzliches Dankeschön für Ihre Mitarbeit.

**Tabelle 4: Ermittelte Anforderungen**

<b>Benutzerspezifische Anforderungen</b>	Übersetzung	<i>DGS</i>
	Übersetzungsrichtung	<i>DGS ⇒ dt. Schriftsprache</i>
	Grammatik	<i>Zunächst einfache Regeln (AVP ⇒ APV)</i>
	Persönliche Erwartungen	<i>- Portabel - Praktikabel - Einfache Nutzung - Annähernd Echtzeitübersetzung</i>
<b>Technische Anforderungen</b>	Produktionstempo/ Erfassungsgeschwindigkeit	$\leq 500 \text{ ms/Gebärde (DGS)}$
	Auflösung	<i>Die Auflösung wird bestimmt durch die Erfassungsgenauigkeit und die Erfassungsgeschwindigkeit. Je höher die Auflösung, desto genauer können Erfassungen durchgeführt werden.</i>
	Bewegungsrichtung/ Bewegungsfreiheit	<i>- Frontal zum Anwender - Gebärdenraum muss vollständig erfasst werden</i>
	Empfohlener Sensortyp	<i>RGB-Tiefensensor - Asus Xtion - PrimeSense Carmine</i>

## Literatur

- [1] P. Boyes-Braem. *Einführung in die Gebärdensprache und ihre Erforschung*, volume Bd. 11. Signum, Hamburg, 2. korrigierte auflage edition, 1992.
- [2] K.-U. Carstensen, C. Ebert, C. Ebert, S. Jekat, R. Klabunde, and H. Langer. *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Spektrum, Akad. Verl., Heidelberg, 3., überarb. und erw. aufl edition, 2010.
- [3] X. Chai, G. Li, X. Chen, M. Zhou, G. Wu, and H. Li. Visualcomm: A tool to support communication between deaf and hearing persons with the kinect. In *ASSETS'13*, pages 1–2.
- [4] X. Chen, L. Hanjing, T. Pan, S. Tansley, and M. Zhou. Kinect sign language translator expands communication possibilities, 2013.
- [5] Deutscher Gehörlosen-Bund e.V. *Faq - gehörlosigkeit*, 2014.
- [6] M. Dos Santos Anjo, E. Brigante Pizolato, and S. Feuerstack. A real-time system to recognize static gestures of brazilian sign language (libras) alphabetu sing kinect. In C. Maciel and Interaction, ACM Special Interest Group on Computer-Human, editors, *IHC'12*,

- pages 259–268, [S.l.], 2012. Brazilian Computer Society.
- [7] H. Eichmann, M. Hansen, and J. Heßmann. *Handbuch deutsche Gebärdensprache: Sprachwissenschaftliche und anwendungsbezogene Perspektiven*, volume Bd. 50 of *Internationale Arbeiten zur Gebärdensprache und Kommunikation Gehörloser*. Signum, Seedorf, 2012.
- [8] S. Falahati. *OpenNI Cookbook*. Packt Publishing, Birmingham, 2013.
- [9] S. Fünfer. *Mensch oder Maschine? Dolmetscher und maschinelles Dolmetschsystem im Vergleich*, volume Bd. 2 of *Transkulturalität - Translation - Transfer*. Frank & Timme, Berlin, 2013.
- [10] M. Hillenmeyer. *Dgs - deutsche gebärdensprache 1: Grundkurs für anfangler*, 2000.
- [11] W. Jiangquin, G. Wen, S. Yibo, and L. Wei. A simple sign language recognition system based on data glove. In Institute of Electrical and Electronics Engineers, editor, *International Conference on Signal Processing (ICSP)*, pages 1257–1260. I.E.E.E. Press, 1998.
- [12] Leap Motion, Inc. *Leap motion - produktionformationen*, 2014.
- [13] MAKATON Deutschland. *Was sie über makaton wissen sollten*, 2014.
- [14] J. Muhs. *Die geschichte der gebärdensprache - von der römerzeit bis heute*, 2014.
- [15] T. D. Pahlke. *Interpretation der studie „hörscreening“ aus sicht des deutschen schwerhörigenbundes*, 2001.
- [16] PrimeSense. *Primesense 3d sensors: Produkt- und technologiebeschreibung*, 2014.
- [17] P. A. Reich and M. Bick. An empirical investigation of some claims made in support of visible english. *American annals of the deaf*, 121(6):573–577, 1976.
- [18] W. Roddeck. *Einführung in die Mechatronik: Mit 9 Tabellen*. Lehrbuch. Teubner, Wiesbaden, 3., überarb. und erg. aufl edition, 2006.
- [19] SparkFun Electronics. *Leap motion tear-down*, 2014.
- [20] Technabling Ltd. *PsIt - portable sign language translator*, 2014.
- [21] A. P. Vicente and A. A. Faisal. Calibration of kinematic body sensor networks: Kinect-based gauging of data gloves “in the wild”. In IEEE, editor, *2013 IEEE International Conference on Body Sensor Networks (BSN)*, pages 1–6, 2013.
- [22] C. Zagel, J. Süßmuth, and F. Bodendorf. *Automatische rekonstruktion eines 3d körpermodells aus kinect sensordaten: Wirtschaftsinformatik proceedings 2013*, paper 35, 2013.