

## **Karlsruhe Reports in Informatics 2017,4**

Edited by Karlsruhe Institute of Technology,  
Faculty of Informatics  
ISSN 2190-4782

# **Ubiquitäre Systeme (Seminar) und Mobile Computing (Proseminar) WS 2016/17**

Mobile und Verteilte Systeme  
Ubiquitous Computing

Teil XIV

Herausgeber:  
Martin Alexander Neumann, Antonios Karatzoglou,  
Anja Exler, Erik Pescara,  
Long Wang, Michael Beigl

2017



# Fakultät für Informatik

**Please note:**

This Report has been published on the Internet under the following  
Creative Commons License:

<http://creativecommons.org/licenses/by-nc-nd/3.0/de>.

**Ubiquitäre Systeme (Seminar)  
und  
Mobile Computing (Proseminar)  
WS 2016/17**

---

Mobile und Verteilte Systeme  
Ubiquitous Computing  
Teil XIV

---

**Herausgeber**

*Martin Alexander Neumann*

*Antonios Karatzoglou*

*Anja Exler, Erik Pescara*

*Long Wang, Michael Beigl*

**Karlsruhe Institute of Technology (KIT)  
Fakultät für Informatik  
Lehrstuhl für Pervasive Computing Systems (PCS) und TECO**

**Interner Bericht 2017-04  
ISSN 2190-4782**



## Vorwort

Die Seminarreihe Mobile Computing und Ubiquitäre Systeme existiert seit dem Wintersemester 2013/2014. Seit diesem Semester findet das Proseminar Mobile Computing am Lehrstuhl für Pervasive Computing System statt. Die Arbeiten des Proseminars werden seit dem mit den Arbeiten des zweiten Seminars des Lehrstuhls, dem Seminar Ubiquitäre Systeme, zusammengefasst und gemeinsam veröffentlicht.

Die Seminarreihe Ubiquitäre Systeme hat eine lange Tradition in der Forschungsgruppe TECO. Im Wintersemester 2010/2011 wurde die Gruppe Teil des Lehrstuhls für Pervasive Computing Systems. Seit dem findet das Seminar Ubiquitäre Systeme in jedem Semester statt. Ebenso wird das Proseminar Mobile Computing seit dem Wintersemester 2013/2014 in jedem Semester durchgeführt. Seit dem Wintersemester 2003/2004 werden die Seminararbeiten als KIT-Berichte veröffentlicht. Ziel der gemeinsamen Seminarreihe ist die Aufarbeitung und Diskussion aktueller Forschungsfragen in den Bereichen Mobile und Ubiquitous Computing.

Dieser Seminarband fasst die Arbeiten der Seminare des Wintersemesters 2016/17 zusammen. Wir danken den Studierenden für ihren besonderen Einsatz, sowohl während des Seminars als auch bei der Fertigstellung dieses Bandes.

Karlsruhe, den 01. Oktober 2016

Martin Alexander Neumann  
Antonios Karatzoglou  
Anja Exler  
Erik Pescara  
Long Wang  
Michael Beigl



# Inhaltsverzeichnis

<i>Quy Think Nguyen</i> Virtuelle intelligente Assistenten .....	1
<i>Yannick Meny</i> Reverse Geocoding APIs .....	14
<i>Robin S. Schnaidt</i> Sensory Augmentation with Tactile Interfaces .....	31
<i>Niklas Keller</i> Überblick über den Einfluss von Smartphones auf den Nutzer .....	48
<i>Felix Rittler</i> Sensorische Augmentation durch taktile Schnittstellen .....	63
<i>Maximilian Stemmer-Grabow</i> Text to Skin – Using Your Skin Senses to Read .....	79
<i>Nicolas Schreiber</i> Electroactive Polymers (EAPs) for Wearable Computing and Exoskeletons .....	96
<i>Stefan Christian Lamp</i> Matrix Factorization for Next Place Prediction .....	114
<i>Adrian Jablonski</i> Artificial Neural Networks for Semantic Location Prediction .....	135

<i>Harun Şentürk</i> Artificial Neural Networks for location prediction.....	157
<i>Mengying Yang</i> In Home Sleep Monitoring.....	178

# Virtuelle intelligente Assistenten

Quy Thinh Nguyen\*

Advisor: Antonios Karatzoglou<sup>†</sup>

Karlsruhe Institute of Technology (KIT)  
Pervasive Computing Systems – TECO

\*ubeky@student.kit.edu

<sup>†</sup>antonios@teco.edu

**Abstract.** Diese Arbeit wird einen Überblick über virtuelle intelligente Assistenten (VIA), die im deutschen auch oft als Sprachassistenten bezeichnet werden, geben. Es werden die verschiedenen Arten von VIA gezeigt und in welchem Zusammenhang sie momentan schon genutzt werden. Als Beispiel werden einige aktuelle Assistenten dienen. Zudem werden neben den Möglichkeiten der Programme, auch noch die Grenzen dieser Technik dargestellt und mögliche Verbesserungen und Optimierungen für die Zukunft besprochen. Ebenfalls wird das empfindliche Thema Datenschutz mit Bezug auf die VIA behandelt. Am Ende des zweiten Teils gehen wir auf die aktuelle Beliebtheit und Nutzung der Anwendungen ein. Im dritten Abschnitt wird dann die Funktionsweise der Software im Hintergrund der Assistenten betrachtet. Es wird der allgemeine Aufbau eines VIA skizziert. Zudem werden zwei Modelle für die Arbeitsweise der Sprachassistenten betrachtet. Dazu werden auch einige theoretische Konzepte wie Ontologien, Knowledge Graphen und POMDP (partially observable Markov decision process ) beleuchtet. Zum Schluss kommt ein Fazit über die weitere Entwicklung der VIA mit Überlegungen für Verbesserungen und Optimierungen.

**Keywords:** Virtuelle intelligente Assistenten - Sprachassistenten - Ontologien - POMDP - Datensicherheit

## 1 Einleitung

Virtuelle intelligente Assistenten sind eine neue Technologie, die den Menschen viele Dienste ermöglichen wollen: Man kann mit einem Sprachassistenten interagieren, Gespräche führen und ihm bestimmte Aufgaben, die er ausführen soll, auftragen. Die Assistenten können dann z.B. in einem Smart Home Musik abspielen lassen, die Temperatur im Raum regeln oder das Licht an- und ausschalten. Ebenfalls gehören zu den Aufgaben, die von den Assistenten übernommen werden können, die Terminplanung, das Bestellen von Waren und Reservierungen. Diese Vorstellungen von einem ubiquitären Assistenten versuchen die großen IT-Unternehmen in der realen Welt umzusetzen. Mittlerweile hat jeder der großen Technologiefirmen einen eigenen virtuellen intelligenten Assistenten vorzuzeigen: Google hat Google Now und seit September 2016 Google Assistant[1], Microsoft

hat Cortana, Apple hat Siri, Amazon hat Alexa, Facebook hat M, der im Vergleich zu den anderen Sprachassistenten ein textbasierter Assistent ist, und IBM hat Watson. Auch asiatische Firmen wie Chinas Baidu und Samsung, das im Oktober 2016 den intelligenten Assistenten Viv übernommen hat[2], investieren in diese Technik. Wir schauen uns an in welchen Branchen(Smart Home, Automobilindustrie usw.) die Assistenten schon aktiv sind, wie groß der Nutzen und wie gut die Performance der Assistenten ist. Ebenfalls wird es auch einen Blick auf "kleinere" VIA geben, die sich auf einzelne Tätigkeiten spezialisiert haben. Ein großes Thema unserer Zeit ist die künstliche Intelligenz und auch hinter der Technologie der Sprachassistenten stecken Themen wie Deep Learning, Ontologien und Knowledge Graph. Ein Schaubild soll einen allgemeinen Blick auf die Funktionsweise und den Aufbau geben. Als spezielles Beispiel dient dann ein Schema von Alexa.

## 2 Allgemeine Übersicht

Man kann virtuelle intelligente Assistenten in zwei Kategorien unterteilen. Auf der einen Seite gibt es die schon erwähnten VIA der großen Weltkonzerne von Google, Apple und so weiter. Diese Assistenten haben die Absicht den Nutzern für den allgemeinen Zweck zur Verfügung zu stellen. Das heißt, dass sie möglichst alle Dienste, die man mit digitalen Medien ausführen kann, nach Aufruf durchführen und dem Nutzer als alltäglichen und zu jeder Zeit verfügbaren Butler zur Verfügung stehen. Zu den Anwendungen gehören z.B. das Eintragen von Terminen in den Kalender, das Abspielen von Liedern oder auch das Anzeigen vom Nahverkehrsplan. Einige der VIA sind auf Smartphones und Tablets integriert oder vorinstalliert. Dazu gehören Siri auf den tragbaren Apple Produkten, Google Now (auf den neuen Modellen wird dann Google Assistant verfügbar sein,) auf vielen Android-Geräten und Bixby[18], das auf dem Samsung Galaxy 8 erhältlich sein soll. Es ist noch nicht bekannt ob die Entwickler von Viv auch an diesem Sprachassistenten für Samsung mitgearbeitet haben. Weitere Anwendungszwecke finden sie auf Audiogeräten wie z.B. Amazon Echo, das Alexa als Sprachassistenten besitzt und das Konkurrenzprodukt Google Home mit Google Assistant als intelligenten Assistenten. Bei vielen Assistenten funktioniert die Anwendung sowohl als Textübergabe als auch als Sprachaufruf (z.B. IBM Watson9, aber einige sind auch auf eine Nutzungsart beschränkt (z.B. Facebooks M ist nur über die Spracherkennung nutzbar). Es gibt aber auch noch eine zweite Kategorie von VIA, die nur für bestimmte Aufgaben entwickelt wurden. Sie sind nicht als allgemeiner Begleiter gedacht, und haben sich auf eine oder wenige Dienste beschränkt. Dazu gehören die App Dragon Dictation von Nuance und die Dom-App von der Restaurantkette Dominos Pizza. Während Dragon Dictation Sprachaufnahmen selbständig in Texte umwandelt, hilft die App von Dominos Pizza den Kunden bei der Bestellung der Pizza.

Assistenten für die allgemeine Nutzung	Assistenten für eine spezifische Anwendung
<ul style="list-style-type: none"> <li>- ein Begleiter im Alltag, der ihm wie ein Butler als Hilfe zur Seite steht</li> <li>- diesen Dienst versuchen die Assistenten der großen Technologie Firmen wie zum Beispiel Amazon, Google und Apple ( Alexa, Google Assistant, Siri) anzubieten</li> </ul>	<ul style="list-style-type: none"> <li>- unterstützt den Nutzer bei einer speziellen Aufgabe</li> <li>- mehrere Firmen wollen damit den Service für ihre Kunden verbessern</li> <li>- ein Beispiel ist Dominos Pizzas App</li> </ul>

Abbildung 9 : Kategorisierung der Assistenten

### 2.1 Leistungsvermögen und technische Grenzen

Die interessanteste Frage ist wohl: Wie gut und nützlich sind die intelligenten Assistenten momentan? Die Technologie ist noch sehr jung und noch lange nicht ausgereift und deswegen funktioniert auch noch nicht jede Anwendung wie man sich es wünscht. Das ist natürlich besonders bei den ubiquitären Assistenten, die viele Dienste anbieten, noch problematisch. Einige Assistenten kann man noch gar nicht bewerten, da sie noch in der Entwicklung sind und bisher nur von wenigen ausgewählten Personen getestet wurde. Beispiele wären da zum Beispiel M und Viv[3]. Von den aktuellen verfügbaren Assistenten gibt es natürlich viele Tests und Berichte, die auch die VIA untereinander vergleichen. Beispiele für Testberichte gibt es unter anderem in der New York Times oder auf der Seite von Financial Times[5]. In den vielen Berichten ist erkennbar, dass sich die Sprachassistenten bei einigen Tätigkeiten schwer tun. Eines der großen Hindernisse für die Assistenten sind noch etwas komplexere Fragen oder auch umgangssprachliche Aussagen, da sie die Semantik der Aussagen nicht richtig interpretieren können. Fehler oder keine Ergebnisse gibt es aber auch bei einfachen Anfragen. Zum Beispiel kann Alexa die Verkehrslage nicht so gut wiedergeben. Siri und Cortana zeigen keine Flüge als Reiseoptionen an. Für Google Assistant ist es problematisch eine Email laut vorzulesen.[4] Das ist der Stand im Jahr 2016, da sich die Assistenten stetig weiterentwickeln, könnte sich die Leistung über die nächsten Jahre stark verbessern. Es gibt natürlich auch jetzt schon Disziplinen in denen die Assistenten schon sehr gut abschneiden. Google Assistant und Siri sind z.B. sehr gut beim Suchen und Anzeigen von Restaurants in der Nähe, Alexa beim Abspielen von gewünschten Liedern und Cortana beim Wiedergeben der Ergebnisse von Sportereignissen. Eine ziemlich außergewöhnliche Leistung hat Watson erreicht. Die Software von IBM hat bei der Quizshow "Jeopardy!" zwei ehemalige Kandidaten, die Rekordsummen gewannen, geschlagen. [7]

### 2.2 Datenschutz

Damit die ganzen Assistenten gute Ergebnisse liefern und sich im Laufe der Zeit noch mehr auf jeden einzelnen Nutzer einstellen können, brauchen sie Daten. Das ist eine kritische Sache für die Datensicherheit und die Privatsphäre der

Menschen, denn die Sprachassistenten speichern sämtliche Anfragen ab. Alexa sammelt die Anfragen dauerhaft mit Hilfe der Amazon-Cloud ein. Es ist möglich die Einträge zu löschen, aber dann muss man auch mit schlechteren Leistungen rechnen. Siri speichert laut Apple die Daten zwei Jahre ab und die ersten sechs Monate davon mit ID. Im Vergleich zu vielen Messenger-Diensten wie WhatsApp oder dem Facebook-Messenger verzichtet Google Allo, das auch Google Assistant implementiert hat, auf eine end-to-end Verschlüsselung, da die künstliche Intelligenz von Allo für seine Arbeit den Konversationen folgen muss. Man könnte seine Privatsphäre sichern in dem man auf einen "Inkognito Modus" schaltet[6], aber auch hier gilt: Wer eine gute Leistung möchte, muss mit seinen Daten bezahlen.

### 2.3 Popularität

Die Wirtschaft schaut natürlich gespannt auf die Entwicklung der Technologie. Mit den Assistenten können die Firmen bessere Dienste anbieten, sowie Dominos Pizza es jetzt schon versucht. Ebenso kann man mit virtuellen Assistenten Arbeitskraft und Kosten einsparen. Ein japanisches Versicherungsunternehmen hat Anfang 2017 30 Angestellte mit dem virtuellen Assistenten von IBM ausgetauscht.[8] In der Bevölkerung haben die Sprachassistenten noch keine große Popularität gewonnen, trotzdem ist die Anzahl der Nutzer schon beachtlich hoch. Nach Umfragen haben in Deutschland ungefähr die Hälfte der Internetuser zumindest einen Sprachassistenten schon genutzt.[9] Ein großes Interesse an den Sprachassistenten ist also durchaus vorhanden, auch wenn die Häufigkeit der Nutzung noch ausbaufähig ist. Das liegt natürlich zum Einen daran, dass die Technik noch am Anfang ihrer Entwicklung ist und noch nicht alles wie gewünscht angeboten werden kann. Ein weitere Ursache ist aber auch, dass es für die meisten Menschen ungewöhnlich ist in der Öffentlichkeit einen virtuellen Assistenten anzusprechen. Der Einsatz von den Menschen, die Sprachassistenten nutzen, ist meistens viel häufiger in einem privaten Raum, z.B. zu Hause oder im Auto.[10] Auch sind einige nicht bereit die vielen Daten, die man zur Verfügung stellen muss, und die Kontrolle über wichtige persönliche Angelegenheiten, abzugeben. Es ist ihnen befremdlich, wenn eine Software Nachrichten an ihre Freunde schreibt oder ihre Termine plant.



## 2.4 Optimierungen und Verbesserungen

Die Sprachassistenten können einige Tätigkeiten schon gut ausführen. Damit die Assistenten sich stetig weiter entwickeln, wollen sich die großen Technologie Firmen Apple, Siri und Amazon für unabhängige Entwickler öffnen. Diese können eigene Erweiterungen und Funktionen, die sie als nützlich ansehen, für die Sprachassistenten erstellen und dadurch zum Wachstum der Software beitragen. Amazon hat es mit Alexa vorgemacht [12]. Apple hat sich jetzt auch nach Jahren für externe Entwickler geöffnet, damit Siri davon profitieren kann. Seit Dezember 2016 erlaubt Google dritten Anbietern über die Plattform API.ai ihre Produkte oder Services in den Google Assistant zu integrieren[11]. Die Kommunikation mit den Assistenten soll besser werden. Erste Fortschritte hat angeblich Viv schon gemacht. Die Entwickler behaupten, dass ihre Software noch bessere Optionen anbietet und sich auch bei komplexen Anfragen besser schlägt als die Konkurrenz. Gespannt kann man auf die Einführung hinschauen, ob der neue Sprachassistent wirklich einen Schritt weiter gehen kann als die anderen Anbieter. [13] Das Ziel der großen Technologiefirmen ist es in der Zukunft mit ihren Assistenten so viele gute Dienstleistungen anbieten zu können, dass das Nutzen von virtuellen intelligenten Assistenten zum Alltag gehört, so wie es das Smartphone auch geschafft hat. Die Notwendigkeit von Apps und Dinge wie das Schreiben von Mails wird sich dann etwas altmodisch anhören, da es ja Assistenten gibt, die die Aufgaben übernehmen können. Das hätte auch eine große Bedeutung für den eCommerce, weil Kunden nicht mehr so erreichbar wären wie heute, aber dazu müssen Siri, Alexa und co. erstmal noch den nächsten Schritt machen.

## 3 Funktionsweise

In diesem Kapitel schauen wir uns an wie die Sprachassistenten aufgebaut sind und welche Technologie hinter ihnen steckt. Dafür schaffen wir uns erst einmal

einen Überblick mit Abbildung 1 und 2. In Abbildung 1 sieht man wie eine Anfrage (Input) bearbeitet wird. Zuerst übermittelt der Nutzer seinen Wunsch, der dann vom Speech Recognizer wahrgenommen wird. Der Input wird dann in eine Wortreihenfolge übertragen und damit vorbereitet bevor er in den NLU (natural language understanding)- Status kommt. Im nächsten Schritt gibt es eine semantische Interpretation der Anfrage. Falls ein Element für eine Interpretation der Aussage fehlen sollte, versucht das Dialogue Management durch eine Interaktion mit dem Nutzer die fehlenden relevanten Informationen zu ergänzen. Wenn die Informationen komplett sind, wird ausgewertet welche Antwort/Aktion am besten gegeben/ausgeführt werden soll und diese wird dann dem User als Output übergeben.

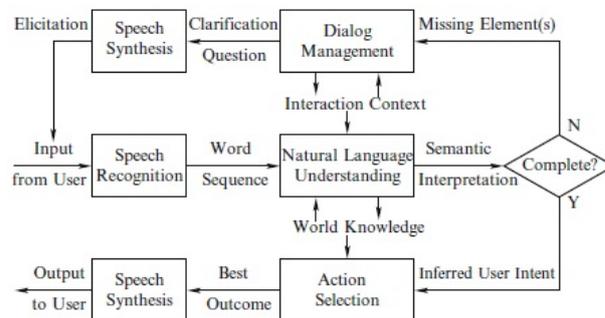


Abbildung 1:  
Model von der Interaktion  
intelligenten Assistenten  
Quelle: J.R. Bellegarda

Wie die Auswertungen von Sprachanfragen oder Texteingaben funktionieren, schauen wir uns mit Hilfe der zweiten Abbildung an. Für die Sprachanfragen gibt es einen automatischen Spracherkenner (Speech Recognition), der für die Textanfragen natürlich nicht notwendig ist. Nach der Erkennung wird der Input mit Hilfe von zwei Komponenten verarbeitet. Zum einen startet die datengetriebene Komponente eine semantische Interpretation der Anfrage über ein statistisches Framework z.B. mit Hilfe von POMDP (partially observable Markov decision process). Hier wird der Prozess als Intent Detection dargestellt. Parallel startet die zweite Komponente, die Entity Extraction. Dieses Konzept ist regelbasiert. Als Grundlage zur Auswertung wird meistens eine Wissensbasis (knowledgebase) genommen. Die beiden Systeme liefern dann Ergebnisse, die nochmal bewertet werden, um das optimale Ergebnis auszuwählen. Danach werden die nötigen Funktionen aufgerufen und eine Antwort wird generiert.

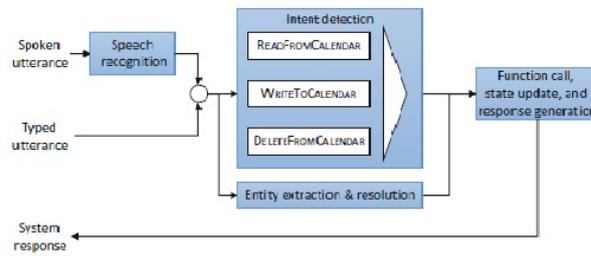


Abbildung 2: Funktionsweise der Sprachassistenten

Quelle JD.Williams

Die unterschiedlichen Frameworks werden wir uns in den folgenden Abschnitten genauer anschauen, da sie von allen großen Sprachassistenten genutzt werden. Teilweise sind beide Modelle in einem Assistenten implementiert, aber es gibt auch Entwickler die sich nur für eine Methode entschieden haben. Für das regelbasierte Framework schauen wir uns Ontologien und den Knowledgegraph an und für das statistische Framework werden wir uns POMDP , das auch von Siri verwendet wird, anschauen.

### 3.1 Ontologien

Ontologien sind eine formale Wissenspräsentation und verwalten ein Wissenssystem. Sie sind eine explizite Spezifikation einer Konzeptualisierung. [19] Das System unterliegt bestimmten Axiomen, die nur Modellierungen, welche legitime Aussagen für die darstellende Welt machen, erlauben soll. In der Wissensbasis werden einzelne Domänen und ihre Beziehungen untereinander modelliert. Daraus entsteht ein semantisches Netzwerk, das Regeln unterworfen ist und sich logisch nachverfolgen lässt. Für Ontologien gibt es Programme, die einem helfen eine Wissensbasis zu modellieren, ein Beispiele wäre Protegé. Ähnlich wie bei Datenbanken gibt es auch eine Abfragesprache wie SQL, für Ontologien ist es z.B. SPARQL.

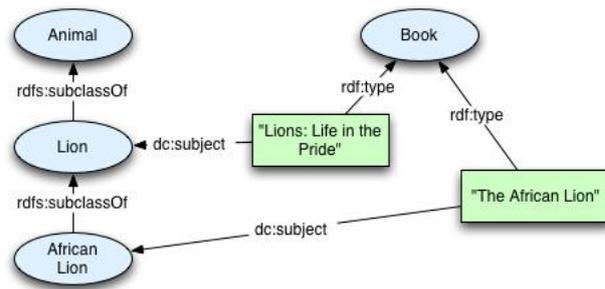
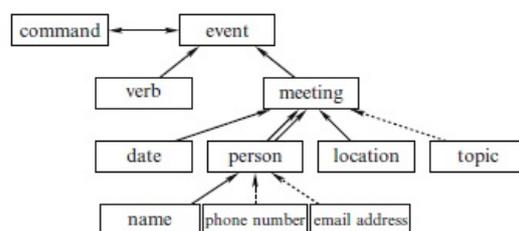


Abbildung 3 Beispiel[16]

Zu den Vorteilen gehören gehören bessere Ergebnisse beim Suchen von Informationen durch die semantische Vernetzung der Daten und das daraus resultierende bessere Verständnis für die Anfragen der Nutzer. Deswegen sind Ontologien eine sehr wichtige Rolle für das Semantic Web, aber auch Sprachassistenten greifen auf die Dienste der Wissensbasis zurück. Im Vergleich zu den klassischen Ontologien, die man in Abbildung 3 z.B. sehen kann, nutzen virtuelle Assistenten wie bspw. Siri aktive Ontologien. Denn die konventionellen Ontologien sind eine statische Datenstruktur und eine formale Repräsentation mit eindeutigen Klassen, Attributen und Beziehungen. Aktive Ontologien sind dynamischer und beinhalten auch noch ein Netzwerk an Konzepten. Das Netzwerk beschreibt zum einen die klassische Datenstruktur und die zu den Domänen zulässigen Regeln für mögliche Aktionen.



**Abbildung 4: Beispiel**  
für eine Ontologie von  
Sprachassistenten  
Quelle: J.R. Bellegarda

Abbildung 4 zeigt uns eine mögliche aktive Ontologie, die von virtuellen Assistenten genutzt wird, für eine Termineintragung. Man sieht, dass es natürlich auch eine große Ähnlichkeit zu den klassischen Ontologien hat, aber es wird mit dieser Modellierung ganz klar definiert wie die Termineintragung aussehen soll. Die Struktur ähnelt einer Baumstruktur. Es gibt die Möglichkeiten von mehreren Pfeilen zwischen zwei Objekten wie z.B. meeting und person, da hier die Kardinalitäten dargestellt werden (zwei Personen sind für ein Meeting nötig). Ebenfalls sind gestrichelte Pfeile möglich, die zeigen, dass bestimmte Attribute optional sind wie z.B. Topic.

### 3.2 Knowledgegraph

Eine weitere Möglichkeit für eine regelbasierte Wissensbasis, die auch von Sprachassistenten implementiert wird, sind Knowledgegraphen. Die Knowledgegraphen sind eine mächtigere und allgemeinere Wissensbasis als die Ontologien, die mehrere

Vorteile mit sich bringt.[20] Ontologien sind für eine gewisse Spezifikation modelliert und sind eine Teilmenge der Knowledgegraphen. Nicht jeder Knowledgegraph ist aber eine Ontologie, da sie meistens das große ganze darstellen. Sie gewannen an großer Bekanntheit durch Googles Vorstellung ihres Knowledgegraphen im Jahr 2012. Neben Google haben auch andere Firmen mittlerweile solche Graphen entwickelt, z.B. Microsoft, Yahoo, LinkedIn und Baidu. Auch in Deutschland wurde unter anderem von deutschen Universitäten mit DBpedia ein bekannter Knowledgegraph entwickelt. Ein Knowledgegraph ist ähnlich wie eine Ontologie ein Wissenssystem, in dem die Informationen semantisch vernetzt sind, um die Intentionen von Suchanfragen besser bearbeiten zu können. Im Vergleich zu Ontologien, die etwas mehr ein Schema repräsentieren, sind Knowledgegraphen faktenorientiert. Ontologien modellieren ihr bestimmtes System mit ihren Domänen auf eine neutrale Art und Weise. Knowledgegraphen können ihre Informationen unter anderem von Ontologien ableiten, aber es ist auch möglich faktenreichere Quellen wie Wikidata und Wikipedia zu nehmen. [17]

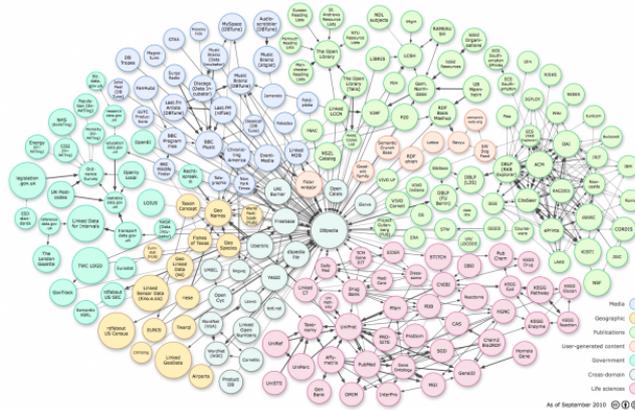


Abbildung 6: Schaubild eines Knowledgegraphen

### 3.3 POMDP

Neben dem regelbasierten Konzept, das aus einer Wissensbasis besteht, nutzen virtuelle intelligente Assistenten oft auch noch das statistische Konzept, das datengetrieben ist. Das ist auch eines der Nachteile des Modells, da es von Daten abhängig ist, dafür lernt es seinen Nutzer bei häufiger Nutzung besser kennen und versteht mit der Zeit besser, welche Wünsche und Vorstellung er hat. Das Modell versucht mit einer empirischen Betrachtung der Informationen die Sprachanfragen zu bearbeiten. Es wurden mehrere Wahrscheinlichkeitsmodelle ausprobiert, damit Programme die Intentionen von natürlicher Sprache nachvollziehen können. Ein Modell, das sich durchgesetzt hatte und auch heute noch verwendet wird, ist POMDP (partially observable Markov decision process). Mit dem Konzept werden verschiedene Entscheidungsprozesse des echten Leben modelliert.

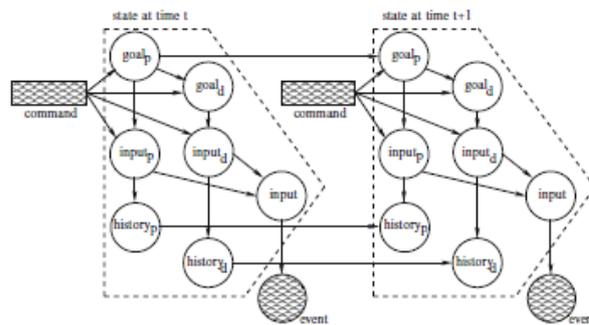


Abbildung 8:  
POMDP Beispiel

In der Abbildung 8 können wir ein POMDP Beispiel für die Terminvereinbarung (ähnlich wie bei der Ontologie in Abb.4), wie sie von Sprachassistenten implementiert wird, betrachten. Man erkennt zeitliche Abschnitte eines dynamischen Netzwerks. Das schraffierte eckige Feld steht für die Anfrage(command) des Nutzers und das rechteckige schraffierte Feld(event) ist das gewünschte Treffen, das eingetragen werden soll. Die Knoten repräsentieren die nötigen Infos des Nutzers, also seinen Input und sein mögliches Ziel (input und goal): d steht für Day(tag) und p für person(Personen).Es ist nur ein vereinfachtes Modell, welches auch noch mit Orten und Themen für die Terminplanung erweitert werden kann. Ein weiterer Knoten ist der History-Knoten,der die Informationen speichert. Bei einer erneuten Anfrage mit gleichem Ziel werden die History-Daten weiter gegeben.

### 3.4 Aufbaumodelle von virtuellen intelligenten Assistenten

Sowohl das regelbasierte Konzept(Ontologien, Knowledgegraphen), als auch das statische Konzept haben ihre Vor- und Nachteile. Einige Sprachassistenten nutzen nur eins der Konzepte , aber es gibt auch welche die beide Methoden implementieren. Ein mögliches Schaubild sieht man dafür in Abbildung 5. Hier wird der Input parallel vom regelbasierten und vom statistischen Modell(POMDP) bearbeitet. Die jeweiligen Daten von Input und möglichen Output werden zudem für die Verbesserung des statistischen Modell im "POMDP Training" erfasst. Im letzten Schritt werden die beiden Outputs verglichen und das bestmögliche ausgewählt. Die Auswahl hängt vom jeweiligen Vertrauen und Glauben an die Qualität von den beiden Systemen an den Output.[14]

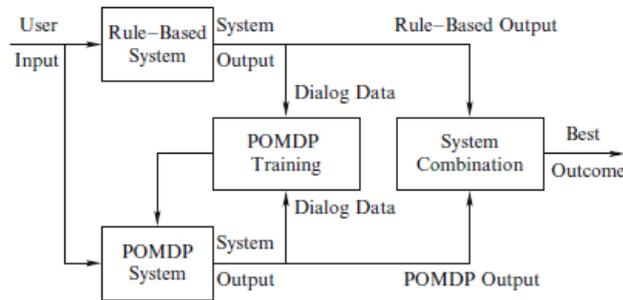


Abbildung 5:  
Quelle: J.R. Bellegarda

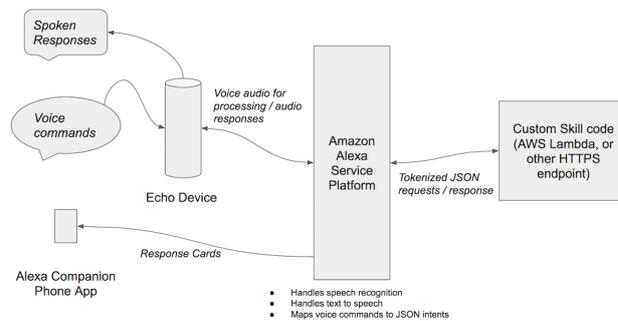


Abbildung 7: Aufbau von Amazon Echo

Bisher haben wir uns den Aufbau der Sprachassistenten etwas theoretisch und abstrakt angeschaut. Als konkretes Beispiel schauen wir uns jetzt Amazons Amazon Echo, das Alexa integriert hat, an und sehen uns an wie der theoretische Aufbau umgesetzt wurde. Die Abbildung 7 dient als Hilfe für eine bessere Vorstellung. Bei Alexa wird die Spracherkennung aktiviert wenn man seine Anfrage mit "Alexa" beginnt, ähnlich wie bei Google und Siri auch mit einem Schlagwort. Die Informationen werden an Alexas Service Platform gesendet, damit sie verarbeitet werden können. Diese Verarbeitungsform ist auch bei anderen Sprachassistenten vorhanden und für solch eine Nutzung ist auch bei allen Assistenten ein Internetzugang eine Voraussetzung. Die Methoden für die Verarbeitung der Informationen sind durch die vorherigen Kapitel erklärt. Danach kommt die Anfrage zum custom skill code, um die passende Methode aufzurufen. Der Weg geht dann über die Platform mit einer Sprachantwort zurück zum User. Ebenfalls gibt Alexa auch eine Nachricht an die gekoppelte Applikation auf dem Smartphone zurück.

## 4 Fazit

Wie schon erwähnt, ist die Technologie der virtuellen intelligenten Assistenten noch sehr jung und noch am Wachsen. Viele Assistenten sind noch in der Entwicklung, wie z.B. M und Viv. Das Interesse an der neuen Technologie ist in der Bevölkerung vorhanden, deswegen versuchen alle großen IT-Unternehmen den virtuellen Assistenten der Zukunft zu entwickeln, sodass die Sprachassistenten zum festen Bestandteil des alltäglichen Lebens gehören. Das werden die Konzerne schaffen, wenn ihre Assistenten mit ihren Leistungen so gute Dienste leisten, dass die Menschen nicht auf diese Dienste verzichten wollen. Ein ähnliches Beispiel für die Entwicklung einer Technologie, die für die Bevölkerung unverzichtbar wurde, ist das Smartphone in den letzten Jahren. In Zukunft soll es nicht mehr nötig sein auf seinem Smartphone Applikationen aufzurufen, da die Assistenten die Dienste nach Anfrage alle ausführen können. Neben dem Wachstum der Sprachassistenten für die allgemeine Nutzung, der von den großen Firmen der IT-Branche vorangetrieben wird, wird wahrscheinlich auch die Anzahl der Assistenten steigen, die nur für eine spezifische Anwendung gedacht ist. Die speziellen virtuellen intelligenten Assistenten können Firmen nutzen, um ihren Kunden einen besseren Service zu bieten, indem sie zum Beispiel die Kundschaft beim Einkauf helfen oder beraten.

## 5 Literatur

1. <https://www.googlewatchblog.de/2016/07/google-assistant-hinweise-assistenten/>
2. <http://www.zdnet.de/88280159/google-assistant-entwickler-koennen-sich-ab-anfang-dezember-integrieren/>
3. <https://www.nytimes.com/2016/01/28/technology/personaltech/siri-alexand-other-virtual-assistants-put-to-the-test.html>
4. <https://www.nytimes.com/2016/01/28/technology/personaltech/siri-alexand-other-virtual-assistants-put-to-the-test.html>
5. <https://www.ft.com/content/4f2f97ea-b8ec-11e4-b8e6-00144feab7de> 6. <http://www.theatlantic.com/technology/archive/2016/05/the-privacy-problem-with-digital-assistants/483950/>
7. <http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html>
8. <http://www.faz.net/aktuell/wirtschaft/japan-versicherer-ersetzt-mitarbeiter-durch-ki-ibm-watson-14605854.html>
9. <https://www.pocketpc.ch/magazin/news/siri-google-now-cortana-sprachassistenten-beliebter-als-gedacht-42451/>
10. <https://www.heise.de/mac-and-i/meldung/Marktforscher-Siri-wird-in-der-Oeffentlichkeit-kaum-genutzt-3228490.html>
11. <http://www.zdnet.de/88280159/google-assistant-entwickler-koennen-sich-ab-anfang-dezember-integrieren/>
12. <https://www.amazon.de/b?node=11242735031>
13. <https://techcrunch.com/2016/05/09/siri-creator-shows-off-first-public-demo-of-viv-the-intelligent-interface-for-everything/> 14. Chapter 1 Spoken Language

- Understanding for Natural Interaction: The Siri Experience Jerome R. Bellegarda 15. Chapter 1 Rapidly Scaling Dialog Systems with Interactive Learning Jason D. Williams, Nabal B. Niraula, Pradeep Dasigi, Aparna Lakshmiratan, Carlos Garcia Jurado Suarez, Mouni Reddy, and Geoff Zweig  
16. <https://www.w3.org/TR/swbp-classes-as-values/fig1.jpg>  
17. <https://www.researchgate.net>  
18. <https://www.heise.de/newsticker/meldung/Bixby-Samsung-Assistent-bekommt-eigenen-Button-bei-Galaxy-S8-3660136.html>  
19.A Translation Approach to Portable Ontology Specifications by Gruber  
20. Knowledge Graph Theory and Structural Parsing by Lei Zhang

# Reverse Geocoding APIs

Yannick Meny\*

Betreuer: Antonios Karatzoglou†

Karlsruher Institut für Technologie (KIT)  
Pervasive Computing Systems – TECO

\*[uceav@student.kit.edu](mailto:uceav@student.kit.edu)

†[antonius@teco.edu](mailto:antonius@teco.edu)

**Zusammenfassung.** Reverse Geocoding beschreibt das Zuordnen von semantischen Informationen zu gegebenen Koordinaten. In dieser Arbeit wurde eine Übersicht und ein Vergleich kostenloser Reverse Geocoding APIs erstellt. Hierzu wurden Koordinaten zu verschiedenen Plätzen festgelegt und getestet, ob die Schnittstellen den korrekten Namen zurückgeben. Es zeigten sich Unterschiede zwischen Sehenswürdigkeiten und lokalen Einrichtungen in der Innenstadt. Insgesamt schnitt Foursquare vor Google und OpenStreetMap am besten ab.

**Schlüsselwörter:** Reverse Geocoding APIs, Übersicht, Evaluation

## 1 Einleitung

Standortbezogene Dienste gewinnen durch das Internet of Things und der zunehmenden Vernetzung von Geräten wie beispielsweise Autos fortlaufend an Bedeutung. Moderne Navigationssysteme sind nicht nur in der Lage, die schnellste Route von einer Start- zu einer Ziel-Adresse zu ermitteln, sondern auch die nahe Umgebung nach bestimmten Orten abzusuchen. So kann im Umkreis nach Restaurants gesucht werden, falls der Nutzer Hunger verspürt.

Um dieses Filtern zu ermöglichen, reichen nicht nur einzelne Koordinaten der Restaurants aus, sondern es sind weitere semantischen Informationen notwendig. So muss das Restaurant auch als solches kategorisiert werden. Für den Nutzer sind daher Informationen wie der Name des Ortes oder die Kategorie (Restaurant, Kino, Bar, Spielplatz etc.) notwendig.

Die Funktionalität, zu festgelegten Koordinaten semantische Informationen abzurufen, bewerkstelligen Reverse Geocoding APIs. Diese nehmen als Eingabe eine Koordinate (Längen- und Breitengrad) entgegen und geben eine Adresse oder einen Namen zurück. Darüber hinaus beinhalten die Antworten der meisten Schnittstellen zusätzliche Informationen wie beispielsweise Öffnungszeiten, Preisniveau oder eine durchschnittliche Nutzerbewertung.

Thema dieser Arbeit ist es, kostenlose Reverse Geocoding APIs zu vergleichen. Zunächst zeigt eine Übersicht auf, welche zusätzlichen Informationen die behandelten Anbieter bereitstellen. Zum Vergleich der Endpunkte wird getestet, ob die Schnittstellen zuvor festgelegte Koordinaten von internationalen sowie

deutschen Sehenswürdigkeiten korrekt in deren Namen umsetzen können. Außerdem wurden Koordinaten von POIs in Karlsruhe vor Ort mit einem Smartphone bestimmt, für die ebenfalls die richtige Zuordnung überprüft wird. Dadurch wird eine Übersicht ermöglicht, in der nicht nur bekannte Sehenswürdigkeiten, sondern auch kleine städtische Unternehmen berücksichtigt werden. Dies bedeutet eine bessere Praxisnähe.

## 2 APIs

Im Nachfolgenden werden Reverse Geocoding APIs behandelt, die kostenlos verfügbar sind. Obwohl diese kostenlos sind, unterscheiden sie sich dennoch darin, wie viele Anfragen in einer bestimmten Zeitspanne getätigt werden können. Zudem setzt die Evaluation dieser Arbeit voraus, dass die getesteten Schnittstellen nicht nur eine Adresse, sondern auch mindestens einen Namen zurückgeben und als Parameter nur die Koordinate und gegebenenfalls ein Suchradius notwendig ist. Das letzte Kriterium erfüllen nicht alle Anbieter; die Benutzung erfordert oft einen zusätzlichen Suchbegriff.

Wie beschrieben geben alle getesteten Endpunkte einen Namen des Ortes als Antwort zurück. Bei der Adresse hingegen stellt GeoNames eine Ausnahme dar. Abgesehen von diesem Anbieter beinhaltet die API-Response immer die Adresse. Im nachfolgenden werden die Datenstrukturen und Inhalte der Requests und Responses der jeweiligen APIs behandelt. Dabei werden nicht alle Parameter und Rückgabewerte einzeln erwähnt, sondern nur solche, die eine Relevanz zum Thema Reverse Geocoding und semantische Informationen aufweisen.

### 2.1 GeoNames

Die Daten der über 10 Millionen geographischen Namen stammen unter anderem von der National Geospatial-Intelligence Agency's (NGA), dem U.S. Board on Geographic Names und weiteren internationalen Ämtern für Statistik.[14] Bei der getesteten API ließ sich keine Datenstruktur erkennen. So wurden Ergebnisse keine Kategorien zugeteilt. Die komplette Datenbank steht auf der Website kostenlos zum Download verfügbar. Zusätzlich kann für eine jährliche Gebühr von 720 Euro ein Premium-Zugang erworben werden. Dieser bietet u. A. die Vorteile, dass die Datenbank Konsistenztests unterzogen und die Änderung der Daten dokumentiert wird.[15]

Für die „find nearby toponym“-API ist neben den Koordinaten ein Username, der bei einer kostenlosen Registrierung festgelegt wird, als Parameter erforderlich. Zusätzlich kann ein Radius in km angegeben werden, in dem die Ergebnisse liegen sollen. Als Antwort liefert GeoNames den Namen des nächstgelegenen Point of Interest ausgehend von den angegebenen Koordinaten. Zusätzlich werden die Koordinaten des POIs, sowie die Distanz in Kilometern und das Land zurückgegeben.[16]

## 2.2 Here

Der Datensatz von Here kann Unternehmen nach vorheriger Absprache zur Verfügung gestellt werden.[21] Dessen Datenstruktur zeichnet sich durch eine Kategorisierung der POIs aus.

Als Parameter für eine gültige Abfrage sind die Koordinaten sowie ein API-Key notwendig, der nach einer kostenlosen Registrierung erstellt werden kann. Durch zusätzliche Parameter können mögliche Ergebnisse weiter eingegrenzt werden. So kann nach bestimmten Kategorien gefiltert oder nur Ergebnisse in einem bestimmten Bereich (Kreis, Bounding Box, ...) angezeigt werden. Außerdem lässt sich die maximale Ergebnisanzahl festlegen. Nach einer Abfrage werden Name, Adresse, Koordinaten und Distanz zum POI zurückgegeben. Darüber hinaus enthält die API-Response eine durchschnittliche Bewertung und die zugeordneten Kategorien.[22]

## 2.3 OpenCage

OpenCage bezieht seine Daten von OpenStreetMap, Yahoo! GeoPlanet, Natural Earth Data, Thematic Mapping und weiteren Diensten.[26] Dabei unterliegen die Daten keiner Strukturierung. Es existieren also keine Kategorien für die Points of Interests.

Zu den erforderlichen Parametern zählen die Koordinaten und ein nach Registrierung erhaltener API-Key. Zusätzlich ist die Festlegung des Ausgabeformats notwendig. Dazu gehören neben XML und JSON auch eine HTML-Seite, auf der der erhaltene Ort auf einer Karte angezeigt wird. Neben den erforderlichen lassen sich auch zusätzliche Parameter festlegen. So ist es möglich, die Ergebnisse auf solche in einer Bounding Box zu begrenzen. Außerdem ist die Ergebnisanzahl limitierbar; Standard sind zehn Ergebnisse. OpenCage erlaubt es zudem, eine minimale Confidence festzulegen. Es handelt sich hierbei um eine Zahl zwischen 0 und 10, die die maximale Distanz zwischen den ausgehenden Koordinaten und denen des Ergebnisses beschreibt.

In der Response der API finden sich Name, Adresse und Koordinaten des POI. Zudem gibt der Typ an, ob es sich um einen POI handelt oder es eine Straße, Stadt oder Ähnliches ist. Der Confidence-Wert, der auch im Request festgelegt werden kann, wird auch bei der API-Antwort zurückgegeben. Darüber hinaus enthält diese ortsspezifische Informationen wie Vorwahl, Währung, Zeitzone und Sonnenauf- und -Untergangszeit. Da zu den Datenquellen unter anderem OpenStreetMap zählt, wird zusätzlich die OpenStreetMap-URL für den entsprechenden POI angegeben.[27]

## 2.4 Mapbox

Mapbox basiert grundlegend auf den Daten von OpenStreetMap.[29] Die POIs werden mithilfe verschiedener Kategorien gegliedert.

**Tabelle 1.** Confidence-Ebenen

Wert	Genauigkeit
0	keine Bounding Box
1	25+ km Distanz
2	< 25 km Distanz
3	< 20 km Distanz
4	< 15 km Distanz
5	< 10 km Distanz
6	< 7,5 km Distanz
7	< 5 km Distanz
8	< 1 km Distanz
9	< 0,5 km Distanz
10	< 0,25 km Distanz

Erforderliche Parameter zur Nutzung des Reverse Geocoding - Endpunktes sind lediglich Koordinaten und ein API-Key. Zusätzlich wird erlaubt, einen nahegelegenen Ort in Form von Koordinaten festzulegen, anhand dessen sich die Ergebnisse orientieren sollen. Des Weiteren ermöglicht das Festlegen auf bestimmte Ergebnistypen das Filtern der zurückgegebenen Orte. Zu besagten Typen zählen unter anderem Land, Postleitzahl, Adresse und POI. Auch kann eine Bounding Box festgelegt werden, in der sich mögliche Ergebnisse befinden müssen. Standardmäßig folgt auf eine Anfrage genau ein Ergebnis. Diese Anzahl lässt sich ebenfalls anpassen.

Neben Informationen wie Name, Adresse, ISO 3166-1 Land- und ISO 3166-2 Regioncodes wird ein Wikidata Identifier zurückgegeben. Außerdem beinhaltet die API-Response eine Liste der Kategorien, die dem jeweiligen Ergebnis zugeordnet sind. Diese Kategorien werden dafür verwendet, ein Icon des Maki-Projekts auszuwählen und mitzuliefern. Zudem benutzt Mapbox einen booleschen „Landmark“-Wert, der anzeigt, ob ein Ort ein Wahrzeichen ist. Dazu gehören zum Beispiel Schulen, Parks und Museen.[24]

## 2.5 OpenStreetMap

Die Daten von OpenStreetMap sind nutzerbasiert und werden von diesen überprüft. Eine neue Version des Datensatzes wird wöchentlich veröffentlicht und kann kostenlos heruntergeladen werden.[32] Strukturiert werden Orte durch „Nodes“, die durch aus einem Identifier und den jeweiligen Koordinaten bestehen. Zusätzlich hat jede Node typischerweise mindestens einen Tag, der zur Beschreibung von geographischen Eigenschaften, so genannten „Map Features“, dient.[30]

Die Reverse Geocoding – Funktionalität auf den OpenStreetMap-Daten stellt das Tool Nominatim bereit. Erforderlich für eine Abfrage sind lediglich Koordinaten, auf einen API-Key wird verzichtet. Durch weitere Parameter kann das Ausgabeformat festgelegt werden. Default ist XML, aber auch JSON ist möglich.

Außerdem erlaubt ein Zoom-Wert die erforderliche Genauigkeit der Ergebnisse festzulegen. Dabei handelt es sich um einen Wert zwischen 0 und 18, wobei 0 der API signalisiert, dass der Name des Landes hinreichend ist, steht der Wert 18 für ein bestimmtes Gebäude. Durch weitere Flags können sowohl zusätzliche Informationen wie Wikipedia-Link oder Öffnungszeiten, als auch Namensdetails angefordert werden. Zu Namensdetails gehören alternative Namen des Ergebnisses sowie verschiedene Sprachvariationen und ähnliches.

Bei einer gültigen Abfrage enthält die API-Response den Namen, die Adresse und den Identifier des nächstgelegenen Ortes. Falls die entsprechenden Parameter gesetzt wurden und die jeweiligen Informationen verfügbar sind, werden die jeweiligen Zusatzinformationen und Namensdetails ebenfalls zurückgegeben.[31]

## 2.6 Yelp

Die Gliederung der Daten, die Yelp zur Verfügung stellt, erfolgt durch eine Vielzahl an hierarchisch aufgebauter Kategorien. Zudem handelt es sich vorwiegend um eine Datenbank von Unternehmen und nicht Points of Interest. Ein Teil des Datensatz steht zum kostenlosen Download bereit. Dieser beinhaltet unter anderem Karlsruhe als einzige deutsche Stadt.[38]

Vorausgesetzt werden bei einem Request ein Koordinatenpaar und ein API-Key, der nach der Registrierung erstellt wird. Zusätzlich erlaubt Yelp die Festlegung eines Umkreises, in der potentielle Ergebnisse liegen sollen. Außerdem können Ergebnisse gefiltert werden. Zu den angebotenen Filtern zählt ein Kategorie-, ein Preisniveau-, „gerade geöffnet“- sowie ein Attribut-Filter. Als Attribut kann beispielsweise „waitlist\_reservation“ angegeben werden, um ausschließlich Unternehmen anzuzeigen, die eine Online-Reservierung anbieten. Neben dem Filtern der Ergebnisse nach solchen, die gerade geöffnet sind, wird auch die Suche nach Unternehmen, die zu einem bestimmten Zeitpunkt geöffnet haben, ermöglicht. Dies entspricht einem weiteren optionalen Parameter. Zudem lässt sich die maximale Ergebnisanzahl festlegen. Darüber hinaus stehen verschiedene Sortierarten zur Verfügung: Höchste Übereinstimmung (bei Textsuche), Bewertung, Anzahl an Reviews und Distanz.

Die API-Response beinhalten einen Identifier, den Namen, die Adresse und die Koordinaten der nahegelegenen Unternehmen sowie die Distanz zu der ausgehenden Koordinate. Außerdem stehen zusätzliche Informationen wie Preisniveau, Telefonnummer, Bewertung und Reviewanzahl zur Verfügung. Zur korrekten Verlinkung enthält die Response die URL des Yelp-Eintrags und eine URL zu einem festgelegten Foto des Unternehmens.[36]

## 2.7 Foursquare

Die bei Foursquare hinterlegten Daten werden in verschiedene Kategorien unterteilt. Diese sind hierarchisch aufgebaut. So werden Top-Level Kategorien in Sub- und Sub-Sub-Kategorien unterteilt.[9] Außerdem erlaubt Foursquare die Abfrage, welche Sub-Kategorien von einer gegebenen Top-Level Kategorie

abstammen.[13] Datensätze der Datenbank werden offiziell nicht verbreitet. Allerdings können veraltete und nicht vollständige Datensätze von anderen Webseiten kostenlos bezogen werden.[1]

Zu den erforderlichen Parametern bei Nutzung der Foursquare Serach API zählen ein Koordinatenpaar und ein API-Key. Darüber hinaus können die potentiellen Ergebnisse mit einem bestimmten Begriff verglichen werden, der per optionalem Parameter übergeben wird. Außerdem kann die Ergebnisanzahl oder der geographische Radius, in dem sich die gefundenen Orte befinden sollen, eingegrenzt werden. Die Vielzahl an Kategorien erlaubt ein weiteres Filtern, indem die Ergebnisse auf in der Abfrage festgelegte Kategorien reduziert werden können. Ein besonderes Feature der Foursquare-API ist ein zusätzlicher Parameter, der es erlaubt, eine URL bereitzustellen, mit der die Ergebnisse abgeglichen werden. Um diese Funktionalität umsetzen zu können, existiert serverseitig eine Abbildung, die einem Ort eine URL von Foursquare zuordnet.

Als Response werden Name, Adresse und Koordinaten des Ortes zurückgegeben. Zusätzlich wird die Distanz zur ursprünglichen Koordinate berechnet. Enthalten ist zudem ein Array von Kategorien, die dem POI zugeordnet sind. Ein boolescher Wert repräsentiert außerdem, ob der Besitzer des Unternehmens die Informationen verifiziert hat. Weitere zusätzliche Informationen sind:

- Webseite
- Öffnungszeiten
- beliebte Zeiten
- Preisniveau
- Bewertung
- Beschreibung
- Tags
- Anzahl an Besuchen
- Fotos
- Likes und Dislikes
- Attribute wie Reservierung- und Parkmöglichkeiten

Darüber hinaus sind Statistiken enthalten, die zeigen, wie viele Check-Ins getätigt wurden, wie viele verschiedene Nutzer ein Check-In getätigt haben und wie viele Tips ausgesprochen wurden.[11] Check-In bezeichnet das Feature von Foursquare, das dem Nutzer ermöglicht, seinen Standort zu ändern. So kann er beispielsweise das Restaurant auswählen, in dem er sich gerade befindet und diesen Besuch kommentieren oder ein Foto einfügen.[7] Als Tips bezeichnet Foursquare Empfehlungen, die Nutzer nach ihren Erfahrungen aussprechen. Zum Beispiel gibt ein Nutzer den Tipp, ein Café zu bestimmten Öffnungszeiten zu meiden, weil der Ansturm zu groß ist.[12]

## 2.8 Google Places

Google benutzt für die Struktur seiner Daten Ortstypen, die den Kategorien anderer Anbieter ähneln. Ein hierarchischer Aufbau ist allerdings nicht implementiert. Einem Ort werden folglich die Typen zugeordnet, die ihn näher beschreiben.[18]

Die erforderlichen Parameter unterscheiden sich bei einer Abfrage abhängig davon, welche Sortierung der Ergebnisse gewählt wird. Allerdings sind in beiden Fällen ein API-Key und die Koordinaten erforderlich. Zudem muss der Datentyp der Antwort festgelegt werden. Zur Auswahl stehen XML und JSON. Google sortiert standardmäßig nach „Prominence“, was einen Radius, in dem die Ergebnisse liegen sollen, als Parameter fordert. Dabei wird sich an der Beliebtheit der Orte orientiert, die „durch die Bewertung eines Ortes im Index von Google, die weltweite Bekanntheit und weitere Faktoren beeinflusst“ [19] wird. Wird hingegen nach Distanz zur ursprünglichen Koordinate sortiert, so muss mindestens einer der zusätzlichen Parameter „keyword“, „name“ oder „types“ gesetzt werden. Keyword und Name sind jeweils ein Begriff, der mit dem Ergebnis abzugleichen ist. Allerdings wird der Name nur mit dem Ortsnamen und das Keyword mit allen verfügbaren Informationsinhalten abgeglichen. Types dient zur Einschränkung der Orte auf bestimmte Ortstypen. Die weiteren zusätzlichen Parameter erlauben, das Preisniveau einer Location einzuschränken oder nur diejenigen anzuzeigen, die zum Zeitpunkt der Abfrage geöffnet sind.

Eine Antwort beinhaltet den Namen, die Adresse und Koordinaten der Orte. Zudem stehen Öffnungszeiten und die Information, ob zum jetzigen Zeitpunkt geöffnet ist, zur Verfügung, falls diese hinterlegt wurden. Auch das Preisniveau, das durch einen Wert zwischen 0 und 4 wiedergespiegelt wird, und die durchschnittliche Bewertung lassen sich abrufen. Die anfangs genannten Ortstypen werden ebenfalls zurückgegeben. [19]

## 2.9 Limitationen

Auch wenn als grundlegende Voraussetzung für das Betrachten einer API in dieser Arbeit die kostenlose Verfügbarkeit ist, unterscheiden sich die Nutzungseinschränkungen je nach Anbieter stark.

**Tabelle 2.** Limitationen in Form von Anfragen im Monat

Anbieter	kostenlos	kostenpflichtig
GeoNames	900.000 [15]	
Here	15.000	100.000 - 1.000.000+
OpenCage	75.000 [28]	300.000 - 30.000.000 [28]
Mapbox	25.920.000	25.920.000 - 103.680.000+
OpenStreetMap	„keine schwere Nutzung“ [6]	
Yelp	750.000 [37]	auf Anfrage [37]
Foursquare	3.600.000	auf Anfrage
Google	30.000 - 4.500.000	benutzerdefiniert [20]

Die Angaben der Tabellen sind auf den Zeitraum eines Monats bezogen. Da nicht alle API-Anbieter diesen Zeitraum als Limitation nutzen, wurde, um einen

besseren Vergleich zu ermöglichen, die Angaben umgerechnet. So erlaubt Mapbox beispielsweise kostenlose 600 Anfragen pro Minute[25], was bei maximaler, konstanter Auslastung 25.920.000 Anfragen pro Monat entspricht.

Bei maximaler, konstanter Auslastung ermöglicht folglich Mapbox die meisten Anfragen im Monat. Danach folgen Google mit bis zu 4.500.000[17] und Foursquare mit 3.600.000[10] Anfragen monatlich. Die wenigsten Anfragen erlaubt die API von Here. Allerdings bezieht sich die Angabe von 15.000 Anfragen im Monat[23] auch tatsächlich auf den Zeitraum eines Monats. Folglich wird nicht vorausgesetzt, eine konstante Auslastung zu verwenden. Diese Variabilität ermöglicht das Szenario, alle Anfragen an einem Tag des Monats zu benutzen, während an den restlichen Tagen keine Anfrage gestellt wird.

Werden die kostenpflichtigen Angebote der jeweiligen Unternehmen betrachtet, lässt sich keine maximale Anfragenanzahl ausmachen, da diese meist durch individuelle Anfragen an die Bedürfnisse der Kunden angepasst werden.

## 2.10 Nicht berücksichtigte APIs

Neben den bereits beschriebenen Reverse Geocoding APIs waren weitere verfügbar, die allerdings nicht getestet werden konnten. So wird man beim Anbieter Factual[5] beim Erstellen eines API-Keys lediglich auf das Kontaktformular verwiesen und keine automatische Zuweisung findet statt.[4] Außerdem wurde die manuell gestellte Anfrage für einen API-Key nicht beantwortet, was das Testen nicht ermöglichte.

Beim Anbieter Yahoo BOSS Geo Services[2] konnte ebenfalls kein API-Key erstellt werden. Auf der dafür vorgesehenen Seite wird „The requested page was not found.“ gemeldet.[35]

Eine weitere Schnittstelle, die nicht berücksichtigt wurde, wird von Whitepages Pro[33] betrieben. Dieser Anbieter verwendet keine weltweite Daten, sondern nur solche aus den USA und Kanada. Daher erscheint nach dem Versuch der Registrierung für einen API-Key die Meldung „Sorry, Whitepages Pro is not yet available in your country“.[34]

## 2.11 Übersicht

**Tabelle 3.** Übersicht der behandelten APIs

Anbieter	Zusatzparameter	Zusatz- informationen	Limits	Datentyp
GeoNames	Radius	Distanz	900.000	XML
Here	Umkreiseingrenzung	Bewertung	15.000	JSON
	Kategorien-Filter	Typ		
	Ergebnisanzahl	Kategorien		

Anbieter	Zusatzparameter	Zusatz- informationen	Limits	Datentyp
OpenCage	Bounding Box	Confidence	75.000	XML
	Ergebnisanzahl	Vorwahl		JSON
	Confidence	Währung		HTML
		Zeitzone Sonnenauf- und untergang		
Mapbox	Orientierung	Wikidata ID	25.920.000	JSON
	Filter	Kategorien		
	Bounding Box	Maki Icon		
	Ergebnisanzahl	Landmark		
OpenStreetMap	Ausgabeformat	Wikipedia-Link	„keine	XML
	Zoom	Öffnungszeiten	schwere	JSON
	alternative Namen	Namensdetails	Nutzung “	
	Extrainformationen			
Yelp	Radius	Foto-URL	750.000	JSON
	Filter	Öffnungszeiten		
	Ergebnisanzahl	Preisniveau		
	Sortierungsart	Telefonnummer		
		Bewertung		
		Distanz Kategorien		
Foursquare	Keyword	Kontaktdaten	3.600.000	JSON
	Ergebnisanzahl	Distanz		
	Radius	Kategorien		
	Kategorien-Filter	Verifikation		
	URL-Abgleich	Statistiken		
Google	Radius	Öffnungszeiten	30.000	JSON
	Keyword	Fotos	bis	XML
	Filter	Preisniveau	4.500.000	
	nur geöffnete Orte	Bewertung		
	Sortierung	Ortstyp		

### 3 Evaluation

Um die zuvor vorgestellten Reverse Geocoding APIs zu vergleichen, wurden drei Testverfahren festgelegt. Dabei wurden zunächst Koordinaten festgelegt, die den Endpunkten übergeben wurden. So wurde bestimmt, wie oft ein API-Anbieter einer Koordinate den korrekten Ort zuordnet. Falls mehrere Ergebnisse zurückgegeben wurden, wurde stets das erste als Vergleichswert benutzt.

Zunächst werden von zehn Großstädten weltweit jeweils zwei verschiedene Orte ausgesucht. Danach folgt die Auswertung der Ergebnisse zu Koordinaten in deutschen Städten. Hier wurden in fünf Städten jeweils drei Orte ausgewählt. Hierbei erfolgte die Auswahl der Orte mithilfe der Rangliste von TripAdvisor. So wurden die jeweiligen höchst-platzierten Aktivitäten übernommen, falls sich für diese Koordinaten bestimmen lassen. Folglich wurden etwa Rundreisen oder ähnliches nicht berücksichtigt. Als Referenzwert wurden die Koordinaten dieser Orte mithilfe von Google Maps erstellt. Somit wird Googles Places API bei den ersten beiden Tests nicht berücksichtigt.

Das Ergebnis zu einer Koordinate wird nur dann als korrekt gewerten, wenn neben dem Ortsnamen auch die erhaltene Kategorie mit dem vorgegebenen Ort übereinstimmt. So ist die Zuordnung Kolosseum-Touren zum eigentlichen Ort des römischen Kolosseums nicht als zutreffend zu werten, da es sich beiden Touren um eine Dienstleistung, aber beim Kolosseum selbst um eine Sehenswürdigkeit handelt.

Sowohl Google als auch Foursquare erlauben es, ihre Ergebnisse nach Distanz zur ursprünglichen Koordinate oder Popularität der einzelnen Vorschläge zu ordnen. Im Folgenden wird der Durchschnitt beider Optionen als Vergleichswert herangezogen.[8] [19]

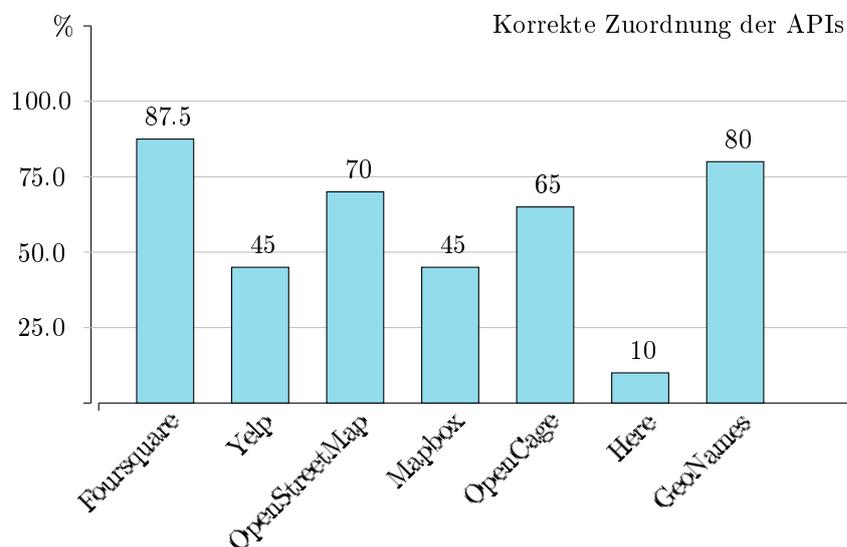
#### 3.1 Internationale Orte

##### Top-Aktivitäten

1. New York
  - (a) Central Park
  - (b) 9/11 Memorial
2. Rio de Janeiro
  - (a) Zuckerhut
  - (b) Christus-Erlöser-Statue
3. Sydney
  - (a) Bondi to Coogee Beach Coastal Walk
  - (b) Sydney Opera House
4. Hongkong
  - (a) Star Ferry
  - (b) Victoria Peak
5. Moskau
  - (a) Roter Platz
  - (b) Basilius-Kathedrale
6. Rom

- (a) Kolosseum
- (b) Pantheon
- 7. Madrid
  - (a) Museo del Prado
  - (b) Retiro-Park
- 8. Paris
  - (a) Musée d'Orsay
  - (b) Eiffelturm
- 9. London
  - (a) British Museum
  - (b) National Gallery

Die Aktivität an erster Stelle in Hongkong ist die Hongkonger Skyline. Diese kann offensichtlich nicht als spezifischer Ort festgelegt werden. Folglich wurden die darauf folgenden Aktivitäten als Vergleichsobjekt ausgewählt.



Das beste Ergebnis erzielte Foursquares Endpunkt mit 87,5% korrekt zugeordneten Koordinaten. Danach folgt GeoNames, das 80 Prozentpunkte erzielte. Abgeschlagen auf dem letzten Platz liegt die Here-API mit 10%.

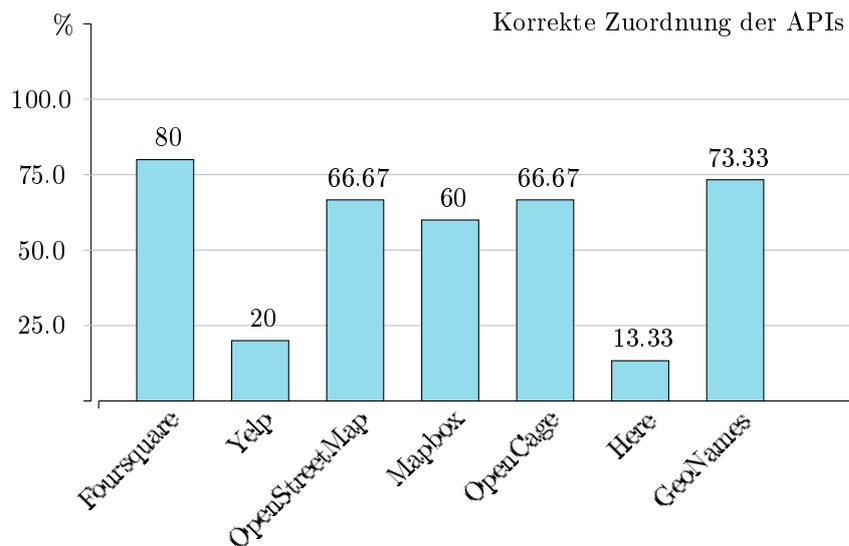
### 3.2 Orte in deutschen Großstädten

#### Top-Aktivitäten

1. Berlin
  - (a) Reichstagsgebäude
  - (b) Brandenburger Tor
  - (c) Gedenkstätte Berliner Mauer
2. München

- (a) Englischer Garten
- (b) Marienplatz
- (c) BMW-Museum
- 3. Stuttgart
  - (a) Mercedes-Benz-Museum
  - (b) Porsche-Museum
  - (c) Wilhelma
- 4. Köln
  - (a) Kölner Dom
  - (b) Kölner Zoo
  - (c) KölnTriangle
- 5. Dresden
  - (a) Frauenkirche
  - (b) Zwinger
  - (c) Grünes Gewölbe

Wie bereits bei den internationalen Aktivitäten waren in Deutschland ebenfalls einige nicht messbar. So war der erste Platz in Berlin „Private Touren“ und Top-Aktivität zwei in München Sightseeing-Touren.



Die Auswertung zeigt, dass Foursquare mit 80% die meisten korrekten Ergebnisse zurückgab. Während fünf von sieben APIs 60% und mehr erreichten, liegen Yelp mit 20% und Here mit 13,33% auf den letzten Rängen. Im Falle von Yelp liegt die Begründung darin, dass sich der Service vorrangig an Unternehmen wie Restaurants und Dienstleistern orientiert. Somit werden Sehenswürdigkeiten im Datensatz meist vernachlässigt.

### 3.3 Orte in Karlsruhe

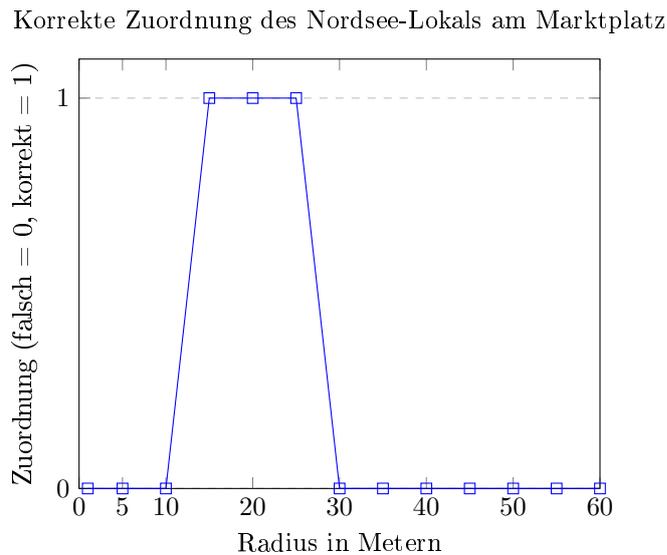
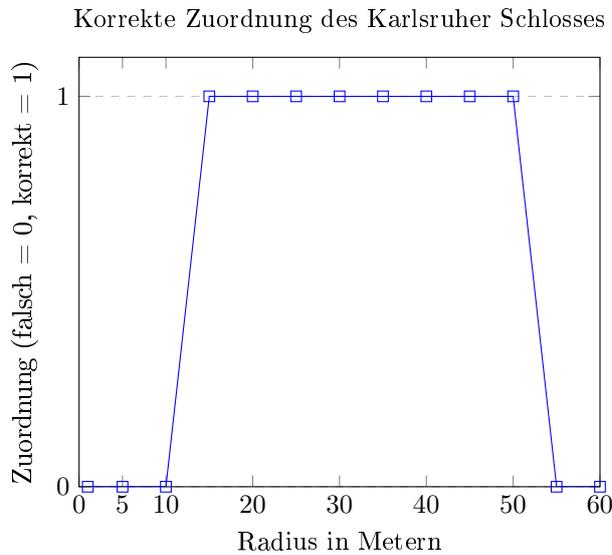
#### Gewählte Points of Interests

1. C&A am Marktplatz
2. McDonald's am Marktplatz
3. Nordsee am Marktplatz
4. Volksbank am Marktplatz
5. Peak & Cloppenburg
6. Karstadt
7. Deutsche Bank
8. Breuninger
9. Talia
10. Starbucks am Europaplatz
11. Fasanengarten
12. Schloss
13. Hauptbahnhof
14. Badisches Staatstheater
15. Wildparkstadion

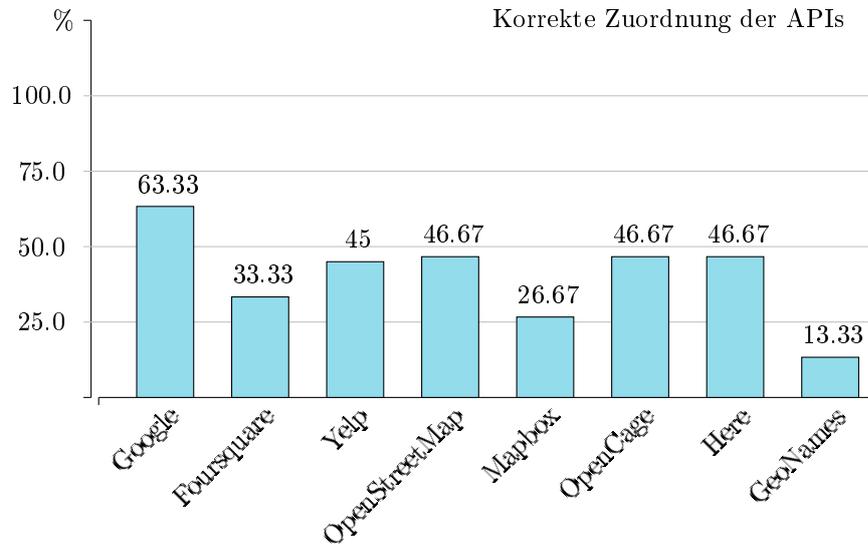
Die Orte 1-10 befinden sich in der Kaiserstraße in Karlsruhe. Folglich sind diese direkt von einer Vielzahl an anderen Unternehmen umgeben. Im Gegensatz dazu wurden die POIs 11-15 so ausgewählt, dass die Dichte der Locations in der direkten Umgebung niedriger ist. Außerdem lassen sich Orte wie Parks oder Stadien, die sich über Areal strecken, nicht ohne Weiteres einer einzelnen Koordinate zuordnen wie ein Restaurant.

Die Bestimmung der Koordinaten erfolgte mithilfe eines Samsung Galaxy S5 Smartphones. Ferner wurde die Android-App „GPS Status & Toolkit“<sup>[3]</sup> verwendet, laut der die Abweichung der GPS-Daten bei den Messungen zwischen drei und fünf Metern betrug. Ermittelt wurden die Koordinatenpaare für die Läden in der Kaiserstraße vor deren Eingang. Bei Orten, die sich hingegen über eine größere Fläche strecken, fand die Messung in einem zentral gelegenen Punkt statt.

Da die Prominence-Suche von Google einen Radius erfordert, bei Foursquare allerdings nur optional ist, wurde er bei beiden APIs auf denselben Wert gesetzt.



Die obigen Abbildungen zeigen, dass bei Orten, die von wenigen anderen Places umgeben werden, ein größerer Bereich von Radien zum Erfolg führt, während in dichteren Umgebungen ein Radius von 10-30 Metern die besten Ergebnisse erzielt. Folglich wurde der Radius für das Schloss und den Fasanengarten auf 50 Meter und der des Stadions aufgrund dessen Größe auf 150 Meter festgelegt. Alle anderen Anfragen begrenzten sich auf 25 Meter.



Zunächst ist hervorzuheben, dass die Foursquare-Suche nach Popularität lediglich einer der Koordinaten den korrekten Ort zuordnet und somit eine Trefferquote von 6,67% aufweist, was das durchschnittliche Ergebnis stark beeinträchtigt. So beträgt die Quote bei der Distanzsuche 60%. Folglich liegt der Durchschnitt bei lediglich 33,33%.

Die Google Places - API konnte die meisten Koordinaten mit durchschnittlich 63,33% richtig zuordnen. Dabei konnte die Distanzsuche 53,33% und die Suche nach Prominenz sogar 73,33% korrekte Ergebnisse verzeichnen. Danach folgen OpenStreetMap, OpenCage und Here mit jeweils 46,67%. Am schlechtesten schnitt GeoNames ab. Ordneten die anderen Endpunkte den Koordinaten bei einem falschem Ergebnis meist einen anderen, nicht zutreffenden Ort zu, so wurde bei GeoNames größtenteils kein Ort in der Nähe gefunden und lediglich die Stadt Karlsruhe ausgegeben.

Auffällig ist das Ergebnis von Here, da die Quote deutlich höher ist, als bei den bisherigen beiden Test. Eine denkbare Begründung ist, dass die intern verwendeten Koordinaten der API zuvor oft zu stark von den festgelegten Referenz-Koordinaten von Google abwichen. Denn die richtigen Locations befanden sich meist erst an zweiter oder dritter Stelle des Rückgabe-Arrays, was zu einem inkorrekten Wert in der Evaluation führte. Bei den Orten in Karlsruhe hingegen zeigte sich, dass im Umkreis einer Koordinate nur wenige bis gar keine Ergebnisse liegen, sodass an erster Stelle öfter die korrekte Zuordnung positioniert war.

## 4 Fazit

Das Testen der verfügbaren, kostenlosen Reverse Geocoding APIs zeigte, dass diese sich in ihrer Konstanz und ihrem Einsatzbereich stark unterscheiden. So

ordnete GeoNames auf der Welt verteilten Koordinaten von Sehenswürdigkeiten in 80% der Fälle den richtigen Ort zu, während der Wert bei der Eingabe von Orten in Karlsruhe auf 13,33% sank.

Die Ermittlung der Koordinaten in Karlsruhe zeigte sich als unpräzise. So wichen diese teilweise zehn Meter von dem Referenz-Standort von Google ab, obwohl die Android-App eine Abweichung von höchstens fünf Metern anzeigte. Zu zeigen wäre folglich, ob sich die Ergebnisse der getesteten APIs signifikant verändern würden, falls ein präziserer GPS-Empfänger eingesetzt wird.

Zusammenfassend ergibt sich das Bild, dass die APIs zwar einen Großteil, aber nicht die vollständige Reverse Geocoding - Funktionalität ausmacht. So hängt es ebenfalls davon ab, wie präzise die vom Endgerät bestimmten Koordinaten sind. Demzufolge sind zuverlässige GPS-Empfänger ebenso wichtig einzustufen wie zuverlässige Programmierschnittstellen.

## Literaturverzeichnis

1. Archive, T.I.: Umn/sarwat foursquare dataset, [https://archive.org/details/201309\\_foursquare\\_dataset\\_umn](https://archive.org/details/201309_foursquare_dataset_umn)
2. BOSS, Y.: Yahoo boss geo services, <https://developer.yahoo.com/boss/geo/#overview>
3. EclipSim, M.: Gps status & toolbox, <https://play.google.com/store/apps/details?id=com.eclipsim.gpsstatus2&hl=de>
4. Factual: Get started, <http://developer.factual.com/get-started/>
5. Factual: Global data. local context., <http://www.factual.com/>
6. Foundation, O.: Nominatim usage policy, <https://operations.osmfoundation.org/policies/nominatim/>
7. Foursquare: Check-ins, <https://support.foursquare.com/hc/en-us/articles/201065340-Check-ins>
8. Foursquare: Explore recommended and popular venues, <https://developer.foursquare.com/docs/venues/explore>
9. Foursquare: Foursquare category hierarchy, <https://developer.foursquare.com/categorytree>
10. Foursquare: Rate limits, <https://developer.foursquare.com/overview/ratelimits>
11. Foursquare: Search venues, <https://developer.foursquare.com/docs/venues/search>
12. Foursquare: Tips, <https://support.foursquare.com/hc/en-us/articles/201065360-Tips->
13. Foursquare: Venue categories, <https://developer.foursquare.com/docs/venues/categories.html>
14. GeoNames: About geonames, <http://www.geonames.org/about.html>
15. GeoNames: Geonames data, <http://www.geonames.org/export/>
16. GeoNames: Geonames web services documentation, <http://www.geonames.org/export/web-services.html>
17. Google: Nutzungsbeschränkungen und abrechnung, <https://developers.google.com/places/web-service/usage>
18. Google: Ortstypen, [https://developers.google.com/places/supported\\_types?hl=de](https://developers.google.com/places/supported_types?hl=de)

19. Google: Places api web service, <https://developers.google.com/places/web-service/search?hl=de>
20. Google: Premium plan usage rates and limits, <https://developers.google.com/maps/premium/usage-limits>
21. Here: Here map data, <https://here.com/en/products-services/data/here-map-data>
22. Here: Places (search) api developer's guide, [https://developer.here.com/rest-apis/documentation/places/topics\\_api/resource-here.html](https://developer.here.com/rest-apis/documentation/places/topics_api/resource-here.html)
23. Here: Public plans, <https://developer.here.com/plans>
24. Mapbox: Mapbox api documentation, <https://www.mapbox.com/api-documentation>
25. Mapbox: Plans & pricing, <https://www.mapbox.com/pricing/>
26. OpenCage: Credit is due, <https://geocoder.opencagedata.com/credits>
27. OpenCage: The opencage geocoder api, <https://geocoder.opencagedata.com/api>
28. OpenCage: Pricing, <https://geocoder.opencagedata.com/pricing>
29. OpenStreetMap: Mapbox, <http://wiki.openstreetmap.org/wiki/Mapbox>
30. OpenStreetMap: Node, <http://wiki.openstreetmap.org/wiki/Node>
31. OpenStreetMap: Nominatim, <http://wiki.openstreetmap.org/wiki/Nominatim>
32. OpenStreetMap: Planet.osm, <http://wiki.openstreetmap.org/wiki/Planet.osm>
33. Pro, W.: Global identity verification data, <https://pro.whitepages.com/>
34. Pro, W.: Start your free api trial, [https://pro.whitepages.com/developer/lp/start-api-trial/?plan\\_param=reverse\\_address&utm\\_campaign=pro-api-trial-reverseaddress](https://pro.whitepages.com/developer/lp/start-api-trial/?plan_param=reverse_address&utm_campaign=pro-api-trial-reverseaddress)
35. Yahoo: The requested page was not found., <https://developer.yahoo.com/dashboard/createKey.html>
36. Yelp: /businesses/search, [https://www.yelp.com/developers/documentation/v3/business\\_search](https://www.yelp.com/developers/documentation/v3/business_search)
37. Yelp: Faq, <https://www.yelp.com/developers/faq>
38. Yelp: Yelp dataset challenge, [https://www.yelp.com/dataset\\_challenge](https://www.yelp.com/dataset_challenge)

# Sensory Augmentation with Tactile Interfaces

Robin S. Schnaidt\*

Advisor: Erik Pescara†

Karlsruhe Institute of Technology (KIT)

Pervasive Computing Systems – TECO

\*[robin.schnaidt@student.kit.edu](mailto:robin.schnaidt@student.kit.edu)

†[pescara@teco.edu](mailto:pescara@teco.edu)

**Abstract.** This survey paper’s purpose is to present a conclusive look at tactile sensory augmentation research as well as its applications, with a HCI and ubiquitous computing perspective in mind. Beginning with early studies in the related field of sensory substitution, basic ideas and concepts are explained with equal focus on interface technology and the cognitive characteristics of man that allow it to function. Taking the leap from substituting lost senses to the claim of creating entirely new ones, the promise itself is discussed from different viewpoints in the interdisciplinary field of cognitive science. Resulting in the question, if after all sensory augmentation might even in the strictest sense be a more tangible approach than trying to generally substitute modalities. An extensive review of several projects in four categories that take advantage of tactile sensory augmentation concludes the file.

**Keywords:** New senses, sensory augmentation, sensory addition, sensory substitution, tactile displays, human-computer devices, ubiquitous computing interfaces

## 1 Introduction

The idea that man’s mind is limited by its number of senses dates back to antiquity. And modern philosophy has discussed this issue imagining the “Brain in a vat”. While changing human physiology to allow for artificial ports into the brain remains unfeasible in its earliest days [21], the fascination of creating new senses persists. Sensory augmentation attempts to use existing human sensory modalities to feed the brain with inputs that are supposed to be interpreted by the nervous system in a new desired way. With computing playing an essential role in contemporary life and the ubiquity of mobile computing devices, this approach to adding senses appears highly interesting for new intuitive and non-distracting concepts in human-computer interaction, even posing questions about human cognition itself and enabling related research [17].

## 2 Roots of the field: Sensory Substitution

### 2.1 Characteristics of the brain and its perception

This preceding section attempts to explain select aspects of human cognition that will be important for our purposes.

In the context of human physiology, the term of “brain plasticity” refers to adaptive capacities of the central nervous system [17]. This characteristic allows the brain to change its structure and functioning based on demand or training while retaining previous abilities [17]. These effects are not limited to a certain age group or special attributes of anatomy (e.g. loss of a sensory modality) and can be demonstrated in healthy human beings through directed training [11]. Neural plasticity can for example be induced by cross-modal and sensorimotor activation [11]. Sensory substitution – and sensory augmentation as an extension of the former, as will be discussed later – is only possible because of brain plasticity [17].

It is important to understand the way our perception through senses originates. Bach-y-Rita emphasises that “*We see with the brain, not the eyes*” [18], stating that every human sense (applied to vision in this case) is not merely based on analysis of input through a sensory organ, e.g. the eye’s retina, but most importantly depends on memory, learning and contextual interpretation [18]. Only therefore can an activity like Braille reading activate the primary visual cortex [13]. This viewpoint on cognition is another cornerstone to understanding the plausibility of sensory substitution and addition.

“Sensorimotor contingencies” connect certain patterns of activation in the nervous system to motor actions that produce them [13]. For example, the human retina is not homogeneous but we normally do not experience irregular color distribution or areas missing from the picture where it lacks corresponding photoreceptors [13]. When moving towards a straight line, our brain receives highly different patterns of activity yet the line is perceived as the same entity [13]. Research indicates that it is possible for humans to learn new sensorimotor contingencies [13] (see 4.1).

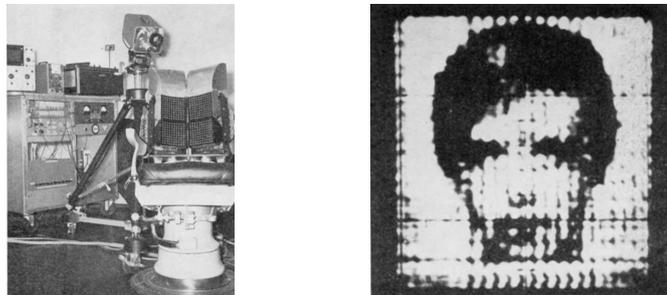
Another important term regarding perception is the concept of “distal attribution”. Loomis explains: “*The perceptual world created by our senses and nervous system is so functional a representation of the physical world [...] The phenomenon, which has been referred to as [...] ‘distal attribution’, is this—that most of our perceptual experience, though originating with stimulation of our sense organs, is referred to external space beyond the limits of the sensory organs.*” [12]. In the context of sensory substitution, achieving distal attribution means that a stimulus is not attributed to the skin, but the location in space that corresponds to the stimulation while moving the sensor attached to the substitution system [20].

In terms of sensor-coupled physical stimulation, the Weber-Fencher law states that as stimulus intensity increases, a greater absolute increase in stimulus intensity is required to create representative perception in humans of the prior [1]. Last, under normal circumstances, the brain does not tend to overload, but is

able to select the information important for its current context successfully and effectively. On the contrary, some sensory modalities thrive when flooded with input, such as vision when reading text [17]. Sensory overload on deliberate stimulation therefore strongly indicates wrong choice of interface [17]. The challenge will also be to create interfaces that humans are able to focus their attention on naturally instead of not perceiving them [19].

With these most fundamental elements of knowledge about the working of the human nervous system and perception in mind, we will take our first look at approaches that facilitate existing sensory modalities in order to create other ones.

## 2.2 An introduction to Bach-y-Rita and other founding works



**Fig. 1.** The TVSS system, 400 point sample as seen on a connected oscilloscope [16]

Early research on brain plasticity led to the idea of directly substituting senses, in order to restore lost sensory modalities. In 1963, Paul Bach-y-Rita et al. began to work on a tactile-visual substitution system (TVSS), transmitting visual information using a stimulator array on abdomen, back and thigh [18]. The final setup was based on a dental chair with a 20x20 tactile stimulator matrix, each vibrator tip having a diameter of 1mm, spaced 12mm apart, touching the back of a user. This array was driven at 60Hz field rate [16]. Bach-y-Rita described his results with TVSS in 2003: “*After sufficient training with the TVSS, our subjects reported experiencing the image in space, instead of on the skin [...] They learn to make perceptual judgments using visual means of analysis, such as perspective, parallax, looming and zooming, and depth judgments*” [18] and “*Although the TVSS systems have only had between 100 and 1032 point arrays, the low resolution has been sufficient to perform complex perception and ‘eye’-hand coordination tasks. These have included facial recognition, accurate judgment of speed and direction of a rolling ball with over 95% accuracy in batting a ball as it rolls over a table edge, and complex inspection-assembly tasks*” [18]. Though this approach opened the field of research into sensory substitution and led to

publications by Collins, Kaczmarek, Szeto, Riso, Mann and others [18], it had many practical problems [18], beginning with its stationary character.



**Fig. 2.** Tongue Display Unit system by Bach-y-Rita et al., 144 points resolution [18]

Bach-y-Rita intended to overcome these limitations by switching to an electrotactile display unit (further discussion in 2.3), meant for stimulation of the tongue through voltage applied to 12x12 contacts [18]. It is also known as the “BrainPort” [17]. Importantly, the TDU design allows for much higher resolution without conceptual changes thanks to its simple nature, the display itself being only a sheet with conducting paths and contacts while externalising controller electronics.

Bach-y-Rita et al. chose the tongue as the place for this kind of stimulation over finger tips in part, because it requires only about 3% the voltage (5-15V) and much less current (0.4-2.0mA) in comparison [18].

While TVSS permitted display of “black and white” information only, the tongue-based system also supports gray-scaled pattern information. To achieve this result, six signal parameters could be varied independently: Current level, pulse width, interval between pulses, number of pulses in a so called burst, burst interval and frame rate [18].

### 2.3 Tactile display technology approaches

We have already distinguished between two kinds of tactile display systems used when implementing sensory substitution systems: Those based on vibrotactile stimulation and such that are purely electric. In the following, a number of angles on their implementation are introduced and discussed.

**Mechanism of stimulation** Kaczmarek et al. consider three kinds of tactile display approaches for sensory substitution purposes [8]:

- **Slowly-varying tactile display**

Persistently slow-varying tactile displays are not suitable for application

in sensory substitution systems, because the tactile sensory modality does adapt to static stimuli too quickly for most applications.

- **Vibrotactile display**

Vibrotactile displays solve the issue present with slowly-varying stimulators by continuously oscillating while carrying information encoded in other signal parameters, such as amplitude or frequency. Vibrotactile displays can be realised using round or linear motors accelerating adequately sized weights.

- **Electrotactile stimulation**

It is also possible to stimulate parts of the human body directly through electric means, for example sensitive skin sites such as finger tips or the tongue. Depending on voltage, current and AC frequency, the feeling of electric stimulation reaches from a subtle vibratory feeling to pain.

Electrotactile displays can easily reach high stimulator density since only contacts and wiring is needed on the display itself while every element in a vibrotactile display requires its own motor with fixture and clearance close to the point of stimulation.

Appropriate measures have to be taken regarding contact material and signal characteristics in order to prevent skin irritation or even burns.

**Display layout** The stimulator elements mentioned above in 2.3 can be composed in various layouts, most importantly forming [8]:

- **Single-element displays**

One point stimulating the body can only vary in intensity, frequency or both. These displays are hence limited to low-bandwidth applications. It is also important to keep in mind, that one single stimuli might be error-prone due to fixture and pressure variations or even effects of perspiration while spatio-temporal patterns, which might relieve effects of such failure, are not possible to implement.

As cited by Spence, distal attribution (see 2.1) can be achieved even with a single stimulator display [19].

- **One-dimensional displays**

Displays spreading multiple points in one direction allow for the utilisation of spatio-temporal patterns, multi-point code, but may also improve perception of static ones because more surface can be covered and stimulated at once.

- **Two-dimensional displays**

Two-dimensional displays can make stimulated areas work like the eye when presenting originally visual information [16] and are therefore even more powerful than displays working in one dimension. It does not seem plausible that the human brain can handle temporal “multiplexing” between rows or columns when dealing with multi-dimensional information. However, Spence points out that this apparent similarity between vision and touch in two-dimensional displays might be misleading since they vary in almost any characteristic other than dimensionality [19]. In general, two-dimensional

displays enable spatio-temporal patterns in more directions than its rows or columns working independently, because of diagonals.

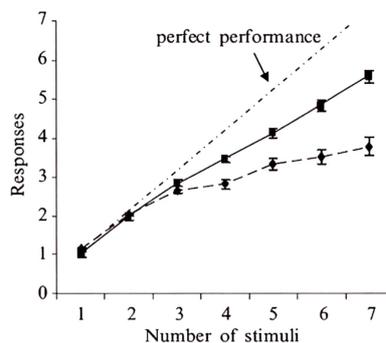
Note: One and two-dimensional displays as characterised above do not have to be placed in a plane. For example, a belt with stimulators worn on the waist can be interpreted as a bent and closed one-dimensional display. In favour of classification simplicity, only these three manifestations of dimensionality variation have been discussed.

**Protocol types** The following protocol classes can be considered when controlling one and multi-dimensional tactile displays:

– **Single and multi stimulator**

With only a single stimulator active at a time, the combinatory coding advantage of multi-element displays can't be exploited fully. Multi-stimulator patterns are only possible with one and two-dimensional displays on the other hand.

Interestingly, Nagel et al. report annoyance of test subjects caused by multiple concurrently active stimulators on the feelSpace belt [13] (see 4.1). Such effects might highly depend on the specific application since other authors don't report these phenomena [10] but in contrary build on multi-stimulator approaches in their work [14].



**Fig. 3.** Results of the tactile numerosity judgement study by Gallace et al. 2006, single (solid line) and repeated (dashed line) stimulus presentation [19]

It is generally hard for humans to perceive complex multidimensional patterns through the tactile sense and it might also be challenging to differentiate between closely placed stimuli. Spence highlights that in a study by Gallace et al. in 2006, participants were not able to count more than two or three vibrotactile stimuli on the skin surface even though displayed once a

second over 10 seconds [19].

Therefore it seems highly application specific whether to choose the first or second approach, if limited to spatially-only encoded patterns.

– **Spatio-temporal patterns**

Only possible with one and two-dimensional displays.

Generally, patterns encoded in space and time or intensity both far exceed performance of only spatially encoded ones [14]. There is also an interesting observation regarding the minimal distance between two stimuli: Using spatio-temporal patterns, the limitations of vibrotactile two-tacton resolution can be overcome [14]. Using vibrotactile display matrices and spatio-temporal patterns, Novich and Eagleman estimate a theoretical maximum throughput of about 600-925bit/s on the human torso [14].

Spatiotemporal patterns that appear as motion may be most intuitively perceived, as some studies have indicated [10].

However, concrete stimulator layout and protocols have to take into consideration many application-specific practical aspects such as compensation for the Weber-Fenchner law (see 2.1) that will be discussed later when reviewing different projects in the field of sensory augmentation.

### 3 From Substitution to Augmentation

When sensory substitution can be defined as “[...] *the use of one human sense to receive information normally received by another sense*” [18], it is intuitive to declare sensory augmentation as the use of one human sense – the tactile one in this paper – to receive information normally not received by any sense. Therefore, we can interpret sensory substitution for individuals with sensory loss as sensory augmentation that isn’t necessary without loss of a certain modality (at least under the assumption, that there is no conceptual difference between re-learning a sense that a person once might have had and entirely new ones). Yet, this apparent analogy combined with the deferred question, what substitution in real-world applications means, will open another perspective on the subject.

#### 3.1 Physiological feasibility

Bach-y-Rita, without reservation, extrapolates from his research in sensory substitution to the possibility of creating new senses: *“In addition to sensory substitution, a whole world of opportunities is opening for sensory augmentation, such as for night vision, sensate robots, sensate Internet and telecommunications, and many others.”* [15].

Furthermore, a number of experimental studies [13]/[11]/[9] have confirmed the hypothesis that new sensorimotor contingencies (see 2.1) can be learned, as will be discussed in 4.1.

There seems to be no problem deterring us from extending the aspects learned on cognitive attributes of man and interface technology to creating entirely new senses instead of only restoring ones lost.

### 3.2 Departing from general restoration

There has been reasoned criticism regarding the general type of sensory substitution, Bach-y-Rita approached from the beginning in his studies. He explains in 2003: *“Are blind persons using tactile-vision sensory substitution (TVSS) actually seeing? Heil [...] and Morgan [...] suggest that because blind subjects are being given similar information to that which causes the sighted to see and are capable of giving similar responses, one is left with little alternative but to admit that they are seeing [...]”* [17].

But the broad claim of fully substituting senses has to be challenged based on the following consideration that points out specific deficiencies. For example, empirical studies have reported that a number of well-trained users of tactile-visual substitution systems report missing emotional content to human images they can recognise, resulting in disappointment [19]. There seem to be aspects of modalities like vision that cannot be achieved through sensory substitution. These basic issues might reveal a deeper problem with the substitution systems in place, leading to the question if they can even recreate the most characteristic attributes of sensory modalities.

Furthermore, Spence states concerning all tactile sensory substitution systems: *“I would like to argue that a number of fundamental perceptual, attentional and cognitive limitations constraining the processing of tactile information mean that the skin surface is unlikely ever to provide such general-purpose sensory substitution capabilities [...] Instead, the use of the skin will likely be restricted to various special-purpose devices that enable specific activities such as navigation, the control of locomotion, pattern perception, etc.”* [19].

This notion is mainly based on the limits of spatial and temporal resolution of the tactile sense, the ability to attend to tactile stimuli at skin sites we are not used to focus attention on, the vast difference in brain utilisation by different senses (e.g. generally more than 50% for the visual sense and less than 10% of the brain is given to the tactile sense) possibly resulting in processing bandwidth problems and the partially open question how real-world conditions of multisensory stimulation impair the capacity to process tactile information [19]. Deroy et al. further argue, that sensory substitution might not be equivalent to other sensory modalities but facilitate skills such as reading in new ways [4].

However, it is important to note, that Spence and other authors don't argue that sensory substitution (and in extension augmentation) does not work per se, but that the tactile modality favours low-bandwidth special purpose concepts that do not suffer under the restrictions of perception mentioned above.

Or, to put it in different terms: Tactile applications that can make the most of man's cognition are likely simple ones that process inputs smartly if needed and display them in an often highly abstracted and task-related yet therefore useful manner. Ultimately, this even means that general sensory substitution needs might be better served by well-designed augmentation systems: Not trying to create a canonical approach to seeing for the blind or to hearing for the deaf, but picking – one at a time – important aspects and solving them appropriately. For example, it could be possible to design a navigation system with collision

detection using a vibratory belt and appropriate sensors, to solve the problem of navigating the world for blind people. While reading is best served by systems like Braille, in a vastly different way.

In the following, we will keep this special-purpose consideration in mind, when reviewing different approaches to sensory augmentation.

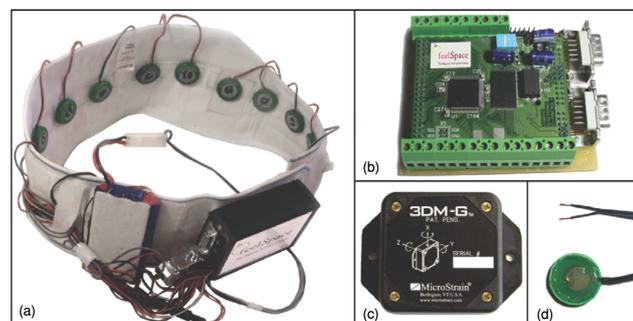
## 4 Creating New Senses

After introducing cognitive and technical characteristics that guide tactile sensory substitution and augmentation systems, we are going to take a look at four promising categories of “new senses” exemplarily for many projects in the field, that will further explain and discuss many practical aspects and important results.

### 4.1 Orientation and navigation: The feelSpace belt

The feelSpace project was initiated in 2005 by Nagel et al. and has resulted in a number of papers exploring sensory contingencies (see 2.1). It has been developed as a commercial spin-off product since November 2015 [11].

feelSpace in its original form is a waist-worn belt equipped with 13 vibrators placed around the perimeter, all driven by a compass controller unit [13] (see Fig. 4). It was designed to either continuously display the direction to the magnetic north pole or manipulated substitute data [13]. At any time, at most one motor is active, as early tests indicated annoyance of test subjects using other configurations [13].



**Fig. 4.** The feelSpace belt device (a), its controller unit (b), the compass used (c) as well as one vibrator element (d) [13]

**Original experimental findings** In 2005, Nagel et al. conducted experiments in order to test the following four hypotheses after extended use of the belt [13]:

- **Weak integration hypothesis:** The sensory information provided by the belt can be processed and improves performance in some tasks.
- **Strong integration hypothesis:** Information on orientation of the belt is firmly integrated in human perception and sensory signals of the belt, inconsistent with sensory inputs provided by the other modalities produce measurable responses.
- **Subcognitive processing hypothesis:** Sensory information provided by the belt can be processed at least partly without attention.
- **New modality hypothesis:** Use of the belt-imposed sensorimotor contingencies results in qualitatively new perceptual experiences.

Four subjects and four controls took part in these experiments after continuously wearing the feelSpace belt for six weeks during their waking hours. They also participated in weekly outdoor training sessions. It is important to note that one experimental subject is the author of the paper [13] which is generally problematic. The authors argue that in this case, it is excusable because of the exploratory nature of their work.

Nagel et al. report the following results [13]:

- **Experiment 1 (Weak integration hypothesis):** Blindfolded subjects are placed at a starting point and then led along the edges of shapes with varying complexity before being asked to return to their origin.  
**Results:** Before training, there was no performance difference between experimental subjects and the control group, whether or not the belt provided correct or manipulated information. For trained subjects however, correct belt information resulted in fewer errors than with incorrect information. As long as the belt was switched on, experimental subjects were better than control subjects. Confirms the hypothesis.
- **Experiment 2 (Strong integration hypothesis):** Subjects have to complete time competitive tasks in a virtual environment.  
**Results:** The trained subjects did not improve their ability to use correct belt information. A disturbing effect of wrong information was only observed in one experimental subject. Does not support the hypothesis.
- **Experiment 3 (Subcognitive processing hypothesis):** Balance tests and testing the stabilisation of visual stimuli on the retina during head movements.  
**Results:** Measurement of the sway path increased after training in one of the experimental subjects and one control subject. Also, the number of fast eye movements increased after training in three of the four experimental subjects. Confirms the hypothesis.

- **Experiment 4 (New modality hypothesis):** The subjects’ experience is assessed using questionnaires, diaries and through face-to-face feedback.  
**Results:** Two of four experimental subjects reported profound changes of sensory experience. North was perceived as an orientation point by one subject, all directions as equally important by others. Yet, no subject reported to perceive a local magnet field as could have been assumed based on the nature of the sensory modality “created”. Lends support to new modality hypothesis.

Nagel et al. conclude: “*The results indicate that the sensory information provided by the belt (1) is processed and boosts performance, (2) if inconsistent with other sensory signals leads to variable performance, (3) does interact with the vestibular nystagmus and (4) in half of the experimental subjects leads to qualitative changes of sensory experience*” [13].

Overall, the experiments point to the assumption that new sensorimotor contingencies can be learned and integrated into human behaviour while they also affect perceptual experience [13].

**Further results and observations** Kaspar et al. in 2014 supported the original results of Nagel et al. After seven weeks of training with the feelSpace device, participants reported substantial changes in perception of space and the belt vibration [9].

Also, König et al. confirm that physiological processes and subjective experience correspond to the hypothesis that new sensory contingencies can be learned: “*Our EEG results indicate that training with the belt leads to changes in sleep architecture early in the training phase, compatible with the consolidation of procedural learning as well as increased sensorimotor processing and motor programming*” [11].

Both Kaspar and König used improved versions of the feelSpace belt, carrying 30 vibratory elements [9]/[11]. Tactile displays mounted to a belt have also been used in other projects, for example to create an obstacle warning system for the blind [7]. Interestingly, Spence points out that the feelSpace belt has not turned into a viable commercial product offering, despite the vast media attention the device has received [19].

## 4.2 Tactile language: Early and contemporary approaches

One of the earliest attempts to create a tactile language, while not generally trying to substitute hearing through another sensory modality, dates back to Geldard in 1957 [5]. Hence, even earlier than Bach-y-Ritas research on sensory substitution. He describes a system called “Vibratese”, consisting of 45 elements transmitted through a vibrotactile display, limited to three different levels of intensity and with temporal discrimination in the range of 0.5-1.5s [5].

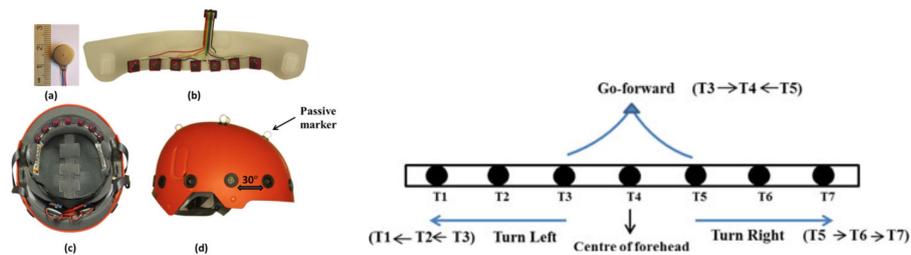
Spence points out: “*A number of early studies of tactile information processing in humans [such as Vibratese] gave the distinct impression that people could*

*be trained to perceive large amounts of information by means of their sense of touch, after, that is, having learned a tactile codification system [...]* [19]

Vibratense didn't turn out to be a successful approach and is no longer discussed as a feasible solution in contemporary publications, but ultimately appears noteworthy in historical terms.

At least with electronically controlled Braille output devices in mind, the idea of Braille (developed in 1825) being a tactile language in terms of the sensory augmentation described in this paper, needs further discussion. It will only be mentioned briefly due to its open character of implementation in interface devices, but remains interesting to sensory augmentation nevertheless.

However, there are many other proposals following Geldards initial idea, reducing the problem of communication in highly application-specific ways. Kerdegari et al. describe a "language" of navigation commands created for a head-mounted vibrotactile sensory augmentation device [10]. *"A head-mounted display may [...] be intuitive for navigation since a relative straightforward mapping can be created between sensed objects [...] and stimulation of the head in the direction of that object."* [10], Kerdegari explains her intentions. The navigation commands designed consist of spatiotemporal patterns, since their apparent motion is generally perceived as most intuitive [10].



**Fig. 5.** Experimental "Mark 2" device by Kerdegari et al., pattern sketches for turn-left, turn-right and go-forward commands [10]

Kerdegari et al. used an adapted version of the Tactile Helmet (discussed in 4.3), equipped with seven vibrotactile stimulators placed around the wearer's forehead [10]. It was important not to display too much information, since vibration at the head is also picked up by the ears as a buzzing noise.

The authors set out to test continuous and discrete modes of command presentation in single and recurring types while the series of stimulator activation in each case were the same ones, as described in Fig. 5 [10]. In the continuous presentation mode, stimulator activation overlaps 75% (300ms) of the active time per stimulator (400ms) in order to take full advantage of the effect that a traveling stimulus is perceived. While in the discrete mode, 75% of the time select for single stimulator activity (also 400ms) has to elapse after deactivation before

the next vibratory element becomes active [10]. With the recurring command type, a command is repeated with delays of 500ms until the next navigational command arrives [10].

In a 15m<sup>2</sup> environment with virtual walls, the authors observed 18 test subjects in 72 navigation tasks each. They came to the result that continuous command presentation led to faster reaction times than discrete ones [10]. Also, measures show that recurring commands were the most effective and easiest to distinguish which was reported by the experimental subjects, too [10]. Recurring stimuli were not perceived to be meaningfully distracting [10].

Kerdegari's work points out that there is potential for tokenised tactile languages in contexts such as navigation and that the head as a position of vibrotactile stimulation might be interesting for further research in sensory augmentation.

### 4.3 Long range touch: The Tactile Helmet project

In 2013, Bertram et al. set out to design a helmet-based system to augment the wearer with long-range touch – to empower users with a rat-like capability to quickly acquire environment information with their synthetic “whiskers” [1]. Research by Prescott has shown that whisking is an effective way to gain sensory environment information in robots [1], therefore it seems reasonable to test similar perception in humans. Naturally, application seems to fit in an emergency response context such as firefighters working their way in a smoke filled or otherwise low-visibility environment. Bertram et al. explain the overall concept and choice of form factor as follows: “[...] it is head mounted, which provides a balance of sensitivity and response time to tactile stimulation whilst also leaving the hands-free for other tasks including direct haptic exploration of surfaces” [1]. The head is also an interesting choice due to the fact that a wearer can react faster to signals that arrive there, because tactile response latencies are about linear in distance from the brain [10]. Using ultrasonic sensors, measured distances to nearby obstacles are signalled to the wearer using a vibrotactile display on the helmet's inside headband [1].

A context sensitive algorithm controls the four stimulator vibrotactile display based on input from eight ultrasonic sensors. At walking speed and faster, the helmet functions as a unidirectional obstacle warning system while at slower speeds, the stimulators are turned off. However, an alternative mode for these latter situations is being developed that provides information on the area of user focus [1].

Objects closer than 50cm result in a 80Hz warning signal, while normal stimulation frequency rests at 150Hz [1]. The requested signal amplitude ( $a$ ) for each of the four stimulator elements is determined from the average measurement of two sensors adjacent to a stimulator. Initially, formula  $a'$  was used for standard mode, where  $x$  is the average distance measured by two sensors,  $m$  is the maximal distance set to measure (200cm). After testing,  $a'$  was select for  $x < 100$  and  $a''$  for  $x \geq 100$ . At slower speeds, the rear stimulators continue to function



**Fig. 6.** The Tactile Helmet with its sensors on the outside and stimulators within, adjacent to a single stimulator element [1]

according to  $a'$ , while the frontal ones are set to zero if  $x \geq 100$  and to  $a'''$  otherwise [1].

$$a' = \left(1 - \frac{x}{m}\right)^3 \quad a'' = 2\left(1 - \frac{x}{m}\right)^3 \quad a''' = \left(1 - \frac{x}{0.5m}\right)^3$$

**Experimental testing** A 4m long corridor with exits to the left and right was built in order to experimentally evaluate the Tactile Helmet. In each run, a blindfolded test subject was asked to escape from the corridor while one was narrowed in order to make passing unfeasible. The authors report: “[...] trials suggests that, although participants used their hands in both conditions, they were relying on their hands much less in the helmet on condition than when no information was available [...]” [1].

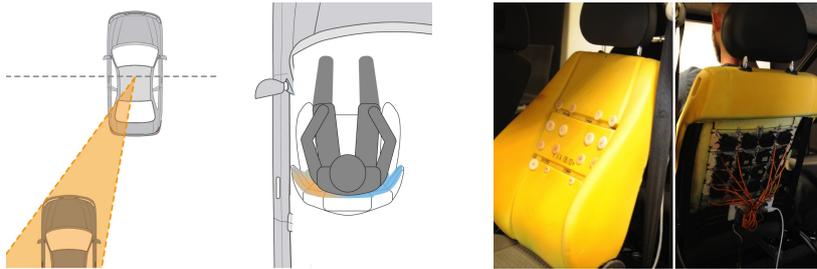
**Comparable projects** Carton et al. developed a similar glove-based system to support firefighters when navigating in low-visibility environments [3].

#### 4.4 Continuous periphery perception: A Shape-changing Car Seat

Vibratory warning systems have long been used in cars to signal certain dangerous situations to the driver. But while those haptic interfaces have been widely adopted for simple applications like vibration on/off corresponding to danger/clear, only little innovation has seen the light of day regarding other more advanced use cases of tactile displays in the automotive space.

Grah et al. have developed a seat-based sensory augmentation system that continuously delivers information of a car’s rear periphery [6]. The authors explain: “When walking, we receive additional information about our periphery from audible and sensorimotor stimuli. This peripheral extension of the human perceptions vanishes to a certain extent when driving a car because the area of non-visual stimuli is more or less limited to the interior of the car. [...] A non-visual extension of the human field of view could help to decrease the visual and mental

*workload.*” [6]. While many cars offer sensory systems that monitor their surroundings, no adequate displays exist as of today that allow drivers to perceive their environment while keeping focus on the road ahead of them. Also, the approach described in the following has the distinct advantage in comparison to mirrors or screen based that the driver is never forced to change perspective of the car’s orientation.



**Fig. 7.** Mapping of an obstacle to driver’s back, prototype car seat with 16 motors [6]

The prototype system that has been built consists of a car seat equipped with a 4x4 servomotor matrix [6]. The motors act as linear actuators that create pressure points on the user’s back, spread by small wooden plates [6]. Though, slow-varying tactile displays have been criticised earlier in 2.3, the specific use case seems to support the interface choice due to its dynamic nature, as experimental results will also show in the following.

The designers decided to focus on the display of obstacle distance, size, pace and position [6]: *“As objects visually appear larger when closer, we map the distance of the object to the total volume of the deformation [...] The lateral position of deformation corresponds to the angle from the obstacle to the cars center [and] The sensitivity levels of the actuation should correlate with the relevance of the situation.”* [6]. When speaking about object relevance, the authors refer to the combined meaning of approach path and speed – in worst case danger displayed in a way the driver can’t ignore and just subtle if no imminent danger seems likely [6].

Interestingly, when reviewed closely, the described system forms a closed sensorimotor loop because driver action (steering, acceleration or deceleration) has direct effects on the information displayed, which may previously have informed driving decisions [6].

**Experimental testing** Using a virtual driving simulator equipped with the prototype car seat, Grah et al. designed two tasks on a straight five-lane road, one with and another one without user input [6]. Both tasks are limited to one stimulus displayed at a time [6].

In the first task without user input, about 63% obstacle detection rate was achieved, while inner lanes had a much higher detection rate than outer ones [6]. In the active test, circa 51% of obstacles were detected [6]. False detections occurred when the lane of an approaching object was crossed, in which case the stimulation move from one side to the other on the user's back [6]. Interestingly, a left-tendency in false detections was discovered, possibly related to the projection of obstacles based on the center of the car [6].

## 5 Conclusion

Sensory substitution and addition are ideas far from new at their core, as this paper has shown, sketching the long way from Bach-y-Rita and Geldard to contemporary applications. In general, it is hard to clearly differentiate the two fields, because this task highly depends on defining aspects that characterise a certain sense and which characteristics are therefore needed to substitute it. The full promise of tactile sensory substitution has not come true over 50 years after its inception in research in the eyes of some authors. But if we take a step back from man's complex high-bandwidth sensory modalities like vision or hearing that might never be matched through tactile interfaces, modern mobile computing technology enables simple yet interesting special-purpose sensory augmentation devices that can be of immediate value. Any sensor can be connected to powerful mobile computing hardware and tactile displays, allowing for rapid prototyping [2] and therefore effective experiments for the evaluation of many use cases.

As the research projects reviewed in this file have demonstrated, there is great promise in the use of tactile displays for low-bandwidth applications in contexts that could not be more different from each other, reaching from firefighters in emergency situations to drivers on their daily commute – but that all make the most out of abandoning ordinary screens in order to empower users to focus modalities such as their vision on other tasks.

Only success of widely available commercial applications can tell, but there is no reason to assume that tactile interfaces will not be able to create useful ubiquitous computing experiences equipping us with “new senses”. And maybe even ones, we cannot imagine today.

## References

1. Bertram, C., Evans, M.H., Javaid, M., Stafford, T., Prescott, T.: Sensory augmentation with distal touch: The tactile helmet project. In: *Biomimetic and Biohybrid Systems*, pp. 24–35. Springer Nature (2013)
2. Bird, J., Marshall, P., Rogers, Y.: Low-fi skin vision: A case study in rapid prototyping a sensory substitution system. *HCI 2009 – People and Computers XXIII – Celebrating people and technology* pp. 55–54 (2009)
3. Carton, A., Dunne, L.E.: Tactile distance feedback for firefighters. In: *Proceedings of the 4th Augmented Human International Conference on - AH '13*. Association for Computing Machinery (ACM) (2013)

4. Deroy, O., Auvray, M.: Reading the world through the skin and ears: A new perspective on sensory substitution. *Frontiers in Psychology* 3 (2012)
5. Geldard, F.A.: Adventures in tactile literacy. *American Psychologist* 12(3), 115–124 (1957)
6. Grah, T., Epp, F., Wuchse, M., Meschtscherjakov, A., Gabler, F., Steinmetz, A., Tscheligi, M.: Dorsal haptic display: A shape-changing car seat for sensory augmentation of rear obstacles (2015)
7. Johnson, L.A., Higgins, C.M.: A navigation aid for the blind using tactile-visual sensory substitution. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. Institute of Electrical and Electronics Engineers (IEEE) (aug 2006)
8. Kaczmarek, K.A., Webster, J.G., Bach-y Rita, P., Tompkins, W.J.: Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38(1), 1–16 (1991)
9. Kaspar, K., König, S., Schwandt, J., König, P.: The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness and Cognition* 28, 47–63 (aug 2014)
10. Kerdegari, H., Kim, Y., Prescott, T.J.: Head-mounted sensory augmentation device: Designing a tactile language. *IEEE Transactions on Haptics* 9(3), 376–386 (2016)
11. König, S.U., Schumann, F., Keyser, J., Goeke, C., Krause, C., Wache, S., Lytochkin, A., Ebert, M., Brunsch, V., Wahn, B., Kaspar, K., Nagel, S.K., Meilinger, T., Bühlhoff, H., Wolbers, T., Büchel, C., König, P.: Learning new sensorimotor contingencies: Effects of long-term use of sensory augmentation on the brain and conscious perception. *PLOS ONE* 11(12), e0166647 (dec 2016)
12. Loomis, J.M.: Distal attribution and presence. *Presence: Teleoperators and Virtual Environments* 1(1), 113–119 (jan 1992)
13. Nagel, S.K., Carl, C., Kringe, T., Martin, R., König, P.: Beyond sensory substitution—learning the sixth sense. *Journal of Neural Engineering* 2(4), R13–R26 (nov 2005)
14. Novich, S.D., Eagleman, D.M.: Using space and time to encode vibrotactile information: toward an estimate of the skin’s achievable throughput. *Exp Brain Res* 233, 2777–2788 (2015)
15. Bach-y Rita, P.: Tactile sensory substitution studies. *Annals of the New York Academy of Sciences* 1013(1), 83–91 (jan 2006)
16. Bach-y Rita, P., Collins, C.C., Saunders, F.A., White, B., Scadden, L.: Vision substitution by tactile image projection. *Nature* 221, 963–964 (1969)
17. Bach-y Rita, P., Kercel, S.W.: Sensory substitution and the human–machine interface. *TRENDS in Cognitive Sciences* 7(12), 541–546 (2003)
18. Bach-y Rita, P., Tyler, M.E.: Seeing with the brain. *International Journal of Human–Computer Interaction* 15(2), 285–295 (2003)
19. Spence, C.: The skin as a medium for sensory substitution. *Multisensory Research* 27(5-6), 293–312 (nov 2014)
20. Visell, Y.: Tactile sensory substitution: Models for enactment in HCI. *Interacting with Computers* 21(1-2), 38–53 (jan 2009)
21. Warwick, K., Gasson, M., Hutt, B., Goodhew, I.: An attempt to extend human sensory capabilities by means of implant technology. In: 2005 IEEE International Conference on Systems, Man and Cybernetics. Institute of Electrical and Electronics Engineers (IEEE)

# Überblick über den Einfluss von Smartphones auf den Nutzer

Niklas Keller\*

Betreuer: Anja Exler<sup>†</sup>

Karlsruher Institut für Technologie (KIT)  
Pervasive Computing Systems – TECO

\*niklas.keller@student.kit.edu

<sup>†</sup>exler@teco.edu

**Zusammenfassung.** Smartphones haben unsere Welt innerhalb weniger Jahre deutlich verändert. Das Smartphone ist unser ständiger Begleiter geworden. Über den Tag hinweg wird das Smartphone für verschiedene Zwecke benutzt, von der Kommunikation mit Freunden und Kollegen über den Konsum von Nachrichten bis hin zum mobilen Zahlungsmittel. Für viele Menschen ist es dabei die erste Beschäftigung am Morgen und die letzte Beschäftigung am Abend.

Die breite Anwendung der Smartphones, in allen Bereichen, hat sowohl positive als auch negative Folgen. Dabei sind viele positive Aspekte direkt ersichtlich, während negative Effekte oft nicht direkt auffallen.

Diese Arbeit gibt einen Überblick über den Einfluss des Smartphones auf den Nutzer und diskutiert dabei beide Seiten. Sie betrachtet den Einfluss auf das gesellschaftliche Leben, zeigt problematische Verhaltensweisen sowie deren Ursachen auf und betrachtet die Folgen von ständiger Erreichbarkeit. Weiter wird der Einfluss in den Bereichen Gesundheit und Sicherheit diskutiert. Das Ergebnis zeigt, dass Smartphones zwar durchaus nützlich sind, die Nutzung jedoch bewusst erfolgen und der Nutzer sich auch der negativen Folgen bewusst sein sollte. Insgesamt besteht auch ein Verbesserungspotential auf der Seite der Softwarehersteller, die durch technische Lösungen einige negative Seiten abmildern könnten.

**Schlüsselwörter:** Smartphones; Ubiquitous Computing; Social Impact; Communication; Addiction; Health; Security

## 1 Einführung

Smartphones haben begonnen das digitale Leben der Menschen zu verändern. Seit der Einführung des ersten iPhones im Jahr 2007 [30, Folie 7] gewinnen Smartphones ständig an Bedeutung. Im Jahr 2014 entfielen laut Digital Consult statistisch fast 1,5 SIM-Karten auf jeden Menschen in Deutschland. [5, Seite 24]. Doch es steigt nicht nur die Zahl der Smartphone-Nutzer. Die durchschnittliche weltweite Nutzungsdauer pro Tag lag im Jahr 2016 bei vier Stunden. [25, Seite

109] Smartphones haben Desktop-Computer als führende Plattform beim Internetzugriff abgelöst. [8] Smartphones sind Mobiltelefone, die neben den traditionellen Funktionen „Telefonieren“ und „Versenden von Textnachrichten“ weitere Funktionen bieten. Die ständigen Begleiter erlauben unter anderem den Internetzugriff von fast überall, fungieren als Taschenrechner, Notiz- und Kontaktbuch, ersetzen den Fotoapparat und Musikspieler, teilweise sogar bereits das Bargeld bzw. die Kreditkarte. Das Smartphone hat unseren Alltag in wenigen Jahren stark verändert und hat heute Einfluss auf fast alle Bereiche des täglichen Lebens. Die schnelle Verbreitung der vergleichsweise noch jungen Technologie zeigt, dass viele Menschen das Smartphone als sinnvoll erachten, allerdings werden gerade beim schnellen Wandel die negativen Folgen außer Acht gelassen. Diese Arbeit legt daher sowohl die positiven Seiten der Smartphone-Nutzung als auch die negativen Folgen dar. Da das Smartphone in fast allen Bereichen des Lebens präsent ist, beschränkt sich diese Arbeit auf wichtige Themen des Alltags. Zuerst wird die Geschichte des Smartphones betrachtet und wie es zu seiner Verbreitung kam. Anschließend wird der Einfluss auf das gesellschaftliche Leben, mögliche problematische Verhaltensweisen von Nutzern und deren Ursachen, der Einfluss der ständigen Erreichbarkeit, die Nutzung im Gesundheitsbereich sowie die mit der Nutzung des Smartphones verbundenen Aspekte der Sicherheit betrachtet.

## 2 Geschichte des Smartphones

Das Smartphone, wie man es heute kennt, wurde von Apple im Jahr 2007 in den Verbrauchermarkt eingeführt. [30, Folie 7] Auch vorher gab es schon Smartphones, diese wurden aber vor allem von Unternehmern genutzt. Zum einen wurden diese primär für den unternehmerischen Nutzen entwickelt und damit auf dessen Ansprüche zugeschnitten, zum anderen wurden sie für Verbraucher als zu teuer angesehen. [33] Das erste Smartphone war das „IBM Simon“ aus dem Jahr 1993. Es kombinierte die Funktionen eines Telefons, eines Personal Digital Assistant (PDA) sowie eines Faxgeräts. Es bot schon damals einen Touchscreen, der zum Wählen von Rufnummern benutzt werden konnte. [30, Folie 2] Obwohl selbst kein Smartphone, wird der PDA „Palm Pilot“ aus dem Jahr 1996 als wegweisend für den Formfaktor von heutigen Smartphones gesehen. Er machte die Nutzung mobiler Daten für Unternehmen populär. [30, Folie 3] [26] 2002 stieg RIM in den Mobilfunkmarkt ein und brachte das „BlackBerry 5810“ heraus, mit dem man E-Mails abrufen und im Internet surfen konnte. [30, Folie 5]. Schließlich veröffentlichte Apple das „iPhone“ im Jahr 2007, das als „revolutionär“ bezeichnet wurde, und begründete damit den Smartphone-Massenmarkt. [30, Folie 7] Noch im selben Jahr stellte Google sein mobiles Betriebssystem „Android“ vor, welches sich inzwischen mit 84,8% Marktanteil bei neu verkauften Geräten als führendes mobiles Betriebssystem durchgesetzt hat. Apple’s „iOS“ kommt auf 14,4% Marktanteil, andere Betriebssysteme sind heute nur noch Randerscheinungen. [7]<sup>1</sup> Die anfangs noch üblichen Hardwaretastaturen gibt es heute fast

<sup>1</sup> Für die Berechnung des Marktanteils wurde die Anzahl der verkauften Geräte im Jahr 2016 betrachtet.

nicht mehr, stattdessen wird der Touchscreen für Eingaben benutzt. [24] Insgesamt ist der Markt stark gewachsen und mit der zunehmenden Nutzung der Einfluss der Smartphones auf den Nutzer gestiegen.

### 3 Gesellschaftliches Leben

Smartphones bieten gegenüber herkömmlichen Mobiltelefonen neue Wege der Kommunikation. Sie bieten unter anderem die Möglichkeit der Videotelefonie sowie des Instant Messagings. Letzteres beschränkt sich dabei nicht wie die SMS auf per Nachricht zu bezahlende Textnachrichten, sondern erlaubt außerdem Audio- sowie Videonachrichten und das Versenden von Bildern und Dateien. Des Weiteren sorgen Push-Benachrichtigungen dafür, dass Inhalte direkt dem Kommunikationspartner zugestellt werden und dieser benachrichtigt wird. Herkömmlich wurden Inhalte explizit angefragt, man sprach z. B. von „E-Mails abholen“. Die neuen Kommunikationsmöglichkeiten erleichtern den Kontakt mit anderen Menschen, verändern aber auch die bisherigen Wege der Kommunikation.

#### 3.1 Vereinfachte Kommunikation

Während in der Vergangenheit nur Telefonate auf weiten Strecken in Echtzeit realisierbar waren, sind heute Videochats und weitere Formen der Kommunikation möglich. Diese Möglichkeiten werden durch die Verbreitung des Internets geschaffen, doch erst das Smartphone erlaubt die Nutzung von nahezu überall. [33]

Die mobile Kommunikation vereinfacht das gegenseitige Finden auf großen Veranstaltungen, das Verlegen von Terminen, macht kurzfristig abgesprochene Treffen und vieles weitere möglich. Mit der Verbreitung von Smartphones ist die Kommunikation heute ortsunbeschränkt möglich. Wichtig sind diese verbesserten Kommunikationsmöglichkeiten auch für Flüchtlinge. Häufig werden Familienmitglieder oder Freunde in der Heimat zurückgelassen.

Im Zuge der syrischen Flüchtlingskrise zeigte sich in den sozialen Medien Verwunderung darüber, dass Flüchtlinge Smartphones besitzen. Laut dem CIA World Factbook besitzt allerdings rund 90% der syrischen Bevölkerung ein Smartphone. [4] Smartphones erleichtern Flüchtlingen nicht nur die Kommunikation mit Menschen, sie können auch helfen humanitäre Hilfe zu bekommen oder allgemein, um an Informationen zu gelangen. [23]

Aufgrund der geringen Größe und des geringen Gewichts kann das Smartphone leicht überall hin mitgenommen werden, auch auf einer Flucht. Mit seinen vielen hilfreichen Funktionen kann das Smartphone ein gutes Hilfsmittel sein. Besonders wenn Menschen vor Gewalt und Krieg anstatt vor Armut fliehen, ist der Besitz von Smartphones daher nicht verwunderlich.

Zu untersuchen bleibt, wie sich der persönliche Kontakt zwischen weit entfernt lebenden Menschen durch den häufigen Kontakt mittels Smartphones verändert. Einerseits könnte er sich erhöhen, da durch die gestiegene digitale Kommunikation der Kontakt eher aufrecht erhalten bleibt, andererseits ist es möglich,

dass der digitale Kontakt als ausreichend empfunden wird und somit etwa ein Besuch ausbleibt bzw. die Häufigkeit von gegenseitigen Besuchen sinkt.

### 3.2 Weniger persönliche Kontakte

Das Internet und Smartphones erleichtern die Kommunikation nicht nur für Menschen in großer Entfernung. Auch Menschen in geringer Entfernung profitieren von einer einfacheren Kommunikation. Während bei großer Entfernung die Technologie die Kommunikation erst ermöglicht, ist diese bei geringer Entfernung auch ohne Technologie möglich. Diese schafft hier eine alternative Form der Kommunikation und verändert dadurch die persönliche Kommunikation.

Bei einer Umfrage durch Coupofy aus dem Jahr 2016 gaben 15% der Befragten geminderte persönliche Kontakte zu Familie und Freunden als größte Negativfolge der Smartphone-Nutzung an. Weitere 17% nannten, dass ihre Verbindungen zu neuen Kontakten und Freunden durch die Smartphone-Nutzung weniger bedeutsam sind. [14] Dies bedeutet, dass eine große Zahl an Nutzern mit der Verschiebung von persönlichen Kontakten in die digitale Welt unzufrieden sind.

Der persönliche Kontakt zwischen Menschen leidet jedoch nicht nur unter der Verschiebung in soziale Netzwerke. Um nichts aus der digitalen Welt zu verpassen, verwenden viele Nutzer ihr Smartphone auch während persönlichen Treffen. Dieses Verhalten wird oft als unfreundlich empfunden. Es zeigt Desinteresse und sorgt durch die Ablenkung dafür, dass dem eigentlichen Geschehen nicht mehr mit voller Konzentration gefolgt werden kann. [10, Punkt 1] [36] Die Angst etwas zu verpassen und das darauf folgende Nutzerverhalten wird neben anderen problematischen Verhaltensweisen im nächsten Abschnitt näher betrachtet.

## 4 Problemverhalten von Nutzern

Laut der Umfrage durch Coupofy sind 29,5% befragten Nutzer unzufrieden mit ihrer häufigen Smartphone-Nutzung, bei Nutzern mit eigenen Kindern liegt dieser Anteil sogar bei 42%. [14] Im Folgenden werden die Gründe für diese überhöhte Nutzung und weitere problematische Verhaltensweisen von Nutzern dargelegt.

### 4.1 Abhängigkeit und Technostress

Für eine zwanghafte Nutzung von Smartphones gibt es eine Reihe von Einflussfaktoren. Nach Lee et al. haben die Art der Kontrollüberzeugung, das Maß an sozialer Angst und das Bedürfnis nach Berührung einen Einfluss. [22] Weitere Faktoren sind soziale Netzwerke und die Angst, etwas zu verpassen. [32] [15] Die zwanghafte Nutzung sorgt in der Folge für Technostress. Technostress bezeichnet Stress, der durch die Nutzung von technischen Geräten erzeugt wird. [22]

Kontrollüberzeugung bezeichnet nach Rotter die Auffassung eines Menschen, in wie weit er selbst durch seine Handlungen sein eigenes Leben bestimmen kann.

Dabei wird zwischen „intern“ und „extern“ unterschieden. Bei interner Kontrollüberzeugung geht eine Person davon aus, selbst für die eigenen Taten und das eigene Schicksal verantwortlich zu sein, während bei externer Kontrollüberzeugung die Person davon ausgeht, dass ihr Leben von externen Kräften bestimmt wird, etwa durch Glauben, Glück oder andere Menschen. [31]

Weiter bezeichnet soziale Angst nach Schlenker et al. die Angst, in sozialen Situationen im Zentrum der Aufmerksamkeit zu stehen, etwa anderen Menschen vorgestellt zu werden oder eine Rede zu halten. [34] Für Menschen mit sozialer Angst ist es einfacher, Interaktionen mit anderen Menschen online durchzuführen. Aufgrund dessen ist eine erhöhte Internet-Nutzung und damit erhöhte Smartphone-Nutzung für diese Personengruppe wahrscheinlich. [22]

„Need for Touch“, das Bedürfnis nach Berührung, bezeichnet nach Peck et al. die Vorliebe für die Nutzung von sensorischen Informationen bei Berührung. Menschen zeigen ein unterschiedlich starkes Bedürfnis danach. Während manche Menschen beispielsweise Produkte beim Einkauf rein für das Einpacken in die Hand nehmen, beziehen andere die sensorischen Informationen in ihre Kaufentscheidung mit ein. [27] Der Touchscreen eines Smartphones ist eine weitere Möglichkeit, dieses Bedürfnis zu befriedigen. [22] [15]

Nach Lee et al. zeigen Menschen, die zu einer eher externen Kontrollüberzeugung neigen, höhere soziale Angst oder ein größeres Bedürfnis nach Berührung haben, eher eine zwanghafte Nutzung von Smartphones. [22, H1, H2, H3]

Die Nutzung von sozialen Netzwerken ist nach Salehan et al. ein signifikanter Prädiktor für eine Abhängigkeit von Smartphones. Die Anzahl der Kontakte in einem sozialen Netzwerk und die Nutzungsintensität beeinflussen dabei die Häufigkeit der Nutzung auf mobilen Geräten. [32]

Nach Cheever et al. kann der Entzug des Smartphones zu einem Gefühl der inneren Unruhe führen. Bei höherer täglicher Nutzungsdauer ist dieses Gefühl statistisch signifikant stärker. Dieses Gefühl tritt bereits bei einer kurzen, vorher bekannten Dauer auf, unabhängig davon, ob das Smartphone abgegeben wird oder außer Sicht und lautlos beim Nutzer bleibt. Es verstärkt sich, je länger Smartphone und Nutzer getrennt bleiben. Mögliche Erklärungen für die innere Unruhe sind einerseits die reine Trennung vom Gerät und damit der Kommunikations- und Informationsmöglichkeit, andererseits die „Fear of Missing Out“ (FoMO), die Angst etwas zu verpassen. [12]

FoMO kann zu einer überhöhten Nutzung des Smartphones führen, da der Nutzer dauerhaft das Gefühl verspürt etwas zu verpassen. Dies kann sich beispielsweise durch ein häufiges Überprüfen des Smartphones äußern, ohne dass dieses den Nutzer benachrichtigt. Nach Elhai et al. bestimmt der Grad an FoMO allerdings nicht zwingend die Viel- oder Wenig-Nutzung, sondern eher, ob die Nutzung als problematisch einzustufen ist. [15]

## 4.2 Datenabhängigkeit

Neben der Abhängigkeit vom Gerät selbst kann ein Nutzer auch abhängig von seinen Daten sein. Neben der reinen Kommunikation lagern Nutzer von Smartphones viele Informationen und Daten auf das Smartphone aus. Es wird als

Notiz- und Kontaktbuch, als Kalender, zur Navigation und für viele weitere Aktivitäten benutzt, die hauptsächlich auf Daten basieren. Vertraut der Nutzer dabei rein auf das Smartphone, macht er sich abhängig.

Einerseits speichern Nutzer eine Vielzahl eigener Daten auf ihrem Smartphone, die ohne regelmäßige Backups bei Verlust, Diebstahl oder technischem Versagen des Geräts verloren gehen. Dazu gehören Notizen, Kontakte, Termine, Bilder und Musik. Andererseits greifen Nutzer auf eine Reihe von externen Daten zu, etwa Suchmaschinen, soziale Netzwerke und Kartendienste. Bei einer durchschnittlichen Akkulaufzeit von einem bis zu wenigen Tagen, benötigt der Nutzer zudem regelmäßigen Zugriff zu Strom.

Gerade Standort- und Navigationsdienste sind bei Nutzern beliebt. 73% der durch Coupofy befragten Nutzer gaben an, dass sie ihr Smartphone aufgrund der jederzeitigen Möglichkeit der Standortfindung und Navigationshilfe mögen. [14] Das ständige und teilweise exklusive Vertrauen auf diese Dienste kann dafür sorgen, dass eigene Navigationsfähigkeiten nicht genutzt und damit gemindert werden. [10, Punkt 4] Dies zeigte sich unter anderem, als mehrere Autofahrer „Apple Maps“ folgten, mehrere Schilder ignorierten und über das Rollfeld sowie Start- und Landebahn zu einem Passagierterminal eines Flughafens fuhren. [11]

## 5 Ständige Erreichbarkeit

Als ständiger Begleiter bietet das Smartphone fast überall die Möglichkeit sowohl andere Menschen zu erreichen als auch erreichbar zu sein. Die Raum-Zeit-Bindung von Kommunikation entfällt. Die ständige Kommunikationsmöglichkeit erlaubt es Nutzern auf dem neuesten Stand bezüglich der aktuellsten Neuigkeiten aus Politik, Wirtschaft und dem eigenen Sozialleben zu bleiben. Sie ermöglicht Nutzern etwa das Warten an der Bushaltestelle zu Nutzen, um mit Freunden oder der Familie zu kommunizieren. [33]

### 5.1 Familienleben

Im Familienleben erlaubt die ständige Erreichbarkeit Eltern, Termine zu organisieren und Aufgaben einfach zu delegieren. Weiterhin gibt sie Eltern ein Sicherheitsgefühl, wenn ihre Kinder anfangen selbständiger zu werden. Andererseits können auch Kinder jederzeit die Eltern erreichen, etwa bei Notfällen. Beide Gruppen bewerten die ständige Erreichbarkeit grundsätzlich als positiv, jedoch bewerten Jugendliche sie als negativ, wenn sie zur Kontrolle genutzt wird. Während Jugendliche bereits 2003 mit ihrer Handynummer sehr offen umgingen, teilten Eltern ihre private Handynummer eher nur mit dem engen Kreis der Familie oder guten Freunden, während die Nummer des Festnetzanschlusses meist öffentlich im Telefonbuch stand. Durch die Geheimhaltung der Handynummer kann die eigene Privatheit erhöht werden. [16]

Ob dies heute noch der Fall ist, ist höchst fraglich. Ein Drittel der Haushalte in der Europäischen Union hatte im Jahr 2015 ausschließlich einen Mobiltelefonanschluss. [6, Seite 51] Zusätzlich muss die Handynummer für verbreitete

Kommunikationsdienste wie „WhatsApp“ geteilt werden, um diese nutzen zu können.

## 5.2 Arbeitsleben

Im Arbeitsleben erlaubt die ständige Erreichbarkeit Unternehmen und Arbeitnehmern in Kontakt zu bleiben, sowohl während sich der Arbeitnehmer außerhalb des Büros aufhält, als auch außerhalb der Arbeitszeiten. Eine Kontaktmöglichkeit war auch vor der Verbreitung des Smartphones per Festnetz- oder Mobiltelefon möglich, allerdings musste dazu explizit ein Telefonat geführt oder eine Textnachricht an das Telefon gesendet werden. Heute erlauben Smartphones den Zugriff auf E-Mails und andere Internet-basierte Kommunikationskanäle, die unabhängig vom benutzten Gerät sind. Damit lässt sich die gewöhnliche Kommunikation mobil abhandeln, ohne dass der Kommunikationspartner den Nutzer zielgerichtet mobil erreichen muss.

In Besprechungen können Smartphones einerseits benutzt werden, um damit Notizen aufzuzeichnen oder Fakten zum aktuell besprochenen Thema zu finden, andererseits können sie zur Kommunikation genutzt werden. Einige Nutzer argumentieren, dass sie für Klienten jederzeit erreichbar sein und damit auch in Besprechungen E-Mails lesen und darauf antworten müssen. Für Anwesende ist dabei häufig nicht ersichtlich, ob das Smartphone als Hilfsmittel für die Besprechung oder rein als Nebenbeschäftigung benutzt wird. Unabhängig davon lenkt die Nutzung des Smartphones vom eigentlichen Geschehen ab und wird als häufig unhöflich gesehen. [36]

Die Nutzung von Smartphones und allgemein des Internets lässt die Grenzen zwischen Arbeits- und Privatleben verschwimmen. Vor der starken Verbreitung von Smartphones war eine Antwort auf Nachrichten außerhalb der Arbeitszeit vor allem aufgrund der fehlenden Benachrichtigung bzw. des fehlenden Wissens über die Existenz der Nachricht nicht möglich. Heute nutzen viele Arbeitnehmer ihr Smartphone um auf Arbeits-E-Mails zuzugreifen. [13]

## 6 Gesundheit und Nutzung in der Medizin

Im Folgenden wird mit dem Gesundheitsbereich ein Einflussbereich des Smartphones genauer betrachtet. Es wird einerseits der Einfluss auf Patienten und somit den privaten Nutzer und andererseits die Nutzung durch Versorger und damit die Auswirkungen im Arbeitsbereich betrachtet. Die Erkenntnisse sind teilweise auf andere Bereiche übertragbar, teilweise aber auch spezifisch für den Gesundheitsbereich.

Die Versorgung im medizinischen Bereich läuft traditionell über persönliche Treffen, etwa mit Ärzten oder Psychologen. Dabei sucht üblicherweise der Patient den Versorger auf. Traditionell wird daher die Zugangsmöglichkeit eines Menschen zum Gesundheitssystem über die damit verbundenen Merkmale gemessen, z. B. über die Zeit, die der Patient bis zum nächstgelegenen Versorger benötigt. Das digitale Zeitalter sowie das Smartphone im Besonderen eröffnen

in diesem Bereich neue Möglichkeiten, die nach Fortney et al. zu einer Rekonzeptionalisierung der Gesundheitssysteme führen sollten. Das System müsse sich den aktuellen Möglichkeiten anpassen und dabei die Einschränkungen einzelner Individuen, wie z. B. einen ländlichen Lebensraum oder Armut, nicht außer Acht lassen. [19]

### **6.1 Wege der virtuellen Gesundheitspflege**

Fortney et al. unterscheiden zwischen vier möglichen Typen der virtuellen Gesundheitspflege aus Sicht des Patienten, welche sich durch den Kommunikationspartner und die Reaktionszeit des Partners unterscheiden. Allen gemein ist jedoch, dass durch die digitale Kommunikation der Weg zwischen Patient und Kommunikationspartner entfällt und damit deren Entfernung irrelevant wird. [19]

#### **Synchrone digitale Patient-zu-Versorger Kommunikation**

Eine synchrone Kommunikation liegt dann vor, wenn sowohl Patient als auch Versorger zur gleichen Zeit miteinander interagieren. Mit Hilfe von Smartphones ist dies einerseits über Videotelefonie möglich, andererseits durch herkömmliche Telefongespräche, wobei letztere eher dazu genutzt werden Seiteneffekte zu besprechen oder Behandlungseffekte zu beurteilen. [19, Tabelle 1]

#### **Asynchrone digitale Patient-zu-Versorger Kommunikation**

Eine asynchrone Kommunikation liegt vor, wenn Patient und Versorger nicht notwendigerweise zur gleichen Zeit miteinander kommunizieren. Der Patient kann etwa seine Informationen als Audio- oder Videonachricht oder in Textform auf eine Plattform hochladen, während der jeweilige Versorger diese häufig zeitverzögert zur Kenntnis nimmt und gegebenenfalls darauf reagiert. [19, Tabelle 1] Dies hat den großen Vorteil, dass Patient und Versorger keinen gemeinsamen Termin für die Kommunikation finden müssen, und ist besonders dann sinnvoll, wenn keine Konversation nötig ist, z. B. für das Teilen von Überwachungsdaten. In der anderen Richtung kann eine solche asynchrone Kommunikation u.a. dazu genutzt werden, Patienten daran zu erinnern verordnete Medikamente einzunehmen. [19, Tabelle 1]

#### **Digitale Peer-to-Peer Kommunikation**

Peer-to-Peer Kommunikation meint die direkte Kommunikation zwischen Patienten. Diese können über soziale Medien und Onlineforen Informationen und praktischen Rat austauschen und sich damit gegenseitig unterstützen. Die Empathie von Mitleidenden wird dabei häufig authentischer empfunden als die von Ärzten oder anderen Versorgern. Diese Art der Kommunikation ist sowohl synchron als auch asynchron möglich und hängt von der genutzten Technologie ab. [19, Tabelle 1]

### **Synchrone digitale Interaktion des Patienten mit Computeranwendungen**

Der Patient kann mit Computeranwendungen interagieren, ohne dass dabei eine andere Person auf der Gegenseite reagiert. Aufgrund der hohen Verfügbarkeit und starken Präsenz wird hierzu immer häufiger das Smartphone anstatt von herkömmlichen Desktop-Computern benutzt. Das Angebot an Anwendungen reicht von der einfachen Suche zum Einholen von Informationen über Symptom-Checker bis zur Selbstüberwachung. [19, Tabelle 1] Gemessen an der Anzahl der verfügbaren Anwendungen, ist die Kategorie „Health & Fitness“ mit 85864 Anwendungen (Stand: 30.03.2017) eine vergleichsweise große Kategorie im Google Play Store. [3]

Insgesamt wird ein digitales Gesundheitssystem vor allem durch das Internet ermöglicht. Das Smartphone im Speziellen sorgt für die ständige Erreichbarkeit, die einfache Kommunikation von überall und kann Basis für Sensoren zur Selbstüberwachung sein.

#### **6.2 Probleme der Smartphone-Nutzung durch Versorger**

In Krankenhäusern werden heute oft Pager zum Rufen von Personal eingesetzt. Ein Ersatz dieser Pager durch Smartphones scheint attraktiv. Diese bieten die Funktionalität eines Pagers, bieten darüber hinaus allerdings viele weitere Funktionen. Ein Umstieg bringt neue Möglichkeiten der Kommunikation und andere neue Funktionalität, allerdings haben Smartphones gerade in diesem Bereich auch Nachteile. Die Kommunikation per Textnachrichten auf Smartphones erfolgt eher flüchtig, wodurch die Kommunikation weniger präzise und fehlerhafter werden kann. Mögliche Fehler beinhalten Rechtschreibfehler, die möglicherweise von einer genutzten Autokorrektur verstärkt werden, da dann nicht nur einzelne Buchstaben, sondern direkt ganze Wörter verfälscht werden. Dies ist kritisch, wenn es um die Gesundheit von Menschen geht. Ein weiteres Problem kann die Menge der unwichtigen Nachrichten sein. Einerseits produzieren Funktionen wie „Allen Antworten“ Nachrichten an viele Empfänger, die diese Informationen nicht zwingend benötigen, andererseits beschränken sich viele Textnachrichten nicht auf das Nötigste. Dies resultiert darin, dass das Personal Zeit mit dem Lesen unwichtiger Nachrichten verbringt, die anders besser genutzt werden kann. Heutige Smartphones sind im Allgemeinen für den Verbrauchermarkt entworfen und damit nicht auf die Bedürfnisse im Gesundheitssystem zugeschnitten. Sie bieten eine kürzere Akkulaufzeit, weniger Schutz vor Fall- und Flüssigkeitsschäden und sind durch ihre Komplexität weniger zuverlässig als herkömmliche Pager. Zudem lassen sie sich aufgrund der häufig fehlenden Dichtigkeit nicht einfach desinfizieren. [29]

Bei der Umstellung von einer existierenden, zuverlässigen Technologie auf eine neue sollten alle Aspekte des Umstiegs berücksichtigt werden. In diesem Fall sollte besonders auf Wasserdichtigkeit und Stoßfestigkeit geachtet werden um die negativen Folgen zu mindern.

### 6.3 Folgen eines digitalen Gesundheitssystems

Während der Zugang zum Gesundheitssystem durch das Smartphone und verwandte Technologien für viele Menschen immer einfacher wird, dürfen arme und alte Menschen nicht außer Acht gelassen werden. Es muss sichergestellt werden, dass ein fehlender Zugang zu Technologien nicht in fehlendem Zugang zum Gesundheitssystem resultiert. Weiterhin erhöht das Smartphone zwar den Zugang zum Gesundheitssystem, doch damit steigt nicht zwangsläufig die Qualität und Effektivität. Wichtig ist es, die möglichen Folgen zu erkennen und entsprechend zu reagieren. Durch den einfacheren und bequemeren Zugang kann es zu einer Überflutung der Versorger kommen. Diese kann durch geeignete Filter- und Priorisierungsmaßnahmen abgemildert werden, wie sie auch heute schon in manchen Krankenhäusern angewendet werden. [19] Ebenso kann es dazu kommen, dass Patienten den digitalen Weg wählen, obwohl ein persönlicher Arztbesuch dringend zu empfehlen wäre.

Je mehr sich die Gesundheit des Nutzers in die digitale Welt verlagert, desto mehr Daten werden auf dem Smartphone ein- und ausgegeben. Diese werden nicht nur dort, sondern auch auf den Servern der Versorger zumindest teilweise gespeichert. Die Masse an Daten öffnet dabei nicht nur neue Wege für die Forschung, sondern schafft auch mögliche Angriffspunkte und Sicherheitsprobleme.

## 7 Sicherheit

Das Thema Sicherheit teilt sich in zwei Bereiche. Im Englischen werden diese durch zwei unterschiedliche Begriffe unterschieden, „Security“ und „Safety“. „Security“ bezeichnet dabei die Sicherheit vor Angriffen wie Datendiebstahl oder absichtlichem Löschen von Daten durch fremde Programme, während „Safety“ die Sicherheit bei natürlichen Ereignissen wie Unfällen bezeichnet. Die im Folgenden dargelegten Auswirkungen auf die Privatsphäre sind Teil des „Security“-Bereichs, während die Auswirkungen auf die Verkehrssicherheit Teil des „Safety“-Bereichs sind.

### 7.1 Privatsphäre

Mit der steigenden Zahl an Smartphones sowie der steigenden Zahl an Apps steigt auch das Datenaufkommen. Teile der Daten werden über das Internet übertragen und auf Servern gespeichert. Diese Daten können an einer Reihe von Orten abgegriffen werden und somit die Privatsphäre des Nutzers verletzt werden. Angriffe auf Server und Übertragungswege sind allgemeine Gefahren des Internets, weshalb im Folgenden nur der Schutz auf dem Gerät selbst betrachtet wird.

Während sich die Forschung anfangs vor allem mit der Privatsphäre des aktuellen Standorts eines Nutzers beschäftigte, wird diese heute von Nutzern nur als mittleres Risiko eingeschätzt. Als sehr hohes Risiko wird dagegen zum Beispiel das Teilen der eigenen Textnachrichten mit der Öffentlichkeit oder mit Freunden gesehen. [17]

### Rechtebeschränkung zur Datenzugriffskontrolle

Auf Desktop-Computern gilt oft noch das Prinzip „*If it runs, it has to be trusted.*“ Zwar gibt es normalerweise einen Administrator-Account, jedoch schützt dieser nur das System selbst vor Anwendungen, nicht aber die Daten des Benutzers bzw. Daten von Anwendungen voreinander.

Mobile Betriebssysteme sind einen Schritt weiter und gehen in Richtung des „*Principle of least privilege*“. Anwendungen haben demnach nicht direkten Zugriff auf alle Daten, sondern müssen Berechtigungen vom System anfordern. Unterschiedliche Betriebssysteme bieten solche Berechtigungen jedoch zu einem unterschiedlichen Grad an Granularität. Während Android 4.0 zwischen 165 Rechten unterscheidet, bietet Windows Phone 7 nur 16 unterschiedliche Berechtigungen an, iOS 5 sogar nur zwei. [17, Seite 1]

### Prozessisolation zur Trennung von Anwendungen

Damit ein solches Berechtigungssystem funktioniert, müssen Anwendungen voneinander abgeschottet werden. Passiert dies nicht, können Anwendungen über andere Anwendungen an Bereiche gelangen, für die sie keine Berechtigung besitzen. Ein weiteres Risiko besteht grundsätzlich darin, dass Anwendungen auf gespeicherte Daten anderer Anwendungen zugreifen können.

Da Android auf Linux basiert, erbt es dessen Basis-Rechtesystem für Dateien mit „Benutzer“, „Gruppe“ und „Anderer“. Dies wird genutzt, um jede Anwendung mit einem eigenen Benutzer und einer eigenen Gruppe auszuführen. Dadurch sind die einzelnen Anwendungen voneinander isoliert und können nicht auf Daten der jeweils anderen Anwendung zugreifen. [2]

### Kontrolle des Nutzers über eigene Daten

Die Betriebssystemhersteller nutzen unterschiedliche Wege, um den Nutzer über die Rechte einer Anwendung zu informieren. Android zeigt die Rechte einer Anwendung zur Installationszeit an, während iOS über einen Systemdialog den Nutzer beim ersten Gebrauch der Rechte durch die jeweilige Anwendung zum Einverständnis auffordert. Seit Version 6.0 fordert auch Android beim ersten Rechtegebrauch einer Berechtigung aus einer riskanten Gruppe den Nutzer zum Einverständnis auf. Riskante Berechtigungen sind solche, die auf private Daten des Nutzers zugreifen oder potentiell dessen Daten löschen können. [2] [1]

Wie Felt et al. zeigen, sind Nutzern die Rechte einer Anwendung häufig nicht bewusst. Einerseits wissen einige Nutzer nicht einmal von der Existenz von Berechtigungen, andererseits bestehen bei vielen Nutzern Schwierigkeiten, die angezeigten Berechtigungen zu verstehen und ihren Umfang zu erkennen. Weiterhin sind Berechtigungen in Android auf Ressourcen bezogen und nicht auf damit verbundene Risiken, was diese Abwägung dem Nutzer überlässt. Felt et al. schlagen in ihrer Arbeit verbesserte Berechtigungskategorien, einen Bezug auf Risiken statt Ressourcen, das Entfernen von Warnungen mit geringem Risiko sowie optionale Berechtigungen vor. [18] Einige dieser Vorschläge sind heute in

aktuellen Android-Versionen implementiert, beispielsweise gibt es die eingeführten Systemdialoge nur für riskante Berechtigungen und der Nutzer hat hier die Möglichkeit die Anfrage abzulehnen.

Ein weiterer Ansatz zur Kontrolle der eigenen Daten bietet „MockDroid“. Bei diesem Projekt können einzelne Ressourcen ersetzt und simuliert werden. Dies erlaubt dem Nutzer bei jeder Anwendung selbst zu entscheiden, ob diese auf die richtigen Daten zugreifen darf, simulierte Ersatzdaten bekommt oder einen „nicht verfügbar“-Status. Beresford et al. kritisieren in ihrer Arbeit, dass Android keine „Jedes mal fragen“-Option besitzt. Auch sie merken an, dass ein Nutzer nicht unbedingt versteht, warum bestimmte Rechte gebraucht werden und was mit seinen Daten passiert. [9]

Unter Berücksichtigung der Erkenntnisse von Felt et al. erscheint „MockDroid“ als zu technisch für einen durchschnittlichen Nutzer. Als technischer Ansatz im Hintergrund ist ein breiterer Einsatz durchaus denkbar, allerdings muss hier eine gute Präsentation für den Nutzer gefunden werden, damit dieser von der Funktionalität und damit dem erhöhten Schutz der Privatsphäre profitieren kann.

Während Privatsphäre und Datenkontrolle wichtig sind, sollten diese nicht auf Kosten der Nutzerfreundlichkeit gehen. Berechtigungen müssen klar abgegrenzt, für den Nutzer verständlich und der Grund für die Nutzung einer Berechtigung durch eine Anwendung dem Nutzer bekannt sein, damit dieser die Kontrolle über seine Daten behalten und eine bewusste Entscheidung treffen kann. [18]

Um sich vor potentiellem Datenmissbrauch zu schützen sind das Nutzen von mehreren Faktoren auf Authentifizierung bei Onlinediensten und eine allgemeine Datensparsamkeit ratsam. Das Smartphone wird häufig als zweiter Faktor bei der Authentifizierungen genutzt, indem entweder Einmalpasswörter über SMS versendet werden oder basierend auf einem kryptographischen Schlüssel solche auf diesem generiert werden. Insgesamt können weniger Daten missbraucht werden, wenn weniger Daten dem Gerät bzw. den Anwendungen bekannt sind, dies schränkt jedoch nicht nur einige Dienste sondern auch den Zugriff auf die eigenen Daten durch den Nutzer ein.

## 7.2 Verkehrssicherheit

Im Straßenverkehr ist die Nutzung von Smartphones sowohl für Fußgänger, als auch für Führer von Fahrzeugen ablenkend. Während Nebenbeschäftigungen wie Essen nur für unerfahrene Fahrer zu einer starken Ablenkung führen, ist die Bedienung eines Smartphones sowohl für unerfahrene als auch für erfahrene Fahrer mit einer starken Ablenkung verbunden. Die Nutzung lenkt den Blick vom Straßengeschehen auf das Gerät und verlängert damit die Reaktionszeit bei ungewöhnlichem Straßengeschehen. In der Folge steigt die Gefahr von Unfällen und Beinahe-Unfällen. [21]

In Deutschland ist die Benutzung von Mobiltelefonen während des Führens eines Fahrzeuges nach §23 StVO untersagt. Andere Länder haben ähnliche Vorschriften. In Kalifornien zeigte sich ein deutlicher Rückgang der Unfälle durch Smartphone-Ablenkungen nach der Einführung eines Verbots von Mobiltelefonen am Steuer. [28]

Abgelenkte Fußgänger überqueren Straßen langsamer, ignorieren eher Ampeln und beachten weniger den Verkehr von beiden Seiten. [35]

Während die generelle Nutzung eines Smartphones im Straßenverkehr lediglich ablenkend wirkt, haben neue Phänomene wie Augmented-Reality-Spiele weitere Effekte. Das Spiel „Pokémon Go“ sorgte im Jahr 2016 für Unfälle, bei denen das Spiel explizit als Ursache genannt wurde. Spieler müssen sich bei diesem Spiel an bestimmten Orten befinden um „Pokémon“ zu fangen. Sie lassen sich dabei teilweise nicht vom Straßenverkehr abschrecken und stellen sich z. B. mitten auf eine Straße. [20] Hier sorgt das Smartphone nicht nur für Ablenkung, sondern lenkt den Nutzer direkt in seiner Navigation und damit in potentiell gefährliche Situationen.

## 8 Fazit

Das Smartphone hat sich innerhalb weniger Jahre rasant verbreitet. Heute hat es einen großen Einfluss auf fast alle Bereiche unseres Lebens, sowohl privat als auch im Beruf. Es macht das Internet von fast überall aus zugänglich. Es bietet eine ständige Kommunikationsmöglichkeit, fungiert als Kamera, Musikspieler, Kontaktbuch, Kalender und hat viele weitere Funktionen. Ohne Frage haben diese nützlichen Funktionen zur Verbreitung des Smartphones beigetragen. Insgesamt hat die Nutzung von Smartphones allerdings neben positiven auch negative Seiten. Diese reichen von geminderten persönlichen Kontakten über Ablenkungen bis hin zur Abhängigkeit vom Gerät bzw. den damit erreichbaren Daten und Informationen. Viele dieser negativen Folgen hängen mit einer zu starken Nutzung des Smartphones zusammen. Sie lassen sich mindern, indem das Smartphone weniger und bewusster eingesetzt wird, doch gerade das Abschalten fällt häufigen Nutzern besonders schwer wie Cheever et al. zeigen. Im Bereich der Sicherheit sind Smartphones deutlich weiter als viele Desktop-Betriebssysteme, doch es gibt weiteres Verbesserungspotential wie eine „Jedes Mal Fragen“-Option. Es müssen weitere Ansätze für digitale Rechte erprobt und validiert werden, um deren Wirkung beim Nutzer zu erreichen und zu zeigen, was gerade bei der weiteren Digitalisierung von persönlichen Daten (z. B. im Gesundheitswesen) nötig ist.

Insgesamt sollte nicht nur die Nutzung des Smartphones bewusst erfolgen, sondern gerade auch der Verzicht darauf. Der bewusste Umgang mit Smartphones sollte auch an Kinder und Jugendliche der nächsten Generation weitergegeben werden, die von Anfang an in einer Welt mit Smartphones aufwachsen und diese gar nicht ohne Smartphones kennen.

## Literatur

1. Requesting Permission - Interaction - iOS Human Interface Guidelines. <https://developer.apple.com/ios/human-interface-guidelines/interaction/requesting-permission/>, Zugriff: 31.03.2017
2. Requesting permissions. <https://developer.android.com/guide/topics/permissions/requesting.html>, Zugriff: 31.03.2017
3. Top categories on Google Play. <https://www.appbrain.com/stats/android-market-app-categories>, Zugriff: 31.03.2017
4. The world factbook - telephones - mobile cellular compares the total number of mobile cellular telephone subscribers. <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2151rank.html>, Zugriff: 31.03.2017
5. 17. TK-Marktanalyse Deutschland 2015. [http://www.vatm.de/uploads/media/VA\\_TM\\_TK-Marktstudie\\_2015\\_211015.pdf](http://www.vatm.de/uploads/media/VA_TM_TK-Marktstudie_2015_211015.pdf) (10 2015), Zugriff: 31.03.2017
6. Elektronische Kommunikation und digitaler Binnenmarkt. <http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/ResultDoc/download/DocumentKey/72565> (5 2016), Zugriff: 31.03.2017
7. Global smartphone sales to end users from 1st quarter 2009 to 3rd quarter 2016, by operating system (in million units). <https://www.statista.com/statistics/266219/global-smartphone-sales-since-1st-quarter-2009-by-operating-system/> (11 2016), Zugriff: 31.03.2017
8. Mobile and tablet internet usage exceeds desktop for first time worldwide. <http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide> (11 2016), Zugriff: 31.03.2017
9. Beresford, A.R., Rice, A., Skehin, N., Sohan, R.: Mockdroid: trading privacy for application functionality on smartphones. In: Proceedings of the 12th workshop on mobile computing systems and applications. pp. 49–54. ACM (2011)
10. Bolluyt, J.: 13 Bad Smartphone Habits That Annoy Everyone Around You. <http://www.cheatsheet.com/gear-style/bad-smartphone-habits-annoy-every-one.html/?a=viewall> (10 2016), Zugriff: 31.03.2017
11. Böhm, M.: Navi schickt Autofahrer auf Rollfeld. <http://www.spiegel.de/netzwelt/apps/iphone-navi-apple-maps-schickt-autos-auf-flughafen-rollbahn-a-924612.html> (9 2013), Zugriff: 31.03.2017
12. Cheever, N.A., Rosen, L.D., Carrier, L.M., Chavez, A.: Out of sight is not out of mind: The impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users. *Computers in Human Behavior* 37, 290–297 (2014)
13. Collins, E.I.M., Cox, A.L.: Out of work, out of mind? smartphone use and work-life boundaries. *International Journal of Mobile Human Computer Interaction* (2014)
14. Coupofy: The 2016 Smartphone User Behavior Report: Millennials & Their Smartphone Habits. <http://www.coupofy.com/millennials-smartphone-behavior-report-2016/> (2016), Zugriff: 31.03.2017
15. Elhai, J.D., Levine, J.C., Dvorak, R.D., Hall, B.J.: Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use. *Computers in Human Behavior* 63, 509–516 (2016)
16. Feldhaus, M.: Die Folgen von Mobilkommunikation für die Privatheit. Empirische Ergebnisse zur Beurteilung ubiquitärer Erreichbarkeit in der Familie. *M&K Medien & Kommunikationswissenschaft* 51(1), 24–37 (2003)
17. Felt, A.P., Egelman, S., Wagner, D.: I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In: Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices. pp. 33–44. ACM (2012)

18. Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E., Wagner, D.: Android permissions: User attention, comprehension, and behavior. In: Proceedings of the eighth symposium on usable privacy and security. p. 3. ACM (2012)
19. Fortney, J.C., Burgess, J.F., Bosworth, H.B., Booth, B.M., Kaboli, P.J.: A reconceptualization of access for 21st century healthcare. *Journal of general internal medicine* 26(2), 639 (2011)
20. Joseph, B., Armstrong, D.G.: Potential perils of peri-pokémon perambulation: the dark reality of augmented reality? *Oxford Medical Case Reports* 2016(10) (2016)
21. Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A.: Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine* 370(1), 54–59 (2014)
22. Lee, Y.K., Chang, C.T., Lin, Y., Cheng, Z.H.: The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior* 31, 373–383 (2014)
23. Loh, T.: Digitizing Refugees: The Effect of Technology on Forced Displacement. <http://www.gnovisjournal.org/2016/04/29/digitizing-refugees-the-effect-of-technology-on-forced-displacement/> (4 2016), Zugriff: 31.03.2017
24. Martin, T.: The evolution of the smartphone. <http://pocketnow.com/2014/07/28/the-evolution-of-the-smartphone> (7 2014), Zugriff: 31.03.2017
25. Meeker, M.: Internet Trends 2016. <http://www.kpcb.com/file/2016-internet-trends-report> (6 2016), Zugriff: 31.03.2017
26. Niccolai, J., Gohring, N.: A Brief History of Palm. <http://www.pcworld.com/article/195199/article.html> (4 2010), Zugriff: 31.03.2017
27. Peck, J., Childers, T.L.: Individual differences in haptic information processing: The “need for touch” scale. *Journal of Consumer Research* 30(3), 430–442 (2003)
28. Ragland, D.: Descriptive analyses of traffic fatalities and injuries before and after California’s law banning hand-held cell phone use while driving was implemented on July 1, 2008. [http://www.ots.ca.gov/Media\\_and\\_Research/Press\\_Room/2012/doc/CA\\_Cell\\_Phone\\_Law\\_Study.pdf](http://www.ots.ca.gov/Media_and_Research/Press_Room/2012/doc/CA_Cell_Phone_Law_Study.pdf) (3 2012), Zugriff: 03.02.2017
29. Redelmeier, D.A., Detsky, A.S.: Pitfalls with smartphones in medicine. *Journal of general internal medicine* 28(10), 1260–1263 (2013)
30. Reed, B.: A brief history of smartphones. <http://www.networkworld.com/article/2869645/network-security/a-brief-history-of-smartphones.html> (6 2010), Zugriff: 31.03.2017
31. Rotter, J.B.: Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs* 80(1) (1966)
32. Salehan, M., Negahban, A.: Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior* 29(6), 2632–2639 (2013)
33. Sarwar, M., Soomro, T.R.: Impact of smartphone’s on society. *European Journal of Scientific Research* 98(2), 216–226 (2013)
34. Schlenker, B.R., Leary, M.R.: Social anxiety and self-presentation: A conceptualization model. *Psychological bulletin* 92(3), 641 (1982)
35. Thompson, L.L., Rivara, F.P., Ayyagari, R.C., Ebel, B.E.: Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. *Injury prevention* 19(4), 232–237 (2013)
36. Williams, A.: Mind your blackberry or mind your manners. <http://www.nytimes.com/2009/06/22/us/22smartphones.html> (6 2009), Zugriff: 31.03.2017

# Sensorische Augmentation durch taktile Schnittstellen

Felix Rittler\*

Betreuer: Erik Pescara†

Karlsruher Institut für Technologie (KIT)  
Pervasive Computing Systems – TECO

\*felix.rittler@student.kit.edu

†pescara@teco.edu

**Zusammenfassung.** Bei der sensorischen Augmentation kann man den menschlichen Sinnesapparat durch neue Sinne erweitern, was möglich wird durch die Gehirnplastizität. Sensoren senden die gesammelten und vorverarbeiteten Informationen über taktile Schnittstellen an das Gehirn. Besonders oft werden neue Sinne in der Navigation, sowie bei Personen eingesetzt, bei denen andere Sinne beschädigt sind. Besonders als Navigationshilfe wurden schon einige Geräte getestet, jedoch können sie auch helfen Fehlfunktionen anderer Sinne auszugleichen oder diese zu ersetzen. Es werden bei der sensorischen Augmentation Daten über Sensoren gesammelt, weiterverarbeitet, an den Zweck des Systems und an die Übertragung auf die Haut angepasst und gefiltert. Im Anschluss übertragen taktile Displays bspw. mittels schwacher elektrischer Stöße, Vibrationen oder Verformungen von Substanzen diese Signale an die Haut.

Damit solche Systeme auch kommerziell erfolgreich sein können, gibt es jedoch noch einige Probleme. So wurden noch keine Konzepte für eine Integration der Geräte erstellt und v.A. für komplexe Informationen lässt die Qualität der Geräte und der Informationen stark nach.

**Schlüsselwörter:** sensorisch, Augmentation, taktil, Substitution, neue Sinne, taktile Displays, Navigation

## 1 Einleitung

Mit technischen Geräten agieren wir heutzutage hauptsächlich auf eine bewusste Art und Weise. Sie liefern uns über den Sehsinn oder das Gehör Informationen mittels Bilder, Text, Geräuschen oder Sprache durch elektronische Geräte wie Smartphones, PCs oder Tablets. Diese Informationen nehmen wir auf und verarbeiten sie. Taktile Schnittstellen werden eher selten verwendet. Wenn diese verwendet werden, dann haben die taktilen Reize meistens eine Warn- oder Benachrichtigungsfunktion, bspw. die Vibrationsfunktion bei Benachrichtigungseingang, welche in Smartphones verwendet wird, oder haptisches Feedback nach Aktionen, das über Erfolg oder Misserfolg informiert. Komplexere Informationen werden fast nie über diesen Weg transportiert. Beim haptischen Feedback oder

der Benachrichtigungsfunktion wird lediglich zwischen Intensität der Vibration und teilweise auch zwischen Vibrationspattern unterschieden. So sind länger andauernde Vibrationen für Anrufe und zwei kürzere, schnell aufeinanderfolgende, Vibrationen für sonstige Benachrichtigungen bekannt. Die Möglichkeiten, den Vibrationen eine bis zu dreidimensionale räumliche Information mitzugeben, indem wir die Vibrationen an unterschiedlichen Körperstellen spüren oder, dass diese verschiedene Typen von Rezeptoren ansprechen können, werden bislang nicht genutzt.

Darüber hinaus werden Informationen in der Regel bewusst angefordert und verarbeitet. So schauen wir bspw. bei der Navigation auf einen Kompass, der uns anzeigt, wohin wir laufen müssen. Stattdessen könnte uns mitgeteilt werden, wo Norden ist, ohne dass wir Aufmerksamkeit darauf richten müssen.

Dieser Ansatz führt zur Entwicklung von Geräten, die unsere Sinne erweitern. Deren Qualität ist ausschlaggebend, inwiefern die sensorische Augmentation erfolgreich sein wird. Diese Seminararbeit wird sich daher mit diesen Geräten, Konzepten, deren Qualität und ungelösten Fragen beschäftigen.

Bei der sensorischen Augmentation werden bestimmte Eigenschaften des menschlichen Körpers genutzt. Genauer: Die Gehirnplastizität[5].

## 2 Grundlagen

### 2.1 Signalverarbeitung des Körpers und Gehirnplastizität

Die neuronale Plastizität spielt eine wichtige Rolle in der Erforschung von sinneserweiternden Schnittstellen, da sie überhaupt erst ermöglicht, dass neue Sinne entstehen können.

Das Gehirn „modifiziert die eigene strukturelle Organisation und Funktionsweise“[3] und passt sie an die Anforderungen an, was neurochemische, synaptische, rezeptorische und neuronale strukturelle Veränderungen beinhaltet[5].

Außerdem wichtig ist, dass unsere Sinnesindrücke nicht in den Sinnesorganen entstehen. So sieht man nicht mit dem Auge, im Auge werden lediglich die Signale empfangen und an das Gehirn gesendet. Dieses konstruiert dann daraus ein Bild, wobei persönliche Erfahrungen miteinbezogen werden. Es „sammelt“ alle Informationen, die verfügbar sind, und konstruiert daraus den Eindruck, den wir sehen, hören, riechen oder fühlen. Auf diese Weise können auch neue Sinne entstehen. Da Erfahrungen nicht im Sinnesorgan entstehen, sondern im Gehirn, können in Kombination mit der Gehirnplastizität neue Sinnesorgane durch Training etabliert werden. Dadurch werden diese neuen Übertragungswege in die Bilder und Eindrücke, die das Gehirn erstellt, unbewusst miteinbezogen[29].

### 2.2 Konzept der sensorischen Augmentation

Ziel der sensorischen Augmentation ist es, den menschlichen Sinnesapparat zu erweitern. Dazu nehmen Sensoren bestimmte Signale aus der Umwelt auf, die dann über bestehende Sinne an das Gehirn weiter transportiert werden.

Die sensorische Augmentation ist aus der **sensorische Substitution** entstanden. Diese wurde in den 1960er Jahren von Paul Bach-y-Rita begründet, wobei er sich mit Geräten beschäftigte, die blinden Menschen helfen, wieder sehen zu können. Später wurde die Forschung auch auf weitere Einsatzzwecke erweitert. Kaczmarek et al. beschreiben die sensorische Substitution so, dass bei diesem Konzept ein Sinn verwendet wird um Informationen zu empfangen, die normalerweise von einem anderen Sinn empfangen werden. Für den Tastsinn bedeutet das, dass an einem Teil der Haut Informationen empfangen werden, die sonst an einem anderen Teil der Haut ankommen[17]. In der neueren Entwicklung werden hierfür zumeist elektronische Geräte genutzt. In Abgrenzung zu herkömmlichen Maßnahmen, wie dem Tragen von Brillen oder Hörgeräten, werden Mängel des bestehenden Sinnesorgans nicht ausgeglichen, sondern eine Ersetzung des Sinns durchgeführt.

Die sensorische Augmentation oder Substitution muss sich jedoch nicht auf elektronische Geräte beschränken. So ist Braille („Blindenschrift“) bereits eine Form der (taktilen) sensorischen Substitution. Informationen, die sonst das Auge wahrnimmt, werden übersetzt, sodass sie ertastet werden können. Die sehbehinderte Person nimmt jedoch nicht mehr die Punkte wahr, die sie ertastet, sondern lediglich die Buchstaben und Wörter, ähnlich dem wirklichen Lesen. Sogar das Lesen kann als sensorische Substitution gesehen werden, da es lediglich eine Substitution für das Gespräch ist. Auch hier werden nicht mehr die Linien auf dem Papier, sondern die Buchstaben, Laute oder Wörter wahrgenommen[5].

Ein anderes Beispiel ist das Tasten mit einem Stock, bspw. einem Blindenstock. An der Spitze des Stocks hat man keine Rezeptoren, die Signale an das Gehirn senden. Lediglich die Rezeptoren in der Hand detektieren Veränderungen der Lage des Stocks. Dennoch fühlen wir nicht dieses Detektieren, sondern bekommen ein Gefühl für das Ende des Stocks[5].

In Erweiterung zur sensorischen Substitution wird bei der **sensorischen Augmentation** nicht nur ein bestehender Sinn und ein dazugehöriges Sinnesorgan ersetzt, sondern ein neuer Sinn geschaffen. Die sensorische Augmentation entwickelte sich, als man nicht mehr nur versuchte schlecht oder gar nicht funktionierende Sinne zu ersetzen, sondern diese mangelbehaftete Funktionalität durch kreative neue Lösungen auszugleichen. Inzwischen geht die Forschung so weit, dass auch versucht wird neue Sinne zu schaffen, die auch für Menschen mit vollständig funktionierendem Sinnesapparat eine Erweiterung darstellen.

Auch wenn es in dieser Seminararbeit vorwiegend um sensorische Augmentation gehen soll, so stellt die sensorische Substitution die Grundlage der Forschung zur Augmentation dar. Die entwickelten Geräte lassen sich auch weitgehend in die sensorische Augmentation übertragen. Dies bedeutet, dass nur die transportierte Information, d.h. der Sinn als solches, ersetzt werden muss, jedoch die Geräte ansonsten gleich aufgebaut sind. Aufgrund dessen sind die Erkenntnisse aus der sensorischen Substitution für die sensorischen Augmentation sehr wichtig und hier auch zu behandeln.

Des Weiteren soll es hier im Besonderen um Geräte und Konzepte gehen, bei denen Informationen von den Sensoren über den Tastsinn, also auf taktile Art und Weise, an das Gehirn übertragen werden.

### 3 Entwickelte Sinne und deren Einsatzzwecke

Wissenschaftler auch abseits von Begründer Paul Bach-y-Rita [4] bzw. denen, die mit ihm zusammen arbeiteten, entwarfen eine ganze Reihe an Geräten und Konzepten, die eine sensorische Augmentation möglich machen. Die dabei entwickelten Sinne kann man in verschiedene Kategorien einteilen: Die, die bereits im menschlichen Sinnesapparat vorhanden sind, womit eine sensorische Substitution stattfindet und die, die so vom menschlichen Körper nicht bereitgestellt werden, womit man von einer sensorischen Augmentation spricht.

#### 3.1 Sensorische Substitution

Ein großer Teil der Forschung über die sensorische Substitution beschäftigt sich mit dem Ersatz des Sehsinns. Bach-y-Rita et al. [4][2] entwickelten sogenannte TVSS-Geräte (Tactile Vision Sensory Substitution), die einem blinden Menschen helfen die Umgebung besser wahrzunehmen, quasi zu sehen.

Des Weiteren gibt es einige Konzepte, die sich damit beschäftigen den Hörsinn zu ersetzen und Tauben zu helfen wieder zu hören, wobei diese Konzepte auch kommerziell getestet wurden[30][10].

Ein weiterer Ansatz ist die Ersetzung des Tastsinns. Diese kann für querschnittsgelähmte Personen hilfreich sein. Sinneseindrücke an Stellen, wo diese Personen nichts fühlen können an Stellen übertragen werden, an denen sie etwas fühlen. Eine medizinische Anwendung könnte darin bestehen, der Druckstellenbildung vorzubeugen. Querschnittsgelähmte Menschen sitzen und liegen oft den ganzen Tag, können jedoch gleichzeitig nicht fühlen, wenn sich Druckstellen bilden. So können sie bspw. an der Zunge wahrnehmen wenn sich Druckstellen an den Beinen bilden[25].

Den Tastsinn zu ersetzen kann jedoch auch für Leute interessant sein, die Schutzanzüge tragen, die den Kontakt mit Gegenständen beeinträchtigen, und diese so weniger fühlbar machen. Dies können bspw. Astronauten sein, man kann es aber auch auf Weltraumroboter übertragen, die auf diese Weise Informationen sammeln, diese an die Basisstation übertragen und so Gegenstände für die Menschen an der Basisstation fühlbar werden, die weit entfernt sind[1].

Es können auch im Körper tiefer integrierte und unbewusst benutzte Sinne substituiert werden. So ersetzen Tyler et al. den Gleichgewichtssinn. Dessen Ersetzung kann sinnvoll sein, da dieser Sinn durch Krankheiten oder Vergiftungen, bspw. durch eine Überdosis Drogen, eingeschränkt sein kann.

#### 3.2 Sensorische Augmentation

Mehr Möglichkeiten an möglichen Sinnen erlaubt die sensorische Augmentation, da sie sich nicht auf die Sinne beschränkt, die ein Mensch normalerweise besitzt.

Ein großer Teilbereich ist die Sinneserweiterung um die Navigationsfähigkeiten zu verbessern. Man kann hier eine Art Sonar entwickeln, womit man Hindernisse und deren ungefähre Lage detektieren kann[19]. Wenn man sich schnell bewegt können die Informationen jedoch zu einer kognitiven Überlastung führen, andererseits jedoch das Gehirn, wenn man steht, detailreichere Informationen verkraften. Bertram et al. schlagen hier vor, dass bei schnellerer Geschwindigkeit praktisch nur noch Kollisionswarnungen angezeigt werden. Diese werden schwächer, bis man steht. Bei langsamer Bewegung schaltet sich darüber hinaus noch ein „Scanning“-Modus ein, der genauere Informationen über die Umgebung liefert[6]. Eine andere Möglichkeit um die Navigation zu verbessern ist dem Träger anzuzeigen, in welcher Richtung „Norden“ liegt.

Einsatzmöglichkeiten für solche Navigationshilfen sind vielfältig. Mit einem Sinn für „Norden“ kann Blinden geholfen werden Abkürzungen zu nehmen oder über längere Strecken die Richtung beizubehalten[18]. Letzteres ist auch für Sehende interessant, da der Effekt bekannt ist, dass man in Terrain ohne Anhaltspunkte zum Orientieren, bspw. in der Wüste, nur im Kreis läuft. Wenn einem jedoch die Richtung des Nordpols immer angezeigt wird, kann so etwas vermieden werden. Ähnliches gilt für Segler und Wanderer[36] oder für Umgebungen, bei denen herkömmliche Displays oder ein herkömmlicher Kompass nur bedingt geeignet sind.

Van Erp et al. beschreiben, dass bei einem Pilot, dessen Sehsinn durch viele Displays ausgelastet ist, ein Kompass nicht so gut geeignet ist. Hier kann ein zusätzlicher Sinn, der nicht so viel Aufmerksamkeit benötigt, Abhilfe schaffen. Für Fahrer von schnellen Motorbooten sind visuelle Displays nicht gut geeignet, da diese sich bei schneller Fahrt durch das Wackeln des Bootes und evtl. Spritzwasser nicht gut ablesen lassen[35]. Dabei beschreiben sie zwar ein Gerät, das nicht die Richtung des Nordpols sondern des nächsten Wegpunktes auf einer Route anzeigt, die dadurch gewonnenen Erkenntnisse lassen sich jedoch ohne weiteres übertragen.

Sonare mit einer Kollisionswarnung sind sowohl für Sehbehinderte interessant, als auch für Sehende in Umgebung mit schlechter Sicht[6]. Bertram et al. führen hier das Beispiel eines Feuerwehrmanns an, der in einem unbekanntem verrauchten Haus navigieren muss.

An Gehörlose richtet sich ein Sinn, der das Lippenlesen vereinfacht. So können Vibrationen in der Nase oder am Kehlkopf des Sprechers sowie die Amplitude während dem Sprechen taktil an den Gehörlosen übertragen werden[24][23].

## 4 Arten von Taktilem Displays

Ein wichtiges Teilgebiet der sensorischen Augmentation ist die Übertragung von Informationen. Diese können bereits vorverarbeitet oder gefiltert sein. Bei taktilem Displays werden diese auf die Haut und weiter über den Tastsinn übertragen. Gelingt die Übertragung nicht, so gehen mindestens einige Informationen verloren, wobei bei groben Verfehlungen im Design des taktilem Displays die Haut sogar Schäden davon tragen kann.

Diese taktilen Displays bestehen aus Signalsendern, wobei zumeist die räumliche Lokalität des Signals auf der Haut entscheidend für die Wahrnehmung und für die überbrachte Information ist. Codes aus Kombinationen von gleichzeitig sendenden Signalsendern werden zumeist nicht verwendet. Somit kann vermieden werden, dass die überbrachten Informationen erst noch dekodiert werden müssen. Bei den Signalsendern sind alle Signale möglich, die in irgendeiner Weise von der Haut detektiert werden können, ob mechanisch, elektrisch oder thermisch. Es kann ein einzelner Signalsender verwendet werden, wobei alternativ auch mehrere Sender ein- oder zweidimensional auf der Haut angeordnet werden können. Hier sind sowohl engere, als auch weiter auseinander liegende Anordnungen möglich. Die Signalsender können des Weiteren oft in verschiedenen Intensitäten und Frequenzen senden. Hieraus ergeben sich viele verschiedene Möglichkeiten, die einen Vergleich schwierig machen. Es haben sich jedoch Bewertungsmaßstäbe etabliert.

Bei der Bewertung taktiler Displays spielen mehrere Aspekte eine Rolle. Einerseits natürlich Komfort, ein potentieller Nutzer muss sich wohl fühlen mit einem solchen Display auf der Haut. Ein Gerät, das einen ausreichenden Komfort nicht bietet wird kaum auf dem Markt erfolgreich sein. Weitere Kriterien sind die Qualität der Informationsübertragung auf die Haut und wie gut die Displays sich steuern lassen. Der Energieverbrauch ist auch ein wichtiger Aspekt, v.A. da es sich bei der sensorischen Augmentation oft um mobile Geräte handelt, die ihre Energie von einem Akku beziehen.

Hier werden mehrere verschiedene Arten von Displays bei der Übertragung von Informationen auf die Haut unterschieden.

#### 4.1 Vibrotaktile Displays

Vibrotaktile Displays bestehen aus angeordneten Vibrationsmotoren, die haptische Informationen mittels Vibrationen auf die Haut transportieren. Sie können durch stärkere Vibrationen (also Erhöhung der Amplitude) auch ihre Intensität und durch schnellere Vibrationen ihre Frequenz steigern.

Probleme mit der Sicherheit während der Benutzung bei vibrotaktilen Displays können Verbrennungen sein, die durch zu stark aufgeheizte Displays entstehen. So beschreibt LaMotte[21], dass bereits 3 Minuten bei 49°C mit einem Display mit 7,5 mm Durchmesser ausreichen, um eine kleinere Verbrennung auszulösen.

Der Tragekomfort hängt von verschiedenen Faktoren ab. Van Erp schrieb, dass es wichtig ist, dass der Komfort auch über längere Zeiträume gesichert ist und, dass das Display den Nutzer nicht stören sollte. 15-20 dB über der Schwelle, an der man etwas spürt sollten nicht überschritten werden. Amplituden über 0.6-0.8 mm führen sogar zu Schmerzempfindungen. Collins schreibt, dass bei einer Amplitude von 0,5 mm für einen Stimulator mit 1 mm Durchmesser der Komfort am höchsten ist[9]. Die Abwärme sollte maximal  $62 \text{ mW/cm}^2$  betragen um ein Hitzegefühl und damit einhergehendes Schwitzen zu vermeiden.

Im angemessenen Komfort zu gewährleisten sollte man außerdem aufpassen, dass die Vibrationen des Displays sich nicht auf den ganzen Arm ausdehnen.

Die Vibrationen des Arms sollten auf max. 12 Hz begrenzt werden und max.  $1\text{-}5\text{ m/s}^2$  stark sein[34].

Die Empfangsqualität scheint zumindest bis zu einer Stimulatorgröße von  $0.05\text{cm}^2$  bei einer Frequenz von 250 Hz gut zu sein[17].

## 4.2 Elektrotaktile Displays

Bei elektrotaktile Displays handelt es sich um Displays, die aus mehreren angeordneten Elektroden bestehen. Wenn ein Signal an die Haut gesendet werden soll, so wird eine Spannung angelegt, was im Vergleich zu vibrotaktile Displays zu ganz eigenen Schwierigkeiten führt.

Die Sicherheit ist ein Problem, das besonders bei elektrotaktile Displays auftritt. So können verschiedene Arten von Verbrennungen entstehen, wenn die elektrischen Impulse zu lang, zu frequent oder wenn Spannung und Stromstärke zu hoch sind[17]. Ein weiteres Problem ist die Möglichkeit eines elektrischen Schocks. So schreibt Olson, dass der Strom durch den Körper zu jedem Zeitpunkt kleiner als 0,1 mA sein sollte, um diesem vorzubeugen.[26]

Der Tragekomfort kann sehr hoch sein, Kaczmarek et al. kommen jedoch zu dem Schluss, dass er nicht nur stark zwischen verschiedenen Personen variiert. Auch bei einer Person kann der Komfort stark variieren, je nachdem zu welchen Zeiten man das Display trägt. Um das zu verbessern, sollte der Nutzer die Stärke der Stimulation schnell und einfach einstellen, aber auch komplett ausschalten können[17]. Außerdem sollten Display und Haut immer in Kontakt sein, da sonst ein ausreichender Tragekomfort nicht gewährleistet ist[15].

Bezüglich des Energieverbrauchs benötigt eine Elektrode mit 3 mm Durchmesser 1,2 mW/Pixel bei einem komfortablen Stimulationslevel von 6 mA. Der reale Verbrauch verringert sich jedoch deutlich und die Effizienz wird erhöht, wenn man bedenkt, dass in einem praktischen System in der Regel nicht alle Elektroden gleichzeitig stimulieren[17].

## 4.3 Weitere Arten

**Skin Stretch:** Ein besonderes Konzept im Bereich der taktilen Displays ist Informationen, insbesondere Richtungsinformationen, über „Skin Stretch“-Feedback zu übertragen. Dabei wird die Haut in eine bestimmte Richtung mittels eines kleinen herausstehenden Stabs „bewegt“. Skin Stretch eignet sich besonders gut für Richtungsinformationen oder um die Steifheit eines Objektes herauszufinden [31][27].

**Displays mittels fokussiertem Ultraschall:** Mittels Ultraschall ist es möglich Rezeptoren direkt zu stimulieren. Gavrilov et al. erforschten die Wirkung von Ultraschall auf neurale Rezeptorstrukturen. Sie kamen zu dem Schluss, dass Ultraschall eine sehr hohe Genauigkeit zulässt[12][8]. Ultraschall lässt ein breites Frequenzband zur Stimulation zu, ohne Schäden zu verursachen. Außerdem kann akustische Druck auf die Haut sehr leicht kontrolliert werden. Akustische

Stimulation eignet sich ebenso, wenn man nur bestimmte Rezeptoren ansprechen will. Die Rezeptoren der Haut liegen in unterschiedlichen Schichten und so kann trotz eigentlich zweidimensionaler Haut ein dreidimensionales Bild erzeugt werden[8].

**Elektro- und magnetorheologische Displays:** Elektro- und magnetorheologische Displays funktionieren mittels einer Flüssigkeit. In Abhängigkeit einer angelegten Spannung oder eines Magnetfelds verändern sich Eigenschaften wie die Viskosität, sowie bei magnetorheologischen Displays zusätzlich die Festigkeit und die Oberflächenstruktur der Flüssigkeit. Die Haut erföhlt diese Veränderungen, sodass sich Informationen übertragen lassen[22].

Bei elektrorheologischen Displays verändert sich das Fluid von flüssig zu halbfest. Wenn eine Spannung angelegt wird, werden in der Flüssigkeit Dipole induziert, die sich in Feldrichtung ausrichten und die Flüssigkeit wird fester.

Magnetorheologische Displays bestehen aus Metallpartikel in einer Flüssigkeit. Wenn ein Magnetfeld angelegt wird richten sich die Metallpartikel an den Feldlinien aus und die Flüssigkeit wird fester und verändert ihre Struktur[8].

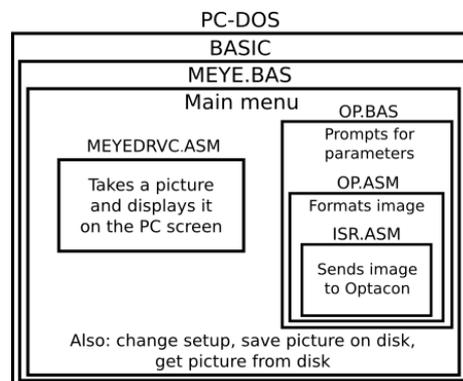
## 5 Verarbeitung der Signale

Die Signalverarbeitung ist das Herzstück eines Geräts zur sensorischen Augmentation. Wenn sie nicht gut funktioniert, dann gehen entweder wichtige Informationen verloren, oder der Empfänger wird mit unnötigen Informationen überlastet.

Deshalb muss eine Selektion stattfinden, welche Informationen verworfen werden und welche gesendet. Dies ist umso wichtiger, da die Signale, die über die Haut empfangen werden zwar schnell verarbeitet werden, die Bandbreite jedoch (im Vergleich zum Auge oder dem Gehör) sehr begrenzt ist[8].

Ebenso wichtig ist, in welcher Form die Informationen gesendet werden. Wenn der Empfänger die Informationen erst noch aufwendig dekodieren muss, dann verliert man Bandbreite und Geschwindigkeit und es ist eine längere Trainingszeit notwendig.

Es gibt Zwecke, für die eine aufwendige Signalverarbeitung und Anpassung an die Rezeptoren der Haut nicht notwendig ist. Dies ist bei sehr einfach strukturierten Signalen wie



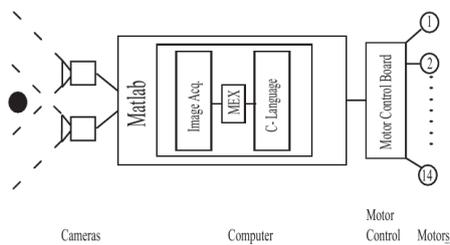
**Abb. 1.** Architektur der Software des Bildprozessors [16]

die, eines Geräts, das in Richtung Norden zeigt, der Fall[18]. Wenn, wie oben beschrieben, der Tastsinn ersetzt werden soll, z.B. für einen Astronauten oder bei querschnittsgelähmten Personen, so müssen die Signale, abgesehen von einer Kalibrierung des Displays, nicht verändert werden, da es sich auf beiden Enden um taktile Signale handelt[25]. Handelt es sich jedoch um komplett verschiedene Signaltypen, die auch noch komplexer sind als eine einfache Richtungsanzeige, so ist die Verarbeitung dieser aufwendiger.

Während in ihrer ersten Arbeit im Jahr 1969 Bach-y-Rita et al. den Kontrast der Pixel, die von der Kamera aufgenommen werden mehr oder weniger direkt mit den Pixel des taktilen Displays verknüpfen[4], schreiben schon Kaczmarek et al. über einen, wenn auch einfachen, Bildprozessor (s. Abb. 1), der die aufgenommenen Bilder verarbeitet[16]. Bei Bach-y-Rita et al. müssen die Informationen zwar auch natürlich verarbeitet werden, jedoch scheint diese Verarbeitung nicht maßgeblich für die Studie, sodass man sich ihr nicht im Artikel widmet.

Bei Johnson und Higgins aus dem Jahr 2005 wird die Bildverarbeitung komplexer. Sie entwarfen eine Navigationshilfe, die zwei Kameras verwendet, welche Bildinformation aufnehmen und diese an ein Notebook senden. Diese Bilder werden durch einen Algorithmus, der in Matlab läuft, in räumlich angeordnete, vibrotaktile Stimuli umgewandelt und an die einzelnen Vibrationsmotoren gesendet (s. Abb. 2). Johnson und Higgins beschreiben den Algorithmus als „computational stereo algorithm“ (deutsch etwa: „rechnergestützter Stereobild-Algorithmus“), bei dem aus zwei Bildern (ähnlich der Funktionsweise des menschlichen Auges) eine dreidimensionale Information gewonnen wird. [14].

Eine Besonderheit, die dieses System mit sich bringt, ist die Tatsache, dass es lediglich als Navigationshilfe gedacht ist. Daher bekommt der Nutzer nicht die gesamte Bildinformation, sondern nur die Richtung eines Objektes und dessen Entfernung[14], welche die einzigen wirklich relevanten Informationen für diesen Einsatzzweck sind. Formen, Farben etc. werden nicht gesendet. Dies ermöglicht, dass man die Vibrationsmotoren nicht zweidimensional anordnen muss. Stattdessen reicht eine eindimensionale Anordnung als Gürtel aus. Die empfangenen Informationen sind stark vereinfacht, was dazu führt, dass der Nutzer schneller Entscheidungen treffen kann und mögliche Hindernisse schneller erkennt. Dies ist wiederum eine der Grundvoraussetzungen dafür, dass diese Art von Navigationshilfe funktionieren kann. Ursprünglich nötiges Training kann reduziert werden und eine kognitive Überlastung kann durch die geringere Information vermieden werden.



**Abb. 2.** Schema des Prototyps[14]

Die empfangenen Informationen sind stark vereinfacht, was dazu führt, dass der Nutzer schneller Entscheidungen treffen kann und mögliche Hindernisse schneller erkennt. Dies ist wiederum eine der Grundvoraussetzungen dafür, dass diese Art von Navigationshilfe funktionieren kann. Ursprünglich nötiges Training kann reduziert werden und eine kognitive Überlastung kann durch die geringere Information vermieden werden.

Ebenso wichtig für Navigationshilfen ist, dass die Signale in der richtigen Intensität ankommen, damit der Nutzer die Entfernung korrekt einschätzen kann. Für die empfundene Intensität gibt es das Weber-Fechner-Gesetz[6], das, obwohl umstritten, wohl zumindest bei niedrigen bis mittleren Intensitäten korrekte Ergebnisse liefert. Vereinfacht sagt dieses aus, dass, obwohl bspw. eine Entfernung linear wächst, der dafür nötige Stimulus jedoch exponentiell wachsen muss, damit das lineare Wachstum als solches wahrgenommen wird. Ein einfaches Übertragen der Intensität des Stimulus ist als solches nicht möglich. Für mittlere bis hohe Intensitäten gibt es eine Debatte über die Korrektheit[13][11]. Hohe Intensitäten sind jedoch im Bereich der sensorischen Augmentation meist nur für kurze Warnungen wichtig, sodass in diesem Bereich ein genaues Empfinden der Signale nicht unbedingt erforderlich ist.

Tyler et al. entwarfen ein System, bei dem der Gleichgewichtssinn substituiert wird. Um das zu erreichen verwenden sie einen 2-Achsen-Beschleunigungssensor. Der Beschleunigungssensor zeigt jedoch zunächst nur an in welche Richtung man sich bewegt, also ob man das Gleichgewicht verliert und zu stark in eine Richtung gebeugt ist. Wenn man dem Benutzer die Beschleunigung direkt überträgt, dann muss er aktiv registrieren, dass er zu weit in eine Richtung gelehnt ist und Gegenmaßnahmen einleiten. Viel intuitiver wäre es, wenn der Benutzer die Richtung angezeigt bekommt, in die er sich lehnen muss. Tyler et al. integrieren hierzu die vom Beschleunigungssensor gesendeten Signale zweifach und bekommen so eine vorne-hinten und eine mitte-seitlich Information. Diese Signale werden an das taktile Display weitergereicht. Der Nutzer bekommt nun durch einen Stimulus an einer Seite des Displays angezeigt, dass er sich in die entsprechende Richtung lehnen muss, was er dann tut, worauf der Stimulus zur Mitte wandert[33].

## 6 Diskussion und Ausblick

### 6.1 Entwickelte Sinne und deren Einsatzzwecke

Der Kreativität bei der Entwicklung neuer Sinne sind kaum Grenzen gesetzt. So gab es in der Vergangenheit viele Ideen für die Einsatzzwecke und Sinne, bei denen ein praktischer Nutzen vorhanden ist. Besonders im Bereich der Navigationshilfen gibt es sowohl für Menschen mit, als auch ohne Sehbehinderung eine Vielzahl an Studien, die das Ziel haben, den Sinnesapparat zu erweitern. Bisher wird meist versucht Schwächen bestehender Sinne auszugleichen. So wird in etwa auf ein Sonar gesetzt, wenn die Sicht schlecht ist. Ideen für neue Sinne, mit denen Informationen aufgenommen werden können, die bisher überhaupt nicht durch unseren Sinnesapparat abgedeckt sind, sind eher wenige, was auch damit zusammenhängen könnte, dass dieser sehr ausgereift und vielseitig einsetzbar ist. Außerdem sind Menschen durch bereits bekannte Dinge und Sinne aus der Natur unter Umständen voreingenommen und so in ihrer Kreativität beschränkt. Zu nennen wäre hier jedoch ein Sinn für magnetische Felder um einen Kompass zu implementieren.

Auch denkbar ist ein Sinn für radioaktive Strahlung. Dies kann für Arbeiter die bspw. in Tschernobyl tätig sind interessant sein. Diese müssen nicht mehr

bewusst auf den Geigerzähler achten oder auf eine Benachrichtigung warten, sondern nehmen unbewusst zu jeder Zeit die radioaktive Strahlung wahr. Eine weitere Möglichkeit ist ein Sinn für offenes WLAN und dessen Stärke. Viele Menschen sind auch in Zeiten steigender Datenvolumen in Mobiltarifen auf offene Internetzugänge angewiesen und können so darüber informiert werden. Sinnvoll ist dies allerdings nur während sie sich nicht fortbewegen, da dies sonst sehr schnell sehr störend werden kann, vor allem in der Stadt.

## 6.2 Taktile Displays

Wenn man die verschiedenen Arten taktile Displays vergleicht, so stellt sich heraus, dass man nicht sagen kann, dass es ein Konzept gibt, das überall am besten abschneidet. So beschreiben Kaczmarek et al., dass oftmals ein Kompromiss gemacht werden muss, zwischen Leistung und Komfort eines Displays[17]. Andererseits hängt die Art des bestmöglichen Displays auch mit dem Körperteil zusammen, an dem es angebracht werden soll. So ist zum Beispiel ein elektrotaktiler Display auf der Zunge wesentlich besser geeignet als bspw. am Finger[28]. Ein elektrotaktiler Display hat jedoch den Nachteil, dass der Kontakt zwischen Haut und Display immer gewährleistet sein sollte und die verwendete Spannung und Stromstärke an das Individuum angepasst sein sollte, was sie für viele Zwecke, besonders für Massenware unpraktikabel machen können[15][17].

Vibrotaktile Displays wurden besonders in der Anfangszeit der sensorischen Augmentation und Substitution erforscht. Obwohl die anderen Arten viele Vorteile haben, sind sie noch nicht so stark präsent, was mit Schwierigkeiten bei der Herstellung und bei der Integration zusammenhängt[8].

Der Komfort bei den unter „Weitere Arten“ geführten Displayarten scheint am höchsten zu sein. Falls ein elektrotaktiler Display korrekt kalibriert wurde, kann es ebenfalls sehr komfortabel sein. Das Problem bei Vibrationsmotoren ist, dass sie recht schnell sehr warm werden können. Dies hängt mit der mechanischen Energie und der Reibung zusammen. Sicherheitstechnisch scheinen elektrotaktile Displays am problematischsten zu sein, da eine ständig anliegende Spannung die Haut reizen kann und im schlimmsten Fall zu Verbrennungen führen kann. Im Vergleich zu vibro- und elektrotaktile Displays sind die unter „Weitere Arten“ geführten Displayarten noch weniger erforscht und oft noch nicht mit sensorischer Augmentation getestet, was auch an Fertigungsschwierigkeiten liegt[8]. Über Risiken dieser Displayarten ist nur wenig bekannt.

So ist der Bereich der taktilen Displays insgesamt noch ein recht wenig explorierter Forschungsbereich. Zwar ist für einige Körperstellen klar, dass deren taktile Sensitivität geringer ist als andere Körperstellen, so ist bspw. die Zunge besser geeignet für elektrotaktile Stimulation als der Finger[28], grundsätzlich ist jedoch noch nicht eindeutig bestimmt, wie sich taktile Displays verändern müssen, wenn sie an andere Körperstellen transportiert werden[15]. Dies hängt auch mit einer Vielzahl an Konfigurationsmöglichkeiten taktile Displays zusammen. Diese reichen von Größe, Intensitäten, Frequenzen, der Größe und Dichte der Signalsender bis hin zu den Wellenformen der Signale. Hiervon müssen die meisten Parameter angepasst werden, sollten das Display an eine andere Stelle

transportiert werden. So beschreiben dies Bach-y-Rita et al. als sie ein elektrotaktil Display, das an den Finger angepasst war, mit der Zunge testen wollten[28]. Insgesamt lassen sich so wenige generell gültige Aussagen treffen. Dies hängt weniger mit der bereits investierten Forschung zusammen, die Anzahl an Studien zu dem Thema ist durchaus beträchtlich, sondern eher mit der Größe des Bereichs und der großen Anzahl an Möglichkeiten wie ein taktil Display funktionieren soll, wo es angebracht ist etc. Ebenfalls kann man nicht sagen, dass Teilbereiche der taktilen Displays besonders gut erforscht sind. Studien beleuchten oft ganz eigene Aspekte, sodass es oft schwer fällt, diese zu verknüpfen.

Dies bedeutet jedoch nicht, dass in der Vergangenheit nicht sehr gute taktile Displays entstanden sind. Jedoch wirkt es so, dass eher nach dem „Trial and Error“-Prinzip vorgegangen wurde, da noch zu wenig bekannt war. Erst in jüngerer Zeit wurde es möglich Richtlinien für deren Entwicklung und Testen zu veröffentlichen[34][32].

### 6.3 Signalverarbeitung

Ob eine Signalverarbeitung notwendig ist, hängt stark von den benötigten Daten ab. Je komplexer diese sind, desto eher ist sie notwendig. In den letzten Jahren hat sich die Signalverarbeitung durch gestiegene Rechenleistung stark verbessert. So konnte durch die gute Parallelisierbarkeit[20] in der Bildverarbeitung Geschwindigkeit und Qualität der Informationen verbessert werden. Auch in anderen Anwendungsbereichen stellt die Rechenleistung ein zunehmend kleineres Problem dar.

Des Weiteren ist durch die Entwicklung der Smartphones ein Trend zu immer kleineren, energieeffizienteren Platinen, Prozessoren und Computern erkennbar. Auch in der sensorischen Augmentation werden so die Hürden für einen mobilen Einsatz bei Verwendung von Signalverarbeitung gesenkt.

Vor allem in der Bildverarbeitung werden die Algorithmen stetig verbessert. So investieren auch große Unternehmen wie Google und Microsoft[7] stark in die Erkennung von Gesichtern und Gegenständen, was auch in der sensorischen Augmentation von Nutzen sein kann. So kann man sich vorstellen, dass eine sehbehinderte Person erkennen kann, ob Personen in einem Raum, in den sie eintritt, anwesend sind, oder ob sich bestimmte Gegenstände darin befinden. Dieses Gebiet ist jedoch recht jung und kam daher noch nicht in der sensorischen Augmentation vor.

### 6.4 Integration

Auch wenn es funktionstüchtige Geräte gibt, die eine sensorische Augmentation ermöglichen, so sind diese von einem Einsatz in der Praxis immer noch weit entfernt. In vielerlei Hinsicht sind diese unpraktisch. Sei es bei den taktilen Displays, dass diese so angebracht werden müssen, dass man nicht in seiner Bewegungsfreiheit eingeschränkt ist und diese als nicht störend empfunden werden, als auch bei den Geräten der Signalverarbeitung. Diese dürfen nicht zu groß sein und sollten bestenfalls entweder bei den taktilen Displays oder bei den Sensoren integriert

werden. Auch Smartphones könnten diese Aufgabe übernehmen, was aber zu Lasten deren sowieso schon schwachen Akkus geht. Der Vorteil wäre, dass kein zusätzlicher Prozessor zur Datenverarbeitung nötig ist, womit man die Geräte billiger machen könnte. Auch in der Kommunikation zwischen den Komponenten gibt es noch offene Fragen. Kabel können als lästig empfunden werden, drahtlose Kommunikation bspw. via Bluetooth benötigt jedoch mehr Energie und damit einen größeren Akku, was die Systeme schwerer macht. Eine Lösung könnten vollintegrierte Geräte darstellen, die Sensoren, Signalverarbeitung und taktiles Display in sich vereinen. Bei dieser Methode scheidet jedoch das Smartphone als zentraler Punkt aus. Außerdem sind vollintegrierte Geräte nicht für alle Sensoren sinnvoll. So ist bei einer Kamera oder einem Sonar wichtig wo diese angebracht sind, was aber u.U. an Stellen ist, an denen ein taktiles Display nicht sinnvoll erscheint.

Zwischen all diesen Aspekten muss abgewogen und im spezifischen Fall entschieden werden, was das jeweils Sinnvollste ist. Diese Fragen müssen jedoch beantwortet werden. Bereits in den 80-er Jahren gab es mit dem „Tickle Talker“ und dem „Tacticon“ kommerzielle Systeme der sensorischen Substitution, die sich jedoch nicht durchsetzen konnten, was auch mit der damals jungen Technik zusammen hing. Heute sollte das besser möglich sein. Ein einfaches Gebiet ist es jedoch trotzdem nicht, sodass wahrscheinlich die Lösungen erfolgreich sein werden, die den besten Kompromiss aus all diesen Aspekten finden und darüber hinaus eine gute Sensorik, Signalverarbeitung und ein gutes taktiles Display bieten. Des Weiteren stellt sich die Frage, ob sich Sensoren und taktile Displays überhaupt so entwickeln lassen, dass der Nutzen, auch vermeintlich unwichtiger Sinne, größer ist, als die mit dem Tragen des Geräts einhergehenden Einschränkungen. So könnte es sein, dass sich der Sinn für WLAN-Signalstärke deshalb nicht durchsetzen kann.

Lediglich bei erforderlichen Sensoren ist wenig oder keine weitere Forschung nötig. In den letzten Jahren wurden für Smartphones eine Vielzahl an Sensoren entwickelt und verbaut, sodass diese sowohl sehr klein, als auch kostengünstig und energiesparend geworden sind.

## 7 Fazit

Die sensorische Augmentation bietet viele Möglichkeiten für alle Menschen unserer Gesellschaft. Unsere Navigationsfähigkeiten können verbessert werden und wir können lernen Dinge wahrzunehmen, die wir vorher nicht konnten. Allein das schon macht die sensorische Augmentation zu einem außergewöhnlichen, ein wenig futuristischen, Konzept.

Für viele spezielle Einsatzzwecke, wie Feuerwehrmänner, Arbeitern in radioaktiv verseuchtem Gelände, Astronauten etc., kann die sensorische Augmentation auch heute schon Vieles erleichtern. Ansonsten sind jedoch viele Probleme vorhanden: Bei der Entwicklung von taktilen Displays treten immer wieder Schwierigkeiten auf. So sind sie aktuell noch nicht für komplexe Informationen, wie z.B. Bildinformation, geeignet. Hierbei stellt sich die Frage stellt, ob der Tast-

sinn überhaupt dafür geeignet und das Mittel der Wahl ist diese Informationen zu empfangen. Auch nach Jahrzehnten der Forschung auf diesem Gebiet wird immer noch sehr viel experimentiert und es werden neue Technologien entwickelt, ohne dass ein Durchbruch wirklich gelungen ist. Einfache und eindimensionale Informationen scheinen dagegen prädestiniert für taktile Schnittstellen. Deren Verarbeitung läuft durch die Geschwindigkeit, mit der die Rezeptoren der Haut Reize erkennen können, schneller, als bspw. beim Auge. Das ist eine große Stärke taktile Schnittstellen. Die Forschung sollte sich darauf konzentrieren eher einfache Informationen, wie Warnungen, Richtungsinformationen o.Ä. an das Gehirn zu senden, und die vollständige Substitution des Sehsinns oder des Gehörs nicht mehr weiter zu erforschen, wie dies zu großen Teilen auch geschehen ist.

Nicht zuletzt muss die Integration der Geräte angegangen werden. Es erscheint mühsam diese Geräte die ganze Zeit mit sich zu führen und auf der Haut zu haben. Nur, wenn Nutzer keine Beeinträchtigung im alltäglichen Leben spüren, wenn sie die Geräte am Besten gar nicht wahrnehmen, wird die sensorische Augmentation abseits spezieller Einsatzzwecke und im täglichen Leben nützlich und für die Masse kompatibel. Vorerst ist ein solcher praktischer Nutzen, der deutlich größer ist, als die Einschränkungen, die mit solchen Systeme einhergehen, jedoch nicht vorhanden.

## Literatur

1. Bach-y-Rita, P., Webster, J., Tompkins, W., Crabb, T.: Sensory substitution for space gloves and for space robots (1987)
2. Bach-y-Rita, P.: Tactile vision substitution: Past and future. *International Journal of Neuroscience* (1983)
3. Bach-y-Rita, P.: Brain plasticity, pp. 113–118. Mosby (1988)
4. Bach-y-Rita, P., Collins, C.C., Saunders, F.A., White, B., Scadden, L.: Vision substitution by tactile image projection. *Nature* 221 pp. 963–964 (1968)
5. Bach-y-Rita, P., Kerchel, S.W.: Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences* 7 (2012)
6. Bertram, C., Evans, M.H., Javaid, M., Stafford, T., Prescott, T.: Sensory augmentation with distal touch: the tactile helmet project. In: *Conference on Biomimetic and Biohybrid Systems*. pp. 24–35. Springer (2013)
7. Chansanchai, A.: System trains machines to look at images the way people do – and answer questions about them; letzter Zugriff: 31.03.2017 (2015), <http://blogs.microsoft.com/next/2015/11/24/system-trains-machines-to-look-at-images-the-way-people-do-and-answer-2Dquestions-about-them/#sm.001xjpc57fzdhh114t13i1d351kp>
8. Chouvardas, V.G., Miliou, A.N., Hatalis, M.K.: Tactile displays: Overview and recent advances. *Displays* 29(3), 185–194 (2008)
9. Collins, C.C.: Tactile television-mechanical and electrical image projection. *IEEE Transactions on man-machine systems* 11(1), 65–71 (1970)
10. Cowan, R., Galvin, K., Sarant, J., Millard, R., Blamey, P., Clark, G.M.: Improved electrotactile speech processor: Tickle talker. *Scientific publications*, vol. 8, 1994–1995, no. 724 (1995)
11. Dehaene, S.: The neural basis of the weber–fechner law: a logarithmic mental number line. *Trends in cognitive sciences* 7(4), 145–147 (2003)

12. Gavrilov, L., Tsurulnikov, E., Davies, I.I.: Application of focused ultrasound for the stimulation of neural structures. *Ultrasound in Medicine and Biology* 22(2), 179 – 192 (1996), <http://www.sciencedirect.com/science/article/pii/S0301562996837823>
13. Hecht, S.: The visual discrimination of intensity and the weber-fechner law. *The Journal of general physiology* 7(2), 235–267 (1924)
14. Johnson, L.A., Higgins, C.M.: A navigation aid for the blind using tactile-visual sensory substitution. In: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. pp. 6289–6292. IEEE (2006)
15. Jones, L.A., Sarter, N.B.: Tactile displays: Guidance for their design and application. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50(1), 90–111 (2008)
16. Kaczmarek, K., Bach-Y-Rita, P., Tompkins, W.J., Webster, J.G.: A tactile vision-substitution system for the blind: computer-controlled partial image sequencing. *IEEE transactions on biomedical engineering* (8), 602–608 (1985)
17. Kaczmarek, K.A., Webster, J.G., Bach-y Rita, P., Tompkins, W.J.: Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38(1), 1–16 (1991)
18. Kärcher, S.M., Fenzlaff, S., Hartmann, D., Nagel, S.K., König, P.: Sensory augmentation for the blind. *Frontiers in human neuroscience* 6, 37 (2012)
19. Kuc, R.: Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification. *IEEE transactions on biomedical engineering* 49(10), 1173–1180 (2002)
20. Ladkat, A.S., Date, A.A., Inamdar, S.S.: Development and comparison of serial and parallel image processing algorithms. In: *Inventive Computation Technologies (ICICT), International Conference on*. vol. 2, pp. 1–4. IEEE (2016)
21. LaMotte, R.H.: Intensive and temporal determinants of thermal pain. In: *Sensory Functions of the Skin of Humans*, pp. 327–361. Springer (1979)
22. Lee, C.H., Jang, M.G.: Virtual surface characteristics of a tactile display using magneto-rheological fluids. *Sensors* 11(3), 2845–2856 (2011), <http://www.mdpi.com/1424-8220/11/3/2845>
23. Mártony, J.: Some experiments with electronic speechreading aids. *STL-QPSR* 2, 34–56 (1974)
24. Miller, J.D., Engebretson, A.M., DeFilippo, C.L.: Preliminary research with a three-channel vibrotactile speech-reception aid for the deaf. *The Journal of the Acoustical Society of America* 55(S1), S64–S64 (1974)
25. Moreau-Gaudry, A., Prince, A., Demongeot, J., Payan, Y.: A new health strategy to prevent pressure ulcer formation in paraplegics using computer and sensory substitution via the tongue. *arXiv preprint physics/0610181* (2006)
26. Olson, W.H.: Electrical safety. *Medical Instrumentation Application and Design* 4, 713 (1978)
27. Quek, Z.F., Schorr, S.B., Nisky, I., Okamura, A.M., Provancher, W.R.: Sensory augmentation of stiffness using fingerpad skin stretch. In: *World Haptics Conference (WHC), 2013*. pp. 467–472. IEEE (2013)
28. Bach-y Rita, P., Kaczmarek, K.A., Tyler, M.E., Garcia-Lara, J.: Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *Journal of rehabilitation research and development* 35(4), 427 (1998)
29. Bach-y Rita, P., Tyler, M.E., Kaczmarek, K.A.: Seeing with the brain. *International journal of human-computer interaction* 15(2), 285–295 (2003)

30. Saunders, F.: Tacticon 1600 electrocutaneous vocoder. Concord, MA: Tacticon Corporation (1985)
31. Schorr, S.B., Quek, Z.F., Romano, R.Y., Nisky, I., Provancher, W.R., Okamura, A.M.: Sensory substitution via cutaneous skin stretch feedback. In: Robotics and Automation (ICRA), 2013 IEEE International Conference on. pp. 2341–2346. IEEE (2013)
32. Stafford, T., Javaid, M., Mitchinson, B., Galloway, A., Prescott, T.: Integrating augmented senses into active perception: a framework. In: Royal Society meeting on Active Touch Sensing, Kavlie Royal Society International Centre (2011)
33. Tyler, M., Danilov, Y., Bach-y Rita, P.: Closing an open-loop control system: vestibular substitution through the tongue. *Journal of integrative neuroscience* 2(02), 159–164 (2003)
34. Van Erp, J.B.: Guidelines for the use of vibro-tactile displays in human computer interaction. In: Proceedings of eurohaptics. vol. 2002, pp. 18–22 (2002)
35. Van Erp, J.B., Van Veen, H.A., Jansen, C., Dobbins, T.: Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP)* 2(2), 106–117 (2005)
36. Zelek, J., Holbein, M.: Wearable tactile navigation system (May 22 2008), <https://www.google.com/patents/US20080120029>, uS Patent App. 11/707,031

# Text to Skin – Using Your Skin Senses to Read

Maximilian Stemmer-Grabow\*

Advisor: Erik Pescara†

Karlsruhe Institute of Technology (KIT)  
Pervasive Computing Systems – TECO

\*ubduq@student.kit.edu

†pescara@teco.edu

**Abstract.** This survey aims to provide an overview over the current state of skin reading – that is, using the skin’s senses to convey information such as text or data – with a particular focus on integrating these techniques into the context of mobile and wearable computing. Following a brief tour into the roots of reading with one’s skin, the general physiological quality of the sense of touch and the resulting technical considerations and parameters that define successful skin reading devices are discussed. Applying these underlying properties of haptic perception, multiple projects aiming to implement this in working skin reading devices are presented. These include systems expressly designed to be based on the Braille system as well as new ones featuring specifically designed encoding schemes. The paper discusses their broader application and tries to evaluate the idea that these techniques might once transcend providing assistance for the visually impaired to also enhance mobile and wearable devices in a more generally accessible way.

**Keywords:** skin reading, vibrotactile, haptic, haptic perception, tactile interfaces, assistive technologies, augmented Braille

## 1 Introduction

Visual displays still dominate the way information is presented on computer systems, even with the advent of mobile devices and their natural restrictions relating to their screen or visual display hardware. Simple notifications that work via touch are very common though – usually transmitted with vibrotactile motors – as they provide a rather discrete but attention-grabbing way to notify the user. With their simplicity, they typically are not designed as a real alternative to the visual representation on the devices though [11]. Recent advancements may allow for meaningful expansion of this model.

On the other hand, the idea of non-visual substitution of written text is even an essential one for visually impaired people. Systems such as Braille have existed for a long time, and more recent computer systems complement this with auditory feedback. Still, these systems have obvious drawbacks for non-visual communication, especially in the context of increasingly mobile or wearable devices.

Starting with a short introduction into the prior condition of this field of research, this paper explores general constraints for usable skin reading systems that are rooted in the physiological boundaries of the skin's perception and presents recent suggestions of usable designs for such systems.

This paper primarily focuses on transmission of text or similarly structured information through the skin, the haptic senses acting as a primary means of transferring information. Other systems using similar haptic methods to enhance perception in other ways or transmit different kinds of information are also actively researched but not covered at length here. This also applies to a more general look at the large field of sensory substitution.

## 2 Braille and the Roots of Haptic Reading

Still heavily in use today, the Braille writing system was arguably the most important step in written communication for blind people as of yet [8]. In this context, it seems useful to very briefly outline its inception and the impact it has had.

Braille originated in the first half of the 19th century, developed by Louis Braille. Blind himself after an accident in his youth, he developed the system now bearing his name. It is outlined in a revised version in his 1837 edition of the publication "Procédé pour écrire les paroles, la musique et la plain-chant au moyen de points, à l'usage des aveugles et disposés par eux" (after originally being published as a first version in 1829). It uses a matrix of 2 (horizontal) by 3 (vertical) possible locations for embossed (raised) dots, with different combinations signifying letters, numbers and punctuation.

Braille is based on another system devised by Charles Barbier (also called the Night Writing system) [8]. Barbier's alphabet was also based on a dot matrix with 6 instead of 3 rows, but encoded French using phonemes instead of letters. It was originally designed for military applications, as a means of transmitting information in situations that do not present conditions suitable to read in. Subsequently, it was picked up on for the use with blind individuals. The large height of single characters impedes the ability to feel them as a whole with one finger, explaining why Braille opted for the use of three rows instead [8]. The original Braille system designed for use with French was later generalized for other languages, starting in other European countries [8].

Both Braille and the Night Reading system are based on patterns formed by simple dots. Another possible design for a haptic language includes more intricate shapes, for example embossing the original letter shapes intended for visual perception into paper. This kind of approach was sometimes used before the popularization of Braille, albeit not to a great extent [8]. A variant of this idea is the basis for the Moon Alphabet, a system also developed in the 19th century. It consists of embossed letters that represent the Latin alphabet with simplified shapes to allow for better discrimination by touch. It is still being actively used, but has never reached the prevalence of Braille (especially for the

congenital or early blind who have never learned visual reading and thus do not profit as much from the resemblance to visual lettering).

With computer systems and dynamic displays arises the question of representing letters in a more dynamic way than Braille letters embossed in a piece of paper. Dynamic, electronically controlled Braille displays, typically representing one line of a varying number of Braille symbols, are commercially available. Still, these systems typically suffer drawbacks such as not easily being adapted to a mobile setting. The idea of evolving Braille code to address this will be further discussed in the course of this paper, specifically in section 5.1.

More recently, the availability of text-to-speech systems now allows for another channel of outputting textual information without the need for a visual display, at least for those individuals without an impaired sense of hearing. This is widely associated with a decline in Braille literacy rate as this may be perceived as a more efficient way of transmitting text when compared to Braille reading, which typically only allows for significantly lower throughput than visual reading [10].

Today, many devices have the capability of relaying simple haptic information, usually by means of a single source of vibration installed in the chassis. Still, these are used mainly for very simple messages such as the notification that an event occurred [11].

In this context, it does not only seem interesting to use the new technologies to improve tactile reading for blind individuals dependent on these kinds of techniques, but to make it accessible for a wider user base. This shall be explored in the rest of this survey.

### 3 General Physiological Considerations

Many considerations that influence the design of skin reading devices are based on more general physiological properties of the skin that can serve as a guideline for the design of effective reading systems [11]. This also requires a basic understanding of the physiological structure of perception in the skin.

#### 3.1 Introduction to the Perception of Touch

The following shall present a very brief overview over the physiological conditions of the haptic senses. It is based on the far more thorough review of its characteristics and their impact on the sense of touch by Lederman and Klatzky [9]. The factors that contribute to the concrete design and system parameters of skin reading devices are discussed in section 4 in more detail.

The perception of feel on the skin can be broadly grouped into cutaneous and kinaesthetic perception [2], [9]. While the former originates in mechanoreceptors and thermoreceptors in various layers of the skin itself, the latter characterizes perceptions of body part positions and compression sensed by muscles, tendons and joints, hence being responsible for the perception of rough object shapes, for example.

The way in which these two types of perception are balanced in importance for the perception of objects or shapes is also dependent upon the way the observer is allowed to act [9]:

- A *passive observer* commonly has less information originating from haptic perception. In this case, cutaneous sensation is especially important as it is less dependent on movement and the reaction of objects under observation when touched. This kind of passive observation model is e.g. relevant for wearable haptic feedback devices that are fixed to certain parts of the skin.
- For an *active observer*, the kinaesthetic side of haptic perception becomes more relevant as the relevant systems here are especially sensitive during movement. Still, cutaneous perception remains important for some properties of the observed object or surface. Braille can be seen as an example here as it requires spatial exploration of a series of Braille symbols.

Across the human body, different areas are covered in different types of skin differing in structure details and receptor density. The skin on the inside of the hand is not actively researched in terms of its structure and haptic properties. It is not covered with hair and hence also called “glabrous” [2]. In this case, four types of mechanoreceptors can be identified, differing in their adaptation rate as well as the size of the area they respond to stimuli in [9].

The combination of cutaneous and kinaesthetic sense is also responsible for multiple different parameters of the perception of observed objects: these include geometric properties, material properties such as texture and roughness (highly dependent on the shape of and gaps between surface elements), the thermal quality of objects, compliance (deformability under pressure), weight and orientation.

### 3.2 Notable Haptic “Illusions” and Deficiencies

When designing devices that utilize haptic perception, it is important to be mindful of the various oddities that this brings with it. These may be both hints to design restrictions as well as actively usable in the context of the device design. Examples of such effects include [9]:

- *Phi Movement* (and the related effect of *Beta Movement*) denote the effect of apparent movement when presented with multiple stimuli in quick succession that are spatially separated. Originally observed in visual perception, this effect to an extent also transfers to haptic perception. This forms the basis of the effect that encodings based on spatiotemporal impulses (as discussed later in section 4) may be experienced as directed motion even on low-resolution haptic displays [14].
- The *Tau Illusion* describes another effect that highlights the interplay of spatial and temporal perception of haptic stimuli. It can be observed that three taps to the skin are perceived to have different distances from one another even if only the time delay between the taps is varied, with them having the same spatial distance from one another.

- *Cutaneous saltation*, also known as the *Cutaneous Rabbit Illusion* [4], describes an apparent perception that can be achieved by rapidly tapping a skin region multiple times, then tapping another skin region. The perception this evokes is one of multiple taps, each moving closer to the second position (likened to a “hopping rabbit”) [4].  
This effect may potentially be usable to increase the apparent resolution of a tactile display for moving impulses. This idea was explored in 1997 by Han and Pentland [17], using an array of nine vibromotors placed on the back of a chair. Exploiting the saltation phenomenon, this was used to simulate a higher-resolution display.
- *Masking* of stimuli occurs when multiple stimuli are present at once. It has been shown that even with as few as three different locations for haptic stimuli, their detection accuracy is severely limited [14], [11]. This also explains the improvement that can be achieved in recognition accuracy by delaying additional stimuli (an idea further discussed in section 4.3), which has been shown to be effective for multiple actuator configurations [14], [18].

## 4 Consequences for Device Design

The following general considerations for haptic display devices are based on the properties outlined in the previous section.

### 4.1 Sensory Input Method

As mentioned before, the cutaneous sense does not only include perception of vibration and indentation of the skin, but also temperature and pain [3]. Early on, many theoretically possible input methods have been identified as ineffective as a means of quickly transmitting information due to impracticality or unsuitability for prolonged usage (e.g. heat or specifically administered pain sensations). The main approaches that remain can be grouped into the following categories [3]:

**Dynamically Refreshed Static Haptic Displays** A method fairly self-evident when starting out from the concept of Braille or similar languages is the concept of representing a haptic surface with the ability to dynamically change its contents. These displays can optionally also be operated by manually exploring the currently displayed contents.

Without the possibility to actively explore its contents, e.g. by moving single fingers across the display, this kind of display with prominent features is rather unsuitable for haptic perception as the mechanoreceptors in the skin rather quickly adapt to static stimuli, fading the display contents, possibly beyond recognition.

Refreshable Braille displays fall into this category, utilizing movable pins to represent arbitrary Braille characters. These devices have been commercially available for a long time. Common criticisms include their price and insufficient ability to be used for mobile devices due to their size and weight.

Another possible method is the lateral stretching of the skin, which has been examined as a means of representing perception arrays, for example by Hayward and Cruz-Hernandez [6].

**Electrotactile Stimulation** The application of direct electric pulses to the skin, either direct currents or alternating currents with varying frequencies, also produces perceptions of differing quality ranging from tingling to stinging or severe pain, dependent on the type, frequency and amplitude of the impulses [3].

Although technically very simple, electrotactile stimulation seems to be less common for the purposes examined in this paper. This is probably owed to the fact that parameters have to be very tightly controlled for this kind of stimulation as to avoid uncomfortable sensations (which seem to have a wider variance for different individuals than mechanical stimuli) or even damage to the skin under prolonged stimulation [3].

**Vibrotactile Stimulation** This type of stimulation seems to be by far the most common in recent research. The skin is both fairly sensitive to vibrational stimuli and they can be easily and controllably reproduced. Vibrotactile actuators are often constructed with rotating or linear motors accelerating a well-known mass. Vibration can be administered with a wide range of frequencies and amplitudes.

## 4.2 Body Site and Tactile Resolution

The tactile resolution, i.e., the distance between two stimuli that can reliably be differentiated is highly dependent on the position of the stimulation. As previously implied, this is due to the different composition of skin on different parts of the body.

The measured resolution is also dependent on the specific way it is defined and measured. The most well-known means to express and measure spatial acuity are the following (as described according to [9]):

- The *Two-Point Touch Threshold* is maybe the most intuitive definition, marking the minimum distance between two stimulus positions to be distinguished from a single stimulus with sufficient reliability (which has to be defined as a test parameter).
- To measure a value for *Point-Localization*, multiple impulses are given with a time delay between them, having the observer decide whether the stimuli were perceived in the same or in different locations. This method typically yields lower values for the distance (that is, a higher spatial resolution), but seems to be consistent with the results from the two-point touch threshold measurements when compared across body sites [9].

The spatial resolution is highest at the fingers (with a two-point touch threshold of about 5 mm, [11]) and lowest on the back (about 6cm, [14]), with values in between for the palm (circa 1 cm) and forearm (circa 4 cm) [11].

### 4.3 Actuator Layout and Encoding Type

According to the number and layout of the actuators used to generate the haptic impulse, different strategies for combining these actuators are possible. In a preliminary experiment for their attempt to estimate the throughput of a particular skin region, in this case on the lower back, Novich and Eagleman compared the basic types of encoding described below with vibrotactile actuators [14]. Their design was comprised of a vest to be worn by the participants, with a matrix of three by three vibromotors affixed to it (shown in figure 1). The symbols to be differentiated were chosen as similar sets in all cases and can be seen in figure 2.



**Fig. 1.** Vibrotactile vest used by Novich and Eagleman [14] to compare encoding types and analyze tactual resolution on the lower back. The vest features an array consisting of 3x3 vibromotors. It also includes wireless controlling hardware and a power supply.

**Single Actuator** With only a single-point stimulation, the possibilities for encoding different symbols are somewhat limited. In this case, amplitude, duration and frequency of the impulse remain with typical vibrotactile setups. Still, it seems inadvisable to use both amplitude and frequency independently for encoding tactual messages, as both are coupled in a way that makes judging them separately difficult: It can be observed that a decreased amplitude may additionally be perceived as an increase in frequency and vice versa [3]. In the case of Novich and Eagleman, the two parameters were hence linked to one another. The patterns that should be differentiated were eight different intensity levels of a single vibromotor with associated frequency (70 to 240Hz). The results for the different encoding types and otherwise identical parameters is shown in figure 2. The figure shows confusion matrices for the different encoding types and the patterns to be identified. The problem that is endemic with the single motor encoding is clearly visible in the matrix: neighboring values tend to be confused with one another, indicating the difficulty to judge amplitude/frequency without an explicit reference that can be used for comparison. For this reason, in his early experiments, Geldard judged that it might be advisable to not use more than three different amplitude levels [3].

**Spatial** Using more than one actuator naturally allows for more complex types of encoding. It also requires a concept for the layout of the actuators. Possible solutions include one- or two-dimensional arrays of actuators, similar to the pixel grid in a visual display. In the case of Novich and Eagleman, a square grid with 3 rows and 3 columns was used. For the comparison of spatial encoding with the other types of encoding, patterns with three involved actuators were chosen, all three triggered at the same time and for the same duration.

Layouts are not limited to this kind of rectangular shape: Linear formations can be closed, e.g. around the head or torso, producing a special case of the one-dimensional array described above. Further, actuators can be placed in arbitrary locations on one or more body parts deemed far enough from each other to allow for easy discrimination between the sites. This approach is further explored with exemplary skin reading systems in section 5.

**Spatiotemporal** Combining spatial and temporal encoding allows for even more potential ways to represent information, for example with the simulation of “sweeping” movement across a haptic display. In this case, multiple actuators are triggered in sequence.

This type of encoding presents the best performance by a fair margin in the experiment conducted by Novich and Eagleman [14].

For their haptic display, Luzhnica et. al. [11] use a variant of this encoding they call *overlapping* spatiotemporal stimulation (OST), which differs from the system Novich and Eagleman use in that the delays between consecutive actuator impulses are shorter than the duration of each impulse, hence creating an overlap between the different positions involved in the transmission of a symbol. They suggest that this way, the total duration of the transmission can be shortened without compromising the ability to differentiate between the patterns [11].

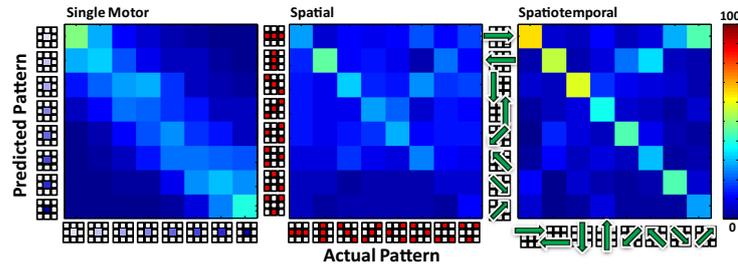
An interesting point is possible interplay between this kind of encoding and the spatial tactile resolution discussed in the previous section. As indicated in testing done by Novich and Eagleman [14], spatiotemporally encoded stimuli may not be subject to the limitations of spatial tactile resolution to the same extent as spatially encoded ones.

#### 4.4 Information Encoding

After looking at the technical constraints a haptic reading device has to follow to obey the physiological principles of human haptic perception, we can take a look at the structure of information that is useful to be transmitted in this manner.

The way information can be represented in haptic interfaces is highly dependent upon the structure of the information to be transmitted. For text, letter (e.g. in languages with a Latin alphabet [11]) as well as phoneme transmission is promising [14].

An important consideration is the dependency of the encoding on the aforementioned parameters. For example, multiple concurrent stimuli in different



**Fig. 2.** Confusion matrices for the three different encoding types tested by Novich and Eagleman [14]: Stimulation with a single vibromotor differentiated by intensity, spatial patterns consisting of concurrent stimulation with three motors for each symbol, and spatiotemporal encoding, combining spatial encoding with delayed stimulation for the different positions. These matrices show the overall results averaged over other parameters. Correct pattern identification corresponds to the diagonal of the matrices.

places mask each other and therefore reduce the accuracy of correctly identified stimuli [11].

This can be used to maximize accuracy by incorporating the expected distribution of symbols in the used encoding (for example, optimizing letter encodings for specific languages), but hence sacrificing performance and possibly accuracy when using the encoding with contents not following the expected distribution the encoding is based on.

As described above, for multipoint tactile feedback, possible encoding modes include simple spatial separation (by using multiple actuators attached to the skin) as well as temporal separation (different stimulus lengths). These modes can of course be combined into spatiotemporal encoding, utilizing both different areas as well as a sequence or delays between impulses [14].

Novich and Eagleman mention that on the torso, the “raw”, roughly estimated theoretical throughput when using their kind of spatiotemporal encoding may reach around 600-925 bit/s [14] when extrapolating from their highest-performing test participant. It should be noted however that this estimate is based on linearly extrapolating from their configuration to vibrotactile arrays covering large portions of the torso, not accounting for possible mutual interactions between them.

*Tactons* In 2004, Brewster and Brown proposed the concept of “structured, abstract messages that can be used to communicate messages non-visually” which they refer to as *Tactons* (short for tactile icons) [2]. As the name implies, they aim to provide a concept for haptic messages similar to visual icons (or the auditory counterpart presented by Brewster earlier labeled “earcons”). Brewster also outlines a model for designing tactons by integrating multiple tactons into one as well as transforming existing and known tactons to express associated concepts.

The meaning when utilizing the word *tacton* has since been somehow expanded by some authors, referring more generally to atomic tactile symbols as a part of an arbitrary encoding (e.g. by Luzhnica et al. [11], Novich and Eagleman [14]). In this case, they can rather be likened to a single visual character than a visual icon representing a more abstract and self-contained concept as it was originally described by Brewster and Brown [2].

#### 4.5 Other Considerations for Usable Designs

As detailed by Luzhnica et al. [11], a usable device does not only need to fulfill the requirements necessary for the transmission of the intended information in an acceptable timeframe (fundamentally influenced by the encoding capacity and the practical throughput of the chosen device design). Instead, additional factors include convenience of use and wearability (at least for devices designed as wearables). As such, a good wearable reading device should interfere as little as possible with common activities. As an example, this discourages a design occupying the whole hand, therefore complicating the handling of other objects.

This is especially important for devices designed for a broader target audience as the willingness to spend time and train with such a device may be reduced by a design that is perceived as unfavorable, even more so for people who are not regularly dependent on or familiar with similar devices and their drawbacks.

One specific design objective for devices that employ multipoint feedback to encode information is achieving a design that allows the device to be worn or oriented in a reproducible manner as to avoid actuators to be placed in different positions on the subject's body between sessions of usage, hence possibly impeding the performance when using the device [11].

## 5 Haptic Reading Devices

Equipped with the fundamentals that should theoretically govern the working of usable devices utilizing the haptic perception to transmit text, we can now take a look at projects trying to implement those and their development approaches.

The idea of devices allowing to encode visual information in a more general sense into information transmitted via touch as assistance for visually impaired people dates back to the middle of the last century, some notable examples including the following (although more similar devices exist): In the 1960s, Bachy-rita et al. [15] worked on an approach they called tactile-visual substitution system (TVSS for short), resulting in a device with a vibrotactor array (20 by 20 stimulation positions) mounted to a chair. It could be used to transmit images from a camera connected to the setup.

An approach that resulted in a commercially available device called *Optacon* was pursued from the 1970s to the 1990s [1], [5]. It consisted of a camera module connected to a unit with a 24 by 6 dot matrix of contact points that could each be vibrated by a piezoelectric actuator, thereby transcribing the image captured by the camera in a monochromatic way to the haptic display. This also allowed

for reading of ordinary visual lettering by moving the camera module across the text.

Even earlier, in 1957, Geldard [3] devised a system called *Vibratese* explicitly to transmit text (see below in section 5.2 for an overview).

Another class of devices employs haptic perception as part of their interface or to enhance the experience, but do not or not exclusively transmit information via touch. One recent example of this is the *FingerReader* device [16], published by Shilkrot et al. in 2014. It can be worn like a ring and features a camera pointing towards the finger tip. Coupling this with OCR (Optical Character Recognition) and text-to-speech software allows reading arbitrary text printed on paper by tracing the lines with one finger. Shilkrot et al. specify the extraction accuracy as 93.9% under optimal conditions. Two vibration motors on the top and bottom of the device are used to provide haptic feedback concerning the alignment of the device. This includes feedback when the finger tracing the line of text is moved too far off the line (indicating the direction the finger should be moved in to compensate), and notifications about reaching the end and beginning of lines when tracing along them. In a way, this shows a more conservative and traditional approach to using haptic feedback in the context of reading, providing more subtle cues but not using the skin for high-bandwidth information transmission, thereby avoiding the need of a more complex encoding to express symbols.

In the following, we will focus on systems both explicitly designed for reading text as well as only using haptic perception as the main way to transmit it. While there is a multitude of projects aimed at developing haptic displays or providing more general visual substitution solutions, the number of projects developing devices designed expressly for the purpose of transmitting text is somewhat more limited. The research projects described in this section yielded functional prototypes for skin reading systems with different characteristics. Note that they were in most cases explicitly declared as devices not ready for widespread usage, requiring more research to tune parameters [11], [13]. Extended studies involving more participants or studies with longer learning and application timeframes were also indicated to be required [11].

These approaches can be grouped in two categories: Devices that use an existing encoding system such a Braille and try to enhance it by leveraging haptic feedback [13], and systems employing both newly developed hardware as well as encoding schemes that are specifically designed for use with the skin reading device. The former allows for synergy effects for a clearly defined target audience (as with blind people already familiar with reading and writing Braille), thus reducing initial learning times and possibly “ported” proficiency with the new system [13]. The latter allows a tighter bond between the hardware and the encoding scheme with fewer design constraints for both. This may increase the performance of the system, but also usually requires more commitment on the part of the perspective user to learn the handling of the new system.

### 5.1 Augmented Haptic Braille Systems

Examples for recently developed reading systems that are based on Braille encoding will be presented in the following. In 2015, Nicolau et al. [13] proposed an approach labeled *HoliBraille*. It ports the Braille encoding to vibrotactile motors dampened with springs (to isolate the vibration sources from each other) and fitted into an enclosure that can be attached to a standard smartphone or similar device. The system allows for both input and output. When using the system, the fingers are placed on the screen with the HoliBraille device mounted to the backside of the smartphone. The six vibration motors are directly mapped to the dots in a Braille letter and similar in terms of their layout, vibrating concurrently to represent raised dots in a given symbol. For Nicolau et al., this prompts the hypothesis that it allows for fairly quick adoption by individuals already proficient in the use of Braille lettering [13], their small-scale experiment with experienced Braille users and allowing for fewer than ten minutes of prior training seemingly supporting this idea.

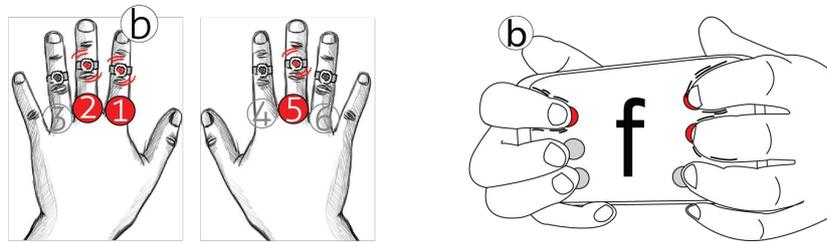
This design is based on a prior, wearable design from 2013 called *UbiBraille* with a similar objective developed by Nicolau et al. which utilizes the same encoding but does so with six vibrotactile rings that are worn on the index, middle and ring fingers of both hands [12]. As with HoliBraille, the rings directly correspond to the encoding of the six dot positions in Braille lettering. Schematic representations of both designs are shown in figure 3.

An advantage of this is the fact that a system like this can easily be mounted to a standard smartphone or similar device. A similar approach designed to not use any additional hardware is *V-Braille*, proposed by Jayant et al. [7]. It works with the vibration motor standard in smartphones, simply translating the Braille characters to the screen by splitting the screen area into six cells, vibrating the device when a cell that represents a raised dot is touched. With only one Braille character shown at any given time, the user is constrained to sense the dots sequentially in this case, limiting the throughput (the original small-scale study cites the time to read a single character as between 4.2 and 26.6 seconds [7]).

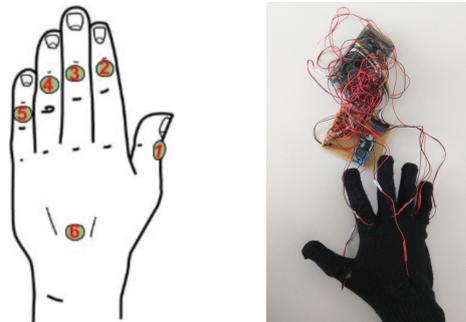
Although the original HoliBraille study only contains experimental testing of the recognition rates for single characters with a stimulus duration of 2 s, the authors indicate that the performance is similar to their previous tests with the UbiBraille configuration.

### 5.2 Devices with Specifically Designed Encodings

One of the first approaches to design a haptic reading device was proposed by Geldard in 1957 [3]. This system labeled *Vibratese* uses vibratory pulses of differing intensity and duration (three distinctive values in each case, with durations between 0.1 s and 0.5 s) in five different locations, totalling 45 different transmittable symbols. The proposed design never reached the stage of actual widespread usage, but has influenced the design of subsequent haptic reading devices.



**Fig. 3.** Basic layout of the UbiBraille [12] (left) and HoliBraille [13] (right) reading devices. The numbers signify the usual order the dots of a Braille symbol are referenced in. Both devices feature a structure closely resembling Braille text, directly translating dots to vibrotactile stimulus positions.

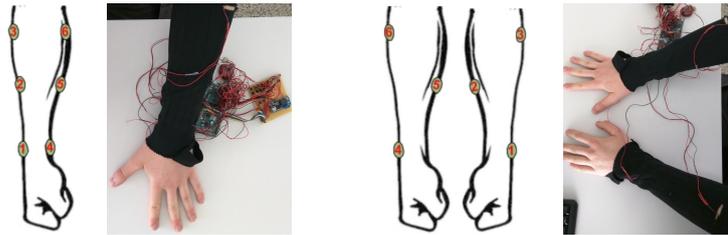


**Fig. 4.** Overview of the haptic display device proposed by Luzhnica et al. [11]. Positions of the six different vibromotors used for encoding tactions on the left, actual image of the device (with cabling and electronics used for controlling the device) on the right. These images show only the device preferred after testing other configurations of the six vibration motors.

In 2016, Luzhnica et al. [11] proposed a display that uses six vibrotactile contact points to transmit text in the English language (92% accuracy on the letter level, about 1.7 (four-letter) words per second after a training period of 5 hours) [11].

As previously mentioned, Novich and Eagleman [14] showed that when comparing impulses varied in intensity, temporally and spatiotemporally separated impulses, spatiotemporal encoding allows for the best differentiation performance [14]. Luzhnica et al. use the previously described modified version of this (overlapping spatiotemporal encoding). In this case, the impulses are started one after another, but only with a time-lag that is smaller than the impulse length, in this case using a delay of 10 ms.

In the final configuration, their device consists of a glove with six included vibrotactile actuators, five of which are located with contact to each finger, plus one on the back of the hand. Other configurations that placed the actuators either on one or two arms were discarded for performance and practicality reasons, respectively. A two-arm configuration seems to improve accuracy at the cost of more complexity in the form of a two-part reading device and increased effort of putting on the device. Both alternative configurations are shown in figure 5.



**Fig. 5.** Rejected configurations for the haptic display device discussed by Luzhnica et al. [11]. The left images show the six actuator positions for a single arm configuration, the images on the right show a configuration utilizing both arms. The single arm configuration was rejected due to poor performance compared to the configuration placed on one hand, the two-arm configuration (despite good performance) because of practical considerations, e.g. pertaining to putting the sleeves on and off.

As for the encoding used for the device, the basic symbols in this case are the letters of the Latin alphabet. As described in section 4, multiple concurrent impulses impair the ability of test persons to discern different tactons. This result is also confirmed in preliminary testing by Luzhnica et al. For this reason, the number of spatially separated actuators involved in each tacton is limited to three while distributing letters to have as few concurrent impulses as possible (to avoid inferior accuracy rates of 85% for three-position tactons [11]).

## 6 Discussion

As different as the approaches to encoding text in haptic information described in the previous section are, they share some lessons that emerge in their experimental testing. Both Nicolau et al. [13] as well as Luzhnica et al. [11] observe a considerable decline in accuracy with an increasing number of concurrent stimuli (even with the spatiotemporal method employed by the latter system). Luzhnica et al. report 85% overall accuracy for symbols whose encoding contains three different positions, compared to 96% for those with one or two positions. In comparison, HoliBraille achieved an accuracy of 55% with four or more concurrent stimuli and 89% with three or less. Also, this result prompts the question of whether the accuracy of HoliBraille could be significantly improved by utilizing a similar method to the one described by Luzhnica et al.

While the approaches provide experimental testing allowing for rough performance estimates, many lack the elaborate studies required to judge the performance for regular users of the systems. Still, some authors explicitly mention the possibility of far better performance than achieved in their studies [13].

Even though the experimental testing done with the different systems has very different conditions and is thus comparable only to a very limited extent, the following shall present a few pointers relating their performance.

When reading Braille texts, reading speeds vary highly with different readers. In testing performed by Legge et al. [10], median reading speed was around 125 words per minute (about 7.5 characters per second, with a span of 24 to 232 words per minute) with standardized reading tests. This compares to a median speed for visual reading of printed text of about 250 words per minute in the same study. Note that reading speeds vary widely when compared across different languages and texts with differing complexity.

In the testing done on the UbiBraille device, reading speeds of around 12 words per minute were possible with some participants, resulting from an individual character duration of 1 s [12]. HoliBraille performance was judged as similar to those results [13].

Luzhnica et al. recognition speeds of about 0.6 s per word and speculate about a sustained feasible transmission rate of about 105.5 words per minute [11].

An important distinction between the two approaches for the reading systems described above is the way they handle the case of learning the system. While approaches to design a system with an entirely new encoding also have to find a way to teach their proposed system (Luzhnica et al. even basing a lot of their experimental testing around their detailed learning procedure [11]), direct translations can make it a point to test the accuracy with minimal prior experience. Still, the availability of well-tested learning procedures rewarding new users with quickly visible progress seems critical for these systems to attract a sizeable user base.

An advantage often mentioned in conjunction with haptic perception is the ability of the human brain to process information coming from these pathways without the need to explicitly focus on them [11]. It is questionable how far this does apply to the task of reading through the means of tactual perception as the

focus required to parse (a more or less intricate) encoding while reading (both in printed text and Braille letter reading) seems to counteract this promise.

## 7 Conclusion

As discussed in this paper, with recent advancements in computing and human-computer interaction, it now seems entirely feasible to design practical systems that allow reading with your skin in a more dynamic, mobile way than Braille letters embossed on a piece of paper – as employed for the better part of two centuries now – can provide you with. For visually impaired individuals, some of the recent approaches allow them to take part in the new possibilities that recent mobile and wearable computing bring to the table with fewer restrictions, while more reasonably being able to combine them with standard devices as opposed to having to resort to more specialized devices.

In any case, the underlying insights these projects provide seem to also apply in a more general way to the design of any interface with a haptic component.

Still, especially for the kind of systems presented in this paper, their true potential only shows after extended usage by a wide variety of people, highlighting the need for research following up on them.

Even though some authors seem hopeful about a broader audience that can profit from the ability to read by using their sense of touch, it remains disputable whether a sufficient amount of users not directly dependent on such techniques are willing to take on the possibly challenging and time-consuming task that is learning to use them. On the other side, it may well turn out that useful haptic feedback may extend to more than only transmitting information in a very structured and limited way, even for users without the explicit need for sensory substitution systems.

## References

1. Bliss, J.C., Katcher, M.H., Rogers, C.H., Shepard, R.P.: Optical-to-tactile image conversion for the blind. *IEEE Transactions on Man-Machine Systems* 11(1), 58–65 (1970), <https://doi.org/10.1109/TMMS.1970.299963>
2. Brewster, S., Brown, L.M.: Tactons: Structured tactile messages for non-visual information display. In: *Proceedings of the Fifth Conference on Australasian User Interface - Volume 28*. pp. 15–23. AUIC '04, Australian Computer Society, Inc., Darlinghurst, Australia, Australia (2004), <http://dl.acm.org/citation.cfm?id=976310.976313>
3. Geldard, F.A.: Adventures in tactile literacy. *American Psychologist* 12(3), 115–124 (1957), <http://dx.doi.org/10.1037/h0040416>
4. Geldard, F.A., Sherrick, C.E.: The cutaneous "rabbit": A perceptual illusion. *Science* 178(4057), 178–179 (1972), <http://dx.doi.org/10.1126/science.178.4057.178>
5. Goldish, L.H., Taylor, H.E.: The optacon: A valuable device for blind persons. *New Outlook for the Blind* 68(2), 49–56 (1974)
6. Hayward, V., Cruz-Hernandez, M.: Tactile display device using distributed lateral skin stretch. In: *Proceedings of the haptic interfaces for virtual environment and teleoperator systems symposium*. vol. 69, pp. 1309–1314. ASME (2000)

7. Jayant, C., Acuario, C., Johnson, W., Hollier, J., Ladner, R.: V-braille: Haptic braille perception using a touch-screen and vibration on mobile phones. In: Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 295–296. ASSETS '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1878803.1878878>
8. Jiménez, J., Olea, J., Torres, J., Alonso, I., Harder, D., Fischer, K.: Biography of louis braille and invention of the braille alphabet. *Survey of Ophthalmology* 54(1), 142 – 149 (2009), <http://dx.doi.org/10.1016/j.survophthal.2008.10.006>
9. Lederman, S.J., Klatzky, R.L.: Haptic perception: A tutorial. *Attention, Perception, & Psychophysics* 71(7), 1439–1459 (2009), <http://dx.doi.org/10.3758/APP.71.7.1439>
10. Legge, G.E., Madison, C.M., Mansfield, J.S.: Measuring braille reading speed with the mnread test. *Visual Impairment Research* 1(3), 131–145 (1999), <http://dx.doi.org/10.1076/vimr.1.3.131.4438>
11. Luzhnica, G., Veas, E., Pammer, V.: Skin reading: Encoding text in a 6-channel haptic display. In: Proceedings of the 2016 ACM International Symposium on Wearable Computers. pp. 148–155. ISWC '16, ACM, New York, NY, USA (2016), <http://doi.acm.org/10.1145/2971763.2971769>
12. Nicolau, H., Guerreiro, J., Guerreiro, T., Carrigo, L.: Ubibraille: Designing and evaluating a vibrotactile braille-reading device. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 23:1–23:8. ASSETS '13, ACM, New York, NY, USA (2013), <http://doi.acm.org/10.1145/2513383.2513437>
13. Nicolau, H., Montague, K., Guerreiro, T., Rodrigues, A., Hanson, V.L.: Holibraille: Multipoint vibrotactile feedback on mobile devices. In: Proceedings of the 12th Web for All Conference. pp. 30:1–30:4. W4A '15, ACM, New York, NY, USA (2015), <http://doi.acm.org/10.1145/2745555.2746643>
14. Novich, S.D., Eagleman, D.M.: Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Experimental Brain Research* 233(10), 2777–2788 (2015), <http://dx.doi.org/10.1007/s00221-015-4346-1>
15. Bach-y Rita, P., Collins, C.C., Saunders, F.A., White, B., Scadden, L.: Vision substitution by tactile image projection. *Nature* 221(5184), 963–964 (1969), <http://dx.doi.org/10.1038/221963a0>
16. Shilkrot, R., Huber, J., Liu, C., Maes, P., Nanayakkara, S.C.: Fingerreader: A wearable device to support text reading on the go. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. pp. 2359–2364. CHI EA '14, ACM, New York, NY, USA (2014), <http://doi.acm.org/10.1145/2559206.2581220>
17. Tan, H.Z., Pentland, A.: Tactual displays for wearable computing. *Personal Technologies* 1(4), 225–230 (1997), <http://dx.doi.org/10.1007/BF01682025>
18. Yanagida, Y., Kakita, M., Lindeman, R.W., Kume, Y., Tetsutani, N.: Vibrotactile letter reading using a low-resolution tactor array. In: Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS'04. Proceedings. 12th International Symposium on. pp. 400–406. IEEE (2004), <http://dx.doi.org/10.1109/HAPTIC.2004.1287227>

# Electroactive Polymers (EAPs) for Wearable Computing and Exoskeletons

Nicolas Schreiber\*

Advisor: Prof. Dr. Michael Beigl†

Karlsruhe Institute of Technology (KIT)  
Pervasive Computing Systems – TECO

\*uzdry@student.kit.edu

†Michael.Beigl@kit.edu

**Abstract.** Electroactive polymers are a new player in the field of robotics. Their ability to contract or bend when a voltage is applied gave them the name artificial muscles. Now, can they also be used in wearable computing, especially connected to exoskeletons?

This paper compares and reviews different EAPs on their potential application in exoskeletons. At first it describes the different properties of EAPs as well as the necessary values for a viable solution. Then two basic approaches for building an exoskeleton are shown. After that the EAPs are categorised into ionic or dielectric EAPs and their general functionality are specified. Collected data on several IEAPs and DEAPs gets listed. Lastly this paper looks further into the most promising and interesting polymers.

The evaluated parameters include the generated force, the needed voltage, other physical properties, but also economic and practical application information.

**Keywords:** Artificial Muscles, Electroactive Polymers, EAPs, Ionic EAPs, Dielectric EAPs, IPMC, Actuators

## 1 Introduction

After an accident or because of an medical issue many people get problems with walking. Usually the muscles get too weak to carry the full weight of a person. This leads to sitting in a wheelchair. Which might be a valid solution for mobility but at the same time a cause for growing problems. By sitting all the time the legs muscles deteriorate. To counteract frequent sessions at a chiropractor or similar are necessary. This is not an optimal solution.

A way to improve this method would be by having the patient train all the time while still giving him or her the possibility to move independently. This could be achieved by exoskeletons. The exoskeleton could support the wearers and give them the ability to walk but would also demand to add some strength themselves. That way it is even be possible to decrease support over time until healed.

To be a viable solution there are certain requirements to be fulfilled. Current exoskeletons are mostly controlled by electronic motors, pneumatic or hydraulic pumps. Those are quite big, need high electric power and are loud. Additionally most of the time they are quite rigid and might hurt the wearer or at least be uncomfortable over time.

A new and promising way to solve those problems might be electroactive polymers, sometimes called artificial muscles because of their similar behaviour. These are substances that react when a electric voltage is applied for example by changing shape or colour. EAPs have a low weight, size and are not as rigid thus reducing risk of hurting. They are already used in classic robotics but not yet used as much on exoskeletons.

This document shows and compares several types of EAPs, certain products and parameters on how viable of a solution they are and how much of a hassle it is to use them in a product.

## 2 Parameters

Of course robots and exoskeletons share some parameters important for consideration. These include the general force of the artificial muscles and the needed power to move. But at the same time the most significant part of a exoskeleton, the human wearing it, creates several new requirements. Most importantly the security of the wearer. Electrocution or too strong movements are to be avoided of course. Additionally it should be relatively comfortable to wear. Which is not a safety issue but would discourage further usage.

The exact parameters are described below. They include a general idea of what the values express, how they influence the application, what unit is used as well as optimal values. These optima are only a rule of thumb to give a relation, exact values depend entirely on the environment and wearer.

### Created Pressure

To see if the polymer would actually be capable of moving and supporting human legs you need to look at the force an EAP can create. For comparing values they need to be relative to surface. Pascal equals to  $N/m^2$  thus it is used as Unit. Depending on the strength of the polymer Pascal is scaled to either megapascal or kilopascal. To distinguish between relative and absolute values the word "pressure" is used to describe relative force.

The created pressure of the human muscle would be up to **0.35MPa** [16]. Because the artificial muscles will have to carry a similar amount of weight this is used as desired value.

### Strain

The possible strain gives information about how much the EAP can change shape. This also influences how far a leg can be moved. Because the change of shape should be considered relative to the shape before activation usually % is used as unit. If the strain is not high enough it is possible to increase movement

with mechanical constructions which would decrease the resulting force because of the law of the lever.

The usual human muscle strain often depends on the persons lifestyle but was measured as 5% [8]. While the exactly necessary value is more difficult to calculate it still gives information about the needed momentum to move.

#### **Dielectric constant**

The higher this constant is, the lower the necessary voltages are to create actuation [13]. But at the same time a high dielectric constant reduces the possible strain.

It is difficult to give a certain needed value, this only gives general information for later applications.

#### **Typical voltages and currents**

A safety risk would be the voltages and currents needed for actuation. Direct skin contact with the actuators can not be avoided. If the voltages or currents are too high an additional isolation is necessary. This though will probably interfere with other properties.

While general perception and discomfort is already possible at 1-5mA [5] the Voltage needed for this current depends on the resistance of the skin which might change depending on several factors including wetness and damages. Resistance of dry skin is between 1000-100000 Ohms.

#### **Electric field**

In some EAPs it is not exactly the voltage that produced the actuation but the electric field that is induced by the voltage. Its usual unit is MV/m.

This also gives mostly general information for later use.

#### **Energy Harvesting**

Energy harvesting is a secondary usage of electroactive polymers. It describes the capability to produce electrical energy out of movement. Meaning that, if for example the exoskeleton does not move actively, the wearer would create energy by moving him- or herself. Depending on the current it might be possible to sense the movements or it could even load the battery.

#### **Reaction Speed**

Another interesting information is the reaction speed of the polymers. Not to make running possible but to have a immediate response to the wearers actions. If he needs to push only for a second until the artificial muscles react movement becomes much more sluggish and the user experience decreases greatly.

Exact values are difficult to produce and find so they are described as either slow, medium or fast. Slow represents reaction speeds in about a second and higher, while fast reaction speeds are closer to 1 millisecond and lower. Medium is to be put in between those two. The desired values are at least medium speed.

### **Flexibility**

While creating a certain force to move is necessary the polymer should still be flexible. This is also called "soft-robotics" because the actuators are not rigid but deformable. Thus it is much more comfortable to wear the exoskeleton but this also avoids hurting the person.

The unit used when measuring stiffness is called **Young's modulus**. It describes the connection between strain and applied external pressure. Another way to ensure that the polymers are still flexible even when activated is to not oversize the strength of the EAP.

### **Adaptability**

Similar to flexibility the actuator should be possible to adapt to the human shape. This includes shaping it before usage to fit the human body. By that the general effort to build the exoskeleton is decreased and it is possible to create actuators for many different shapes.

### **Preparation**

The preparations include all steps between buying the raw materials and the final usage in an exoskeleton. Generally it would be better if it is few effort to create an actuator but as long it is not extremely difficult the properties above are more important.

### **Needed supporting constructions**

Another thing to consider are the necessary constructions around the actuator, including straps to fix the EAP to the body, to keep the Polymer in the needed shape, and the size of the power source.

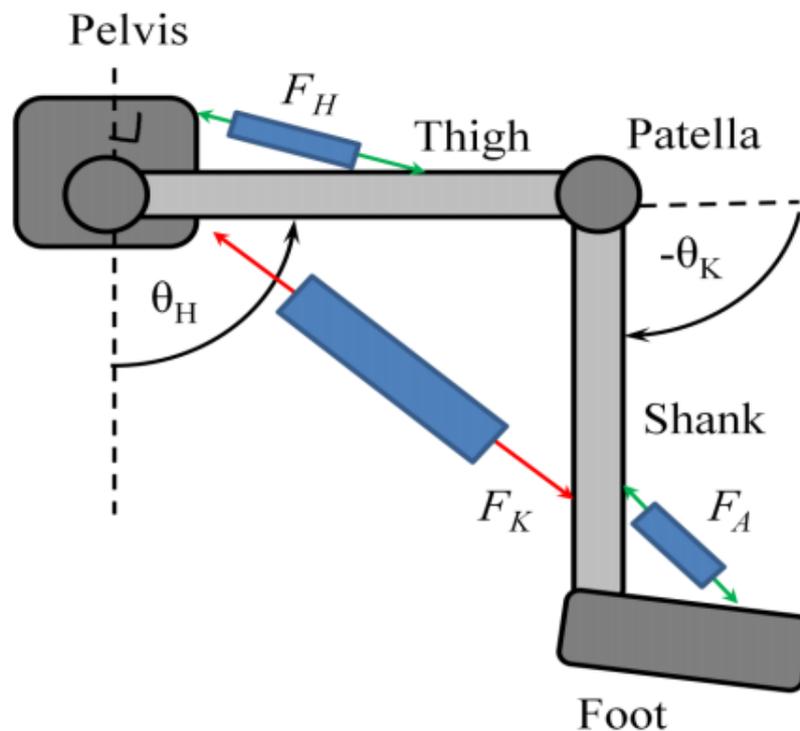
### **Availability**

Besides the properties of EAPs the economical side is not to be forgotten. To create many exoskeletons it is necessary that the used products are easily available and at the same time are still affordable. For this example the commercial availability in Germany is considered. This will be categorised into "good", for easily available, "import" if it needs to be imported, "disc" if discontinued, or "-" if not available or not found.

### 3 Exoskeleton build

In this section a few ways to build the exoskeleton and how to use the different EAPs are described. Two builds will be shown. One more classic exoskeleton and one using knitting to create a useful shape.

#### 3.1 Classic Build



**Fig. 1.** A simple exoskeleton approach with 3 actuators [11]

The classic build of exoskeletons consists of three main actuators. One between the pelvis and the thigh, between the pelvis and the shank, and one between the shank and the foot. This way all the main motions of a leg are covered. In this build the necessary force of the actuators is smaller than the one of a muscle while the necessary difference in length is bigger than usually in a human muscle.

While this build is simple to produce the necessary fixation points need extra construction that would reduce general movability and comfort.

### 3.2 Knitting

A desired build would be a suit that the wearer could put on like normal pants. Knitting artificial muscles seems to be a viable way to achieve that. The knitting gives structure and makes it more rigid. Depending on the way of knitting and the patterns the cloth will react in different ways to current. For bending the leg the lower parts of the pants would need to contract, for stretching, the upper part [24]. Knitting gives further restrictions especially to the shape of the actuator which is not possible with every EAP.

In this build there is a more direct contact between the skin and the EAP so lower voltages or extra isolation is necessary.



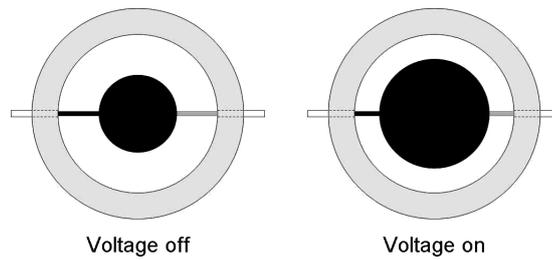
Fig. 2. Photograph of a Lyocell-based knitwear [24]

## 4 Types of EAPs

There are generally two main categories of electroactive polymers. They are differentiated by what physically makes the polymer move. This also creates a difference in how they move and changes the main physical properties.

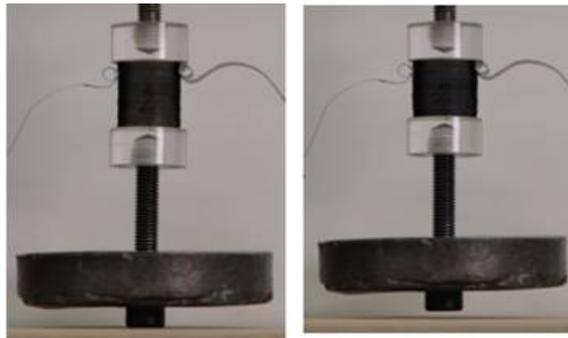
### 4.1 Dielectric EAPs

Dielectric EAPs usually consist of a thin and stretched piece of foil with two attached electrodes that cover opposite parts of the foil. When applying a very high voltage (usually 2-9kV) a electric field is induced which compresses the polymer while expanding in area. This compression is very small but it can be used by stacking many layers of the EAP on top of each other. The created actuator is mainly capable of contracting and relaxing which is the same behaviour as human muscles [21]. By increasing the number of electrodes or their shape many more kinds of movement are possible.



**Fig. 3.** The lateral expansion of an DEAP with carbon powder in the middle. [25]

A problem of dielectric electroactive polymers is that it acts as a capacitor. This combined with the up to 9kV it would endanger everyone touching it. Therefore it is not possible to use Dielectric EAPs without isolating it from the wearer which increases cost and might reduce efficiency. But because the main driving force is the electrical field, the voltage can be reduced by using thinner foils or some with a higher dielectric constant. [26]



**Fig. 4.** A stacked DEAP actuator in active and inactive state. [2]

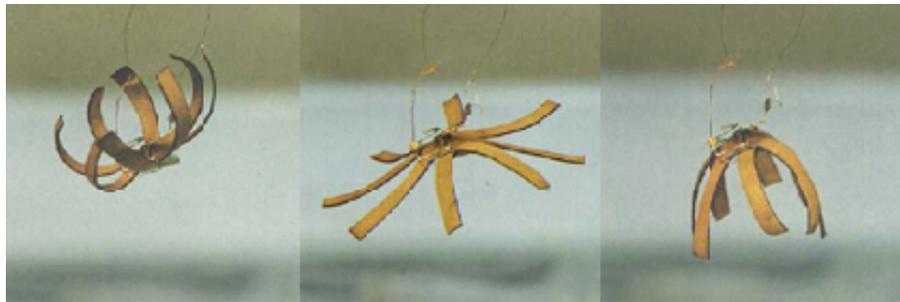
## 4.2 Ionic EAPs

Ionic EAPs consist of a polymeric membrane that is on two sides plated with electrodes to deliver power to the whole actuator. When a voltage is applied the membrane displaces ions. To do that it needs a constant exchange with a surrounding fluid. Often the membrane itself is soaked in this fluid. That way it is temporarily possible to use the actuator outside of the fluid. A way to keep the "wetness" permanently is by encapsulating the actuator. This is already done by some companies [14].

A subclass in IEAPs are the ionic polymer-metal composites (IPMC). These use metal as electrode. IPMCs are some of the most used ionic EAPs. They have been tested with several many different platings and fluids.

Compared to the DEAP an interesting difference is the bending movement of IEAP instead of contracting. Exchanging the anode and cathode it is even possible to make the IEAP bend in both directions. The bending movement the classic exoskeleton build difficult but might be useful for the knitting approach.

Another thing to mention is, that the produces pressure of IEAPs is much more dependent on the dimensions of the actor. This is connected to usually having only one fixed point while DEAPs often use an entire frame. Experiments have shown that the strength to increases exponentially with thickness and linearly with width while it is inversely proportional to length.[17].



**Fig. 5.** Several IPMC used together to simulate behaviour of a starfish [1]

### 4.3 Others

Other EAPs include Ferroelectric EAPs which are, due to their piezoelectric properties, used in sensors but have no real actuation capabilities[35].

Electrostrictive graft EAPs have been developed by NASA but have not yet been researched as much as Dielectric and Ionic Polymers.

Liquid EAPs like "Liquid crystalline EAPs" and "Electrorheological fluid EAPs" can be used in a system similar to pneumatics[33].

Stimuli-responsive gel EAPs are similar to the LCEAPs and EFEAPs not applicable to exoskeletons.

## 5 Values

In this section the collected data is displayed, separated into the two main categories IEAP and DEAP. The tables sometime contains the same material several times but the data always comes from different experiments. It needs to be mentioned that the DEAP-tables use megapascal, the IEAP however kilopascal. Another thing to mention is that with the Ionic Electroactive Polymers some more information about the dimensions is given. This is due to the stronger effect of the dimensions on created pressure.

Symbol	Meaning
eMax	Ultimate Strain
W	Elastic Energy Density
P	Pressure
E	Operational electric field
Y	Youngs Modulus
t	Reaction Speed
Er	Dielectric Constant
d	Thickness
l	Length
w	Width
U	Voltage
I	Current
Av	Availability

Name	eMax [%]	W [J/cm <sup>3</sup> ]	P [MPa]	E [MV/ml]	Y [MPa]	t	Er	Av	Source
Mammalian Skeletal Muscle	40	0,08	0,35	-	10 to 60	medium	-	-	[12]
1 Silicon proprietary	422	-	-	80	0,25	-	2,4	-	[4]
2 Prestrained Acrylic	380	3,4	7,2	-	-	medium	-	-	[6]
3 Dielectric elastomer	300	3,4	-	-	0,1 to 10	m. – fast	-	-	[12]
4 Prestrained Silicone	63	0,75	3	-	-	fast	-	-	[6]
5 Silicone – Nominal Prestrain	32	0,2	1,36	-	-	fast	-	-	[6]
6 Electrochemo-mechanical Conducting Pol.	20	1	-	-	200 – 3000	m. – slow	-	-	[12]
7 Electrostrictive Polymer	7	1	-	-	400 – 2000	fast	-	-	[12]
8 Polyelectrolyte	40	0,06	-	-	-	slow	-	-	[12]
9 Polymer Electrostrictor (zhang)	4	-	0,8	150	-	fast	-	-	[3], [23]
10 Thermoplastic dielectric elastomer	250	0,15	-	-	-	fast	-	-	[20]
11 A-00 (CN 9021)	134	-	-	179	0,16	-	5	-	[13]
12 A-05 (CN 9021)	172	-	-	220	0,28	-	6,1	-	[13]
13 A-10 (CN 9021)	90	-	-	157	0,38	-	7,2	-	[13]
14 A-15 (CN 9021)	56	-	-	140	0,53	-	8,4	-	[13]
15 A-CN Rubber	-	0,084	0,3	50	4	-	14	-	[6]
16 Bayfol EA102	300	-	-	130	1,44	-	7,1	-	[4]
17 Chlorosulfonatedpolyethylene(CSM)	-	-	-	-	-	-	7 – 10	-	[18]
18 DEA Acrylic	380	3,4	-	O(100)	1,0 to 3,0	-	-	-	[30]
19 DEA Silicone	120	0,75	-	O(100)	0,1 to 3,0	-	-	-	[30]
20 Deerfield PT6100S	11	0,087	1,6 – 1,9	160	17	fast	7	-	[6], [28], [29]
21 Deerfield PT6100S	32	-	0,2	144	-	fast	-	-	[3], [29]
22 Dow Corning 730	28	0,055	-	80	0,5	-	6,9	-	[28]
23 Dow Corning 730	-	0,0055	0,39	80	0,5	-	3,9	-	[6]

Table 1. First table comparing the different properties of DEAP

	Name	eMax [%]	W [J/cm <sup>3</sup> ]	P [MPa]	E [MV/m]	Y [MPa]	t	Er	Av	Source
	Mammalian Skeletal Muscle	40	0,08	0,35	-	10 to 60	medium	-	-	[12]
<b>24</b>	Dow Corning HS3	-	0,026	-	72	0,125	-	2,8	disc	[28]
<b>25</b>	Dow Corning HS3	63	0,75	-	-	0,13	fast	-	disc	[20]
<b>26</b>	Dow Corning HS3	69	0,034	0,13	72	-	-	2,8	disc	[6]
<b>27</b>	Dow Corning HS3	93	0,098	0,3	110	0,1	-	2,8	disc	[6]
<b>28</b>	Dow Corning HS3	-	0,026	0,13	72	0,135	-	2,8	disc	[6]
<b>29</b>	Dow Corning HS3	117	0,16	0,4	128	-	-	2,8	disc	[6]
<b>30</b>	Dow Corning Sylgard	32	0,034	0,21	-	0,7	-	-	good	[29]
<b>31</b>	Dow Corning Sylgard 186	32	0,082	0,51	144	0,7	-	2,8	good	[6], [28]
<b>32</b>	Dr. Scholl's Gelactive Tubing	-	-	0,0037	28	-	-	-	good	[6]
<b>33</b>	EPDM	420	-	-	-	2	-	3,0 - 3,5	-	[18]
<b>34</b>	EPR	-	-	-	-	-	-	3,17 - 3,34	-	[18]
<b>35</b>	Estane TPU 588	-	0,0025 - 0,0043	0,14	8	-	-	6	disc	[6]
<b>36</b>	Ethylene propylene	12	0,013	0,022	-	1,8	-	-	-	[29]
<b>37</b>	Fluorosilicone	28	0,019	0,14	-	0,5	-	-	-	[29]
<b>38</b>	Ionic Polymer Metal Composite 3,3	11	0,006	-	-	50 - 100	medium - slow	-	-	[12]
<b>39</b>	Isoprene	11	0,0052	0,094	-	0,85	-	-	-	[29]
<b>40</b>	Isoprene Natural Rubber Latex	11	0,0059	-	67	0,85	-	2,7	-	[28]
<b>41</b>	Isoprene Natural Rubber Latex	-	0,0059	0,11	67	0,85	-	2,7	-	[6]
<b>42</b>	Lauren L143HC	8	0,0046	0,11	32	2,5	-	12,7	disc	[6], [28]
<b>44</b>	Nusil CF19-2186	32	0,22	1,36	235	1	fast	2,8	-	[28], [6]
<b>45</b>	Nusil CF19-2186	33	0,091	0,6	160	-	-	2,8	-	[6]
<b>46</b>	Nusil CF19-2186	64	0,75	3	350	1	-	2,8	-	[6]

Table 2. Second table comparing the different properties of DEAP

	Name	eMax [%]	W [J/cm <sup>3</sup> ]	P [MPa]	E [MV/m]	Y [MPa]	t	Er	Av	Source
	Mammalian Skeletal Muscle	40	0,08	0,35	-	10 to 60	medium	-	-	[12]
47	Nusil CF19-2186	63 (linear)	0,2	0,8	181	-	-	2,8	-	[6]
48	P(VDF-TrFE)	4,3	0,92	43	-	-	fast	-	good	[6]
49	P(VDF)	0,1	0,0024	-	-	450	fast	-	good	[20], [12]
50	PANI	10	23	-	-	-	slow	-	-	[20]
51	Polybutadiene	12	0,011	0,19	-	1,7	-	-	-	[29]
52	Polyurethane TPU LPT 4210 UT 50	421	-	-	218	3,36	-	6	-	[4]
53	PTBA	335	1,2	3,2	260	0,42	medium	5,4	-	[6]
54	SEBS161 (30 wt.% copolymer)	16,5	0,013	-	27	1,3	-	-	-	[6]
55	SEBS161 (5 - 30 wt.% copolymer)	180 - 30	0,141 - 0,151	-	32 - 133	0,007 - 0,163	-	1,8 - 2,2	-	[6]
56	SEBS217 (30 wt.% copolymer)	18,8	0,013	-	29	1,1	-	-	-	[6]
57	SEBS217 (5 - 30 wt.% copolymer)	245 - 47	0,119 - 0,139	-	22 - 98	0,002 - 0,133	-	1,8 - 2,2	-	[6]
58	SEBS75 (25 - 30 wt.%)	19	0,0263	-	29 - 41	12	-	2,08 - 2,12	-	[6]
59	Solvane 250	525	-	-	250	800	-	14	-	[4]
60	VBH 4905	879	-	-	31	0,04	-	4,5	good	[4]
61	VHB 4905-TMPTMA	146	0,68	1,51	265,4	3,94	-	2,43	good	[6]
62	VHB 4910	40	0,022	0,13	55	-	-	4,8	good	[6]
63	VHB 4910	158	3,4	7,2	412	3	-	4,8	good	[6]
64	VHB 4910	380	3,4	7,2	-	1 - 3	fast	-	good	[6]
65	VHB 4910	215 (linear)	1,36	2,4	239	-	-	4,8	good	[6]
66	VHB 4910-HDDA	233	-	-	300	2,5	-	-	good	[6]
67	VHB 4910-TMPTMA	300	3,5	5,06	418	4,15	-	3,27	good	[6]

Table 3. Third table comparing the different properties of DEAP

Name	d [mm]	l [mm]	w [mm]	A [mm <sup>2</sup> ]	F [mN]	P [kPa]	U [V]	I [mA]	DeltaL [mm]	Y [MPa]	Av	Source
Mammalian Skeletal Muscle	-	-	-	-	-	350	-	-	-	10 to 60	-	[12]
1 Flemion BMI-BF4	0,145	25	3,5	87,5	0,05884	0,67	5	-	-	-	-	[34]
2 IH-immersion EMIBF4	419	20	3	60	0,14	2,33	3	-	-	-	-	[32]
3 IH-immersion LiTFSI	417	20	3	60	0,1	1,67	3	-	-	-	-	[32]
4 MPI-immersion EMIBF4	301	20	3	60	4	66,67	3	-	-	-	-	good [32]
5 MPI-immersion LiTFSI	294	20	3	60	2	33,33	3	-	-	-	-	good [32]
6 MPI-vapor EMIBF4	145	20	3	60	0,5	8,33	3	-	-	-	-	good [32]
7 MPI-vapor LiTFSI	155	20	3	60	4	66,67	3	-	-	-	-	good [32]
8 Nafion Ag-Nafion	0,23	20	5	100	2,177	21,77	6,78	-	8	21540	-	good [9]
9 Nafion Li+ 0,22mm	0,22	25	5	125	1,96	15,68	3	-	11	-	-	good [17]
10 Nafion Li+ 0,5mm	0,5	25	5	125	18,6	148,80	3	-	11	-	-	good [17]
11 Nafion Li+ 1mm	1	25	5	125	63,7	509,60	3	-	11	-	-	good [17]
12 Nafion Pd 120sec	0,5	50	10	500	6	12,00	1	120	1,2	-	-	good [27]
13 Nafion Pd 150sec	0,5	50	10	500	9	18,00	1	120	1,45	-	-	good [27]
14 Nafion Pd 90sec	0,5	50	10	500	3,5	7,00	1	120	0,75	-	-	good [27]
15 Nafion-117	0,18	40	5	200	2,7	13,50	3	-	5,5	900	-	good [19]
16 Nafion-117 AlfaAesar RuO2	0,233	40	8	320	3,3	10,31	2	-	-	-	-	good [27]
17 Nafion-117 Coconut	0,233	40	8	320	3,1	9,69	2	-	-	-	-	good [27]
18 Nafion-117 O2-2 µm	0,18	40	5	200	2,8	14,00	3	-	7,3	1000	-	good [19]
19 Nafion-117 O2-4 µm	0,18	40	5	200	2,6	13,00	3	-	9	880	-	good [19]
20 Nafion-117 O2-6 µm	0,18	40	5	200	2,4	12,00	3	-	6,6	780	-	good [19]
21 Nafion-117 SigmarAldrich RuO2	0,233	40	8	320	3,5	10,94	2	-	-	-	-	good [27]
22 Nafion-117 TIC	0,233	40	8	320	3,6	11,25	2	-	-	-	-	good [27]
23 PEO/NBR/PEDOT IPN EMITFSI	0,294	15	8	120	30	250,00	3	180	1,2	-	-	[10]
24 polypyrrole/gold-coated PVDF	0,152	20	2	40	1,2	30,00	1-4	50	60	-	-	[36]

Table 4. Table comparing the different properties of IEAP

## 6 Most promising EAPs

In this section the most viable solutions for exoskeletons will be discussed and further evaluated.

### 6.1 VHB4910/05

VHB4910 could be called the standard product when thinking about dielectric polymers. While it can create a measured pressure of up to 7.2 MPa it is still very flexible and survives a relatively high strain.

This though comes with the disadvantage that voltages of up to 4000V need to be applied for the actuation, making it dangerous to touch. The main way to solve this problem is to add an isolation layer. This however would reduce the efficiency of VHB4910.

Another difficulty is that, to produce the best results, the foil needs to be stretched before actuating. To keep this stretching permanently a flexible frame has to be used. This creates additional cost, effort and the size of the artificial muscle.

It is still one of the most used EAPs for experiments, robotics and even art [22]. A factor for this is the easiness to buy VHB4910. It is possible to purchase it directly from the German M3 Onlineshop. 20 square meters are available from 2.120EUR (March/2017). Either carbon-grease or carbon-powder is usually used for electrode, both of which are quite cheap.

The possibility to harvest energy from VHB4910 was tested by C. Jean-Mistral et. al.[15]. They created a generator that showed promise to be used in autonomous devices. The general build of the generator is was the same as it would be for a normal generator. So harvesting energy from a VHB4910-actuator is more than possible.

### 6.2 Dr. Scholls Gelactive Tubing

A remarkable and unexpected candidate for a exoskeleton came out of the consumer medical treatment market. The Dr.Scholls Gelactive Tubing are normally used for comfort and healing on toes and fingers but Paul Brochu et. al.[7] found they had potential as EAP. Sadly the created pressure is much less interesting than the one of VHB4919 but is still mentioned here to show that EAPs sometimes are just found by trying different substances.

### 6.3 P(VDF-TrFE)

While P(VDF-TrFE), a ferroelectric copolymer, appears to create by far the highest pressure it happens without big strain. Meaning that indeed it does create a high force but it does not move much while doing it. To convert the created pressure to a bigger movement lots of this force would be lost so in the end it sadly is not a viable solution.

They are still widely used but not as actuators but as sensors.

#### 6.4 Nafion

As the main competitor in ionic EAPs Nafion is still much weaker than most dielectric EAPs. One of its properties though is, that the strength seems to increase exponentially with the thickness and linearly with width while it decreases inversely proportional with length [17].

These properties could be used with the knitting approach of building a exoskeleton. Vertical yarns could give the necessary actuation points to reduce loss of strength by length. Increasing the thickness or polymer strains would increase the general pressure.

While VHB4910 has to be stretched and needs carbon-powder the production is more difficult and expensive with Nafion. It is available either as foil, with  $1.5m^2$  for 2400\$, or as fluid with 500mL for 978\$ in the USA. The foil (Nafion 117) is  $183\mu m$  thick while the fluid allows to choose the thickness.

After buying or creating the membrane the plating of metal is done through a chemical process. In this example palladium is used. Needed are: tetraamineplatinum(II) chloride (446EUR/5g), sodium borohydride (30EUR/5g), and surface electroding. [31]. (Prices from different international stores)

## 7 Conclusion

While the dielectric electroactive polymers are definitely the winners with the VHB4910 as strongest usable actuator the ionic electroactive polymers might be a promising candidate in the near future. Although their created pressures are weaker the much smaller required voltage are a clear advantage and some solutions for the needed "wetness" are also already found. A interesting topic to look further into is to create a fabric out of IPMCs to make it possible to wear the exoskeleton like very expensive pants.

Currently though only the dielectric EAPs were capable of creating a pressure similar to that of a human muscle. With them the high voltages have to be isolated from the wearer. Therefore wearing them like pants is not so easy. With DEAPs it would be necessary to add actuation points on the body. Definitely less comfortable but a viable option.

All in all it is possible to say that EAPs are with some limitations a good alternative to classical motors, hydraulics or pumps in exoskeletons.

## References

1. Electroactive polymers 2: Ionic and conductive polymers, <http://boltent.ca/flight/lessons/research/youDecide/ionicNConducPolym.html>
2. Bar-Cohen, Y.: Eap-in-action session 3 (2009), [http://ndea.jpl.nasa.gov/nasa-nde/lommas/eap/EAPIA/EAP-in-Action\\_2009.htm](http://ndea.jpl.nasa.gov/nasa-nde/lommas/eap/EAPIA/EAP-in-Action_2009.htm)
3. Bar-cohen, Y., Leary, S.: Electroactive polymers as artificial muscles changing robotics paradigms (2000)
4. Biggs, J., Danielmeier, K., Hitzbleck, J., Krause, J., Kridl, T., Nowak, S., Orselli, E., Quan, X., Schapeler, D., Sutherland, W., Wagner, J.: Electroactive polymers: Developments of and perspectives for dielectric elastomers. *Angewandte Chemie International Edition* 52(36), 9409–9421 (2013), <http://dx.doi.org/10.1002/anie.201301918>
5. Bikson, D.M.: A review of hazards associated with exposure to low voltages (2014), [https://www.researchgate.net/publication/242705238\\_A\\_review\\_of\\_hazards\\_associated\\_with\\_exposure\\_to\\_low\\_voltages](https://www.researchgate.net/publication/242705238_A_review_of_hazards_associated_with_exposure_to_low_voltages)
6. Brochu, P.: Dielectric elastomers for actuation and energy harvesting (2012)
7. Brochu, P., Pei, Q.: Advances in dielectric elastomers for actuators and artificial muscles. *Macromolecular Rapid Communications* 31(1), 10–36 (2010), <http://dx.doi.org/10.1002/marc.200900425>
8. Chen, E.J., Novakofski, J., Jenkins, W.K., O'Brien, W.D.: Young's modulus measurements of soft tissues with application to elasticity imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 43(1), 191–194 (Jan 1996)
9. Chung, C., Fung, P., Hong, Y., Ju, M., Lin, C., Wu, T.: A novel fabrication of ionic polymer-metal composites (ipmc) actuator with silver nano-powders. *Sensors and Actuators B: Chemical* 117(2), 367 – 375 (2006), [//www.sciencedirect.com/science/article/pii/S0925400505008944](http://www.sciencedirect.com/science/article/pii/S0925400505008944), transducers '05/Selected Papers from {TRANSDUCERS} '05. The 13th International Conference on Solid-State Sensors, Actuators and Microsystems
10. Festin, N., Maziz, A., Plesse, C., Teyssié, D., Chevrot, C., Vidal, F.: Robust solid polymer electrolyte for conducting ipn actuators. *Smart Materials and Structures* 22(10), 104005 (2013)
11. Henderson, G.C., Ueda, J.: Pneumatically-powered robotic exoskeleton to exercise specific lower extremity muscle groups in humans. In: *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on*. pp. 1506–1511. IEEE (2012)
12. Herr, H.M., Kornbluh, R.D.: New horizons for orthotic and prosthetic technology: artificial muscle for ambulation, <http://dx.doi.org/10.1117/12.544510>
13. Hu, W., Zhang, S.N., Niu, X., Liu, C., Pei, Q.: An aluminum nanoparticle-acrylate copolymer nanocomposite as a dielectric elastomer with a high dielectric constant. *J. Mater. Chem. C* 2, 1658–1666 (2014), <http://dx.doi.org/10.1039/C3TC31929F>
14. Jaakson, P., Aabloo, A., Tamm, T.: Encapsulation of ionic electroactive polymers: reducing the interaction with environment (2016), <http://dx.doi.org/10.1117/12.2219053>
15. Jean-Mistral, C., Basrour, S., Chaillout, J., Bonvilain, A.: A complete study of electroactive polymers for energy scavenging: modelling and experiments. *CoRR abs/0802.3046* (2008), <http://arxiv.org/abs/0802.3046>
16. Kern, T.A.: *Engineering Haptic Devices: A Beginner's Guide for Engineers*. Springer Publishing Company, Incorporated, 1st edn. (2009)

17. Kim, H.I., Kim, D.K., Han, J.H.: Study of flapping actuator modules using ipmc (2007), <http://dx.doi.org/10.1117/12.715633>
18. Kim, K.J., Tadokoro, S.: *Electroactive Polymers for Robotic Applications: Artificial Muscles and Sensors*. Springer Publishing Company, Incorporated, 1 edn. (2007)
19. Kim, S.J., Lee, I.T., Kim, Y.H.: Performance enhancement of ipmc actuator by plasma surface treatment. *Smart Materials and Structures* 16(1), N6 (2007), <http://stacks.iop.org/0964-1726/16/i=1/a=N02>
20. Koo, C.: *Electroactive Thermoplastic Dielectric Elastomers as a New Generation Polymer Actuators*. INTECH Open Access Publisher (2012), <https://books.google.de/books?id=QGmZoAECAAJ>
21. Kovacs, G., Düring, L., Michel, S., Terrasi, G.: Stacked dielectric elastomer actuator for tensile force transmission. *Sensors and Actuators A: Physical* 155(2), 299 – 307 (2009), [//www.sciencedirect.com/science/article/pii/S0924424709004002](http://www.sciencedirect.com/science/article/pii/S0924424709004002)
22. Larkin, N.: *Ilts13 elastomeric escapisms* (2014), <https://www.behance.net/gallery/14401615/Installation-ILTS13-Elastomeric-Escapisms>
23. Li, F., Zhu, W., Zhang, X., Zhao, C., Xu, M.: Shape memory effect of ethylene–vinyl acetate copolymers. *Journal of Applied Polymer Science* 71(7), 1063–1070 (1999), [http://dx.doi.org/10.1002/\(SICI\)1097-4628\(19990214\)71:7<1063::AID-APP4>3.0.CO;2-A](http://dx.doi.org/10.1002/(SICI)1097-4628(19990214)71:7<1063::AID-APP4>3.0.CO;2-A)
24. Maziz, A., Concas, A., Khaldi, A., Stålhund, J., Persson, N.K., Jager, E.W.H.: Knitting and weaving artificial muscles. *Science Advances* 3(1) (2017), <http://advances.sciencemag.org/content/3/1/e1600327>
25. Michel, S., Zhang, X.Q., Wissler, M., Löwe, C., Kovacs, G.: A comparison between silicone and acrylic elastomers as dielectric materials in electroactive polymer actuators. *Polymer international* 59(3), 391–399 (2010)
26. Mirfakhrai, T., Madden, J.D., Baughman, R.H.: Polymer artificial muscles. *Materials Today* 10(4), 30 – 38 (2007), <http://www.sciencedirect.com/science/article/pii/S1369702107700482>
27. Palmre, V., Pugal, D., Leang, K.K., Kim, K.: The effects of electrode surface morphology on the actuation performance of ipmc. In: *Proc. SPIE*. vol. 8687, p. 86870W (2013)
28. Pelrine, R., Kornbluh, R., Joseph, J., Heydt, R., Pei, Q., Chiba, S.: High-field deformation of elastomeric dielectrics for actuators. *Materials Science and Engineering: C* 11(2), 89 – 100 (2000), [//www.sciencedirect.com/science/article/pii/S0928493100001284](http://www.sciencedirect.com/science/article/pii/S0928493100001284)
29. Pelrine, R.E., Kornbluh, R.D., Joseph, J.P.: Electrostriction of polymer dielectrics with compliant electrodes as a means of actuation. *Sensors and Actuators A: Physical* 64(1), 77 – 85 (1998), <http://www.sciencedirect.com/science/article/pii/S0924424797016579>
30. Rosenblatt-Weinberg, F.: *Modelling and Optimisation of Electro-Active Polymer (EAP) Devices*. Imperial College London (2013), <https://books.google.de/books?id=bXzirQECAAJ>
31. Siripong, M., Fredholm, S., Nguyen, Q.A., Shih, B., Itescu, J., Stolk, J.: A cost-effective fabrication method for ionic polymer-metal composites. In: *MRS Proceedings*. vol. 889, pp. 0889–W04. Cambridge Univ Press (2005)
32. Temmer, R., Must, I., Kaasik, F., Aabloo, A., Tamm, T.: Combined chemical and electrochemical synthesis methods for metal-free polypyrrole actuators. *Sensors and Actuators B: Chemical* 166, 411–418 (2012)
33. Veale, A.J., Anderson, I.A., Xie, S.Q.: The smart peano fluidic muscle: a low profile flexible orthosis actuator that feels pain (2015), <http://dx.doi.org/10.1117/12.2084130>

34. Wang, J., Xu, C., Taya, M., Kuga, Y.: A flemion-based actuator with ionic liquid as solvent. *Smart Materials and Structures* 16(2), S214 (2007), <http://stacks.iop.org/0964-1726/16/i=2/a=S03>
35. Wikipedia: Ferroelectric polymers — wikipedia, the free encyclopedia (2016), [https://en.wikipedia.org/w/index.php?title=Ferroelectric\\_polymers&oldid=748319496](https://en.wikipedia.org/w/index.php?title=Ferroelectric_polymers&oldid=748319496), [Online; accessed 7-November-2016]
36. Wu, Y., Alici, G., Spinks, G.M., Wallace, G.: Fast trilayer polypyrrole bending actuators for high speed applications. *Synthetic Metals* 156(16), 1017–1022 (2006)

# Matrix Factorization for Next Place Prediction

Stefan Christian Lamp\*

Betreuer: Antonios Karatzoglou†

Karlsruher Institut für Technologie (KIT)  
Pervasive Computing Systems – TECO

\*stefan.lamp@student.kit.edu

†antonius@teco.edu

**Zusammenfassung.** Next-Place-Prädiktion ist eine wichtige Aufgabe, welche die User Experience verbessern und weiterhin, eingebettet in anderen Diensten, einen Mehrwert bieten kann. Beispielhaft können dem Benutzer bereits proaktiv konkrete Vorschläge für den nächsten Ort gemacht werden, z.B. durch Empfehlung eines Restaurants zu einer Zeit in der der Benutzer häufig Restaurants aufsucht. So können Dienste nicht nur auf den gegenwärtigen Kontext reagieren, sondern auch auf den zukünftigen und im Gegensatz zu bloßen Rekommandern auf Bedürfnisse eingehen, bevor der Benutzer diese äußert. Um diese Aufgabe zu stemmen, wird in dieser Arbeit ein Verfahren für Semantic-Next-Place-Prädiktion vorgeschlagen. Hierfür wird nicht nur ein einzelner Benutzer betrachtet, sondern auch ihm ähnliche. Weiterhin spielen neben den direkt zuvor besuchten Orten auch die Reihenfolge, in welcher diese besucht wurden, eine Rolle. Um dies zu realisieren, wird eine auf eine Kombination aus Matrixfaktorisierung und Markov-Ketten höherer Ordnung zurückgegriffen, namentlich Factorized Personalized Markov Chains (FPMC). Durch die Verwendung von Matrixfaktorisierung wird neben der Reduzierung der benötigten Rechenleistung auch der nötige Speicherplatz deutlich reduziert.

**Schlüsselwörter:** Location Prediction, Markov Chain, Matrix Factorization

## 1 Einleitung

Die Erfassung der Position des Benutzer spielt in Zeiten von Ubiquitous Computing eine immer größere Rolle. Mehr und mehr Menschen besitzen Smartphones: Im Jahr 2015 gab es weltweit 1,86 Milliarden Smartphone-Benutzer [7], in Deutschland waren es 46 Millionen [5], was einem Anteil von etwa 56 Prozent der Gesamtbevölkerung entspricht [17]. Im Zuge des Smartphone kamen auch neue Anwendungen hinzu, welche den Standort des Benutzers verwenden, um verschiedene Funktionen zu ermöglichen und zu unterstützen. Beispiele für solche kontext-bewussten Anwendungen sind Location-based Services, wie Navigation, Suche in der Umgebung, sowie Dienste aus dem Bereich von Sozialen Netzwerken, Sport, Tourismus und Gastronomie.

Eine Online-Umfrage aus dem Jahre 2015 ergab, dass 42 Prozent der Befragten ihr Smartphone für eine Suche in der Umgebung gelegentlich verwenden, 30 Prozent sogar häufig [16]. Dies spiegelt sich auch im Erfolg des Touristikportals TripAdvisor wieder, welches die Suche von Hotels, Freizeitaktivitäten sowie Restaurants in der Nähe ermöglicht und somit viele der genannten Anwendungen abdeckt. Im dritten Quartal 2016 betrug der Umsatz des Portals 421 Millionen US-Dollar [18] und allein in Deutschland hatte TripAdvisor im Jahr 2015 sechs Millionen unterschiedliche Benutzer [1].

Sowohl TripAdvisor, als aber auch andere Rekommendersysteme, versuchen ihre Ergebnisse und Empfehlungen dem Benutzer anzupassen.

Dies passiert einerseits anhand der Position der Benutzer, sodass beispielsweise Empfehlungen in örtlicher Nähe zum Anwender stehen. Andererseits spielen hier auch seine individuellen Interessen eine Rolle. Hierzu wird meist der Ansatz der Matrixfaktorisierung verwendet, welcher seit dem Netflix Price an Bekanntheit gewonnen hat [13]. Ziel des Wettbewerbs war die Entwicklung eines Rekommendersystems, das dem Benutzer der Film-Streaming-Plattform Netflix Filme, passend zu dessen Geschmack, empfiehlt. Beide Gewinner verwendeten dabei Matrixfaktorisierung in ihren Einreichungen.

Im Rahmen dieser Arbeit wird ein semantischer Next-Place-Prädiktor erarbeitet. Er basiert dabei auf einer Kombination von faktorisierten Markov-Ketten höherer Ordnung und Matrixfaktorisierung, um so anhand der bisherigen Orte des Benutzers sowie basierend auf dem Verhalten ähnlicher Benutzer Voraussagen treffen zu können, welchen Ort der Benutzer als nächstes besucht. Um besser das Verhalten verschiedener Benutzer vergleichen zu können, aber auch um Datenschutz zu gewährleisten, wird, anstatt konkrete Koordinaten zu verwenden, auf einer semantischen Ebene gearbeitet. Statt mit konkreten Orten wird somit mit Kategorien, wie zum Beispiel *Zuhause* und *Arbeitsstelle*, gearbeitet.

Im Folgenden soll zunächst ein Blick über das Forschungsfeld gegeben werden und es wird erörtert, wie sich diese Arbeit von anderen abhebt. Anschließend werden für die Arbeit relevante mathematische Ausdrücke definiert, sowie eine Einführung zu Markov-Ketten und Matrixfaktorisierung, welche die Grundlage des Prädiktors bilden, gegeben. Anschließend wird das mathematische Modell des Prädiktors, sowie dessen algorithmische und softwaretechnische Umsetzung näher beleuchtet, um diese schließlich mit verschiedenen Datensätzen zu evaluieren. Die Ergebnisse der Evaluation werden genauer betrachtet. Schlussendlich wird die Arbeit nochmals zusammengefasst und ein Ausblick über mögliche weitere Schritte wird gegeben.

## 2 Ähnliche Arbeiten

Die Kombination von Matrixfaktorisierung und Lokationsprädiktion scheint neu zu sein, da keine wissenschaftlichen Arbeiten zu diesem speziellen Thema gefunden werden konnten. Deshalb beleuchtet dieses Kapitel die beiden Themen

getrennt voneinander. Zunächst werden Arbeiten zu Lokationsprädiktion mit verschiedenen Methoden vorgestellt. Anschließend, um das Thema Matrixfaktorisierung abzudecken, wird ein zu Lokationsprädiktion verwandtes Feld betrachtet, in welchem diese Methode oft verwendet wird: Lokationsrekommender.

Lokationsprädiktion kann als zwei zueinander unterschiedliche Dinge aufgefasst werden. Einerseits, wie in dieser Arbeit verwendet, die Vorhersage des nächsten Ortes (Next Place Prediction), den ein Benutzer besuchen wird, oder die Voraussage, wo sich Orte befinden (Location Learning). Location Learning ist das Thema von [3]. Die vorgeschlagene Methode erlaubt die Vorhersage von Orten anhand von GPS Tagestrajektorien, welche von größerer Bedeutung für den Benutzer sind, wie z.B. Zuhause, Arbeit, oder andere. Darüberhinaus wird in der Arbeit ein Next Place Predictor beschrieben, der eine Markov-Kette ersten Grades verwendet. Dagegen konzentriert sich [8] nur auf Next Place Prediction und verwendet dafür höhergradige Markov-Ketten. Somit werden die letzten  $n$  Orte beachtet, um eine Aussage über den nächsten zu treffen. [9] erweitert die Markov-Ketten um den zeitlichen Kontext. Dies wird erreicht durch die Verwendung einer Wahrscheinlichkeitsverteilung: Es wird berücksichtigt zu welcher Wahrscheinlichkeit ein Ort zu einer bestimmten Tagesstunde besucht wird. Somit betrachtet das Modell sowohl den letzten besuchten Ort, als auch die momentane Uhrzeit. Weiterhin werden in der Arbeit Tagestrajektorien verwendet, welche die Zeitabschnitte beinhalten, in welchen der Benutzer in der Nähe verschiedener Mobilfunkzellen sich befand.

Im Unterschied zu den vorherigen Arbeiten, verwendet [20] Daten aus lokationsbasierten sozialen Netzwerken (LBSN). Die vorgeschlagene Methode verwendet die semantischen Annotationen von Check-Ins beziehungsweise Kategorien anstatt von GPS-Koordinaten. Sie basiert auf einer Mixed Hidden Markov Chain (MHKC), welche, anstatt den konkreten nächsten Ort vorherzusagen, die Kategorie dessen vorhersagt, um die Vorhersagequalität zu erhöhen. Ebenfalls berücksichtigt werden sowohl der örtliche, als auch der zeitliche Kontext des Benutzers. Unter Berücksichtigung des selben Kontexts wurde in [12] ein Lokationsprädiktor erarbeitet, der auf rekurrenten neuronalen Netzen basiert.

Die Modelle der vorherigen Arbeiten berücksichtigen immer nur einen Benutzer. Anstatt für jeden Benutzer ein eigenes Modell zu verwenden, könnte auch ein Modell für mehrere Benutzer Vorhersagen treffen, was, abgesehen von einer geringeren Speicherplatznutzung, dem Prädiktor auch erlauben würde, Ähnlichkeiten und Muster zwischen verschiedenen Benutzern zu finden und so bessere Vorhersagen zu treffen. Hierfür soll Matrixfaktorisierung verwendet werden.

Mehrere verschiedene Arten von Matrixfaktorisierung sind beschrieben in [11]. Sehr hervorgehoben wird dort auch die Arbeit von [10], welche den Netflix Prize [13] gewonnen hat. Der vorgeschlagene Recommender schlägt den Benutzern des Streaming-Dienstes Filme vor. Unter Verwendung von Matrixfaktorisierung werden Vorschläge gemacht, implizit basierend auf Benutzern mit ähnlichem Filmgeschmack.

Wie anfangs erwähnt, wird Matrixfaktorisierung auch in Lokationsrekommandern verwendet, welche einen hohen Verwandtschaftsgrad zu Lokationsprädiktoren aufweisen. [4] stellt eine Methode vor, welche Matrixfaktorisierung mit Markov-Ketten verbindet und Check-In-Daten aus LBSN verwendet. Dies erlaubt personalisierte Vorschläge für mehrere Benutzer, welche zudem auch die Region, in welcher der Benutzer sich aufhält mit einbezieht. Dadurch wird der Suchraum verkleinert und nur Orte in näherer Umgebung vorgeschlagen.

Der in dieser Arbeit vorgestellte Prädiktor verwendet ebenfalls Check-In-Daten von LBSNs und konzentriert sich auf die Vorhersage der Semantik des nächsten besuchten Ortes. Die Vorhersage wird durch Verwendung von Matrixfaktorisierung sowie Markov-Ketten getroffen. Im Gegensatz zur Arbeit von Chen Cheng u.a. [4] werden nicht nur die zuvor besuchten Orte betrachtet, sondern, durch Verwendung von Markov-Ketten höherer Ordnung, auch deren Reihenfolge. Weiterhin ist auch hervorzuheben, dass im Gegensatz zu vielen anderen Arbeiten, welche Check-In-Daten verwenden, es sich hier um einen Prädiktor und keinen Rekommender handelt. Der Prädiktor sagt den nächsten besuchten Ort des Benutzers voraus, während ein Rekommender diesem etwaige nächste Orte vorschlägt.

### 3 Grundlagen

#### 3.1 Formalisierung

In den weiteren Kapiteln werden durchgängig verschiedene Notationen verwendet, die wie folgend definiert sind.

Die Menge  $U = \{u_1, \dots, u_{|U|}\}$  steht hierbei für verschiedene Benutzer und  $I = \{i_1, \dots, i_{|I|}\}$  für eine Menge von unterschiedlichen semantischen Orten.

$H_u$  sei hierbei die geordnete Menge der besuchten Orte von Benutzer  $u$ , sortiert nach dem Datum des Besuchs.  $H_{u,t}$  entspricht dem Ort, den der Benutzer  $u$  zum Zeitpunkt  $t$  besucht hat. Folgerichtig handelt es sich bei  $H_{u,t-1}$  und  $H_{u,t+1}$ , um die beiden Orte die der Benutzer zuvor und danach besucht hat. Der absolute Zeitpunkt ist hierbei nicht entscheidend, sondern die relative Abfolge von Besuchen. Weiterhin ist  $H$  die Menge der Besuchshistorien aller Benutzer  $U$ .

Ziel des hier erarbeiteten Prädiktors ist es, für einen Benutzer  $u$  eine persönliche Rangfolge von möglichen Nachfolgeorten  $H_{u,t+1}$  zum Zeitpunkt  $t$  zu bestimmen:

$$\langle_{u,t} \subset I \times I$$

Die Spitze dieses Rankings dient hierbei als Vorhersage über  $H_{u,t+1}$ . Die Variable  $n$  sei definiert als die Anzahl von Orten die betrachtet werden, um eine Voraussage darüber zu treffen, wohin der Benutzer als nächstes gehen wird. Sie entspricht weiterhin dem Grad der verwendeten Markov-Ketten.

### 3.2 Matrixfaktorisierung

Matrixfaktorisierung ist die Faktorisierung einer Matrix in ein Produkt von Matrizen. Sei  $A$  eine zu faktorisierende Matrix, dann werden die Faktormatrizen  $B_1, \dots, B_n$  so gewählt, dass gilt:

$$A = B_1 \cdot \dots \cdot B_n$$

Die Anzahl der Matrizen und deren Größe hängt dabei davon ab, welche der unzähligen verfügbaren Methoden gewählt wurde. Jede der Methoden ist einer Klasse von Problemen zugeordnet. Im Bereich des Machine Learning werden dabei oft Methoden verwendet, bei welchen das Produkt der Faktorisierung die faktorisierte Matrix lediglich annähert. Folgend soll dies anhand eines Beispiels veranschaulicht werden:

In Recommender-Systemen wie bei Netflix gibt es eine Menge von Benutzern  $U$  und eine Menge von Filmen  $I$ , welche von den Benutzern bewertet werden können. Die Bewertungen werden in einer Matrix gespeichert, wie in Tabelle 1 zu sehen ist. Einige der Bewertungen sind noch unbekannt. Anschließend wird die

**Tabelle 1.** Beispielmatrix eines Recommender-Systems [2]

	I1	I2	I3	I4
U1	5,00	3,00	-	1,00
U2	4,00	-	-	1,00
U3	1,00	1,00	-	5,00
U4	1,00	-	-	4,00
U5	-	1,00	5,00	4,00

Matrix in zwei kleinere Matrizen  $Q$  und  $P$  faktorisiert. Dies geschieht durch eine bewusst ungenaue Faktorisierung, wodurch bei der Remultiplikation die unbekanntenen Werte vorausgesagt werden. Wie in Tabelle 2 zu sehen ist, orientieren

**Tabelle 2.** Remultiplizierte faktorisierte Matrix [2]

	I1	I2	I3	I4
U1	4,97	2,98	<b>2,18</b>	0,98
U2	3,97	<b>2,40</b>	<b>1,97</b>	0,99
U3	1,02	0,93	<b>5,32</b>	4,93
U4	1,00	<b>0,85</b>	<b>4,59</b>	3,93
U5	<b>1,36</b>	1,07	4,89	4,12

sich die vorausgesagten Bewertungen eines Benutzers dabei an Benutzern mit ähnlichem Bewertungsschema. Weiterhin ermöglicht die Faktorisierung die Speicherung einer großen Matrix in mehreren kleinen, welche in der Summe weniger Speicherplatz verbrauchen.

### 3.3 Markov-Ketten

Eine Markov-Kette ist ein stochastischer Prozess mit dem Ziel die Wahrscheinlichkeit für das Eintreten eines zukünftigen Zustands anzugeben. Die Grundannahme ist dabei, dass für eine ausreichend gute Vorhersage nicht alle vergangenen Zustände nötig sind, sondern nur bestimmte Anzahl. So ist bei einer Markov-Kette 1.Ordnung bereits der gegenwärtige Zustand ausreichend. Analog dazu verwendet eine Markov-Kette  $n$ .Ordnung den gegenwärtigen Zustand und  $n - 1$  vergangene. Somit werden Markov-Ketten meist bei Systemen verwendet, deren Zustandsveränderungen keinen oder nur zeitlich begrenzten Einfluss aufeinander haben, um eine gute Vorhersage zu garantieren. Sei  $X_t$  der Zustand eines Systems zum Zeitpunkt  $t$ , dann sind die Wahrscheinlichkeiten der Zustandsübergänge mit einer Markov-Kette 1.Ordnung darstellbar, wenn gilt:

$$P(X_{t+1}|X_t, \dots, X_0) = P(X_{t+1}|X_t)$$

Respektive gilt ein Systemverhalten darstellbar durch eine Markov-Kette  $n$ . Ordnung:

$$P(X_{t+1}|X_t, \dots, X_0) = P(X_{t+1}|X_t, \dots, X_{t+1-n})$$

Im Bereich des Machine Learning werden Markov-Ketten gerne dann verwendet, wenn das Verhalten in jüngster Vergangenheit eine höhere Relevanz haben soll als das davorige. So verwendet Rendle u.a [15] Markov-Ketten für seinen Produkt-Rekommender, da er dem Kaufverhalten des Benutzers im direkten Zeitintervall zuvor eine höhere Bedeutung zumisst. Auch in dieser Arbeit werden Markov-Ketten verwendet, um das jüngste Verhalten der Benutzer, i.e. die zuletzt besuchten Orte, im Vorhersagemodell abzubilden.

## 4 FPMC für Next Place Prediction

### 4.1 Faktorisierte Markov-Ketten

Eine faktorisierte personalisierte Markov-Kette (FPMC) bildet die Grundlage für das Vorhersage-Modell. Dieses wiederum basiert auf den faktorisierten Markov-Ketten. Die Wahrscheinlichkeit, dass der Ort  $i$  nach den Orten  $j_1, \dots, j_n$  von einem Benutzer besucht wird, ist demnach:

$$P(i = H_{u,t+1}|j_1 = H_{u,t}, \dots, j_n = H_{u,t+1-n})$$

Für Markov-Ketten erster Ordnung und somit  $n = 1$  und einer benutzerunabhängigen Betrachtung der Wahrscheinlichkeiten ergibt sich die Transitionsmatrix  $A \in [0, 1]^{I \times I}$  mit den Einträgen:

$$a_{j,i} = P(i = H_{u,t+1}|j = H_{u,t})$$

Um nun auch Markov-Ketten höherer Ordnung in einer Transitionsmatrix abbilden zu können, ohne dabei der Matrix weitere Dimensionen hinzuzufügen zu

müssen, werden die  $n$  zuletzt besuchten Orte  $j_1, \dots, j_n$  als eine Entität  $l \in L$  kodiert:

$$l = (j_1, \dots, j_n) \in I^n = L$$

Demnach gilt für die Transitionsmatrix  $A \in [0, 1]^{L \times I}$  für Markov-Ketten höherer Ordnung:

$$a_{l,i} = P(i = H_{u,t+1} | l = (H_{u,t}, \dots, H_{u,t+1-n}))$$

Durch Matrixfaktorisierung eben dieser erhält man zwei Faktormatrizen  $V^{L,I}$  und  $V^{I,L}$ . Die Elemente von  $A$  lassen sich dann folgendermaßen berechnen:

$$a_{l,i} = \langle v_i^{I,L}, v_l^{L,I} \rangle$$

Man spricht hierbei von faktorisierten Markov-Ketten (FMC).

## 4.2 Modell

Das Modell soll den nächsten besuchten Ort  $H_{u,t+1}$  eines Benutzer  $u$  vorher-sagen, wenn dieser zuvor die Orte  $H_{u,t-n+1}, \dots, H_{u,t}$  besucht hat. Dazu wird für jeden Ort  $i$  in  $I$  und  $l \in L$  die Wahrscheinlichkeit dafür berechnet. Durch Erweiterung der Transitionsmatrix aus dem vorherigen Kapitel um eine Benutzerkomponente, ergibt sich somit der Transitionstensor  $\mathcal{X} \in [0, 1]^{|U| \times |L| \times |I|}$ :

$$x_{u,l,i} = p(i = H_{u,t+1} | l = (H_{u,t-n+1}, \dots, H_{u,t}))$$

Als Grundlage wird die Methode Factorized Personalized Markov Chains (FPMC) aus der Arbeit von Stefan Rendle u.a [15] übernommen. Zwar wird die Methode in ihrer Arbeit für Einkaufs-Rekommender-Systeme verwendet, dennoch ähnelt es stark der hier geschilderten Problemstellung und kann bereits durch geringe Anpassungen dafür verwendet werden. So wird in dieser Arbeit die Methode um  $n$ -gradige Markov-Ketten erweitert, sodass auch die Reihenfolge der Orte in der Besuchshistorie im Modell berücksichtigt werden kann. Durch Verwendung von Tensor-Faktorisierung, genauer gesagt die Tucker-Dekomposition (TD), sowie Parallel Factor Analysis (PARAFAC) lässt sich der Übergangs-Tensor durch paarweise Interaktion zwischen den aufspannenden Dimensionen  $U, I^n$  und  $I$  folgendermaßen annähern [15]:

$$x_{u,l,i} = v_u^{U,I} \cdot v_l^{I,U} + v_l^{I,L} \cdot v_i^{L,I} + v_u^{U,L} \cdot v_i^{L,U}$$

Daraus ergeben sich sechs Matrizen, für jede Dimension zwei, welche alle nötigen Werte enthalten und somit im Vergleich zum Tensor  $\mathcal{X}$  später deutlich weniger Speicherplatz benötigen:

- Für die Wechselwirkung zwischen  $U$  und  $I$  enthält  $V^{U,I} \in \mathbb{R}^{|U| \times k_{U,I}}$  die Eigenschaften des Benutzers  $u$  und  $V^{I,U} \in \mathbb{R}^{|I| \times k_{U,I}}$  die des nächsten Ortes  $i$ . Beide zusammen enthalten somit Informationen darüber, welche Orte ein bestimmter Benutzer gerne besucht.

- Für die Wechselwirkung zwischen  $I$  und  $L$  enthält  $V^{I,L} \in \mathbb{R}^{|I| \times k_{I,L}}$  die Eigenschaften des nächsten Ortes  $i$  und  $V^{L,I} \in \mathbb{R}^{|L| \times k_{I,L}}$  die der vorangehenden Orte  $l$ . Durch Kombination der Matrizen erhält man somit Informationen darüber, welcher Ort  $i$  besucht wird, wenn zuvor die Orte  $l$  besucht wurden. Zu beachten ist dabei, dass diese Informationen nicht personalisiert und somit benutzerunabhängig sind.
- Für die Wechselwirkung zwischen  $U$  und  $L$  enthält  $V^{U,L} \in \mathbb{R}^{|U| \times k_{U,L}}$  die Eigenschaften des Benutzers  $u$  und  $V^{L,U} \in \mathbb{R}^{|L| \times k_{U,L}}$  die der vorangehenden Orte  $l$ . Wollte man die hier enthaltenen Informationen deuten, so könnte man sagen, dass die Orte  $l$  enthalten sind, welche ein Benutzer  $u$  gerne vor einem anderen Ort besucht. Daraus lässt sich bereits schon ableiten, dass diese Werte für das Modell nicht interessant sind, da sie keinen Einfluss haben. Bewiesen wird das in der Arbeit von Rendle u.a. [15].

Bei  $k_{U,I}, k_{I,L}, k_{U,L}$  handelt es sich um die Faktorsierungsdimensionen. Deren Werte beeinflussen später das Ergebnis. Je größer diese gewählt werden, desto mehr Werte müssen berechnet werden, was später zu einer höheren Laufzeit des Lernalgorithmus führt.

Um den nächsten besuchten Ort vorherzusagen, wird für jeden möglichen Folgeort  $i \in I$  folgendes berechnet:

$$x_{u,l,i} = v_u^{U,I} \cdot v_l^{I,U} + v_l^{I,L} \cdot v_i^{L,I}$$

Anschließend werden die Orte anhand ihres Wertes  $x$  absteigend geordnet. Da  $V^{U,L}$  und  $V^{L,U}$  die Reihenfolge nicht beeinflussen, können sie aus der Gleichung herausgelassen werden. Der Ort  $i$  mit dem höchstens Wert  $x_{u,l,i}$  wird schließlich für die Vorhersage von  $H_{u,t+1}$  verwendet:

$$H_{u,t+1} = \arg_i \max x_{u,l,i}$$

### 4.3 Lernalgorithmus

Bei dem nachfolgendem Lernalgorithmus handelt es sich um eine Adaption des Sequential Bayesian Personalized Ranking (S-BPR) aus der Arbeit von Rendle u.a. [15]. S-BPR selbst ist ebenfalls Adaption von Bayesian Personalized Ranking (BPR) aus einer früheren Arbeit von Rendle u.a. [14]. Es berechnet eine Näherung und beruht auf dem Gradientenverfahren.

Im ersten Schritt werden die Matrizen  $V^{U,I}, V^{I,U}, V^{I,L}, V^{L,I}$  initialisiert. Dazu wird unter der Verwendung der Gaußschen Normalverteilung jeder Eintrag zufällig gezogen. In jedem weiteren Schritt wird zunächst zufällig ein Eintrag aus dem Datensatz gezogen. Weiterhin wird zufällig ein zweiter Ort  $j$  gewählt. Für jede der Matrizen  $V^{U,I}, V^{I,U}, V^{I,L}, V^{L,I}$  wird der Gradient berechnet und das Gradientenverfahren angewandt. Die Berechnungen haben die Eigenschaft, dass sie konvergieren. Praktisch heißt dies beim Gradientenverfahren, dass sie sich um ein Optimum eependeln.

**Aufstufung 1.1.** Adaption des Sequential Bayesian Personalized Ranking (S-BPR)

```

1  procedure LearnSBPR - FPMC(D)
2      draw  $V^{U,I}, V^{I,U}, V^{I,L}, V^{L,I}$  from  $N(0, \sigma^2)$ 
3      repeat
4          draw  $(u, l, i)$  uniformly from  $D$ 
5          draw  $j$  uniformly from  $I \setminus \{i\}$ 
6           $\delta \leftarrow (1 - \sigma(x_{u,l,i} - x_{x,l,j}))$ 
7          for  $f \in \{1, \dots, k_{U,I^n}\}$  do
8               $v_{u,f}^{U,I^n} \leftarrow v_{u,f}^{U,I^n} + \alpha(\delta(v_{i,f}^{I,U} - v_{j,f}^{I,U}) - \lambda_{U,I}v_{u,f}^{U,I})$ 
9               $v_{i,f}^{I,U} \leftarrow v_{i,f}^{I,U} + \alpha(\delta v_{u,f}^{U,I} - \lambda_{I,U}v_{i,f}^{I,U})$ 
10              $v_{j,f}^{I,U} \leftarrow v_{j,f}^{I,U} + \alpha(-\delta v_{u,f}^{U,I} - \lambda_{I,U}v_{j,f}^{I,U})$ 
11         end for
12         for  $f \in \{1, \dots, k_{I,L}\}$  do
13              $v_{i,f}^{I,L} \leftarrow v_{i,f}^{I,L} + \alpha(\delta v_{l,f}^{L,I} - \lambda_{I,L}v_{i,f}^{I,L})$ 
14              $v_{j,f}^{I,L} \leftarrow v_{j,f}^{I,L} + \alpha(-\delta v_{l,f}^{L,I} - \lambda_{I,L}v_{j,f}^{I,L})$ 
15              $v_{l,f}^{L,I} \leftarrow v_{l,f}^{L,I} + \alpha(\delta(v_{i,f}^{I,L} - v_{j,f}^{I,L}) - \lambda_{L,I}v_{l,f}^{L,I})$ 
16         end for
17     until convergence
18     return  $V^{U,I}, V^{I,U}, V^{I,L}, V^{L,I}$ 
19 end procedure

```

## 5 Umsetzung

### 5.1 Verwendete Werkzeuge

Zur Umsetzung des Algorithmus wird auf die Programmiersprache Python 2.7.10 und dessen Laufzeitumgebung zurückgegriffen. Für mathematische Operationen, wie das Skalarprodukt von Vektoren, wird das Paket numpy 1.12.0rc2 verwendet. Weiterhin wird für die Ermittlung von Mustern innerhalb der Tagestrajektorien einzelner Benutzer das Tool SPMF von Philippe Fournier-Viger verwendet. Es bietet eine große Zahl an verschiedenen Sequential-Pattern-Mining-Algorithmen. Von diesen ist für die Arbeit der Algorithmus SPAM relevant, da er im Vergleich zu beispielsweise PrefixSpan es auch erlaubt eine Mindestlänge erlaubter Sequenzen zu definieren.

### 5.2 Entwurf

Für die Umsetzung des Algorithmus wurde ein objektorientiertes Programmierparadigma gewählt. Dies hat später den Vorteil, dass durch Erstellung mehrerer Objekte der Methode verschiedene Parameter schnell evaluiert werden können. Darüberhinaus werden das Einlesen und Verarbeiten der Daten von eigentlichen Methode dadurch getrennt, dass sie in voneinander unabhängigen Klassen implementiert werden. Die Klasse FPMC enthält die Implementierung des FPMC-Verfahrens. Der Lern-, sowie die Regularisierungsparameter sind hierbei als Attribute modelliert. Der Konstruktor nimmt den Trainings- und Testdatensatz, die

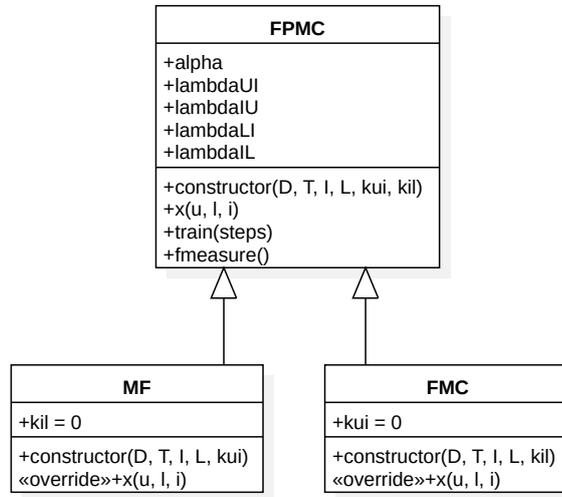


Abb. 1. Klassendiagramm zu den Methoden

beiden Mengen  $I$  und  $L$ , sowie die Faktorisierungsdimensionen  $k_{U,I}$  und  $k_{I,L}$  entgegen. Im Konstruktor werden auch bereits die Matrizen  $V^{U,I}$ ,  $V^{I,U}$ ,  $V^{I,L}$ ,  $V^{L,I}$  initialisiert. Die für die Evaluation relevanten Metriken werden in der Methode `fmeasure` ermittelt. Die Berechnung von  $x_{u,l,i}$  ist einer eigenen Methode gekapselt. Die Verfahren MF und FMC werden ebenfalls durch Klassen modelliert. Mit Ihnen soll FPMC später in der Evaluation verglichen werden. Da FPMC auf MF und FMC basiert, lassen sich diese sehr leicht durch Vererbung modellieren. Lediglich einer der Faktorisierungsdimensionen  $k_{U,I}$  und  $k_{I,L}$  sind in den Konstruktoren der beiden Klassen bereits vordefiniert auf den Wert 0. Außerdem wird die Methode zur Berechnung von  $x_{u,l,i}$  überschrieben. Das Einlesen

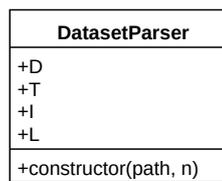


Abb. 2. Die DatasetParser-Klasse

sowie die Konvertierung der Daten in das richtige Format findet in der Klasse `DatasetParser` statt. Der Konstruktor nimmt dabei den Pfad zum Datensatz, sowie den Parameter  $n$ , der angibt, wie viele vorherige Orte betrachtet werden

sollen und somit der Länge der zu verwendenden Markov-Ketten entspricht. Die Daten werden in eine Liste von Tupeln umgewandelt. Diese haben drei Einträge:

- u:** Die ID des Benutzers. Sie entspricht der Position des Elements in der Liste  $U$
- l:** Die ID der zuvor besuchten Orte des Benutzers. Sie entspricht der Position des Elementes in der Liste  $L$
- i:** Die ID des daraufhin besuchten Ortes des Benutzers. Sie entspricht der Position des Elementes in der Liste  $I$

Der Datensatz wird aufgeteilt in einen Trainings- sowie einen Testdatensatz, welche in den Attributen  $D$  und  $T$  abgelegt werden.

## 6 Evaluation

### 6.1 Datensatz

Der Prädiktor dient der Vorhersage des geographischen Verhaltens eines Benutzers. Dementsprechend sollte der Prädiktor mit realistischen Tagestrajektorien von Benutzern trainiert werden. Hierfür gibt es leider keine vorgegebenen, semantisch-annotierte Datensätze. Deshalb wird, wie auch in der Arbeit von Cheng u.a [4], zur Evaluation des Modells stattdessen ein Datensatz verwendet, der das Check-In Verhalten von Benutzern enthält. Foursquare bietet zwei dieser Datensätze an [19]. Die beiden Datensätze enthalten die Check-In-Daten jeweils einer Stadt: New York und Tokyo. In den Check-In-Daten enthalten sind dabei GPS-Koordinaten, der Zeitpunkt, die Ortskategorie (z.B. Office) sowie eine anonymisierte Benutzer-ID, anhand welcher die Check-Ins einzelnen Benutzern zuordenbar sind. In Tabelle 3 sind die wichtigsten Informationen zu den Datensätzen zusammengefasst.

**Tabelle 3.** Rohdaten

Ort	$ U $	$ I $	Check-Ins	∅ Check-Ins pro Benutzer
New York	1083	251	227428	209,00
Tokyo	2293	247	573703	250,20
MIT	77	752	1105004	14350,70

Wie sich später nach Evaluation herausstellt eignen sich die beiden Foursquare-Datensätze nicht sonderlich für die Evaluation des Modell. Deshalb wird noch ein weiterer, dritter Datensatz verwendet. Dabei handelt sich es sich um den MIT-Reality-Mining-Datensatz [6]. Dieser enthält die Tagestrajektorien von 77 MIT Studenten. Dafür wurden die Zeiten, in welchen sich der jeweilige Student beziehungsweise dessen Mobilfunkgerät in einer bestimmten Mobilfunkzelle aufhält, aufgezeichnet. Jede Mobilfunkzelle hat eine eindeutige ID. Weiterhin hat jeder Student jede besuchte Mobilfunkzelle annotiert mit einem Namen oder

der Bedeutung für ihn. Über diese Zuordnungen lässt sich rekonstruieren, wo welcher Student wie lange war. Darüberhinaus enthält der Datensatz eine von allen Studenten ausgefüllte Umfrage, welche hauptsächlich Fragen enthält, die auf die Konsistenz der Tagestrajektorien abzielen. Beispielsweise wird abgefragt, wie konsistent der Student die Daten selbst einschätzt oder, ob dieser oft sein Mobilfunkgerät vergessen hatte und ob dieses manchmal ausgeschaltet wurde. Die Ortsangaben der Studenten ist leider nicht vergleichbar mit den einheitlichen semantischen Annotationen des Foursquare-Datensatzes. So sind diese nicht übergreifend konsistent, da der gleiche Ort von unterschiedlichen Studenten unterschiedlich benannt werden kann. Zudem ist nicht innerhalb der Angaben eines Studenten Einheitlichkeit gegeben, da auch hier der gleiche Ort u.U. unterschiedlich benannt wurde. Dies äußert sich in komplett unterschiedlichen Benennungen, Abkürzungen und Schreibfehlern. Die eigentliche Semantik des Ortes ist weiterhin oft nicht klar, da die Studenten stellenweise nur den Straßennamen angeben oder eigenwillige Abkürzungen verwenden.

## 6.2 Vorverarbeitung des Datensatzes

Da die Foursquare-Datensätze lediglich freiwillige Check-Ins von Benutzern enthalten, sind diese sehr lückenhaft. Nur sehr wenige Benutzer besitzen vollständige Tagestrajektorien, und es wurden oft nur Orte eingetippt, welche für die Benutzer selbst von Interesse sind und welche sie öffentlich sichtbar machen wollen. Dementsprechend weisen viele Benutzer kein durchgängiges Verhalten auf und in ihren Daten ist kein Muster erkennbar.

Aus diesem Grund wird vor der Verwendung im Lernalgorithmus der Datensatz in einem mehrstufigem Verfahren gefiltert (siehe Abbildung 3). Ziel ist es Benutzer mit möglichst vollständigen Tagestrajektorien zu extrahieren.

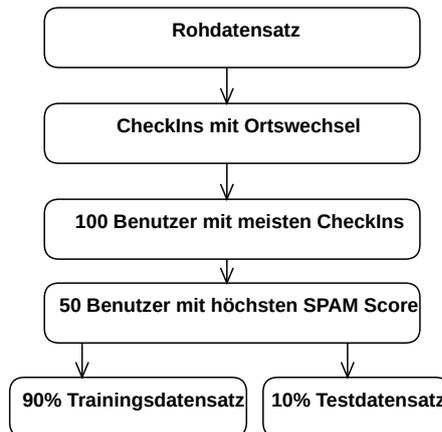


Abb. 3. Mehrstufiges Filterverfahren

Da für den Prädiktor lediglich der Ortswechsel interessant ist, werden zunächst im ersten Schritt an einem Tag mehrfach direkt hintereinander auftretende Check-Ins am gleichen Ort aussortiert. Im zweiten Schritt werden die Benutzer nach der Anzahl ihrer verbleibenden Check-Ins absteigend sortiert. Die einhundert Benutzer mit den meisten Check-Ins werden schließlich im zweiten Schritt auf Verhaltensmuster untersucht.

Hierzu wird der Algorithmus SPAM verwendet. Dieser analysiert Sequenzen auf Subsequenzen und überprüft, wie oft diese wiederkehren. Treten diese Subsequenzen in einem festgelegtem Prozentsatz der Sequenzen auf, so gelten diese als ein Verhaltensmuster. SPAM kann hierdurch einen Score ermitteln, welcher der Anzahl der bei einem Benutzer erkannten Verhaltensmuster entspricht.

Die Tagestrajektorien eines jeden Benutzers werden somit mit SPAM analysiert. Schließlich werden die Benutzer absteigend nach ihrem SPAM-Score sortiert. Die fünfzig Höchsten bilden mit ihren Check-Ins nun den Datensatz für die Evaluation. Die Informationen zu den gefilterten Daten sind in der Tabelle 4 zusammengefasst.

**Tabelle 4.** Nach Filterung der Daten

Datensatz	$ U $	$ I $	Check-Ins	∅ Check-Ins pro Benutzer
New York	50	208	28574	571,48
Tokyo	50	186	29221	584,42
MIT	26	317	14741	566.96

Die Filterung des MIT-Datensatzes verläuft ähnlich. So werden ebenfalls nur Ortswechsel betrachtet. Da außerdem auch die Aufenthaltsdauer bekannt ist, werden lediglich Aufenthalte verwendet, welcher länger als 5 Minuten anhalten. Dadurch werden auch Fehlmessungen herausgefiltert, da es öfters vorkommt, dass Mobilfunkgeräte die falsche Mobilfunkzelle und somit den falschen Ort erkennen. Darüberhinaus werden nur die Benutzer betrachtet, welche bei der Umfrage angaben, dass sie sehr ausführliche Tagestrajektorien aufgezeichnet haben. Da sich nach diesen Einschränkungen bereits die Zahl der potentiellen Benutzer auf 26 reduziert, ist es nicht mehr nötig die beiden weiteren Filterschritte, welche die Anzahl der Check-Ins und die SPAM-Scores betrachten, auszuführen. Stattdessen wurden alle 26 Benutzer für den Trainings- und Testdatensatz verwendet.

### 6.3 Evaluationsmetriken

Zur Evaluation wird jeder Datensatz in zwei Teile aufgeteilt: Einen Trainings- und einen Testdatensatz. Mit dem Trainingsdatensatz wird der Lernalgorithmus durchgeführt und mit dem Testdatensatz werden anschließend die Vorhersagen überprüft. Bei der Aufteilung wird so vorgegangen, dass die Check-In-Historie jedes Benutzers einzeln betrachtet wird. Die 10% neusten Check-Ins landen dabei im Testdatensatz und die restlichen 90% Check-Ins im Trainingsdatensatz.

Für jeden Ort  $i \in I$  wird schließlich berechnet, wie gut dieser vorhergesagt wird. Dazu werden False-Positives, Positives und False-Negatives ermittelt. Als Vorhersage-Güte dient hierbei der *f1score*, welcher sich aus Precision und Recall ableiten lässt:

$$recall_i = \frac{positives_i}{positives_i + falsenegatives_i}$$

$$precision_i = \frac{positives_i}{positives_i + falsepositives_i}$$

$$f1score_i = 2 * \frac{precision_i * recall_i}{precision_i + recall_i}$$

Der *f1score* des Prädiktors im Gesamten wird schließlich durch Mittelung der Werte ermittelt:

$$f1score = \frac{\sum_{i \in I} f1score_i}{|I|}$$

Als weitere Evaluationsmetrik dient der benutzergewichtete *f1score*. Hierfür wird für jeden Benutzer einzeln jeweils ein *f1score* berechnet, indem nur seine Check-Ins aus dem Testdatensatz verwendet werden, dennoch aber der gesamte Trainingsdatensatz. Schließlich wird über all diese Werte gemittelt. Im Gegensatz dazu wird weiterhin auch für jeden Benutzer berechnet, wie sich der *f1score* verhält, wenn der Benutzer ein eigenes Modell erhält, also nur seine Check-Ins aus dem Trainingsdatensatz verwendet werden. Auch diese Werte werden gemittelt. Aus der Differenz zwischen Gesamtmodell und Individualmodell lässt sich ablesen, wie stark die Vorhersagequalität zwischen beiden sich unterscheidet.

#### 6.4 Parameter

Während der Evaluation werden verschiedene Variationen an Parameter getestet. Diese wurden wie folgt gewählt:

- $k_{I,L} = k_{U,I} \in \{8, 16, 32, 64, 128\}$
- $n \in \{1, 2, 3\}$

Weiterhin werden neben der FPMC-Methode auch MF und FMC mit den beiden Datensätzen getestet. Alle 100.000 Iterationen des Lernalgorithmus wurden Precision, Recall und F1-Score ermittelt. Sobald bei diesen eine Konvergenz sichtbar wird, wird das Verfahren abgebrochen.

#### 6.5 Ergebnis

Nach einer Laufzeit von etwa 3 Tagen wurden alle Ergebnisse berechnet. Dabei wurden stets fünfzig Sätze à 100.000 Schritte des Lernalgorithmus durchgeführt. Wie in Abbildung 5 zu sehen ist, erreicht der *f1score* von FPMC nach etwa zehn Sätzen seinen Höchstwert. In den weiteren vierzig Sätzen nimmt der Wert

stetig ab. Bei höheren Werten von  $n$  ist keine Verbesserung bei den Evaluationsmetriken erkennbar. Deshalb wurde auf die Darstellung der Ergebnisse bei  $n > 1$  verzichtet. Während der Evaluation wurden weiterhin die Auswirkung der Faktorisierungsparameter  $k_{I,L}$  und  $k_{U,I}$  getestet. So ist in den Abbildungen 4 und 5 erkennbar, dass ein höherer Wert gleichzeitig auch zu einem höheren  $f1score$  führt, das Wachstum aber sehr schnell abflacht. Gleiches gilt auch bei den Verfahren MF und FMC. Gleichzeitig ist auch klar erkennbar, dass FPMC bei durchgängig einen größeren  $f1score$  als die anderen Verfahren aufweist.

**Tabelle 5.**  $f1score$  der einzelnen Verfahren mit  $k_{I,L} = k_{U,I} = 128$  und  $n = 1$

Datensatz	FPMC	MF	FMC
New York	0,1116	0,07414	0,02959
Tokyo	0,15522	0,09473	0,04955
MIT	0,33837	0,22093	0,2580

In Tabelle 5 sind die jeweiligen  $f1score$ -Höchstwerte der Verfahren nochmals zusammengefasst.

So zeigt sich, dass bei beiden Foursquare-Datensätzen FPMC einen mehr als dreifach so großen  $f1score$  aufweist als FMC. Gegenüber der Matrixfaktorisierung zeigt sich immerhin eine Verbesserung von über 50%. Die Ergebnisse zu FMC und MF geben aber auch Aufschluss über die verwendeten Datensätze. MF betrachtet lediglich die Vorlieben der Benutzer, während FMC-Verfahren benutzerunabhängig betrachtet, welcher Ort meist auf welchen anderen folgt. Da MF bessere Werte aufweist als FMC, kann man daraus schließen, dass diese Informationen bei den verwendeten Datensätzen tatsächlich wichtiger sind und zu besseren Ergebnissen führen. Betrachtet man die Foursquare-Datensätze genauer, erkennt man auch, dass die meisten Benutzer nur eine geringe Anzahl von unterschiedlichen Orten aufsuchen. Diese sind allerdings zwischen den Benutzern unterschiedlich. Hier ist MF klar im Vorteil, da für jeden Benutzer dessen Vorlieben speichert, während FMC für jeden Benutzer die gleiche Vorhersage trifft. Analysiert man desweiteren die ermittelten SPAM-Scores der einzelnen Benutzer, so stellt man fest, dass diese nur sehr weniger Muster aufweisen (siehe Tabelle 7). Dies ist eine weitere schlechte Voraussetzung für FMC und erklärt, warum dieses schlechter abschneidet. Da allerdings FPMC trotzdem noch besser abschließt als MF und FMC, obwohl es eine Kombination der beiden Verfahren ist, lässt sich daraus schlussfolgern, dass sich diese Verfahren gut ergänzen und Vorhersageschwächen von MF mit FMC ausgleichen lassen und andersherum.

Dagegen konnten mit dem MIT-Reality-Mining-Datensatz deutlich bessere Ergebnisse erzielt werden, wie in Tabelle 6 ersichtlich ist. Auch bei diesem schneidet FPMC wesentlich besser ab, als die beiden anderen Verfahren. Allerdings besitzt im Gegensatz zu den Foursquare-Datensätzen FMC hier einen höheren  $f1score$  als MF. Somit ist es bei diesem Datensatz wichtiger die Verhaltensmuster des Benutzers zu kennen, als dessen Vorlieben. Dies spiegelt sich auch den

**Tabelle 6.** Messwerte mit MIT-Datensatz mit  $k_{I,L} = k_{U,I} = 128$  und  $n = 1$ 

Datensatz	FMC	MF	FPMC
accuracy	0,4247	0,3546	0,5280
recall	0,1849	0,1596	0,2593
precision	0,5350	0,3587	0,4867
<i>f1score</i>	0,2748	0,2209	0,3384

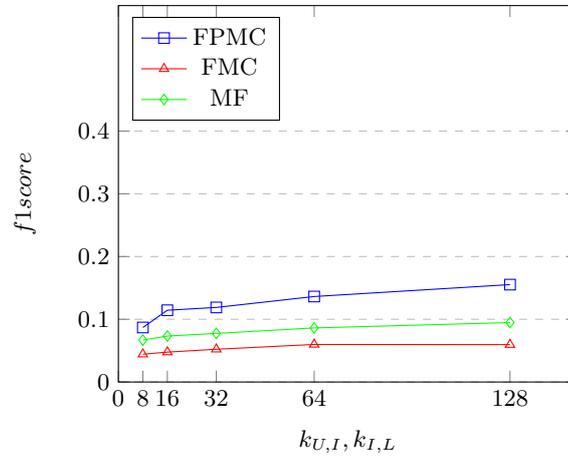
**Tabelle 7.** Durchschnittliche SPAM-Scores der Benutzer der gefilterten Datensätze bei verschiedenen Minimum-Support-Werten. Der Wert gibt an, wie oft (prozentual) eine Sequenz vorkommen muss, damit sie gezählt wird.

Datensatz	10%	30%	50%
New York	30	1	0
Tokyo	190	4	0
MIT	66881	1400	119

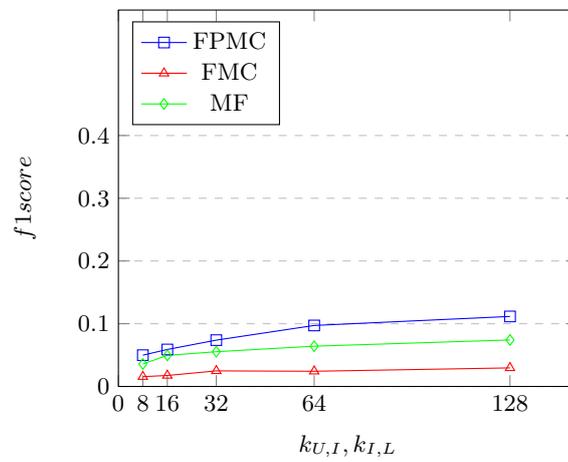
SPAM-Scores der Benutzer wider (siehe Tabelle 7) ersichtlich, welche erheblich höher als bei den Foursquare-Datensätzen ausfallen und somit bescheinigen, dass die Benutzer deutlich konsistenter in ihrem Verhalten sind. Darüberhinaus zeigt sich auch beim MIT-Datensatz, dass sich MF und FMC gut ergänzen.

Weiterhin wurde untersucht, wie sich die Vorhersagequalität zwischen einem Gesamtmodell und Individualmodellen unterscheidet. Hierbei wurde der MIT-Datensatz verwendet, da dieser sich bereits zuvor als geeigneter erwies.

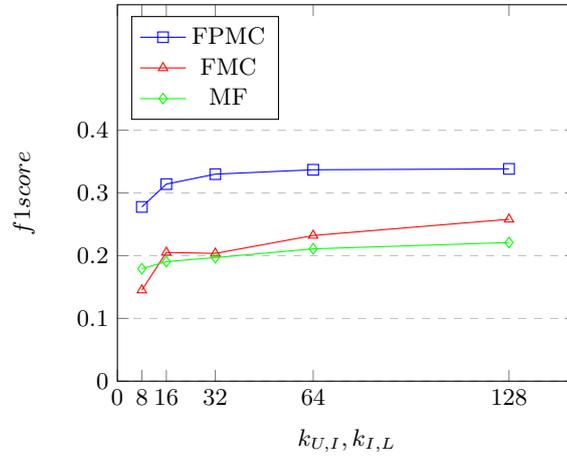
Wie in Abbildung 8 ersichtlich, schneidet das Gesamtmodell, wie erwartet, schlechter ab. Dennoch ist die Differenz geringer als erwartet, immerhin arbeitet der FMC-Teil des FPMC benutzerunabhängig, trotzdem muss das Gesamtmodell in diesem das Verhalten aller Benutzer abbilden. Individualmodelle können im Gegensatz dazu auch hier benutzerspezifisch arbeiten, weshalb hier eine höhere Vorhersagequalität zu erwarten ist. Allerdings benötigen diese auch wesentlich mehr Speicherplatz und Rechenzeit beim Trainieren. Insgesamt ist auch interessant, dass der benutzergewichtete *f1score* wesentlich größer ausfällt, als der zuvor ermittelte ungewichtete. Insbesondere da dieser in möglichen Anwendungen relevanter ist, vor allem, wenn diese benutzerzentrisch arbeiten.



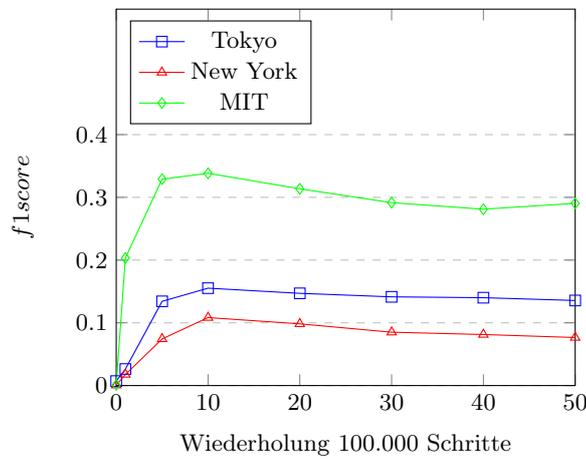
**Abb. 4.**  $f1score$  bei verschiedenen Werten von  $k_{U,I}$  und  $k_{I,L}$  und Datensatz Tokyo sowie  $n = 1$



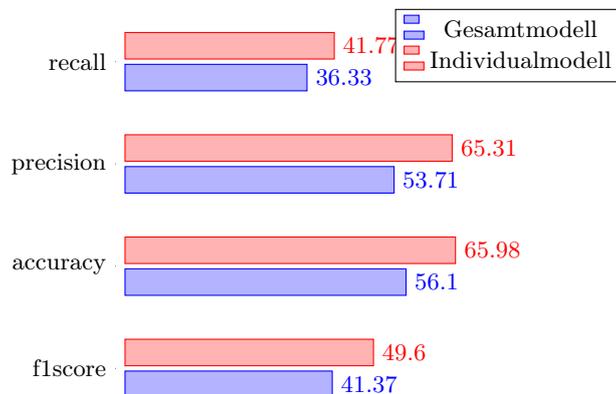
**Abb. 5.**  $f1score$  bei verschiedenen Werten von  $k_{U,I}$  und  $k_{I,L}$  und Datensatz New York sowie  $n = 1$



**Abb. 6.**  $f1score$  bei verschiedenen Werten von  $k_{U,I}$  und  $k_{I,L}$  und Datensatz MIT Reality Mining sowie  $n = 1$



**Abb. 7.** Entwicklung des  $f1score$  während der Lernphase mit  $k_{I,L} = k_{U,I} = 128$  und  $n = 1$



**Abb. 8.** Vergleich zwischen Individualmodell und Gesamtmodell in Prozent mit  $k_{I,L} = k_{U,I} = 128$  und  $n = 1$ .  $f1score$ ,  $accuracy$ ,  $precision$  und  $recall$  sind benutzergewichtet.

## 7 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde ein semantischer Next-Place-Prädiktor entwickelt, basierend auf dem angepassten FPMC-Verfahren. Dieses wurde zusammen mit den beiden anderen Verfahren MF und FMC anhand von drei Datensätzen evaluiert. Es wurde festgestellt, dass die Prädiktion beim FPMC-Verfahren deutlich besser abschneidet als die beiden anderen Verfahren. Weiterhin hat sich herausgestellt, dass die beiden Foursquare-Datensätze sich nicht für die Lokationsprädiktion eignen. Die ermittelten  $f1scores$  liegen unter den Erwartungen. Dies lässt sich darauf zurückführen, dass die Tagestrajektorien unvollständig sind und weiterhin nur wenige Muster aufweisen. Zwar wurden diese vorgefiltert und Benutzer die wenige Muster aufweisen aussortiert, dennoch weisen die verbliebenen Benutzer gemessen an ihrer Zahl an Check-Ins immernoch wenige Muster auf. Aufgrund dessen wurde auf den MIT-Reality-Mining-Datensatz zurückgegriffen. Dieser weist zwar nahezu vollständige Tagestrajektorien auf, besitzt allerdings keine einheitliche Semantik. Dies hat zur Folge, dass der selbe Ort mehrfach vorkommt. Wegen teilweise kryptischer Namensgebung lässt sich auch keine konsistente Semantik konstruieren, die dem entgegen wirkt. Trotzdem lassen sich mit dem Datensatz deutlich höhere  $f1scores$  erzielen und auch hier schneidet FPMC besser als die beiden Vergleichsverfahren ab. Auch die Differenz der Vorhersagequalität zwischen einem Gesamtmodell und Individualmodellen ist sehr gering, obwohl das Gesamtmodell wesentlich weniger Speicherplatz benötigt.

Insgesamt zeigt sich anhand der Ergebnisse, dass Matrixfaktorisierung allein zwar nicht für die Lokationsprädiktion geeignet ist, allerdings bisherige Modelle, die z.B. auf Markov Ketten basieren, sinnvoll erweitern kann und so die Vorher-

sagequalität verbessern kann.

Im nächsten Schritt wäre es sinnvoll das Verfahren noch mit anderen Datensätzen zu testen, welche sowohl vollständige, wie auch einheitlich semantisch annotierte Tagestrajektorien enthalten. Weiterhin wäre es hierbei auch interessant die Semantik weiter zu generalisieren. Das heißt, zu den semantischen Orten Kategorien zu finden, um diese zusammenzufassen, so dass beispielsweise alle Restaurants zu einer gleichnamigen Kategorie zusammengefasst werden. Hierdurch würde sich voraussichtlich die Anzahl ähnlicher Muster zwischen Benutzern erhöhen, aber auch innerhalb der Tagestrajektorien eines Benutzers würden mehr Muster erkennbar werden. Dadurch ließe sich eine höhere Treffer-Quote bei der Prädiktion erzielen.

## Literatur

1. AGOF: Anzahl der Unique User von TripAdvisor.de von November 2014 bis Juni 2016 (in Millionen) (Juni 2016), <https://de.statista.com/statistik/daten/studie/418697/umfrage/online-besucherzahlen-von-tripadvisor-als-zeitreihe/>
2. Albert Au Yeung: Matrix Factorization: A Simple Tutorial and Implementation in Python (September 2010), <http://www.quuxlabs.com/blog/2010/09/matrix-factorization-a-simple-tutorial-and-implementation-in-python/>
3. Ashbrook, Daniel and Starner, Thad: Learning significant locations and predicting user movement with GPS. In: Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on. pp. 101–108. IEEE (2002)
4. Cheng, Chen and Yang, Haiqin and Lyu, Michael R and King, Irwin: Where You Like to Go Next: Successive Point-of-Interest Recommendation. In: IJCAI. vol. 13, pp. 2605–2611 (2013)
5. comScore MobiLens: Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2016 (in Millionen) (Juni 2016), <https://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonennutzer-in-deutschland-seit-2010/>
6. Eagle, N., Pentland, A.S.: Reality mining: sensing complex social systems. vol. 10, pp. 255–268. Springer (2006)
7. eMarketer: Prognose zur Anzahl der Smartphone-Nutzer weltweit von 2012 bis 2020 (in Milliarden) (Juni 2016), <https://de.statista.com/statistik/daten/studie/309656/umfrage/prognose-zur-anzahl-der-smartphone-nutzer-weltweit/>
8. Gambs, Sébastien and Killijian, Marc-Olivier and del Prado Cortez, Miguel Núñez: Next place prediction using mobility markov chains. In: Proceedings of the First Workshop on Measurement, Privacy, and Mobility. p. 3. ACM (2012)
9. Gao, Huiji and Tang, Jiliang and Liu, Huan: Mobile location prediction in spatio-temporal context. In: Nokia mobile data challenge workshop. vol. 41, p. 44 (2012)
10. Koren, Y.: The bellkor solution to the netflix grand prize. Netflix prize documentation 81, 1–10 (2009)
11. Koren, Yehuda and Bell, Robert and Volinsky, Chris and others: Matrix factorization techniques for recommender systems. Computer 42(8), 30–37 (2009)

12. Liu, Qiang and Wu, Shu and Wang, Liang and Tan, Tieniu: Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. In: AAAI. pp. 194–200 (2016)
13. Netflix: Netflix Prize (Oktober 2006), <http://www.netflixprize.com>
14. Rendle, Steffen and Freudenthaler, Christoph and Gantner, Zeno and Schmidt-Thieme, Lars: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. pp. 452–461. AUAI Press (2009)
15. Rendle, Steffen and Freudenthaler, Christoph and Schmidt-Thieme, Lars: Factorizing personalized markov chains for next-basket recommendation. In: Proceedings of the 19th international conference on World wide web. pp. 811–820. ACM (2010)
16. Statista: Wie häufig verwenden Sie diese Location-based Services mit Ihrem Smartphone/Tablet? (August 2015), <https://de.statista.com/statistik/daten/studie/609350/umfrage/nutzungsfrequenz-von-location-based-services-in-deutschland/>
17. Statistisches Bundesamt: Einwohnerzahl - Anzahl der Einwohner von Deutschland von 1990 bis 2015 (in Millionen) (August 2016), <https://de.statista.com/statistik/daten/studie/2861/umfrage/entwicklung-der-gesamtbevoelkerung-deutschlands/>
18. TripAdvisor: Umsatz von TripAdvisor weltweit vom 1. Quartal 2011 bis zum 3. Quartal 2016 (in Millionen US-Dollar) (November 2016), <https://de.statista.com/statistik/daten/studie/543148/umfrage/umsatz-von-tripadvisor-weltweit-nach-quartalen/>
19. Yang, Dingqi and Zhang, Daqing and Zheng, Vincent W and Yu, Zhiyong: Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. vol. 45, pp. 129–142. IEEE (2015)
20. Ye, Jihang and Zhu, Zhe and Cheng, Hong: What’s your next move: User activity prediction in location-based social networks. In: Proceedings of the SIAM International Conference on Data Mining. SIAM. SIAM (2013)

# Artificial Neural Networks for Semantic Location Prediction

Adrian Jablonski\*

Advisor: Antonios Karatzoglou<sup>†</sup>

Karlsruhe Institute of Technology (KIT)  
Pervasive Computing Systems – TECO

\*adrian.jablonski@student.kit.edu

<sup>†</sup>antonios@teco.edu

**Abstract.** A user's location and movement is an important part of the relevant context. Correct prediction of the next visited location allows various applications, especially for the ubiquitous computing environment. For this purpose, a suitable machine learning approach is needed. In this paper, we focus on the applicability of artificial neural networks (ANNs) to this problem - in particular - to the prediction of semantic locations. Firstly, we introduce basic theoretical background concerning ANNs. Secondly, we give an overview on existing approaches in this area of study. Then, we implement our own ANN-based location prediction approach and evaluate it using a real-life dataset.

**Keywords:** Artificial Neural Network, Location Prediction, Long Short-Term Memory, Machine Learning, Context

## 1 Introduction

Location prediction deals with the problem of determining the next location  $l_{n+1}$  of a (moving) user  $u$  based on the user's previous trajectories  $t_u = (l_0, \dots, l_n)$ . The term location can mean both (geographical) coordinates and semantic location labels, e.g., "Restaurant" or "University". Many systems require or benefit from prior knowledge of the user's location to prepare the environment or allocate necessary resources for the user. In particular, ubiquitous systems that are designed for actively assisting users such information can be crucial if location is a critical part of the relevant context. Another common use case for location prediction are mobile communication networks, where the predicted locations are used to minimize the signaling overhead and thus allowing faster handoffs between base stations. Location prediction can be implemented using various statistical modeling techniques such as Markov Chains, Matrix Factorization or (Artificial) Neural Networks (ANNs). Different types of neural networks have shown promising results for particular use cases. Especially since the advance of Deep Learning, utilizing neural networks for location prediction has become a hot topic of contemporary research interest.

In this paper, we will focus on neural networks for predicting location using semantic labels which can be used in ubiquitous computing appliances. First, we explain basic theory of ANNs necessary to understand their application to location prediction. The following section presents recent findings in research in that area. In section 4, we choose a promising approach and implement it in order to evaluate its performance on semantically labeled location histories.

## 2 Artificial Neural Networks

Artificial Neural Networks (ANNs) [1][2] are a statistical modeling approach to "learn" a probability distribution over a set of data samples. ANNs are inspired by natural neural networks that can be found in animal or human brains. An ANN consists of a set of interconnected units that generate an output value  $\mathbf{y} \in \mathbb{R}^n$  based on an input value  $\mathbf{x} \in \mathbb{R}^m$ . The strength of the connections (*weight*) is determined by a process called *training*, where the ANN "learns" to output a desired value given its corresponding input value. Then, the trained network can be used for classification or regression problems using new, unknown input values.

### 2.1 The perceptron

One of the simplest units used for ANNs are *perceptrons*. A perceptron [1][2] consists of the following parts: An input vector  $\mathbf{x} \in \mathbb{R}^n$ , a weight vector  $\mathbf{w} \in \mathbb{R}^n$ , an (optional) bias value  $b \in \mathbb{R}$  and the *activation function*  $o(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ , for example:

$$o(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w}\mathbf{x} - b > 0 \\ -1 & \text{else} \end{cases} \quad (1)$$

Thus, given the weight and bias values, a perceptron allows it to separate (classify) the input space into two parts along a hyperplane. As a result, a perceptron allows it to represent the logical functions AND, OR, NAND and NOR, but not for example XOR, since this would require to separate the hyperplane into four distinct parts.

To determine the weight values of a perceptron, the perceptron - analogous to a natural neuron - has to be trained. For this purpose, we define a *training set*  $S = \{(\mathbf{x}_1, l_1), \dots, (\mathbf{x}_n, l_n)\}$  consisting of the input vectors and their corresponding class labels (1 or -1). Furthermore, for notational simplification, we define  $x_0 = 1$  so that we do not require a separate bias value. Then we initialize  $\mathbf{w}$  with random values and apply the following algorithm called *perceptron training rule* with  $\eta$  being the *learning rate*:

The training rule adjusts the weights for each element of the training set until there are no elements that are being classified incorrectly by the perceptron. This simple algorithm only converges if the training set is linearly separable. To allow learning sets that are not linearly separable, another training rule, the *delta rule*,

```

while  $\exists(\mathbf{x}, l) \in S : o(\mathbf{x}) \neq l_i$  do
  | foreach  $(\mathbf{x}_i, l_i) \in S$  do
  | |  $\mathbf{w} \leftarrow \mathbf{w} - \eta \mathbf{x}_i (l_i - o(\mathbf{x}_i))$ 
  | | end
  | end
end

```

**Algorithm 1:** Perceptron training rule

can be used to learn a best-fit approximation. The delta rule applies a technique called *gradient descent*: First, we define a *loss function*  $E$  to determine the error between the target and current classification obtained from the given weights. A common function for this purpose is

$$E(\mathbf{w}) = \frac{1}{2} \sum_{(\mathbf{x}_i, l_i) \in S} (l_i - o(\mathbf{x}_i))^2 \quad (2)$$

Then, the delta rule consists of determining the gradient  $\nabla E(\mathbf{w})$  and applying (3) until the classification error is minimized:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \nabla E(\mathbf{w}) \quad (3)$$

Since the negative gradient vector always points towards the steepest descent, the delta rule is guaranteed to converge to the local minimum of the loss function. Furthermore, the delta rule requires the loss function to be differentiable. A simple example of such a function is the activation function  $o(\mathbf{x}) = \mathbf{w}\mathbf{x}$ . A perceptron using this activation function is also called *linear unit*. Some other commonly used activation functions are the sigmoid function

$$o(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}\mathbf{x}}} \quad (4)$$

and the hyperbolic tangent  $o(\mathbf{x}) = \tanh(-\mathbf{w}\mathbf{x})$ , which exhibit similar behavior to the step activation function from (1) with the difference of being differentiable.

## 2.2 Multilayer Perceptrons

To overcome the limitation of a single perceptron, several of such units can be combined into a network called *multilayer perceptron* (MLP) [1][2]. A MLP consists of  $N \geq 2$  layers of  $k_i$  ( $1 \leq i \leq n$ ) units with the outputs  $o_{j,i}$  ( $1 \leq j \leq K_i$ ) of layer  $i$  being the inputs of layer  $i + 1$  weighted with  $w_{j,i+1}$  ( $1 \leq j \leq K_{i+1}$ ). Layer 1 is called *input layer*, layer  $N$  is called *output layer*, and layers 2 to  $N - 1$  are called *hidden layers*. The units on the input layer are passive nodes, merely distributing the input values to the next layer. Furthermore, every node in layer  $N$  is required to be connected to a node in layer  $N + 1$ . Thus, a MLP can be interpreted as a directed acyclic graph with its nodes being the neuron units and the edges being the weighted "data flow" from one unit to another. A sample MLP with three layers is shown in Figure 1.

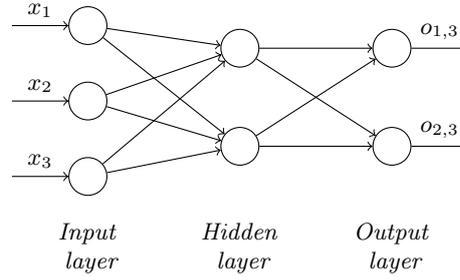


Fig. 1: Visualization of a sample three-layer MLP

A MLP allows it to learn arbitrary decision boundaries as opposed to the single perceptron unit with linear separation. To train a MLP, we use a generalization of the delta rule to multiple layers called *backpropagation*. First, since a MLP can have more than only one output node, it is necessary to introduce a different type of loss function that considers every output  $o_{j,N}$  of the network. Given a training set  $S = \{(\mathbf{x}_1, \mathbf{o}_1), \dots, (\mathbf{x}_n, \mathbf{o}_n)\}$  consisting of pairs of input and output vectors and the randomly initialized weight vector of the output layer  $\mathbf{w} = (w_{1,N} \dots w_{K_i,N})^T$ , we define the loss function as

$$E(\mathbf{w}) = \frac{1}{2} \sum_{(\mathbf{x}_i, \mathbf{o}_i) \in S} \sum (\mathbf{o}_i - o_N)^2 \quad (5)$$

where  $o_i = (o_{1,i} \dots o_{K_i,i})$  is the vector of all outputs  $o_{j,i}$ . Backpropagation consists of applying a modified version of the gradient descent method called *stochastic gradient descent* that updates the weight vectors after considering each training sample from  $S$ . This approach allows faster convergence to a local minimum of the loss function. The backpropagation algorithm works as follows:

```

Randomly initialize  $\mathbf{w}_i \forall i \in \{1, \dots, N\}$ ;
do
  foreach  $(\mathbf{x}_i, \mathbf{o}_i) \in S$  do
    Calculate  $o_N$  by propagating  $\mathbf{x}_i$  through the network;
    foreach  $l \in \{2, \dots, N\}$  do
       $\delta_l \leftarrow \begin{cases} o_N(1 - o_N)(\mathbf{o}_i - o_N) & \text{if } l = N \\ o_l(1 - o_l) \sum_{j \leq K_{l+1}} w_{j,l+1} \delta_{j,l+1} & \text{if } l < N \end{cases}$ 
    end
    foreach  $l \in \{2, \dots, N\}$  do
       $\mathbf{w}_l \leftarrow \mathbf{w}_l + \eta \delta_l \mathbf{x}_i$ 
    end
  end
while  $E(\mathbf{w}_N) > \varepsilon$ ;

```

**Algorithm 2:** The backpropagation algorithm

On each training sample, backpropagation first calculates the network output, then determines the error term  $\delta_l$  for each layer. These error terms can be derived from the loss function gradient  $\nabla E(\mathbf{w})$ . Analogous to the delta rule, the error terms, together with the learning rate  $\eta$ , are used to update the weight vectors of the network. The backpropagation algorithm is repeated until either the error falls below a threshold  $\varepsilon$  or until a defined number of iterations, also called *epochs*.

In order to improve its convergence behavior, numerous improvements to the backpropagation algorithm<sup>1</sup> have been proposed, such as the *momentum method* [3] that alters the weight-update rule as follows, with  $\Delta\mathbf{w}_l(n)$  being the weight update of layer  $l$  during iteration  $n$ :

$$\Delta\mathbf{w}_l(n) = \eta\delta_l\mathbf{x}_i + \alpha\Delta\mathbf{w}_l(n-1) \quad (6)$$

$$\mathbf{w}_l \leftarrow \mathbf{w}_l + \Delta\mathbf{w}_l(n) \quad (7)$$

The parameter  $\alpha$  ( $1 \leq \alpha < 1$ ) is called *momentum* and makes the current weight update dependent on the corresponding update during the previous iteration. Inspired by the physical phenomenon, the momentum method prevents harsh changes in direction during gradient descent, thus keeping the gradient move further towards a local minimum.

### 2.3 Encoding and Representation

The inputs and outputs to an ANN may be various, such as real valued data or class labels. To represent those data in an appropriate way, several methods of encoding may be applied [1][2]. Raw real valued data may usually be fed into a network without encoding, since the learned weights will result in (approx.) the desired values. On the other hand, multi-class classification requires an adequate representation of each class for usage with the neural network. A common approach is assigning each class a class identifier  $id \in \mathbb{N}$  that will be used as network input or output. In this case, each input or output neuron may represent a single bit of the binary encoding of that identifier. For example, given class identifiers ranging from 1 to  $N$ ,  $\lceil \log_2(N) \rceil$  neurons are required to represent these identifiers in the resulting network.

Another method for this purpose is using *one-hot* binary encoding. To one-hot encode an integer  $n \in \mathbb{N}$ , a binary string of length  $m$ ,  $n \leq m$  consisting of all zeros except at index  $n$  is created. For example, given  $n = 3$  and  $m = 5$ , the one-hot representation of  $n$  is 00100. Analogously to the binary encoding, one neuron for each binary digit is required for the network. The advantage of one-hot encoding is the possibility of obtaining a probability distribution from the network. This requires using the *softmax* function as the activation function

<sup>1</sup> Note that backpropagation with (stochastic) gradient descent is merely one (although common) out of several methods of training multilayer perceptrons.

of the network. The softmax function [4] is applied to the outputs of the network and assigns a probability value to each of the outputs  $o_{i,N}$ :

$$\text{softmax}(o_N)_i = \frac{\exp(o_{i,N})}{\sum_{j=1}^{K_N} \exp(o_{j,N})}, \quad i \in \{1, \dots, K_N\} \quad (8)$$

When using the softmax activation function, a different loss function is more suitable for training the network than the function described in 2.2. The *cross-entropy* or *log-loss* function [5] has been shown resulting in maximum likelihood probability estimates of a network trained using this function, defined as ( $\mathbf{t}$  being the classification target):

$$E(\mathbf{t}) = - \sum_{j=1}^{K_N} \mathbf{t}_j \log(o_{j,N}) \quad (9)$$

Using softmax in combination with cross-entropy loss has become a common approach for training artificial neural networks for multi-class classification problems [6].

## 2.4 Recurrent Neural Networks

The output  $o(\mathbf{x})$  of a multi-layer perceptron is only dependent on the current input vector  $\mathbf{x}_n$  ( $n \in \mathbb{N}$ ). However, classification problems such as predicting the next item  $i_{n+1}$  given a sequence of items  $S = (i_1, \dots, i_n)$  would benefit if the next output  $o_{n+1}$  of the network would be determined by the current network state as well. For this purpose, so called *recurrent neural networks*<sup>2</sup> (RNNs) have been proposed. A RNN [1] is an artificial neural network, whose units in layer  $n$  are not only connected to units in layer  $n + 1$ , but may also be connected to any previous layer  $i$  ( $1 \leq i < n$ ). Several basic types of such RNNs have been proposed, such as *Jordan* and *Elman* networks. A Jordan network [7] is an RNN with connections between only the output layer and the input layer, whereas an Elman network [8] has its hidden layers used as input (see Figure 2).

Because of the cyclic dependence from the outputs, a RNN cannot be trained using the standard backpropagation algorithm presented in section 2.2. Instead, a modified version called *Backpropagation Through Time* (BPTT) has been proposed. BPTT [9] uses the idea that a RNN can be interpreted as a cascade of feed-forward networks. Figure 3 shows the result of this process called *unfolding* for a simple Elman network for three time steps. The vectors  $\mathbf{i}_i$  and  $\mathbf{o}_i$  denote the inputs and outputs at a given step  $i \in \mathbb{N}$  of the original recurrent network.

The static network resulting from the BPTT unfolding process can be used to perform a specialized variant of the backpropagation algorithm to determine the weight vector  $\mathbf{w}$  of the initial recurrent network. Further details on this topic can be found in [9].

<sup>2</sup> Non-recurrent ANNs are also called *feed-forward* ANNs to be distinguishable from the recurrent type.

## 2.5 Long Short-Term Memory

A problem of recurrent neural networks as described in 2.4 is that only recent outputs have a significant influence on the current network state. Furthermore, with every unfolding step the network becomes deeper, leading to either vanishing or exploding gradients during error backpropagation. This impedes training such a network with sequences that have input dependencies with large time lags. To approach this problem, Hochreiter et al. [10] introduce a new type of RNN called *Long short-term Memory* (LSTM) network.

Core idea of LSTM is adding an additional unit called *Constant Error Carousel* (CEC). The CEC is a self-recurrent unit with weight  $\mathbf{w} = 1$  that may store a value (called *state*) indefinitely. The state of the CEC is set by a so called *input gate* which determines the value that will be passed to the CEC. Another unit, the *output gate*, controls what value will be propagated to the output of the CEC. Both the input and the output gate have weights that will be set during training of the network. For this purpose a modification of BPTT called *truncated BPTT* is used. The combination of input gate, output gate and CEC is called *memory cell*. A block diagram of such an unit is shown in Figure 4.

The diagram shows the memory cell with the CEC in its center. The values  $\mathbf{i}_c$  and  $\mathbf{o}_c$  denote the inputs to the cell, whereas  $\mathbf{g}_i$  and  $\mathbf{g}_o$  denote the inputs to the input and output gate respectively. The rectangular blocks represent the necessary activation functions:  $o_i$  (cell input),  $o_g$  (input gate),  $o_c$  (cell output) and  $o_h$  (output gate). The circular nodes denote the operation (addition or multiplication) that is performed to combine the particular values. The inputs and outputs of an LSTM memory cell may be connected in arbitrary fashion, e.g., the cell output might be used as input to the input or output gate. Furthermore, several LSTM cells can be stacked forming a deep layer network. Note that the inputs and outputs are listed as vectors, which allows multiple cells being representable by the same block.

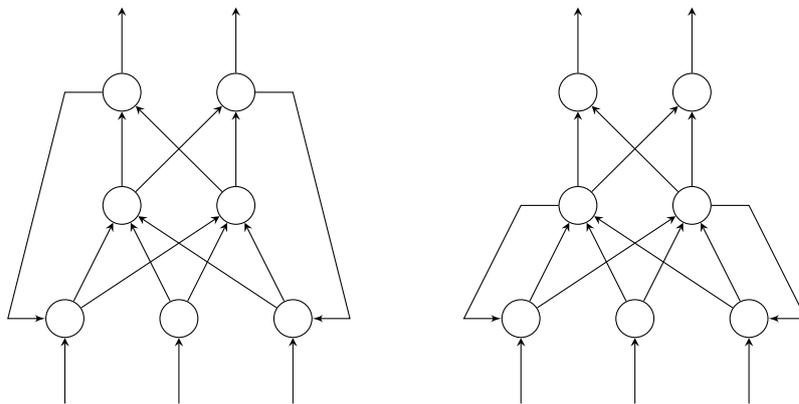


Fig. 2: Exemplary Jordan (left) and Elman (right) RNNs

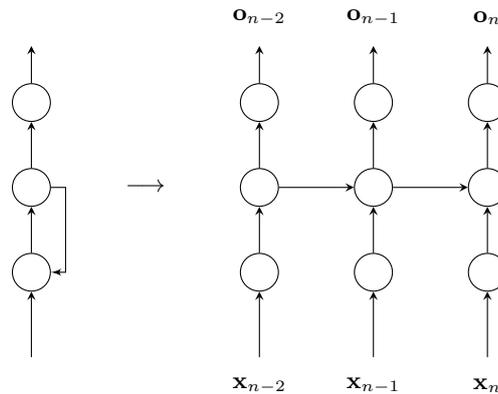


Fig. 3: Unfolding a simple Elman network

Several extensions and improvements to the original LSTM have been proposed. For example, in [11] Gers et al. introduce an additional type of gate called *forget gate* that allows clearing the state of the CEC (see Figure 5). This facilitates training the LSTM with sequences that contain subsequences that are rather unrelated.

### 3 Related Work

In this section, we present a number of existing approaches and studies regarding location prediction using neural networks in order to give an overview over the field and to choose a promising approach for implementing our own neural network architecture for a evaluation using a real-life dataset.

#### 3.1 Early Approach (Besterfeld et. al)

One of the earliest approaches to location prediction using neural networks can be found in [12]. Biesterfeld et al. examine the performance of various neural network types and encoding techniques on location prediction in a mobile

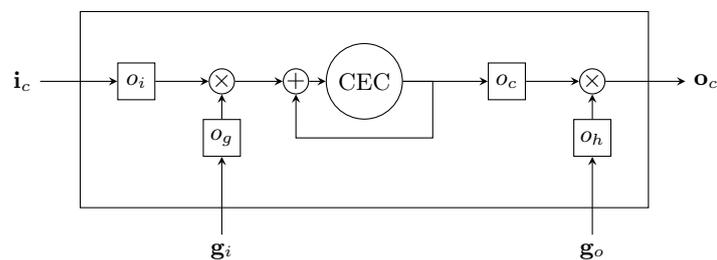


Fig. 4: An LSTM memory cell

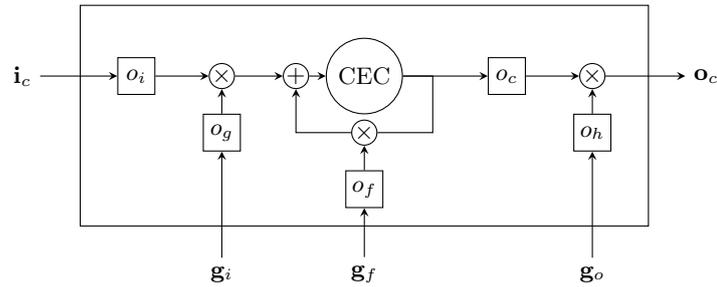


Fig. 5: An LSTM memory cell with forget gate

communication network consisting of 20 base stations with a unique identifier number. The neural networks consisted of three layers: input, hidden and output layer. Depending on the network type, recurrent or feed-forward, different encoding schemas were used: the recurrent networks were trained with input vectors consisting only of the last known location, whereas the feed-forward networks were trained with a binary vector encoding the last  $N$  locations, where  $N$  was a varying parameter during the examination. Their results showed that the feed-forward networks performed better than the recurrent models. Furthermore, the authors concluded that neural networks outperform previous database-based methods when predicting dynamic movement patterns.

### 3.2 Movement in Buildings (Vintan et al.)

As one of the first, Vintan et al. [13] applied the neural network approach to a ubiquitous computing application. To predict user movements in a large office building, every room was assigned a room identifier for each individual user. Similar to [12], Vintan et al. used a three-layer feed-forward network with variable neuron count. For the input layer, the room identifiers were binary encoded (with binary values -1 and 1) and concatenated into an input vector consisting of the last  $K$  visited rooms, with each neuron representing a bit from this vector. The output layer was the binary representation of the following (predicted) location. The examination consisted of determining the best values for the length of the input vector  $N$ , the number of required hidden units  $M$  and the learning rate for the back-propagation algorithm. The networks were trained on trajectory sets of four different users obtained in a time span of a few months, resulting in one model for each individual user. Furthermore, a second, combined model was trained with the trajectories of all users. In this case, a user identifier was added to the input vector. Their results show that the optimal value of the history length  $K$  is  $K = 2$  and that the best number of hidden neurons  $M$  is  $M = N + 1$ . They recommend use of a back-propagation learning rate of 0.1. The results show prediction accuracies of up to 92%. Other metrics were not used for evaluation.

### 3.3 Movement of Mobile Phone Users (Etter et al.)

Another examination of neural networks for location prediction and their performance in comparison to other statistical modeling approaches can be found in [6]. Etter et al. propose a solution to the "next place prediction"-task of the Nokia Mobile Data Challenge [14] using existing mobile phone trajectories from 80 different users. For the prediction task, the locations were represented by integer location identifiers. The following three different modeling methods were examined: three-layer feed-forward neural network, Dynamic Bayesian Networks (DBN) and Gradient Boosted Decision Trees (GBDT). For the neural network, the softmax function was used to obtain a probability distribution over the next predicted locations. The input vector was a one-hot encoded representation of the location identifier  $L$ . The network uses the negative log-likelihood function as the loss function and  $\text{hardTanh}$  as the transfer function between the layers. The authors' examination shows that a hidden unit count of  $N = 50$  shows optimal performance. Furthermore, the neural network is compared to the DBN and GDBT, showing similar accuracy results, with the neural network showing the best performance on the dataset (DBN: 60.17%, neural network: 60.87%, GBDT: 57.63%).

### 3.4 Combining ANNs with other Approaches (Vukovic et al.)

A joint approach combining a neural network with a probability-based model for the mobile network user mobility can be found in [15]. The probability-based model was used to determine regularity of user movements from ninety-five users obtained by MIT's Reality Mining Project [16]. If the movement was irregular, the neural network was used to predict the most likely next location. The authors did not only consider the location identifiers  $Loc$  for prediction, but also the weekday  $Dp$  and the hour interval  $Int$  in which the movement occurred, thus resulting in the tuple  $(Loc, Int, Dp)$ . The probability-based model was obtained by analyzing the movement records to find location histories of size  $N \geq 3$  that occurred more than once. During prediction, if the current location history matches a history that was classified as regular, the information from the probability-based model is used. If it did not, the neural network was used to predict the next location based on the last two location identifiers: hour interval and weekday. The neural network used for this irregularity prediction was a three-layer feed-forward network using one-hot encoding for the locations and binary encoding for interval and day inputs. The number of hidden units was varied depending on the individual user in order to obtain optimal prediction results. For evaluation, the models were trained on 10% of the movement dataset for each user. The evaluation results showed varying accuracies for the individual users' predictors. While some predictors showed very low accuracies in the range of 25% – 30%, other predictors obtained substantially better accuracies of about 65% – 85%. The authors explained this with differing movement habits of the individual users.

### 3.5 Comparison of Prediction Approaches (Petzold et al.)

A recent evaluation of different methods for location prediction in the context of ubiquitous computing has been conducted by [17]. Petzold et al. compare the performance of two neural network types (three-layer feed-forward and Elman networks), Dynamic Bayesian Networks (DBN), a Markov Model and a state predictor using Finite Automata with an evaluation set that was obtained from four individual persons. For the feed-forward network, binary encoding for the location history and output was used. Optimal history length was determined to be two with a hidden neuron count of nine. The Elman network, consisting of two hidden layers, was investigated using a one-hot encoded inputs and outputs. The optimal number of hidden neurons in each layer was determined as five with a history length of one. In their evaluation, both neural network approaches showed highest accuracy scores. However, the authors criticized the high modeling costs compared to the other investigated approaches.

### 3.6 Deep Learning and Big Data (Song et al.)

The recent emergence of Deep Learning [18], i.e., the usage of neural networks with many hidden layers on large datasets, has shown remarkable improvements of prediction and classification performance. Song et al. [19], following the Deep Learning trend, have proposed a Long-Short-Term-Memory (LSTM)-type based neural network architecture to predict trajectories of users based on GPS-Coordinate location records on a large dataset (approx. 1.6 million users). Aside from location, the system also predicts the transportation mode used by the subject, e.g., car or train. Their Deep Learning architecture consists of four LSTM layers: one for input encoding, two for learning the statistical patterns of the movements and one for output encoding. The network had a LSTM cell count of 80 and stochastic gradient descent with a learning rate of 0.01 and momentum 0.95. For evaluation, the authors compared their system to a shallow (one-layer) LSTM network and a Time-Delay Neural Network (TLDM), and other models including a Hidden Markov Model (HMM). Since their system outputs GPS coordinates, Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) was used as an evaluation metric. For transportation mode, precision and recall were computed. The Deep Learning LSTM network performed best over the other systems, showing lowest MAPE and MSE values, and highest precision (83.3%) and recall (81.4%) values in the comparison.

## 4 Evaluation of an ANN for Location Prediction

In this section, we present an implementation and evaluation of a location predictor using semantic labels. We choose the LSTM approach - similar as described in [19] - as the type of our artificial neural network architecture since LSTM networks showed best prediction results. Furthermore, by design of a LSTM network, it seems optimal to be used for training a sequence with time gaps, such as in our case, the semantic location history.

#### 4.1 Design and Implementation

For the implementation of the network, we use Python in combination with the open-source computational library TensorFlow [20] by Google Research. TensorFlow facilitates implementing the LSTM network by providing a set of library functions that implement an LSTM memory cell and the backpropagation algorithm.

The implemented network is a three-layer LSTM network with one hidden layer consisting of  $n_C$  LSTM cells and a softmax activation output layer (see Figure 6). Our implementation allows a varying number of LSTM cells to evaluate different network configurations and their corresponding predictive performance. As activation functions, the provided LSTM implementation in TensorFlow utilizes the sigmoid and tanh functions as described in [10]. For back-propagation, the cross-entropy loss function is used. The input to the network  $x$  is a list of length  $n_I$  of binary encoded location identifiers of bit count  $n_L$  and a binary encoded user identifier of length  $n_U$ , since we intend to train a model based on many different users' location histories. As for the softmax output layer, we use one-hot encoded vectors of length  $n_O$  to represent the prediction target. To obtain the predicted location  $L_{n+1}$ ,  $\text{argmax}_i o_i$  is calculated over the softmax activation outputs  $o_i$ . The final implementation expects a text file of daily location histories where each line of the file represents a single day of a particular user. The histories are provided as a sequence of integer location identifiers, with the first integer being the identifier of the corresponding user (see Listing 1.1 for an example).

```
0  1 0 1 0 2 3 1 3 1 3 2 0
0  2 3 2 0
1  1 6 3 1 2 0
1  2 3 2 0 1 0
1  1 0 1 0 2 3 5 3 2 0 2 5
```

Listing 1.1: Sample training input file

For training, the implementation extracts sub-histories of length  $n_I + 1$  from each daily history and transforms the first  $n_I$  items into binary encoding and the last item into an one-hot encoded representation. Both are stored in an in-memory list for later access during training, together with the binary representation of the user identifier.

After this pre-processing, the implementation randomly selects 10% of the generated histories as test data and stores these separately. Then, the training is performed using the TensorFlow implementation of stochastic gradient descent. TensorFlow expects as input a randomly selected subset (*batch*) of the training set in order to update the weights of the network faster than after only considering one training sample at one time. Thus, during each training epoch, our implementation divides the training set into  $m$  random subsets of  $n_S$  size and feeds each of these subsets into the TensorFlow training function. To evaluate the performance of the network after each training epoch, the test dataset is fed

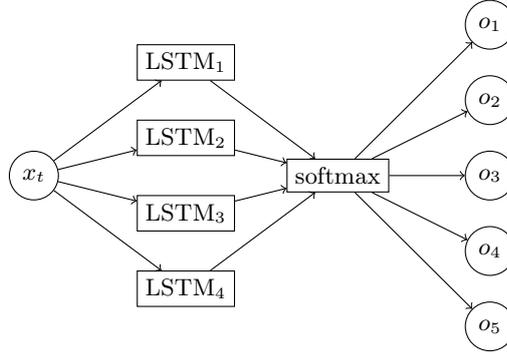


Fig. 6: Architecture of the LSTM implementation (sample configuration)

into the network and the corresponding output layer activation are computed. Then, the evaluation metrics *accuracy*, *precision*, *recall* and *F<sub>1</sub>-measure* are obtained by comparing the predicted output with the expected values from the test subset.

#### 4.2 Evaluation metrics

The evaluation metrics [21] described in the previous section are obtained as follows:

$$accuracy = \frac{\text{number of correctly predicted cases}}{\text{number of all cases}} \quad (10)$$

Thus, accuracy measures how many times the prediction was correct, but does not take into account the frequency of one class<sup>3</sup> compared to another. For example, a predictor could always return an identical class, and score high accuracies if the prevalence of this class is also high. To mitigate the influence of class frequency, *recall* and *precision* are used. These values can be obtained from the confusion matrix  $M$  which contains the number of predicted classes for each true class (see Table 1). The diagonal of a confusion matrix shows the correct classifications, and the remaining entries wrong classifications. There are four cases to distinguish:

**True positive (TP):** A true class  $C_i$  was correctly classified as  $C_i$

**True negative (TN):** A true class  $C_i$  was correctly classified as *not*  $C_i$

**False positive (FP):** A true class  $C_i$  was wrongly classified as  $C_i$

**False negative (FN):** A true class  $C_i$  was wrongly classified as *not*  $C_i$

These values can be derived per-class from the confusion matrix  $M = (m_{ij})$

<sup>3</sup> The class in the case of location prediction is the predicted destination, which in our case is the predicted location identifier.

as follows [22]:

$$\text{TP}_i = m_{ii} \quad (11)$$

$$\text{TN}_i = \sum_{j \neq i} m_{jj} \quad (12)$$

$$\text{FP}_i = \sum_{j \neq i} m_{ji} \quad (13)$$

$$\text{FN}_i = \sum_{j \neq i} m_{ij} \quad (14)$$

Now, the per-class precision and recall values are defined as:

$$\text{precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} = \frac{m_{ii}}{\sum_j m_{ji}} \quad (15)$$

$$\text{recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} = \frac{m_{ii}}{\sum_j m_{ij}} \quad (16)$$

Thus, precision is the fraction of relevant positive outcomes, whereas recall measures the sensitivity of the predictor for a certain class  $C_i$ . To combine these two values into a single metric for overall performance, one may calculate  $F_1$ -measure.  $F_1$ -measure is the harmonic mean of the arithmetic means of each individual precision or recall value from each of the  $n$  classes:

$$\text{precision} = \frac{\sum_i \text{precision}_i}{n} \quad (17)$$

$$\text{recall} = \frac{\sum_i \text{recall}_i}{n} \quad (18)$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (19)$$

		Predicted class		
		$C_1$	$C_2$	$C_3$
True class	$C_1$	14	2	1
	$C_2$	1	12	4
	$C_3$	4	3	10

Table 1: Sample confusion matrix ( $C_n$  denoting a class)

### 4.3 Evaluation Dataset

For training and testing the ANN, it is necessary to obtain suitable datasets for this purpose. There are several datasets available that contain location trajectories, such as the Nokia Mobile Data Challenge dataset [14]. Many of these datasets do not contain semantic labels but merely GPS trajectories or network identifiers which makes them unsuitable for semantic prediction. The Foursquare location-based social network has collected [23] semantic histories of "check-ins," i.e., when a user wanted to share their location with their peers. But these "check-ins" cannot be treated as actual semantic data since a user would not always "check-in" at a certain place, as argued by Ying et. al in [24]. Instead, we follow the authors' approach of using the MIT Reality Mining dataset [16] as the basis of our evaluation.

The dataset contains user trajectories obtained from cell tower identifiers over a period of several months. Each trajectory consists of a sequence of entries

$$(\text{id}_{\text{user}}, t_{\text{start}}, t_{\text{end}}, \text{id}_{\text{tower}})$$

ordered by  $t_{\text{start}}$ . The value  $\text{id}_{\text{tower}}$  is the identifier of the current (i.e. nearest) cell tower. In contrast to pure GPS or mobile network trajectories, the MIT dataset contains also user-defined semantic annotations (e.g., "Home" or "Work") for each cell tower identifier:

$$(\text{id}_{\text{tower}}, \text{annotation})$$

Thus, we can map a cell tower identifier to its semantic label. Similar to Ying et. al, we only use locations where we know that the user has stayed for a longer period of time, avoiding noise from transit movement. For this purpose, we divide each day into one-hour slices and determine the sum of the stay duration for each location during this time-slice. Then, we discard all locations with a stay duration  $< 5$  minutes. If the dataset contains repeating sequences of the same location, we only save this location once. Then we remove all histories from weekends since we believe this would add too much noise to the dataset. Since every user has made their own annotations, it is necessary to find semantically equal labels. After manually examining the annotation dataset, we conclude that the annotations are usually very similar except for minor spelling differences. Thus, we use the phonetic *SOUNDEX algorithm* [25] to obtain spelling-invariant labels. Moreover, to eliminate further annotation errors, we apply the *Levenshtein algorithm* with edit distance of 1 on the previously obtained labels. This procedure results in finding 21 suitable users with 1043 daily histories in total. Finally, we replace the semantic text labels with an integer location identifier (see Listing 1.1).

### 4.4 Experimental Setup

We used our implementation as described in 4.1 to test the performance of a three-layer LSTM network for location prediction. As for the stochastic gradient descent, the momentum optimization is used with momentum parameter  $\alpha = 0.95$  (as used by [19]) and a varying learning rate  $\eta \in \{0.005, 0.01, \dots, 0.025, 0.03\}$ .

Parameter	Learning Rate ( $\eta$ )	LSTM cells ( $n_C$ )	History Length ( $l_I$ )	Batch Size ( $n_S$ )
min	0.005	2	1	50
max	0.03	16	4	300
step	0.005	2	1	50

Table 2: Evaluated model parameters

The length of the input history  $n_I$  was varied between  $n_I = 1$  and  $n_I = 4$ , as conducted in [17]. For the length of the binary encoded input items, we set  $n_B = 8$  since no location identifier exceeds  $2^8 = 256$ . The length of the one-hot encoded output layer  $n_O$  was set to the value of the highest location identifier in the current training set.

Furthermore, we varied the number of LSTM cells  $n_C$  between  $n_C = 2$  and  $n_C = 16$ . In contrast to [19], we chose a small number of cells since our dataset is comparatively small. Moreover, we tested different batch sizes  $n_S$  between  $n_S = 50$  and  $n_S = 300$ . In order to examine the convergence behavior, we performed  $n_E = 20$  epoch iterations and measure the accuracy and F1-measure scores after each epoch. Since we rely on a randomized selection of the batch subsets, we trained 10 different models per parameter configuration  $P_i = (\eta, n_I, n_C, n_S)$  and take the arithmetic mean of each evaluation metric computation to mitigate the influence of the sampling bias. Table 2 gives an overview on the varied parameters during our evaluation. In total, we tested  $6 \cdot 8 \cdot 4 \cdot 6 = 1152$  different parameter configurations.

#### 4.5 Results

For evaluation, we used the common *box-and-whisker plot* [26] to visualize the distributions of the obtained accuracy and F<sub>1</sub>-measure averages. This plot indicates the lower and upper quartiles as a rectangular shape and the minimum and maximum range as a line with a horizontal bar at each end ("whisker"). Furthermore, the median and arithmetic mean are represented by a white (median) and black (mean) horizontal bar.

First, we examined the performance of the model as a function of LSTM cell count. Figure 7 shows the obtained accuracy and F<sub>1</sub>-measure plots for each tested cell count parameter  $n_C$ . The results indicate that the scores constantly increase with additional available LSTM cells. The performance seems to plateau at a cell count of about  $n_C \approx 14$ . For further investigation, we fix  $n_C = 16$  since it shows best overall scores for both accuracy ( $x_{max} = 0.628$ ,  $\tilde{x} = 0.53$ ,  $\bar{x} = 0.502$ ) and F<sub>1</sub>-score ( $x_{max} = 0.483$ ,  $\tilde{x} = 0.375$ ,  $\bar{x} = 0.353$ ).

Next, we investigated the influence of the learning rate  $\eta$ . For reference, in [19],  $\eta = 0.01$  was used. Figure 8 shows the plot of the obtained metrics for the different values of  $\eta$ . Our results confirm that  $\eta = 0.01$  is indeed a good parameter choice, although  $\eta = 0.005$  showed slightly better median and average

scores than  $\eta = 0.01$ . However, due to sample bias,  $\eta = 0.015$  resulted in the best overall  $F_1$ -score ( $x_{max} = 0.483$ ) in our evaluation. For further examination, we fix  $\eta = 0.005$ .

With  $n_C = 16$  and  $\eta = 0.005$  fixed, we looked at the influence of the batch size. Figure 9 shows the obtained distributions. Although there is a slight decrease in performance with increasing batch size; this parameter seems not to

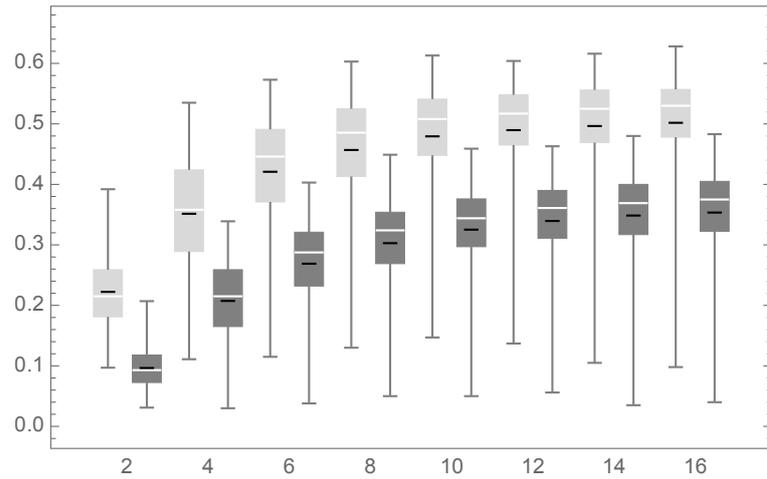


Fig. 7: Accuracy (light gray) and  $F_1$ -measure (dark gray) distributions per LSTM cell count  $n_C$

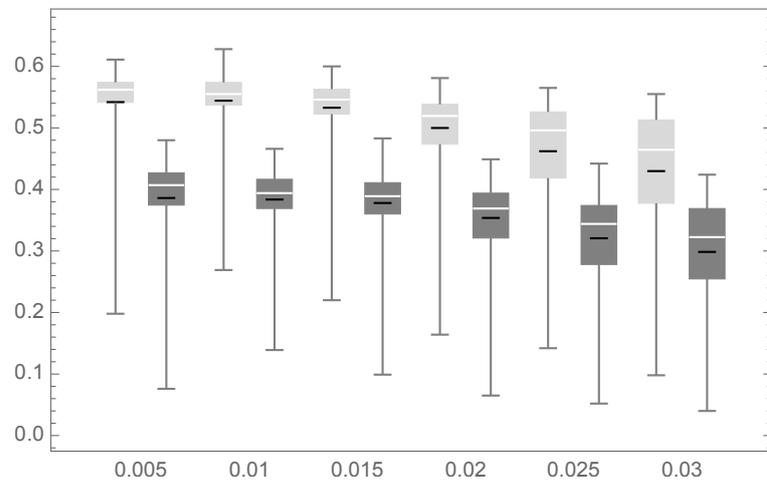


Fig. 8: Accuracy (light gray) and  $F_1$ -measure (dark gray) distributions per learning rate  $\eta$ . LSTM cell count fixed at  $n_C = 16$ .

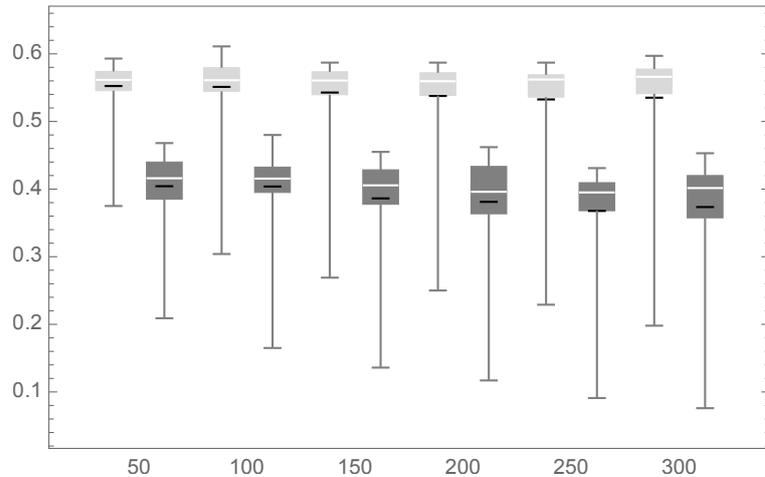


Fig. 9: Accuracy (light gray) and  $F_1$ -measure (dark gray) distributions per batch size  $n_S$ . Other parameters:  $n_C = 16$  and  $\eta = 0.005$ .

affect the performance of the model significantly. Since we see a drop of the minimum performance with increasing batch size, smaller batch sizes seem to make the model converge faster, thus resulting in higher minimal scores.

Furthermore, we examined the convergence behavior by evaluating the model performance at each epoch iteration  $e$  ( $e \leq n_E$ ) with fixed parameters  $n_C = 16$  and  $\eta = 0.005$ . As can be seen in Figure 10, the models converge after about  $e \approx 10$  epoch iterations. This is similar to the results obtained by [19], who report  $e = 8$ .

Subsequently, we investigated the dependency of the models' performance on the history length  $n_I$ . For this purpose, we set  $\eta = 0.005$ ,  $n_C = 16$  and  $e \geq 10$ . Figure 11 shows the result of this evaluation.  $n_I = 2$  seems to perform the best over the remaining three parameter settings. This is confirming the findings obtained by Vintan et al. in [13].

Parameter	$\eta$	$n_C$	$l_I$	$n_S$
Value	0.005	16	2	50

Table 3: Best-performing model parameter values

Since we had obtained the parameters for optimal performance, we conducted a further evaluation by investigating the performance of the model per user. For this purpose, we set the parameters according to our evaluation results (see Table 3). Furthermore, we altered the test holdout parameter to  $h = \frac{1}{3}$  to prevent *overfitting*, i.e. the inability of the model to generalize on new, unseen data

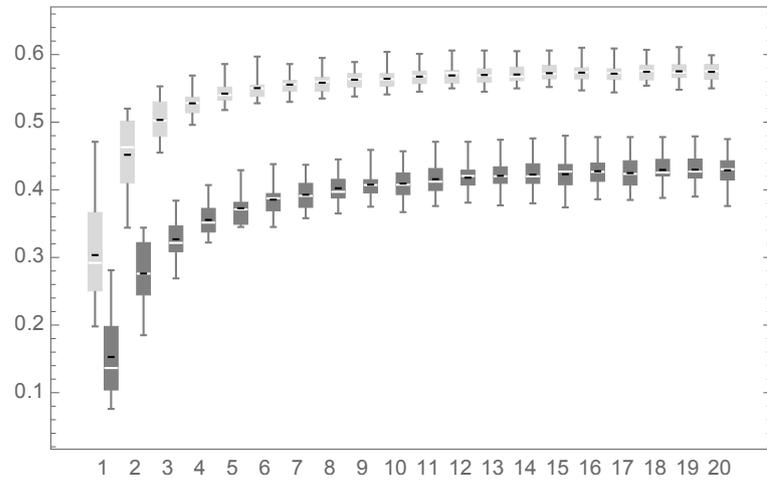


Fig. 10: Accuracy (light gray) and  $F_1$ -measure (dark gray) distributions per epoch  $e$ . Other parameters:  $n_C = 16$  and  $\eta = 0.005$ .

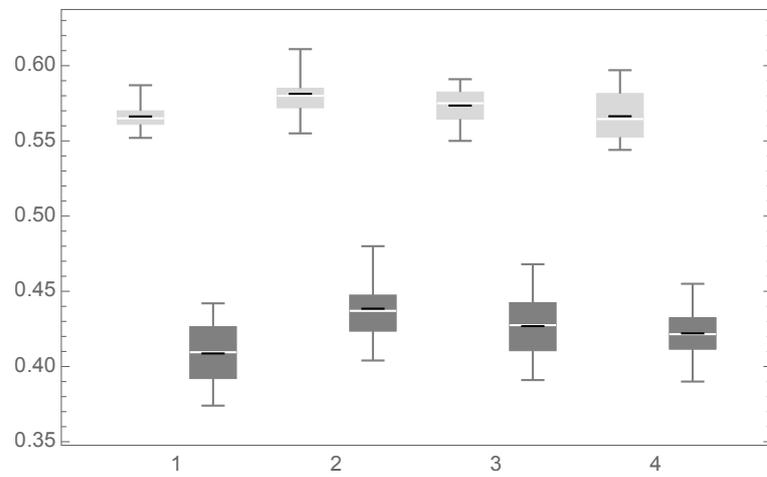


Fig. 11: Accuracy (lightgray) and  $F_1$ -measure (darkgray) distributions per history length  $n_I$ . Other parameters:  $n_C = 16$ ,  $\eta = 0.005$ ,  $e \geq 10$ .

points. With this configuration, we trained 100 models with  $n_E = 20$  epoch iterations. For each trained model, we stored the accuracy and  $F_1$ -measure scores computed on per-user basis. Figure 12 shows the  $F_1$ -measure scores at  $e = 20$  of each user ordered by the arithmetic mean of the  $F_1$  score. As can be inferred from the results, there is a large spread of the obtained averages. We explain this with the difference in regularity of each user’s movement. However, 80.1% of the  $F_1$ -scores were above 0.3. Furthermore, our evaluation showed that 76.2% of the accuracy scores were greater than 0.5.

To summarize our evaluation, we have calculated the average of the arithmetic means and the maximums over all users. Since the number of daily histories was not equal for each user, we have further weighted each average score with the occurrence of the particular user in order to obtain a more accurate result. Overall, our model shows similar performance to previous approaches discussed in Section 3. The resulting scores are shown in table 4.

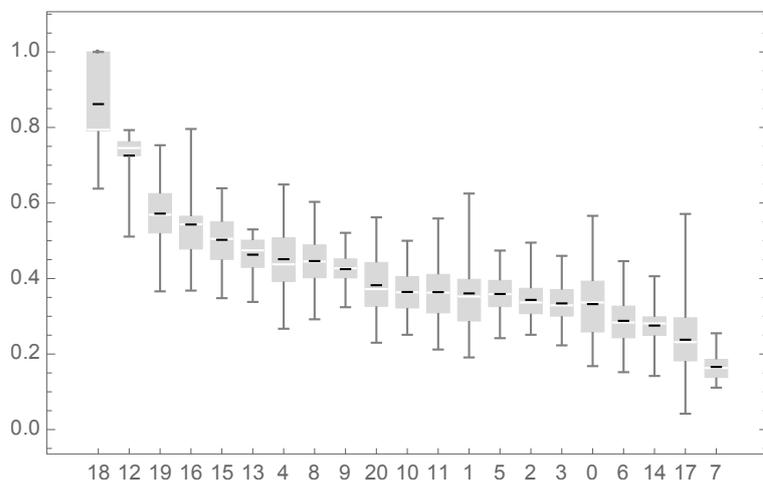


Fig. 12:  $F_1$ -measure distributions per user ordered by arithmetic mean. The x-axis values indicate the respective user identifier.

## 5 Conclusion

In this paper, we selected a promising neural network approach and evaluated its performance on location prediction using semantic labels on a joint. Using location histories from the MIT dataset, we trained a joint model of several users. Our results show that LSTM networks are a suitable method for predicting semantic user locations. The models obtained from our architecture showed a performance comparable to those of previous, non-semantic neural network

	Unweighted	Weighted
Accuracy (mean)	0.605	0.597
Accuracy (max.)	0.730	0.710
F <sub>1</sub> -measure (mean)	0.420	0.402
F <sub>1</sub> -measure (max.)	0.581	0.547

Table 4: Average performance scores of the trained models

implementations. Although we only used low-level location labels without clustering or other semantic categorization, our model performed with average F<sub>1</sub> score of 0.42 and average accuracy of 0.605. For further improvement, adding more features, e.g. location type, to the classificatory might increase the model's performance significantly.

Furthermore, this paper shows that location prediction remains a hot topic of contemporary research, due to the increasing relevance and emergence of location based services. It can be anticipated that this will lead to an overall improvement of location prediction models, especially through the availability of extensive datasets and the potential of distributed computational approaches such as TensorFlow.

## References

1. T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, Inc., 1 ed., 1997.
2. C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995.
3. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," tech. rep., DTIC Document, 1985.
4. R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*. Cambridge, MA, USA: MIT Press, 1st ed., 1998.
5. M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate bayesian a posteriori probabilities," *Neural computation*, vol. 3, no. 4, pp. 461–483, 1991.
6. V. Etter, M. Kafsi, and E. Kazemi, "Been there, done that: What your mobility traces reveal about your behavior," in *Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing*, no. EPFL-CONF-178426, 2012.
7. M. I. Jordan, "Serial order: A parallel distributed processing approach," *Advances in psychology*, vol. 121, pp. 471–495, 1997.
8. J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
9. M. C. Mozer, "A focused back-propagation algorithm for temporal pattern recognition," *Complex systems*, vol. 3, no. 4, pp. 349–381, 1989.

10. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
11. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
12. J. Biesterfeld, E. Ennigrou, and K. Jobmann, "Neural networks for location prediction in mobile networks," 1997.
13. L. Vintan, A. Gellert, J. Petzold, and T. Ungerer, "Person movement prediction using neural networks," in *First Workshop on Modeling and Retrieval of Context*, vol. 114, pp. 1–12, 2004.
14. J. K. Laurila, D. Gatica-Perez, I. Aad, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, M. Miettinen, *et al.*, "The mobile data challenge: Big data for mobile computing research," in *Pervasive Computing*, no. EPFL-CONF-192489, 2012.
15. M. Vukovic and D. Jevtic, "Agent-based movement analysis and location prediction in cellular networks," *Procedia Computer Science*, vol. 60, pp. 517–526, 2015.
16. N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the national academy of sciences*, vol. 106, no. 36, pp. 15274–15278, 2009.
17. J. Petzold, F. Bagci, W. Trumler, and T. Ungerer, "Comparison of different methods for next location prediction," in *European Conference on Parallel Processing*, pp. 909–918, Springer, 2006.
18. J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
19. X. Song, H. Kanasugi, and R. Shibasaki, "Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level," The University of Tokyo, Japan.
20. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015.
21. D. L. Olson and D. Delen, *Advanced Data Mining Techniques*. Springer Publishing Company, Incorporated, 1st ed., 2008.
22. M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
23. D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 1, pp. 129–142, 2015.
24. J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng, "Semantic trajectory mining for location prediction," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 34–43, ACM, 2011.
25. MySQL AB, *MySQL 5.7 Reference Manual*, 2017.
26. R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *The American Statistician*, vol. 32, no. 1, pp. 12–16, 1978.

# Artificial Neural Networks for location prediction

Harun Şentürk\*

Betreuer: Antonios Karatzoglou†

Karlsruher Institut für Technologie (KIT)  
Pervasive Computing Systems – TECO

\*sentuerk@teco.edu

†antonios@teco.edu

**Zusammenfassung.** In dieser Arbeit wird die Predizierbarkeit der nächsten semantischen Lokation einer Person untersucht. Dafür wird mit Bewegungsdaten einer Person gearbeitet, welche sich über einen Zeitraum von zwölf Wochen strecken. Auf diesen Daten werden Feed-Forward Neuronale Netze und Recurrent Neuronale Netze mit jeweils einer mittleren Schicht trainiert. Als Referenzwerte wurden Vorhersagegenauigkeiten von Markovketten betrachtet. Anschließend werden die besuchten Lokationen zu Lokationstypen zusammengefasst, um zu untersuchen, ob und wie sehr die Vorhersagbarkeit auf diesen Lokationstypen besser ist als auf den Lokationen selbst.

**Schlüsselwörter:** Location prediction, Semantic location, Markov Chain, Feed-Forward Neural Net, Recurrent Neural Net

## 1 Motivation

Ubiquitäre Systeme sind eine Idee der Mensch-Maschine-Interaktion, welche die Integration von Computer-Komponenten in alltägliche Dinge propagiert. miteinander und mit der Umgebung interagierende Geräte oder Unsichtbarkeit sind Punkte, welche diese Idee charakterisieren. Die „intelligente Umgebung“ ist eine Welt, in welcher verschiedene intelligente Geräte permanent arbeiten, um die Lebensqualität der Bewohner zu verbessern. Eine intelligente Umgebung beinhaltet die Erfassung, Prediktion, das Entscheiden und proaktive Handeln, um die Nutzererfahrung zu verbessern und sich nach bestimmten Mustern wiederholende Aufgaben zu automatisieren. [22]

Das Verhalten des Menschen kann in zwei Kategorien geteilt werden: Zum einen handelt er in bestimmten Verhaltensmustern. Zum anderen können diese Verhaltensmuster durch äußere und innere Reize unterbrochen bzw. das Verhaltensmuster grundlegend geändert werden. Eine Idee der Next Location Prediction ist, dem Nutzer (wobei „Nutzer“ eine aktiv interagierende Person impliziert, was nicht der Fall sein muss, insbesondere im Zusammenhang mit ubiquitären Systemen) Aufgaben abzunehmen, welche in bestimmten Mustern durchgeführt werden. Diese Muster sollen automatisch gelernt und gegebenenfalls angepasst werden. Auf Klassifikatorebene heißt das: Statisches Trainieren auf vergangenen

Daten und Predizieren künftiger Werte. Bei korrekter und auch falscher Prediktion soll das vom Klassifikator gelernte Modell jeweils „bestärkt“ bzw. „bestraft“ werden.

In dieser Arbeit geht es insbesondere um die Prediktion semantischer Lokationen. Unter der Menge der semantischen Lokationen ist eine Wissensrepräsentation in Form einer Baumstruktur (DAG) zu verstehen. Beispielhaft kann der Wurzelknoten „Ort“ semantische Oberbegriffe anderer Orte als Kindknoten haben. In der Baumstruktur könnte es bspw. einen inneren Knoten „Freizeit“ mit Kindknoten „Bowling“, „Schwimmbad“ und „Kino“ geben. Dabei kann es verschiedene Zwischenabstraktionen geben. Die zu prüfende Behauptung dieser Arbeit ist, dass auf semantisch abstrakteren Stufen (mögliche Orte bspw. „Einkaufen“ und „Freizeit“) Lokationen besser prediziert werden können als auf niedrigeren Stufen (mögliche Orte bspw. „Aldi\_Durlach“ und „Kino\_Innenstadt“).

Ein wichtiges Anwendungsgebiet von Lokationsprediktion stellt die Klimaregulierung von Bürogebäuden dar [19]. In diesen summieren sich das Heizen, Ventilieren und die Klimatisierung auf 70% des Gesamtenergieverbrauchs dieser Gebäude. Ein Großteil der Energie wird unter statisch festgelegten Plänen verbraucht, welche oft immernoch die zur Zeit des Gebäudeentwurfs festgelegten Zeitabschnitte als Grundlage haben. Für eine Verbesserung der Energieeffizienz sollte das Gebäude-Management-System Prediktoren benutzen, um das Klima auf Grundlage der Nutzer-Bewegungsmuster zu regulieren. Da Heiz- und Kühlsysteme eine verzögerte Reaktion vorweisen, muss das Nutzerverhalten eine bestimmte Zeitspanne vorher prediziert werden.

In [19] wird die Behauptung aufgestellt, dass das Nutzerverhalten in Bürogebäuden von Faktoren wie der Jahreszeit, dem Wochentag und der Tageszeit abhängt. Außerdem wird vermutet, dass Verhalten im Laufe des Tages von vorherigem Verhalten innerhalb desselben Tages beeinflusst wird. Außerdem wird auf externe Einflüsse wie beispielsweise dem Verhalten der Kollegen hingewiesen. Manche dieser Überlegungen werden auch in diese Seminararbeit einfließen, wie in späteren Kapiteln zu sehen ist.

Zur Motivation sollen weitere Beispielszenarien erläutert werden:

- Ein intelligentes Steuergerät im Auto lernt, an welchen Wochentagen und zu welcher Uhrzeit der Besitzer das Auto benutzt, sodass die Standheizung eine halbe Stunde vorher eingeschaltet werden kann. Das führt dazu, dass der Autofahrer sich in ein angenehm warmes Auto setzen kann. Gute Voraussetzungen für eine konzentrierte Fahrt.
- Eine ähnliche Anwendung wäre die automatische Durchlüftung des Zuhauses, bevor nach Hause gekommen wird. Das Vorhersagesystem würde anhand vergangener Bewegungsmuster des Einwohners einen Stundenplan über die Ankunftszeiten modellieren, sodass beispielsweise eine halbe Stunde vor dem Nachhausekommen für frische Luft im Haus gesorgt werden kann.
- Es soll mobil eine große Datei versendet werden, während die Signalstärke schwach ist. Das System könnte vorhersagen, dass der Nutzer ein Gebiet mit besserem Empfang betreten wird, sodass das energiezehrende Versenden bei schlechtem Empfang vermieden werden kann. [8]

Als Kontrast zu diesen Anwendungsbereichen müssen auch negative Anwendungsszenarien der Lokationsprediktion betrachtet werden. Bei Zugriff auf die sensiblen Daten könnten böswillige Menschen anhand solcher Modelle ermitteln, zu welchen Zeiten das Haus leerstehen wird. Eine negative Seite dieser Technologie.

### 1.1 Semantische Trajektorien

Betrachtet man die Forschung im Bereich der Bewegungsprediktion mobiler Nutzer, fällt auf, dass dabei meist nur geographische Eigenschaften der Nutzertrajektorien berücksichtigt werden. Darunter ist eine Sequenz von geographischen Punkten, bestehend aus Breiten- und Längengrad, zu verstehen, welche jeweils mit einem Zeitstempel versehen sind. Solche beobachteten Bewegungsmuster sind nur von geographischen Eigenschaften der Bewegungsdaten abhängig.

Die Nachteile einer Lokationsprediktion basierend auf geographischen Trajektorien sind vor allem folgende: Zum Einen können zwei semantisch nahe Trajektorien als unterschiedlich klassifiziert werden, da die geographische Ähnlichkeit der einen Trajektorie zu einer anderen Trajektorie näher ist. Zum anderen berücksichtigen auf geographischen Eigenschaften arbeitende Prediktionsstrategien nur zuvor besuchte Lokationen und können bisher unbesuchte Lokationen nur schlecht vorhersagen. Eine Lokationsprediktion basierend auf allein den geographischen Trajektorien ist also nicht ausreichend. [6]

In Abbildung 1 ist exemplarisch zu sehen, dass die geographische Distanz und Form der Trajektorie 1 und 2 ähnlicher sind als die zwischen Trajektorie 1 und 3. Aus diesem Grund würden einige Vorhersagetechniken zur Prediktion der nächsten Lokation der Trajektorie 2 die Trajektorie 1 als Bezugstrajektorie nehmen. Dazu kommt noch, dass diese Prediktion, welche auf Positionsdaten beruhen, nur zuvor besuchte Lokationen berücksichtigen und deshalb keine guten Ergebnisse liefern, wenn zuvor nicht besuchte Orte berücksichtigt werden. Auch hier ist zu sehen: Die bloße Berücksichtigung geographischer Informationen zur Prediktion der nächsten Lokation ist nicht ausreichend.

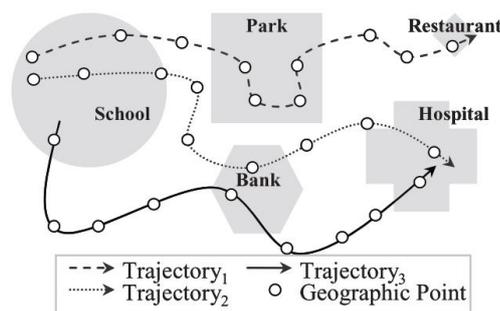


Abb. 1: Beispiel zu semantischen Trajektorien. [6]

## 2 Related work

In der Arbeit [5] wird eine Architektur eines Feed-Forward Neuronales Netzes mit einem Input-, einem Hidden- und einem Output-Layer entworfen, um anhand der zuletzt besuchten Räume den nächsten zu betretenden Raum verschiedener Mitarbeiter in einem Bürogebäude zu predizieren. Dabei wird die Unfähigkeit eines Multi-Layer-Perceptrons, Sequenzen schrittweise einzulesen, umgangen, indem die zuletzt besuchten  $N$  Räume kodiert an die Eingabeschicht übergeben werden. Hier beschränkt man sich auf ein kleines Netz und verringert die Größe der Eingabeschicht, um möglichst wenig Rechen- und Speicherkapazität zu beanspruchen, sodass diese Neuronale Netze mit akzeptablem Energie- und Speicherkonsum ausgeführt werden können. Man hat dabei „lokale“ Prediktoren, welche auf Bewegungsdaten pro Mitarbeiter trainiert werden, mit „globalen“ Prediktoren, welche auf Bewegungsdaten aller Mitarbeitern trainiert werden, verglichen. Hier hat man die Prediktion in einem Szenario eingesetzt, in welchem smarte Türschilder einem Besucher die aktuelle Lokation des jeweiligen Mitarbeiters angezeigt und die vermutliche Rückkehrzeit prediziert hat.

In einer anderen Arbeit [13] wird Lokationsprediktion angewandt, um Kapazität an der Luft-Schnittstelle von mobilen Radionetzwerken zu sparen, welche den Flaschenhals bei dieser Technologie darstellt. Die Ergebnisse weisen darauf hin, dass der Ertrag aus Neuronale Netzen vom Nutzerprofil und der Zahl ungewöhnlicher Bewegungen des Nutzers abhängt. Auch in diesem Fall umgeht man den Nachteil eines fehlenden Speichers, indem die Historie der  $N$  letzten besuchten Lokationen kodiert an die Eingabeschicht geleitet wird.

Weiterhin wurde im Rahmen des „Adaptive House“ Projektes an der Universität Colorado [16] ein intelligentes Haus entwickelt, welches die Bewegungen seiner Einwohner mithilfe von Bewegungsdetektoren aufgezeichnet und mit Feed-Forward Neuronale Netzen den nächsten zu betretenden Raum vorhersagt.

In [17] werden fünf verschiedene Vorhersagemodelle vorgestellt und verglichen, unter welchen sich eine Markovkette und Neuronale Netze befinden. Dabei wird ohne zu Vertiefen ein oberflächlicher Überblick gegeben. In dieser Seminararbeit soll hingegen neben der Markovkette verschiedener Ordnung, welche als Bezugssystem (Baseline) verwendet werden soll, Feed-Forward Neuronale Netze (FFNN) und Recurrent Neuronale Netze (RNN) betrachtet werden. Der vergleichende Charakter des Papers war Anlass dazu, in diesem Kapitel aufgeführt zu werden.

In [15] wird untersucht, wie gut bei Verwendung von RNNs die täglichen Aktivitäten unter Altersdemenz leidender Menschen in ihren eigenen Häusern vorhergesagt werden können. Dabei werden die Bewegungsmuster mit Sensoren ermittelt und diese dann in zeitliche Sequenzen von Aktivitäten umgewandelt. Anschließend werden diese für die Vorhersage der Nutzerbewegung genutzt. Für das Vorhersagemodell werden dabei verschiedene RNNs betrachtet, welche besonders zeitliche Abhängigkeiten in den Bewegungssequenzen gut erschließen können. Dabei werden sowohl durch einen Simulator generierte Daten, als auch Daten echter Umgebungen verwendet.

Weitere Arbeiten, welche Feed-Forward Neuronale Netze zur Bewegungsprediktion nutzen, sind [18] und [23].

### 3 Theorie Markovkette und Neuronale Netze

#### 3.1 Markovketten

In diesem Abschnitt soll kurz das Konzept von Markovketten [9] eingeführt werden, die nur die aktuelle Position zur Vorhersage der nächsten Lokation verwenden. Diese werden auch als Markovkette 1. Ordnung bezeichnet. Anschließend wird kurz darauf eingegangen, wie dieses Konzept auf die Berücksichtigung von  $N$  vorher besuchten Lokationen erweitert werden kann (für  $N \geq 1$ ). Diese werden dann als Markovketten  $N$ -ter Ordnung bezeichnet.

Eine Markovkette modelliert das Bewegungsverhalten einer Person als einen diskreten stochastischen Prozess, in welchem die Wahrscheinlichkeit, sich von einem Ortszustand in einen anderen zu bewegen, ausschließlich von der vorher besuchten Lokation und der Wahrscheinlichkeitsverteilung der Transitionen zwischen den Zuständen abhängt. Genauer besteht eine Markovkette aus:

- Einer Menge von Zuständen  $P = \{p_1, \dots, p_k\}$ , in welchem jeder Zustand eine Lokation modelliert. Diese Zustände haben eine semantische Bedeutung und werden beispielsweise mit „Universität“ oder „Heim“ gelabelt.
- Einer Menge von Transitionen, wie z.B.  $t_{i,j}$ , welche die Wahrscheinlichkeit darstellt, sich von der Lokation  $p_i$  zur Lokation  $p_j$  zu bewegen. Eine Transition aus einem Zustand in denselben Zustand kann in unserem Fall vorkommen, wenn Lokationstypen betrachtet werden. Beispielsweise, wenn die Transition Mensa  $\rightarrow$  Infobau ersetzt wird durch die Transition Universität  $\rightarrow$  Universität.

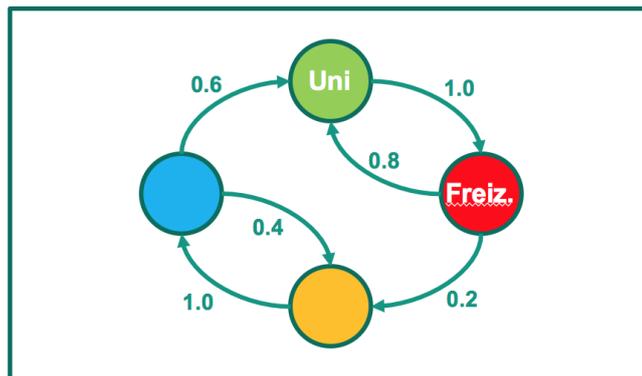


Abb. 2: Markovkette als Graph.

Eine Markovkette kann als Graph oder Transitionsmatrix repräsentiert werden. In der Graphenrepräsentation, wie beispielsweise in Abbildung 2, werden Lokationen durch Knoten und Transitionen mit der zugehörigen Übergangswahrscheinlichkeit zwischen zwei Lokationen durch Pfeile zwischen jeweils zwei Knoten mit einem Gewicht dargestellt.

Markovketten erster Ordnung haben kein Gedächtnis. D.h., dass die Vorhersage einer zukünftigen Lokation nur von der aktuellen Lokation abhängt. Diese Einschränkung, dass nämlich die Markovkette die vorher besuchten Lokationen „vergisst“, kann die Vorhersagegenauigkeit verschlechtern. Um diesem Problem entgegenzuwirken, können Markovketten größerer Ordnung benutzt werden.

In Abbildung 2 ist eine Markovkette erster Ordnung skizziert, deren Knoten Lokationstypen modellieren. Hier beträgt bspw. die Wahrscheinlichkeit, von „Uni“ zu „Freizeit“ zu gelangen, 100% und die Wahrscheinlichkeit, von „Freizeit“ zu „Uni“ zu gelangen, 80%. Die Gewichtssumme der aus einem Knoten wegführenden Kanten muss 100% betragen. Ein Nachteil von Markovketten ist, dass nicht generalisiert werden kann: Es können nur Übergangswahrscheinlichkeiten für Lokationshistorien angegeben werden, die überhaupt irgendwann beobachtet wurden. D.h., eine beobachtete Bewegung muss sich vollkommen mit dem Beginn eines zuvor beobachteten Musters decken. [6]

Auf Markovketten basierende Vorhersagetechniken sind weitverbreitet. Markovketten bieten eine leicht verständliche Sicht auf das System und können angewandt werden, wenn Lokationen in eine endliche Menge nichtüberschneidender Lokationen unterteilt werden können. [8]

### 3.2 Feed-Forward Neuronale Netze

Neuronale Netze sind formale mathematische Modelle, welche biologische neuronale Strukturen imitieren. Mit Beginn der ersten Neuronenmodelle in den 1940er Jahren wurden sie eine der beliebtesten Werkzeuge, Aufgaben mithilfe von künstlicher Intelligenz zu lösen. Die Lernfähigkeit erlaubt es Neuronalen Netzen eine Vielfalt von Problemen wie z.B. Mustererkennung und Funktionsapproximierung zu lösen. [8]

Ein großer Vorteil künstlicher Neuronaler Netze ist ihre Fähigkeit, anhand von Beispielen zu lernen. (Supervised Learning) Um ein Problem auf herkömmliche Weise zu lösen, muss eigentlich ein zugrundeliegendes Modell entworfen und eine Abfolge von Operationen festgelegt werden, welche den Lösungsalgorithmus bilden. Mit zunehmender Komplexität wird es jedoch immer schwerer oder sogar unmöglich, einen deterministischen Algorithmus zu entwickeln. Bei Neuronalen Netzen ist man nicht gezwungen, einen lösenden Algorithmus zu einem bestimmten Problem vorzugeben. Stattdessen wird dem Neuronalen Netz eine Vielzahl verschiedener Beispiele zur Verfügung gestellt, um von ihnen zu lernen und zu generalisieren. Das Netz extrahiert die Information aus den Trainingsdaten und baut so implizit ein bestimmtes Modell des Problems. Anders ausgedrückt entwickelt das Neuronale Netz von alleine einen Algorithmus zum Lösen des Problems. Die Fähigkeit eines Neuronalen Netzes, komplexe Probleme mithilfe einer Menge von Beispielen zu lösen, ist vielversprechend. [5]

Dieser Lernprozess kann im Kontext Neuronaler Netze betrachtet werden als das Problem, die Gewichte der Neuronenübergänge so zu verändern, dass eine bestimmte Aufgabe zufriedenstellend erfüllt werden kann. Das Netz passt die Übergangsgewichte anhand verfügbarer Trainingsdaten an. Die Leistung des Neuronalen Netzes kann durch iterative Veränderung der Gewichte einfach verbessert werden. Gerade diese Fähigkeit Neuronaler Netze, automatisch aus Beispielen zu lernen, macht sie attraktiv. Neuronale Netze (im Folgenden mit NN abgekürzt) lernen zugrundeliegende Regeln, wie z.B. Eingabe-Ausgabe-Abhängigkeiten, aus der zur Verfügung gestellten Menge, im besten Falle, repräsentativer Beispiele. Das ist einer der großen Vorteile von künstlichen NNs gegenüber traditionellen Expertensystemen. NNs können bei gegebener Rechenkapazität diese Lernfähigkeit dazu nutzen, um zu generalisieren. Generalisierung bezieht sich auf die Fähigkeit Neuronaler Netze, passende Ausgaben für Eingaben zu berechnen, welche zur Trainingszeit nicht gesehen wurden. Aus diesem Grund werden künstliche NNs benutzt, um Bewegungen mobiler Nutzer zu predizieren. [21]

Wie in Abbildung 3 gezeigt, sind die Neuronen des Feed-Forward Neuronalen Netzes in Schichten organisiert, mit Verbindungen von jeweils einer Schicht zur nächsten. Ein Eingabevektor aktiviert die Eingabeschicht und breitet sich durch die mittlere Schicht hindurch an die Ausgabeschicht weiter. Da die Ausgabe eines FFNNs nur von der aktuellen Eingabe abhängt und insbesondere keine Abhängigkeit zu vorherigen Eingaben besteht, sind FFNNs von der Architektur her eher als Musterklassifikatoren als für Sequenzvorhersagen geeignet. Ein FFNN mit einer bestimmten Menge von Gewichten stellt eine Funktion dar, welche den Eingabevektor auf einen Ausgabevektor abbildet. Durch das Verändern der Gewichte kann ein FFNN verschiedene Funktionen darstellen. Tatsächlich wurde in [11] bewiesen, dass ein FFNN mit einem einzelnen Hidden Layer ausreichender Größe und nichtlinearer Einheiten (z.B. tanh oder Sigmoid) beliebige stetige Funktionen auf einem kompakten Intervall approximieren kann. Aus die-

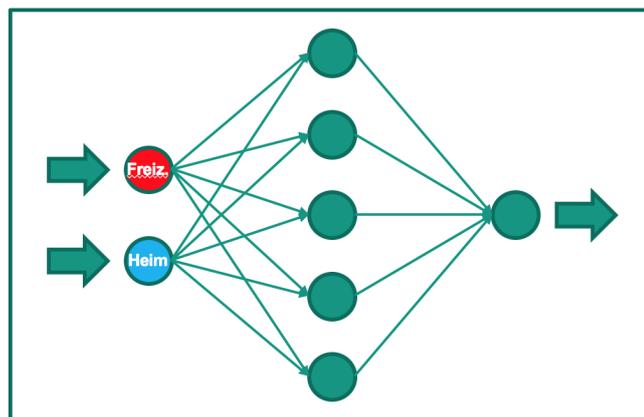


Abb. 3: Feed-Forward Neuronales Netz

sem Grund werden FFNNs auch universelle Funktionsapproximatoren genannt. [10]

Man Betrachte ein Feed-Forward Neuronales Netz (im Folgenden mit FFNN abgekürzt) mit  $I$  Eingabeneuronen, welche mit einem Eingabevektor  $x$  ( $|x| = I$ ) aktiviert werden. Jede Einheit der mittleren Schicht berechnet eine gewichtete Summe der Eingabeneuronen. Für das Neuron  $h$  der mittleren Schicht bezeichne man diese Summe als die Netzwerkeingabe in  $h$  und schreibe  $a_h$ . Anschließend wird die Aktivierungsfunktion  $\Theta$  angewandt, was schließlich zum Aktivierungswert  $b_h$  des Neurons führt. Sei  $w_{ij}$  das Gewicht des Übergangs von Neuron  $i$  zu Neuron  $j$ . Beschrieben durch diese Gleichungen:

$$a_h = \sum_{i=1}^I w_{ih} x_i \quad (1)$$

$$b_h = \Theta(a_h) \quad (2)$$

Einige Aktivierungsfunktionen sind in Abbildung 4 dargestellt. Die beliebtesten Aktivierungsfunktionen sind  $\tanh$  und die Sigmoid-Funktion. Da diese beiden Funktionen durch die Lineartransformation

$$\tanh(x) = 2\sigma(2x) - 1 \quad (3)$$

in Beziehung stehen, kann jede Funktion, welche mithilfe eines NNs mit  $\tanh$  als Aktivierungsfunktion auch mithilfe eines NNs mit  $\text{sigmoid}$  als Aktivierungsfunktion berechnet werden. Sie sind also zum großen Teil äquivalent. Ein Unterscheidungsmerkmal ist jedoch ihr Ausgabebereich. Dieser ist beim Sigmoid  $(0,1)$ , weshalb dieser bei Beschreibung von Wahrscheinlichkeiten eher benutzt werden sollte. Eine wichtige Eigenschaft dieser beider Aktivierungsfunktionen ist ihre Nichtlinearität. Nichtlineare NNs sind mächtiger als lineare, da diese nichtlineare Klassifikationsgrenzen und nichtlineare Gleichungen modellieren können. [10]

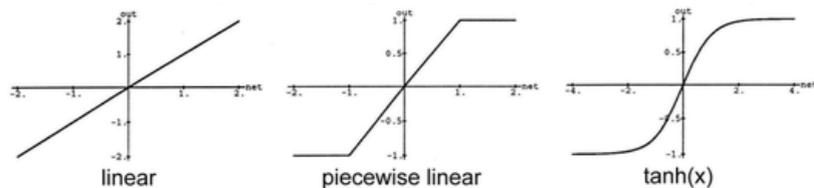


Abb. 4: Aktivierungsfunktionen

### 3.3 Recurrent Neuronales Netz

Die Idee hinter Recurrent Neuronalen Netzen ist es, Nutzen aus sequentieller Information zu ziehen. In einem gewöhnlichen Neuronalen Netz wird angenommen,

dass alle Eingaben (und Ausgaben) unabhängig voneinander sind. Für viele Aufgaben jedoch ist das nicht ausreichend. Wenn das nächste Wort in einem Satz prediziert werden soll, ist eine Kenntniss der vorhergehenden Worte notwendig. RNNs werden als recurrent bezeichnet, da sie denselben Berechnungsschritt für jedes Element einer Sequenz ausführen. Dabei ist die Ausgabe dann abhängig von Berechnungen vorhergehender Schritte. Eine andere Perspektive über die RNNs ist es, ihnen einen „Speicher“ zuzuschreiben, welcher Information über bisher Berechnetes enthält. In der Theorie können RNNs Informationen aus beliebig langen Sequenzen generieren, sind jedoch in der Praxis auf wenige vorhergehende Schritte beschränkt. [7] Ein einfaches RNN ist in Abbildung 5 skizziert. Ein Unterschied zu FFNNs ist, dass Neuronen der mittleren Schicht Kanten zu Neuronen derselben Schicht besitzen.

Der Vorwärts-Schritt eines RNNs ist derselbe wie in einem FFNN mit einer einzelnen mittleren Schicht, außer dass die Aktivierungen der mittleren Schicht von diesmal zwei Quellen stammen: Dem aktuellen Eingabevektor und den Aktivierungen derselben mittleren Schicht aus dem vorherigen Zeitschritt. Betrachte eine Eingabesequenz  $x$  der Länge  $T$  in ein RNN mit  $I$  Eingabeneuronen,  $H$  Neuronen in der mittleren Schicht und  $K$  Ausgabeneuronen. Sei  $x_i^t$  der Wert von Eingabe  $i$  zum Zeitpunkt  $t$  und seien  $a_j^t$  und  $b_j^t$  jeweils die Netzwerkeingabe zu Neuron  $j$  zum Zeitpunkt  $t$  und die Aktivierung von Neuron  $j$  zum Zeitpunkt  $t$ . Für die Neuronen im Hidden Layer gilt dann:

$$a_h^t = \sum_{i=1}^I w_{ih} x_i^t + \sum_{h'=1}^H w_{h'h} b_{h'}^{t-1} \quad (4)$$

Danach werden nichtlineare, differenzierbare Aktivierungsfunktionen genauso wie beim FFNN angewandt:

$$b_h^t = \Theta_h(a_h^t) \quad (5)$$

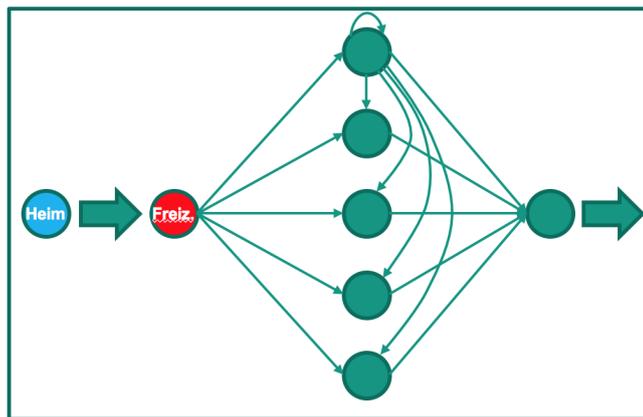


Abb. 5: Recurrent Neuronales Netz

Die komplette Sequenz der Aktivierungen der mittleren Schicht kann berechnet werden, indem bei  $t=1$  begonnen wird und rekursiv die obigen Anweisungen nacheinander ausgeführt werden. [10]

RNNs scheinen besonders gut geeignet zu sein für Sequenzvervollständigung, da Wissen über vorher besuchte Orte im Allgemeinen besonders wichtig sind für eine präzise Vorhersage. Es muss jedoch angemerkt werden, dass RNNs schwerer zu trainieren sind als ihre simplizistischen Alternativen, die FFNNs. [22]

## 4 Implementierung Neuronale Netze

### 4.1 Kodierung der Ein- und Ausgabevektoren

Es gibt eine feste Gesamtanzahl an Lokationen, jede mit einer individuellen Nummer, beginnend bei 1. Es gibt verschiedene Ansätze, die besuchten Lokationen zusammen mit den Tageszeit- und Wochentags-Features auf einen Eingabevektor abzubilden:

- Man leite den numerischen Wert als solches in einen einzelnen Neuron der Eingabeschicht.
- Für jedes Paar von (Lokation, Zeitschritt) soll ein separates Neuron der Eingabeschicht (und der Ausgabeschicht) aktiviert werden.
- Man konvertiere den numerischen Code in eine binäre Darstellung und weise jedem Neuron der Eingabeschicht ein Bit des resultierenden Binärcodes zu.

Die erste der aufgezählten Möglichkeiten wird bei der Prediktion stetiger numerischer Werte verwendet. Dieser Ansatz hat einige Nachteile, wie z.B., dass so das Neuronale Netz nicht richtig generalisieren wird. Wenn z.B. das Netz eine unbekannte Eingabe bekommt, welches als Prediktion dann den Lokationen 1 und 5 dieselbe Wahrscheinlichkeit beimisst, so würde ein Wert zwischen 1 und 5, z.B. ein an 3 naher Wert, als Ausgabe generiert werden. Es ist unwahrscheinlich, dass ein zu 1 oder 5 naher Wert produziert wird. Diese Möglichkeit zerstört also die einzigartige Fähigkeit Neuronaler Netze zu generalisieren und muss deshalb verworfen werden.

Die Auswahl der zweiten der drei Möglichkeiten führt zu einer Größe der Eingabeschicht von  $\langle \text{Anzahl Lokation} \rangle \cdot \langle \text{Anzahl Zeitschritte} \rangle$ . Ein Eingabevektor würde dann so generiert werden, dass ein gegebener Raum durch das Setzen des jeweiligen Neurons auf 1 und aller anderer Neuronen auf 0 repräsentiert werden würde. Der Ausgabevektor würde in einer ähnlichen Art erzeugt werden. Das im vorigen Paragraphen beschriebene Problem ist damit behoben. Allerdings wächst selbst bei geringer Zunahme der Lokationen oder der Zeitschritte die Größe der Eingabeschicht stark an, was inpraktikabel ist.

Verwendet wurde in dieser Arbeit der dritte Ansatz, da dieser einen Kompromiss zwischen den ersten beiden Methoden darstellt. [22] Der verwendete Ansatz wurde konkret folgendermaßen umgesetzt. Der Ausgabevektor enthält die vorhergesagte Lokation in Form einer One-Hot-Kodierung. D.h., dass z.B.

bei insgesamt 10 Lokationen für die Lokation mit der Nummer 3 das 3. Bit eine 1 und alle anderen Bits eine 0 enthalten. Der Ausgabevektor ist dabei nur abhängig von der Anzahl der Lokationen. Somit hat sie sowohl beim FFNN, als auch beim RNN dieselbe Form.

Die Größe des Eingabevektors ist beim FFNN und RNN jeweils unterschiedlich. Diese werden im Folgenden separat betrachtet. Beim FFNN hängt die Größe des Eingabevektors zum Einen von der Gesamtzahl  $z_1$  der Lokationen ab, welche One-Hot-kodiert werden. Hier wird bewusst auf ein platzsparendes (weniger Eingabeneuronen), binäres Durchnummerieren verzichtet. Der Grund dafür ist, den Neuronalen Netzen mit jeweils nur einer mittleren Schicht den Aufwand der Dekodierung des Eingabevektors soweit zu erleichtern, dass jeweils die Aktivierung eines einzelnen Eingabeneurons eine bestimmte Lokation darstellt. So sollen die Modellierungskapazitäten der kleinen Netze ausschließlich für die Herstellung der Zusammenhänge zwischen Ein- und Ausgabevektor aufgebraucht werden. Zum Anderen hängt der Eingabevektor von der beobachteten Anzahl  $z_2$  der zuletzt besuchten Lokationen ab, welche jeweils mit  $z_1$  Bits kodiert werden. Bisher besitzt ein Eingabevektor also  $z_1 \cdot z_2$  Bits.

Wenn der Wochentag der Transition als Feature berücksichtigt wird, so wird dieser in One-Hot-kodierter Form mit einer Länge von 7 Bits an den Eingabevektor angehängt. Wird wiederum die Uhrzeit der Transition als Feature berücksichtigt, so wird dieser in One-Hot-kodierter Form mit einer Länge von  $\langle \text{Anzahl Partitionen Tageszeit} \rangle$  an den Eingabevektor angehängt. Die Maximale Größe des Eingabevektors und somit der Eingabeschicht des FFNNs ergibt sich so zu  $z_1 * z_2 + 7 + \langle \text{AnzahlPartitionenTageszeit} \rangle$ . In dieser Seminararbeit wurde eine Partitionierung in acht Teile vorgenommen.

Beim Recurrent Neuronalen Netz ist wichtig, dass er aufgrund seiner rückwärtsgerichteten Verbindungen in der Lage ist, sich die Reihenfolge zu merken, in welcher die Eingabe während der Trainingsphase betrachtet wird. Aus dieser Fähigkeit des Recurrent Netzes erwächst die Notwendigkeit, die Form des Eingabevektors zu einer Eingabematrix zu verändern, sodass pro Zeitschritt (der insgesamt  $z_2$  Zeitschritte) eine Spalte der Eingabematrix der Größe  $z_1$  betrachtet wird, was der ersten Lokation der Historie entspricht.

Wenn der Wochentag der Transition beim RNN als Feature berücksichtigt wird, so wird dieser in One-Hot-kodierter Form an jede Spalte der Eingabematrix angehängt. Ebenso wird mit der Uhrzeit als Feature vorgegangen. Dabei entspricht Datum und Uhrzeit dem Zeitpunkt der Transition von der aktuellen Lokation der Lokationshistorie zur vorherzusagenden Lokation.

Somit ergibt sich beim RNN eine Maximale Größe der Eingabematrix zu  $(z_1 + 7 + \langle \text{AnzahlPartitionenTageszeit} \rangle) * z_2$ . Dabei ist zu beachten, dass die Größe der Eingabeschicht gleich der Größe einer Spalte der Eingabematrix entspricht.

Beim Entwickeln von Vorhersagemodellen hat man die Freiheit, ein NN nach Belieben zu wählen. Obwohl RNNs sich besonders für eine Sequenzvorhersage bzw. -vervollständigung eignen, können FFNNs benutzt werden, um Selbiges zu tun. Dazu muss allerdings die zeitliche Verzögerung im Eingabevektor explizit

dargestellt werden. Tatsächlich ist nach [23] die Verwendung von FFNNs für diese Art von Aufgaben die gängige Variante.

## 4.2 Verwendete Programmierumgebung

Die praktische Implementierung dieser Seminararbeit wurde in der Programmiersprache Python innerhalb der für interaktive Datenanalyse optimierten Jupyter Notebook-Umgebung [14] umgesetzt. Die benötigten Module wurden selbst von Grund auf oder mithilfe von Open Source Bibliotheken implementiert. Während die Markovketten selbst implementiert wurden, ist für die Umsetzung der Neuronalen Netze die von Google entwickelte Bibliothek TensorFlow [20] benutzt worden. Zur Erleichterung der Datenvorverarbeitung wurde auf die NumPy-Bibliothek [2] zugegriffen, welche eine einfache Handhabung von multidimensionalen Arrays gewährleistet. Die Evaluierung der generierten Daten wurde unter Zuhilfenahme der Datenstruktur-Bibliothek Pandas [3] und die auf Matplotlib [1] basierende Bibliothek Seaborn [4] zur statistischen Datenvisualisierung vorgenommen.

Die Implementierung, welche sich aus der Datenvorverarbeitung, dem Trainieren und Evaluieren der Vorhersagemodelle zusammensetzt, hat das Ausmaß von etwa 1000 Zeilen Code. Das rechenaufwändige Ausführen wurde mittels der Batch-Software HTCondor auf ein Rechnernetz ausgelagert, auf welchem mit 2 Rechenkernen und 16 GB RAM in etwa 24 Stunden trainiert und getestet wurde. Die Ergebnisse wurden in einem Pandas-DataFrame gespeichert, welche sich besonders einfach modifizieren lässt.

## 4.3 Aufzeichnen der Bewegungen

Die Generierung der Bewegungsdaten für die Vorhersagemodelle wurde in einem zeitlichen Rahmen von 12 Wochen in einer Tagebuch-Form vorgenommen: Im Laufe von sechs Wochen ist jeweils abends niedergeschrieben worden, an welchen Lokationen die Person sich an jenem Tag mindestens 15 Minuten aufgehalten hat. Weitere sechs Wochen wurden mithilfe eines detailliert ausgefüllten Kalenders (Google Calendar) rekonstruiert. Ein kurzer Abschnitt dieser Bewegungsdaten ist in Abbildung 6 zu sehen. Ein Tag beginnt mit dem zugehörigen Datum. Die Transitionen dieses Tages belegen jeweils eine Zeile. Diese Zeile beginnt mit der Uhrzeit, welche den Transitionszeitpunkt markiert. Die Transition selbst ist durch eine Abbildung der Form  $\langle \text{Startlokation} \rangle \rightarrow \langle \text{Ziellokation} \rangle$  gegeben.

## 4.4 Datenvorverarbeitung

Im Datenvorverarbeitungsschritt werden aus den Bewegungsdaten Trainings- und Testdaten generiert. In einem ersten Schritt werden dazu die Transitionen zu einer einzigen Transitionskette konkateniert. Es entsteht ein Array, siehe Abbildung 7. Nun wird jeder Lokation in einem Zwischenschritt eine individuelle ID zugeordnet, um diese im letzten Schritt in eine One-Hot kodierte Form

```

27.10.16
08:30 Heim -> TECO
13:30 TECO -> Mensa
14:00 Mensa -> AKK
14:15 AKK -> TECO
18:00 TECO -> Heim

28.10.16
08:40 Heim -> TECO
13:30 TECO -> Mensa
14:00 Mensa -> AKK
14:15 AKK -> Heim
17:00 Heim -> Freunde_HD
    
```

Abb. 6: Aufzeichnen der Bewegungen

umzuwandeln. In Abbildung 7 entspricht jede Zeile des letzten Teilbildes der One-Hot-Kodierung einer Lokation.

In diesem Zustand kann ein Vorhersagemodell trainiert werden: Ein Fenster, dessen Größe der festgelegten Historienlänge entspricht, umfasst die Kodierungen der zuletzt besuchten Lokationen. Gleichzeitig wird die Kodierung der nächsten Lokation dem Vorhersagemodell als Wunschvorhersage präsentiert. Dieses Gesamtfenster, dargestellt als geschweifte Klammern in Abbildung 7 und außerdem hier mit einer Historienlänge von 3 und somit einer Gesamtlänge von  $3 + 1 = 4$ , wird im Trainingsprozess eines Vorhersagemodells entlang der Transitionskette bewegt. Man sieht, dass im Trainingsprozess tagesübergreifende Historien mitberücksichtigt werden.

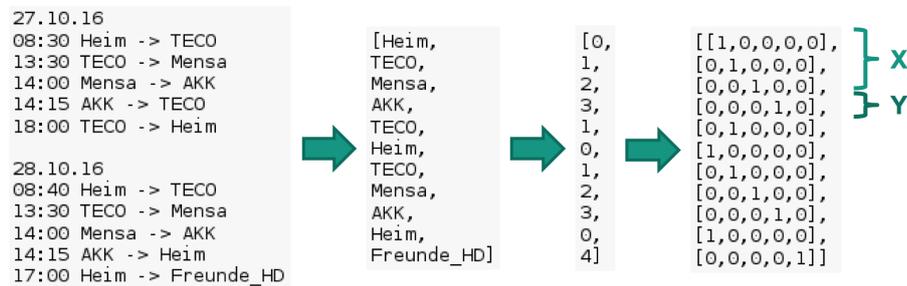


Abb. 7: Datenvorverarbeitung

#### 4.5 Konfiguration Vorhersagemodell

##### Tageszeit

Das Tageszeit-Feature kann optional bei der Evaluierung verwendet werden. Dabei wird der Zeitraum zwischen 5 Uhr und 22 Uhr in acht gleichgroße Zeitintervalle geteilt. Das erste Zeitintervall wird um das Intervall von 0 Uhr

bis 5 Uhr erweitert und das letzte Zeitintervall wird um das Intervall von 22 Uhr bis 24 Uhr erweitert. Jeder Zeitpunkt kann so auf eine One-Hot-Kodierung der Länge acht abgebildet werden. Diese sind die Standardwerte für die in der Evaluierung benutzten Tageszeit-Features. Es können sowohl der Start- und Endzeitpunkt für die Partitionierung, als auch die Anzahl der Teile verändert werden.

### **Wochentag**

Das Wochentag-Feature kann ebenfalls optional bei der Evaluierung verwendet werden. Der Wochentag wird dabei auf einen One-Hot-Vektor der Länge sieben abgebildet.

### **Historienlänge**

Die Historienlänge gibt die Anzahl der zuletzt besuchten Lokationen an. Beispielsweise wird bei einer Historienlänge von 2 die nächste Lokation unter Berücksichtigung der aktuellen und der vorherigen Lokation vorhergesagt. Es muss eine Historienlänge angegeben werden. Bei Fehlen dieser Information wird als Standardwert 2 angenommen.

### **Trainingsanteil**

Der Trainingsanteil gibt an, auf welchem Teil der Bewegungsdaten (beginnend mit der ersten Transition) die Vorhersagemodelle trainiert werden sollen. Der komplementäre Anteil wird als Testdatensatz bezeichnet und zur Auswertung des Vorhersagemodells benutzt. Dieser Wert kann beliebig gewählt werden. In dieser Arbeit wird ein Trainingsanteil von 70% gewählt.

### **Lernrate**

Die Lernrate ist ein Faktor, welcher den Grad der Änderung der Kantengewichte eines Neuronalen Netzes beschreibt. In dieser Arbeit wird als Lernrate 0.01 verwendet.

### **Epochen**

Die Epochenzahl gibt an, mit wievielen Durchgängen über den Datensatzes die Neuronalen Netze trainiert wurden.

### **Batch-Größe**

Eine Batch bezeichnet die Anzahl der Trainingsbeispiele, nach welcher die Parameter angepasst werden. Hier wird die geringste Batch-Größe 1 verwendet, da aufgrund des kleinen Datensatzes keine größeren Performance-Einbußen zu befürchten sind.

### **Größe mittlere Schicht**

In dieser Seminararbeit werden nur Neuronale Netze mit *einer* mittleren Schicht verwendet. Während die Größe der Ausgabeschicht von der Anzahl der Lokationen und die Eingabeschicht zusätzlich von den ausgewählten Features abhängt, kann die Größe der mittleren Schicht festgelegt werden.

### Gütemetrik

Es wird die Genauigkeitsmetrik, englisch Accuracy, verwendet.

## 5 Evaluierung

Die Vorhersagemodelle wurden mit unterschiedlichen Konfigurationen trainiert und ausgewertet. Dabei ist anzumerken, dass manche dieser Parameter nur die Neuronale Netze beeinflussen. In Abbildung 8 ist in einer geschachtelten Schleife zu sehen, welche Werte der jeweilige Parameter annimmt. Es ist zu sehen, dass für die Größe der mittleren Schicht die Werte [32, 64, 128] gewählt wurden, welche sich an Werten in anderen Papers orientieren. Die möglichen Werte für die Epochen sind [10, 20, 50, 100, 150]. Die Features Wochentag und Tageszeit wurden sowohl in aktivierter, als auch in deaktivierter Form betrachtet. Für die Historienlänge wurden die ganzen Zahlen im Intervall von 1 bis 15 berücksichtigt. Es finden also insgesamt  $3 \cdot 5 \cdot 2 \cdot 2 \cdot 15 = 900$  Aufrufe der Funktion `calc_results_for_params` statt.

```

for f_hidden in [32, 64, 128]:
    for f_epoch in [10, 20, 50, 100, 150]:
        for f_weekday in [True, False]:
            for f_time in [True, False]:
                for look_back in range(1,16,1):
                    dict_result = calc_results_for_params(training_percent=0.7,
                                                         look_back=look_back,
                                                         training_epochs_ffnn=f_epoch,
                                                         training_epochs_rnn=f_epoch,
                                                         feature_time=f_time,
                                                         feature_weekday=f_weekday,
                                                         n_hidden=f_hidden,
                                                         n_hidden_1=f_hidden)

                    df_main = df_main.append(dict_result, ignore_index=True)

date_and_time_str = str(datetime.today()).split('.')[0].replace(' ', '_')
df_main.to_pickle('./dataframe_training_70_for_loop_'+date_and_time_str)

```

Abb. 8: Training mit verschiedener Konfiguration.

Bei einmaligem Ausführen von `calc_results_for_params` werden jeweils eine Ausprägung dieser Parameter entgegengenommen, um die Vorhersagemodelle mit den Trainingsdaten (70%) zu trainieren und die Genauigkeitswerte mit den Testdaten (30%) zu generieren. Dabei werden pro Aufruf der Funktion zwei Markovketten trainiert: Einmal auf den Lokationsdaten und einmal auf den Lokationstypdaten. Ebenso werden in diesem Aufruf zwei Feed-Forward Neuronale Netze und zwei Recurrent Neuronale Netze auf den Lokations- und Lokationstypdaten trainiert und ausgewertet. Insgesamt werden also  $900 \cdot 4 = 3600$  Neuronale Netze und  $900 \cdot 2 = 1800$  Markovketten evaluiert. Als Rückgabewert von `calc_results_for_params` werden neben den Genauigkeitswerten die entgegengenommene Konfiguration ausgegeben, welche dann sukzessive in `df_main`

gespeichert werden. Dieses wird nach den Schleifendurchläufen im aktuellen Verzeichnis abgelegt.

In den Abbildungen 9 bis 11 sind die Ergebnisse zusammengefasst. Die Vorhersagegenauigkeit für eine bestimmte Historienlänge ist als 95% Konfidenzintervall in Form eines senkrechten Striches und einem Mittelwert darauf abgebildet. In Abbildung 9 sehen wir beispielsweise, dass die auf den Lokationen trainierten (blaue Kurve) Markovketten bei einer Historienlänge von 2 bei 95% der Konfigurationen eine Vorhersagegenauigkeit zwischen 0.35 und 0.51 erzielt. Die Vorhersage auf den Lokationstypen ist bei allen Historienlängen genauer als die auf den Lokationen.

Dabei ist die mit steigender Historienlänge abnehmende Vorhersagefähigkeit der Markovketten wie folgt zu begründen: Nach dem Trainingsprozess sagt eine Markovkette für einen bestimmten Lokationen-Historienverlauf die am häufigsten beim Trainingprozess zu diesem Historienverlauf zugehörige nächste Lokation vorher. Da bei steigender Historienlänge  $h = 1, \dots, 15$  und einer Lokationsanzahl von  $L = 36$  (bzw. einer Lokationstypanzahl von 6) die Anzahl der möglichen Historienverläufe  $A = L^h$  exponentiell steigt, nimmt (zusammenhängend mit der Beschaffenheit einer Markovkette) auch die Zahl der nicht mehr richtig vorhergesagten Historienverläufe zu. Auch die maximalen Vorhersagegenauigkeiten der Markovketten für jede Historienlänge auf den Lokationstypdaten ist größer als auf den Lokationsdaten, siehe Abbildung 12. Von Abbildung 13, Zeilen 0 und 5 ist zu entnehmen, dass die maximale Vorhersagegenauigkeit durch eine Markovkette sowohl auf den Lokations-, als auch auf den Lokationstypdaten bei einer Historienlänge von 1 erreicht wurden. Die maximale Genauigkeit auf den Lokationsdaten beträgt 94.2%, auf den Lokationstypdaten 100%.

In Abbildung 10 sind die Vorhersagegenauigkeiten der Feed-Forward Neuronalen Netze dargestellt. Auch hier ist zu erkennen, dass die Vorhersage von Lokationstypen größer ist als die von Lokationen selbst. Bei niedrigstelliger Historienlänge nimmt die durchschnittliche Genauigkeit mit steigender Historienlänge zu. Abbildung 12 und Zeile 10 in Abbildung 13 ist zu entnehmen, dass auf den Lokationen die maximale Vorhersagegenauigkeit von 59.4% durch ein FFNN mit einer Historienlänge von 2, deaktivierten Zeit- und Wochentags-Features, 100 Trainingsepochen und 32 Neuronen in der mittleren Schicht erreicht werden. Auf den Lokationstypen wird die maximale Vorhersagegenauigkeit von 78.6% durch ein FFNN mit einer Historienlänge von hier 9, deaktivierten Zeit- und Wochentags-Features, nur 10 Trainingsepochen und wieder 32 Neuronen in der mittleren Schicht erreicht.

Die Ergebnisse der Recurrent Neuronalen Netze werden durch Abbildungen 11, 12 und 13 dargestellt. Ein Blick auf Abbildung 11 zeigt, dass die Vorhersage auf den Lokationstypen auch bei RNNs größer ist als die von Lokationen. Im Folgenden sollen Lokationen betrachtet werden. Bei einem Vergleich der blauen Kurven in den Abbildungen 10 und 11 fällt auf, dass die FFNNs im Durchschnitt besser vorhersagen können als die RNNs. Bei einem Blick auf Abbildung 12 mit den maximalen Vorhersagegenauigkeiten (rote und gelbe Kurven) fällt allerdings auf, dass für jeweils dieselbe Historienlänge das der maximale RNN-Wert durch-

gehend größer ist als der maximale FFNN-Wert. Auf den Lokationstypen (violette und hellblaue Kurven) sind die FFNNs und RNNs gleichauf. Außerdem zu den RNNs: Zeile 20 in Abbildung 13 ist zu entnehmen, dass auf den Lokationen die maximale Vorhersagegenauigkeit von 64.0% durch ein RNN mit einer Historienlänge von 15, deaktivierten Zeit- und Wochentags-Features, 10 Trainingsepochen und 32 Neuronen in der mittleren Schicht erreicht werden. Auf den Lokationstypen wird die maximale Vorhersagegenauigkeit von 78.9% durch ein RNN mit einer Historienlänge von 7, deaktivierten Zeit- und Wochentags-Features, 150 Trainingsepochen und 64 Neuronen in der mittleren Schicht erreicht.

## 6 Zusammenfassung

Es wurde im Rahmen dieser Seminararbeit auf 12 Wochen umfassenden Bewegungsdaten gearbeitet. Diese Bewegungsdaten haben Datum und Uhrzeit mitgeführt und sind in Form von Transitionen zwischen Lokationen aufgezeichnet worden. Es sollte untersucht werden, wie sich die Vorhersagegenauigkeit auf Lokationstypen zu der auf Lokationen unterscheidet.

Dazu wurden zum einen auf den Lokationsdaten, zum anderen auf den daraus generierten Lokationstypdaten Markovketten, FFNNs und RNNs trainiert. Dabei sind wir zum Resultat gelangt, dass auf den Lokationstypen besser prediziert werden kann. Genauer: Im untersuchten Konfigurationsraum liefern FFNNs durchschnittlich bessere Genauigkeiten. Die Maximalgenauigkeit bei vorgegebener Historienlänge wird allerdings zumeist durch RNNs erreicht.

## 7 Ausblick

Die durchgehend deutlich bessere Vorhersagegenauigkeit auf den Lokationstypen kann dazu genutzt werden, die Lokationsvorhersage zu verbessern. In einem durch Mobilitätseigenschaften des Nutzers (Fußgänger, Fahrrad, Auto, öffentliche Verkehrsmittel) beeinflussten Radius kann beispielsweise die nächstgelegene Lokation des vorhergesagten Typs gewählt werden. Eine weitergehende Fragestellung ist, ob und in welchem Ausmaß ein Vorhersagemodell auf diese Art verbessert werden kann.

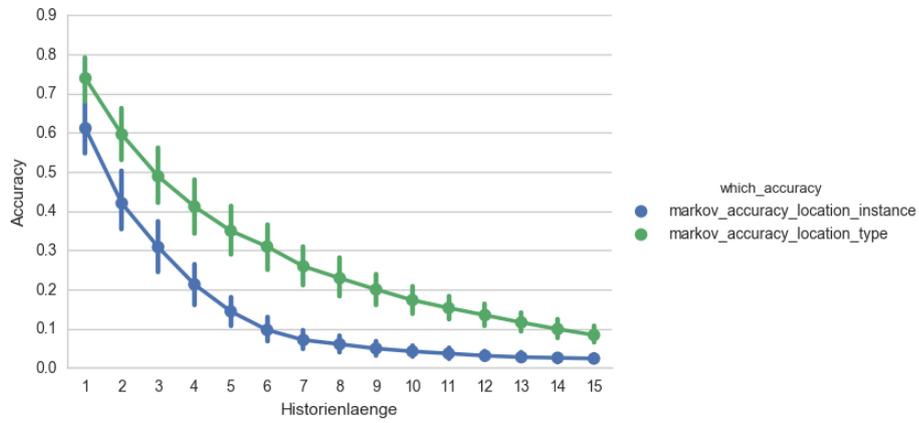


Abb. 9: Vorhersagegenauigkeit Markovketten.

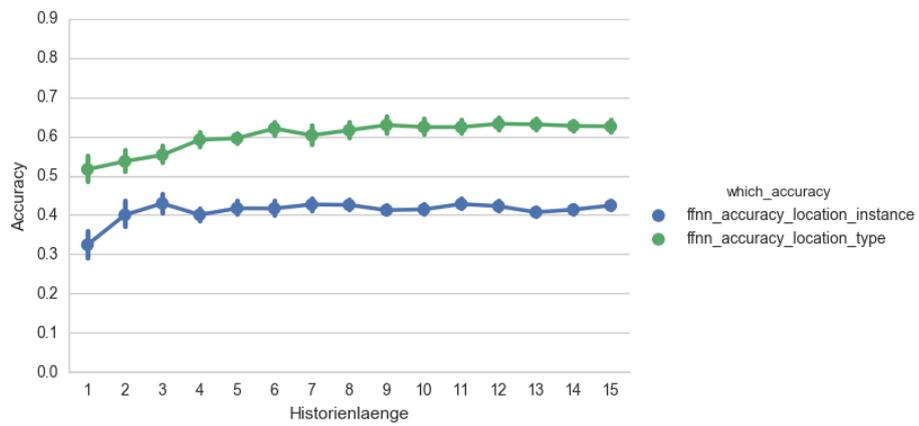


Abb. 10: Vorhersagegenauigkeit Feed-Forward Neuronale Netze.

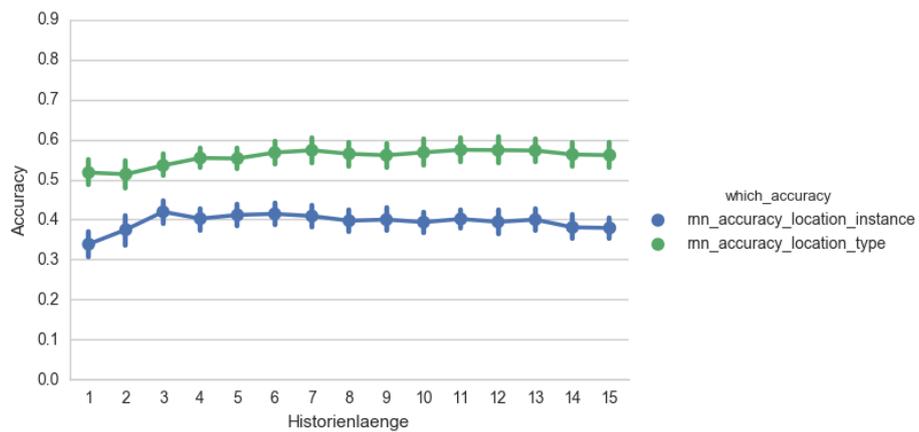


Abb. 11: Vorhersagegenauigkeit Recurrent Neuronale Netze.

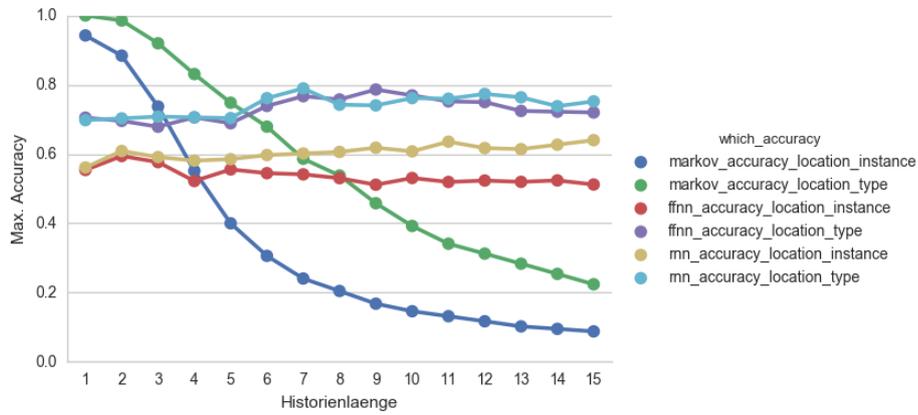


Abb. 12: Maximale Vorhersagegenauigkeiten.

	look_back	feature_time	feature_weekday	training_epochs	n_hidden	accuracy	which_accuracy
0	1.0	False	False	50.0	64.0	0.942446	markov_accuracy_location_instance
1	1.0	False	False	150.0	64.0	0.942446	
2	1.0	False	False	20.0	32.0	0.942446	
3	1.0	False	False	100.0	64.0	0.942446	
4	1.0	False	False	50.0	32.0	0.942446	
5	1.0	False	False	100.0	32.0	1.000000	markov_accuracy_location_type
6	1.0	False	False	150.0	128.0	1.000000	
7	1.0	False	False	50.0	32.0	1.000000	
8	1.0	False	False	150.0	64.0	1.000000	
9	1.0	False	False	10.0	32.0	1.000000	
10	2.0	False	False	100.0	32.0	0.594203	ffnn_accuracy_location_instance
11	2.0	False	False	150.0	32.0	0.586957	
12	2.0	False	False	150.0	128.0	0.579710	
13	2.0	False	False	150.0	64.0	0.579710	
14	2.0	False	False	20.0	32.0	0.579710	
15	9.0	False	False	10.0	32.0	0.786260	ffnn_accuracy_location_type
16	9.0	False	False	20.0	128.0	0.770992	
17	10.0	False	False	10.0	32.0	0.769231	
18	7.0	False	False	10.0	32.0	0.766917	
19	7.0	False	False	150.0	128.0	0.766917	
20	15.0	False	False	10.0	32.0	0.640000	mn_accuracy_location_instance
21	11.0	False	False	10.0	32.0	0.635659	
22	14.0	False	False	10.0	32.0	0.626984	
23	9.0	False	False	10.0	32.0	0.618321	
24	9.0	False	False	10.0	64.0	0.618321	
25	7.0	False	False	150.0	64.0	0.789474	mn_accuracy_location_type
26	7.0	False	False	50.0	32.0	0.774436	
27	12.0	False	False	20.0	64.0	0.773438	
28	12.0	False	False	50.0	128.0	0.765625	
29	13.0	False	False	10.0	32.0	0.763780	

Abb. 13: Jeweils fünf größte Vorhersagegenauigkeiten pro Modell und Konfiguration.

## Literatur

1. Matplotlib - introduction, <http://matplotlib.org>, online; abgerufen am 22.03.2017
2. Numpy for matlab users, <https://docs.scipy.org/doc/numpy-dev/user/numpy-for-matlab-users.html>, online; abgerufen am 22.03.2017
3. Pandas - package overview, <http://pandas.pydata.org/pandas-docs/stable/overview.html>, online; abgerufen am 22.03.2017
4. Seaborn: statistical data visualization, <http://seaborn.pydata.org>, online; abgerufen am 22.03.2017
5. Person Movement Prediction Using Neural Networks, Reports / Technische Berichte der Fakultät für Angewandte Informatik der Universität Augsburg, vol. 2004-10. Universität Augsburg, Augsburg (2006)
6. Agrawal, D.: Semantic trajectory mining for location prediction (2011)
7. Britz, D.: Recurrent neural networks tutorial, part 1 - introduction to rnns, <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>, online; accessed 15-Dec-2016
8. Burstein, F., Brézillon, P., Zaslavsky, A. (eds.): Supporting Real Time Decision-Making: The Role of Context in Decision Support on the Move, Annals of Information Systems, vol. 13. Springer Science+Business Media LLC, Boston, MA (2011), <http://www.ebib.com/patron/FullRecord.aspx?p=645648>
9. Gambs, S., Killijian, M.O., del Prado Cortez, M.N.: Next place prediction using mobility markov chains. In: Haddadi, H., Yoneki, E. (eds.) the First Workshop. pp. 1-6
10. Graves, A.: Supervised sequence labelling with recurrent neural networks, Studies in computational intelligence, vol. Vol. 385. Springer, Berlin (2012), <http://www.loc.gov/catdir/enhancements/fy1409/2011940011-d.html>
11. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks 2(5), 359 - 366 (1989), [//www.sciencedirect.com/science/article/pii/0893608089900208](http://www.sciencedirect.com/science/article/pii/0893608089900208)
12. Ilya Sutskever, Oriol Vinyals, Quoc V. Le: Sequence to sequence learning with neural networks
13. Jens Biesterfeld, Elyes Ennigrou, Klaus Jobmann: Neural networks for location prediction in mobile networks. In: Fabozzi, F.J., Kothari, V. (eds.) Introduction to Securitization, pp. 1-12. John Wiley & Sons, Inc, Hoboken, NJ, USA (2008)
14. Loric, B.: Ipython: A unified environment for interactive data analysis, <https://www.oreilly.com/ideas/ipython-a-unified-environment-for-interactive-data-analysis>, online; abgerufen am 22.03.2017
15. Mahmoud, S., Lotfi, A., Langensiepen, C.: Behavioural pattern identification and prediction in intelligent environments. Applied Soft Computing 13(4), 1813-1822 (2013)
16. Mozer, M.C.: The neural network house: An environment hat adapts to its inhabitants. In: Proc. AAAI Spring Symp. Intelligent Environments. vol. 58 (1998)
17. Petzold, J., Bagci, F., Trumler, W., Ungerer, T.: Comparison of different methods for next location prediction. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006 Parallel Processing, Lecture

- Notes in Computer Science, vol. 4128, pp. 909–918. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
18. Plummer, E.A.: Time Series Forecasting with Feed-forward Neural Networks: Guidelines and Limitations. Master's thesis, Department of Computer Science and The Graduate School of The University of Wyoming (2000)
  19. Ryan, C., Brown, K.N.: Occupant location prediction using association rule mining
  20. Sandjideh, A.M.: Announcing tensorflow 1.0, <https://developers.googleblog.com/2017/02/announcing-tensorflow-1.0.html>, online; abgerufen am 22.03.2017
  21. Subramanian, J.V., Sadiq, M.A.K.: Implementation of artificial neural network for mobile movement prediction. In: Jahanbegloo, R. (ed.) Introduction to Nonviolence, pp. 1–10. Macmillan Education UK, London (2014)
  22. Tomáš MIKLUŠČAK\*, M.G.: Person movement prediction using artificial neural networks with dynamic training on a fixed-size training data set. In: Fabozzi, F.J., Kothari, V. (eds.) Introduction to Securitization, pp. 1–12. John Wiley & Sons, Inc, Hoboken, NJ, USA (2008)
  23. Wong, W., Xia, M., Chu, W.: Adaptive neural network model for time-series forecasting. *European Journal of Operational Research* 207(2), 807 – 816 (2010), [//www.sciencedirect.com/science/article/pii/S0377221710003784](http://www.sciencedirect.com/science/article/pii/S0377221710003784)

# In Home Sleep Monitoring

Mengying Yang\*

Advisor: Long Wang†

Karlsruher Institut für Technologie (KIT)  
Pervasive Computing Systems – TECO

\*mengying.yang@gmx.de

†wanglong@teco.edu

## ABSTRACT

Nowadays sleep monitoring is a hot topic, and home sleep monitoring is valuable because of its benefits comparing to clinical sleep monitoring. It is more suitable for normal users for a long term sleep monitoring. A lot of different devices of home sleep monitoring have been invented. In this paper the recent technologies about home sleep monitoring will be discussed. Four home sleep monitoring technologies will be mentioned as based on wireless signals, audio signals, pressure and accelerometer. All the technologies will be introduced from design to test, including its advantages and disadvantages.

## Keywords

Home sleep monitoring; wireless signals; audio signals; pressure; accelerometer

## 1 INTRODUCTION

Sleep plays an important role in their overall health. Having insufficient amount of sleep can easily cause fatigue and lack of concentration during the day. Besides the amount of sleep, the quality of sleep is also an important factor in maintaining a healthy life style. Clinical studies show that sleep is related to many serious diseases including diabetes, obesity and depression.

But sleep is a dynamic process that varies from day to day, and hence it is important to measure multiple nights of sleep for medical, research, and wellness reasons. Home monitoring devices offer the potential to provide a more realistic platform in which many nights of sleep data can be captured.

The present paper reviews recent developments in area of technology that can be used for home-based sleep assessment.

Since Sleep, a combination of physiologic and behavioral processes, is a period of inactivity and rejuvenation, some physical signals need to be monitored to evaluate sleep, such as human vital signs and movements during sleeping.

Tracking human vital signs such as breathing and heart rates during sleep is important as it can help to assess the general physical health of a person and provide useful clues for diagnosing possible diseases.

## **2 Sleep-Monitoring Based on Wireless Signals**

Traditional approaches (e.g., Polysomnography (PSG)) are limited to clinic usage. Recent radio frequency (RF) based approaches require specialized devices or dedicated wireless sensors and are only able to track breathing rate.

Emerging quantified-self apps and wearable devices allow people to measure and keep track of sleep duration, patterns and quality. However, these approaches are intrusive, placing a burden on the users to modify their daily sleep related habits in order to gain sleep data; for example, users have to wear cumbersome devices (e.g., a headband) or inform the app when they go to sleep and wake up. So, a wireless-signal sleep monitoring device is needed.

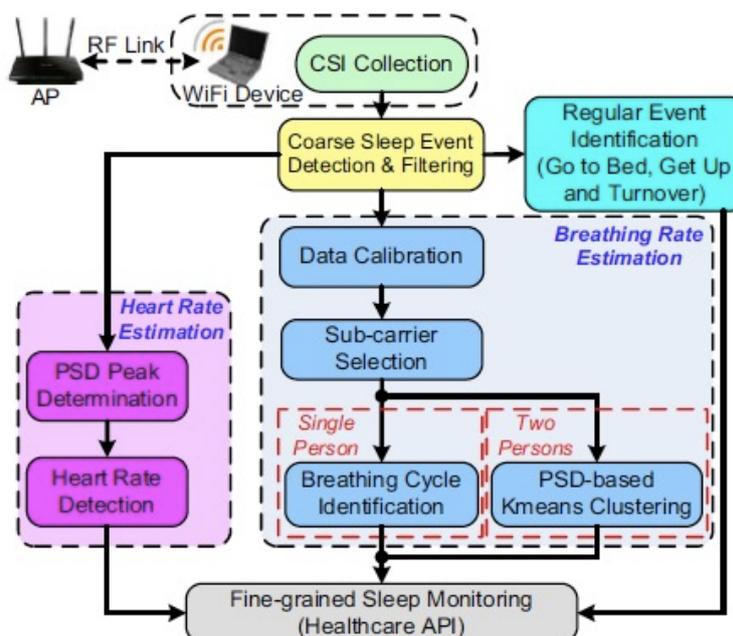
### **2.1 Idea of WiFi-based system**

In this work, they propose to track the vital signs of both breathing rate and heart rate during sleep by using off-the-shelf WiFi without any wearable or dedicated devices[1]. The system reuses existing WiFi network and exploits the fine-grained channel information to capture the minute movements caused by breathing and heart beats. The system thus has the potential to be widely deployed and perform continuous long-term monitoring. The developed algorithm makes use of the channel information in both time and frequency domain to estimate breathing and heart rates, and it works well when either individual or two persons are in bed. The extensive experiments demonstrate that the system can accurately capture vital signs during sleep under realistic settings, and achieve comparable or even better performance comparing to traditional and existing approaches, which is a strong indication of providing non-invasive, continuous fine-grained vital signs monitoring without any additional cost.



**Fig.1 WiFi devices, transmitter and receiver**

In the paper they set up an experiment, in which two MicaZ node are deployed, one as the transmitter (TX) and the other one as the receiver (RX). Now they take CSI into consideration. Channel State Information (CSI) which has gained much attention in wireless communications (i.e. to estimate the number of human beings within a region), and indoor localization. These successful applications do enlighten them on utilizing CSI for tracking a person's respiration, and this will be explored in the next section. In Fig.1, we can see the user is in between an WiFi transmitter and the receiver. The person does not carry any sensor in his body.

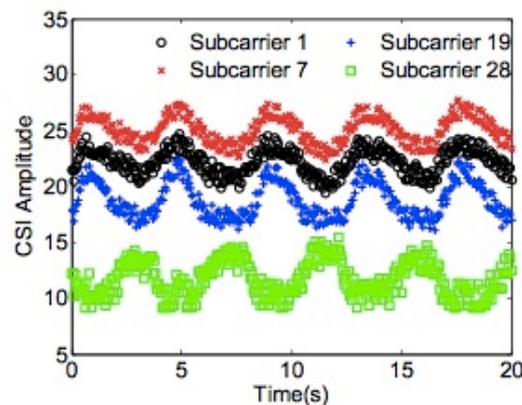


**Fig.2 Overview of system flow**

Fig.2 shows the structure of the system flow. The system receives WiFi signals and then go through three parts, to get the heart rate, breathing and other behaviors during sleeping.

While proliferating WiFi networks are usually used for wireless Internet access and connecting local area networks, such as an in-home WiFi network involving both mobile and stationary devices (e.g., laptop, smartphone, tablet, desktop, smartTV), they have great potential to sense the environment changes and capture the minute movements caused by human body [2]. Indeed, WiFi signals are affected by human body movements at various scales during sleep, such as large scale movements involving going to bed and turn over, minute movements including inhaling/exhaling for breathing and diastole/systole for heart beats. By extracting and analyzing the unique characteristics of WiFi signals, they could capture and derive the semantic meanings of such movements including both breathing rate and heart beats during sleep. they are thus motivated to re-use existing WiFi network to monitor the fine-grained vital signs during sleep as it doesn't require any dedicated/wearable sensors or additional infrastructure setup.

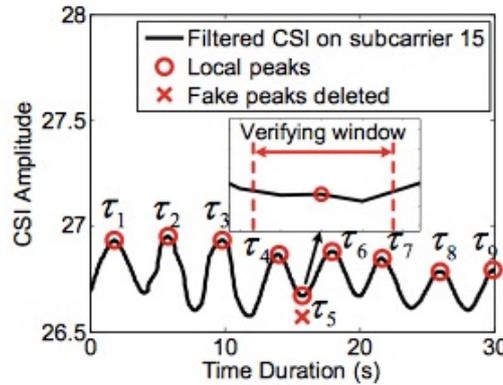
## 2.2 Principles behind WiFi-based system



**Fig.3 CSI amplitude of four subcarriers over time when a person is asleep.**

From Fig.3, it is observed that the CSI amplitude of these four subcarriers exhibits an obvious periodic up-and-down trend. Such a pattern could be caused by the person's breathing during sleep. This observation strongly suggest that they may achieve device-free fine-grained vital signs monitoring by leveraging the CSI from off-the-shelf WiFi devices.

It shows that the existing WiFi network can be reused to capture vital signs of breathing rate and heart rate through using only one AP and a single WiFi device. Such an approach can also be extended to non-sleep scenarios when the user is stationary.

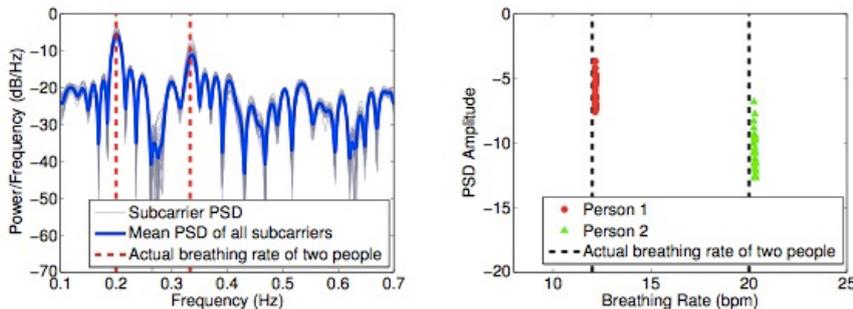


**Fig.4 Illustration of fake peak removal**

The proposed system extracts fine-grained channel state information (CSI) from off-the-shelf WiFi device to detect the minute movements and provide accurate breathing and heart rates estimation concurrently.

In Fig.4, the peak  $\tau_5$  has larger value than its two neighboring samples, yet, it is a fake peak among these nine identified peaks. In order to filter out the fake peaks, a threshold is applied to the minimum distance between two neighboring peaks based on human’s maximum possible breathing rate.

Specifically, adults usually breathe at 10-14 breathes per minute (bpm) [3], while new born babies breathe at around 37 bpm [4]. Therefore the range of breathing rates has being considered in the work to 10-37 bpm which covers a broad range including fast and slow breathing rates.

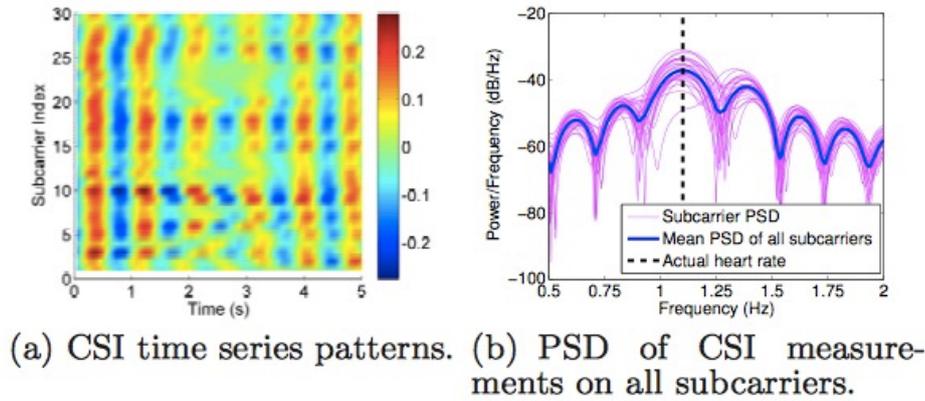


(a) PSD of CSI measurements. (b) K-means clustering of subcarriers’ peaks.

**Fig.5 Illustration of two people breathing at different frequencies (12bpm and 20bpm)**

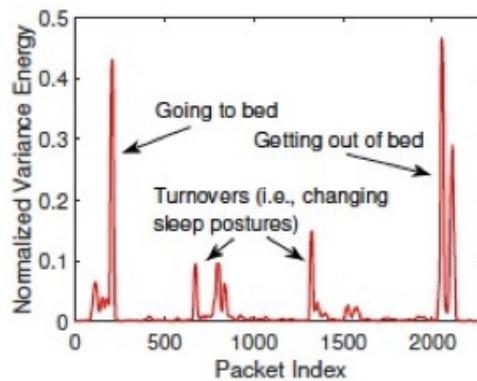
They also develop algorithms that have the capability to track breathing rates of a single person as well as two-person in bed cases, which cover typical in-home scenarios. From Fig.5, it is shown that a strong sinusoidal signal generates a peak at the frequency corresponding to the period of the sinusoidal signal in PSD. Therefore, the CSI amplitude measurements collected when two

persons in bed should present two strong peaks at the frequency corresponding to the breathing rate of two persons, respectively.



**Fig.6 Recovered heart beats by applying pass band filtering and PSD of CSI measurements**

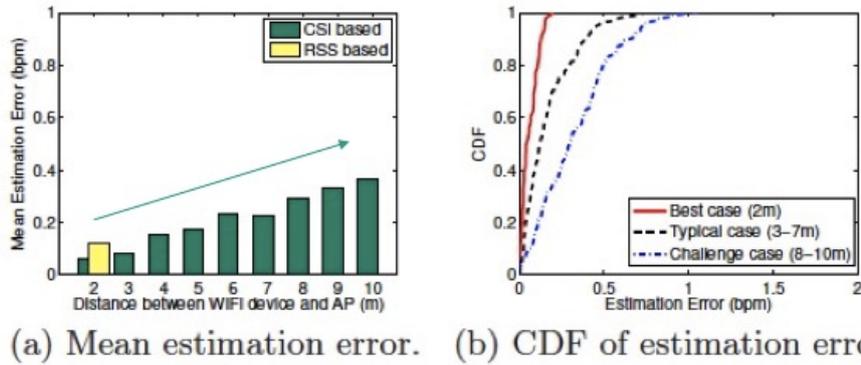
It is shown in Fig.6, that CSI measurements with the frequency range related to normal heart rate range of resting people (i.e., 60bpm to 80bpm which corresponds to 1Hz to 1.33Hz) will be separated and served as input to the Heart Rate Estimation.



**Fig.7 Short time energy of the variance of difference sleep events**

The sleep events or environmental changes involving large body movements (e.g., going to bed and turn over) result in much larger changes of CSI measurement than that of minute movements of breathing and heart beat. The large movements thus can be detected once the variance energy of the corresponding CSI measurements exceeds a particular threshold. Thus like in Fig.7, a threshold-based approach is used to determine whether a segment of CSI measurements contains sleep events/environmental changes or not by

examining the short-time energy of the moving variance of the CSI measurements.



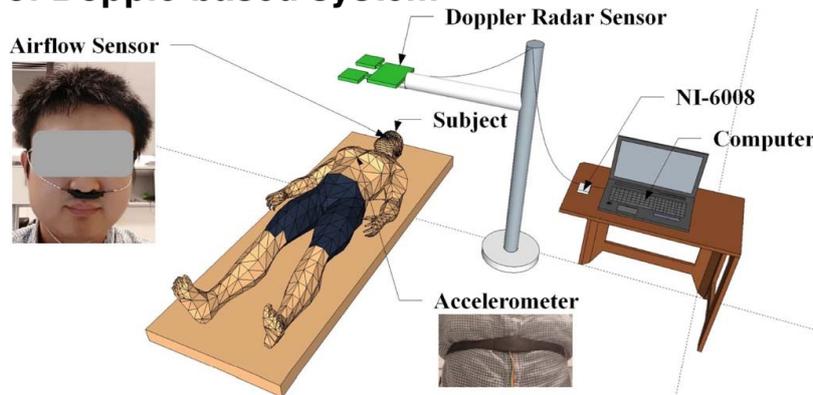
**Fig.8 Performance under different distances between WiFi device and AP**

Besides, extensive experiments in both lab and two apartments over a three-month period show that the system can achieve comparable or even better performance as compared to existing dedicated sensor based approaches. From Fig.8, the results show that, as the distance between WiFi device and AP grows, the error chance is also bigger.

### 2.3 Discussion of WiFi-based system

Comparing to other devices, the wifi-based sleep-monitoring system, is completely contactless, incurs low cost and is robust to low-light environments. But on the other hand, it is also Influenced by setup of wifi devices and sleep positions.

### 2.4 Idea of Dopple-based system

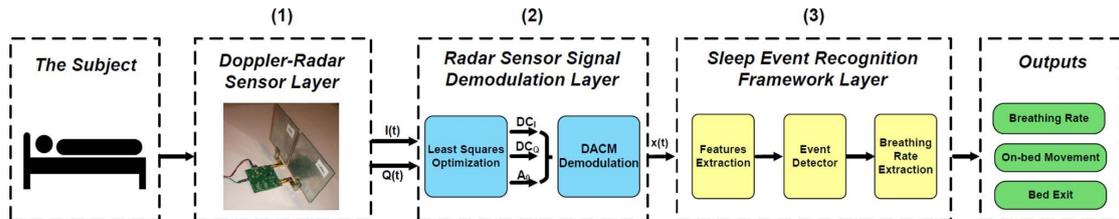


**Fig.9 The experimental setup: the subject lies his back on a mattress, and the radar locates on the top of chest of the subject with the distance of 1 meter. Meanwhile, an airflow sensor and an accelerometer sensor are attached to serve as the ground truth signal.**

A noncontact and cost-effective sleep monitoring system, named SleepSense, is proposed for continuous recognition of the sleep status, including on-bed

movement, bed exit, and breathing section. SleepSense consists of three parts: a Doppler radar-based sensor, a robust automated radar demodulation module, and a sleep status recognition framework[15].

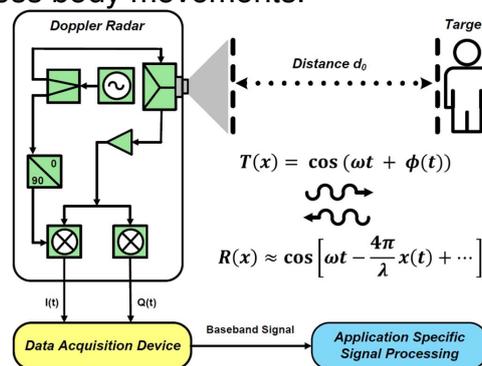
### 2.5 Principles behind Dopple -based system



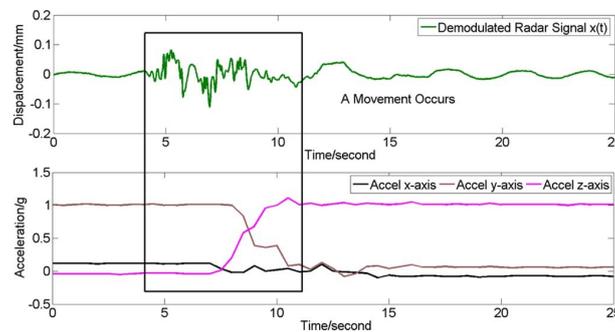
**Fig.10 The function block diagram of Sleepsense system. (1) Doppler radar sensor. (2) Radar signal demodulation. (3) Sleep status recognition framework.**

On the basis of breathing patterns and movement patterns, they define three sleep statuses: on-bed movement event, bed exit event, and breathing section. They focus on event recognition when people are already in the sleep state. The on- bed movement event is a movement such as turn over or arm trembling. The breathing section is when a subject is still on the bed (no observable body movements). The bed exit event refers to the bed exit movement that indicates interruption of the sleeping state.

The status detector categorizes these segments into three classifications: breathing section, on-bed movement, or bed exit. The status detector is a decision-tree based classifier. Based on the classification result, this layer also uses a breathing rate extracting algorithm to calculate the breathing rate when the subject has no gross body movements.



**Fig.11 The principle of Doppler radar sensor for human motion detection.**



**Fig.12** When a movement occurs, there is a large-magnitude and fast-frequency fluctuation in the demodulated signal. The  $x$ -axis, or  $y$ -axis, or  $z$ -axis value of the accelerometer changes correspondingly

According to the Fig.11 and 12, it is true that movements can cause some sounds. So Doppler radar is possible to monitor sleeping.

There are also some errors for on-bed movement and breathing section recognition, which is caused by the windowing issue. Specifically, because they employ the fixed length window in the segmentation, a movement may be incorrectly divided into several windows and lose the on-bed movement features, leading to misclassification.

## 2.6 Discussion of Dopple-based system

Dopple-based system is contactless and Unobtrusiveness.

The system also cares about user privacy, by avoiding recording sounds or videos.

But it costs extra and is not very sensitive to micro on-bed movements.

## 3 Sleep-Monitoring Based on Audio Signal

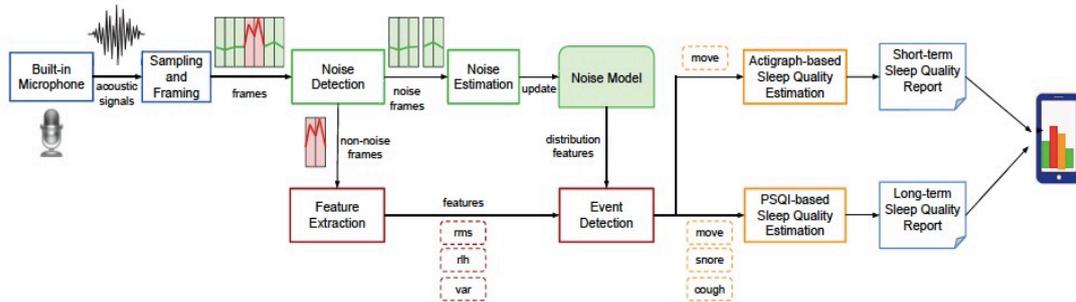
Smart devices, such as smartphones, have allowed us to monitor sleep at home more easily. Here is an audio-signal-based home sleep monitoring APP, iSleep, which is designed to be a sleep diary that provides the user real-time, fine-grained feedback to their sleep quality on a daily basis.

### 3.1 Idea of audio-signal-based system

iSleep – a practical system to monitor an individual's sleep quality using off-the-shelf smartphone[5]. iSleep uses the built-in microphone of the smart- phone to detect the events that are closely related to sleep quality, including body movement, cough and snore, and infers quantitative measures of sleep quality. iSleep adopts a lightweight decision-tree-based algorithm to classify various events based on carefully selected acoustic features, and tracks the dynamic ambient noise characteristics to improve the robustness of classification. By providing a fine-grained sleep profile that depicts details of sleep-related events,

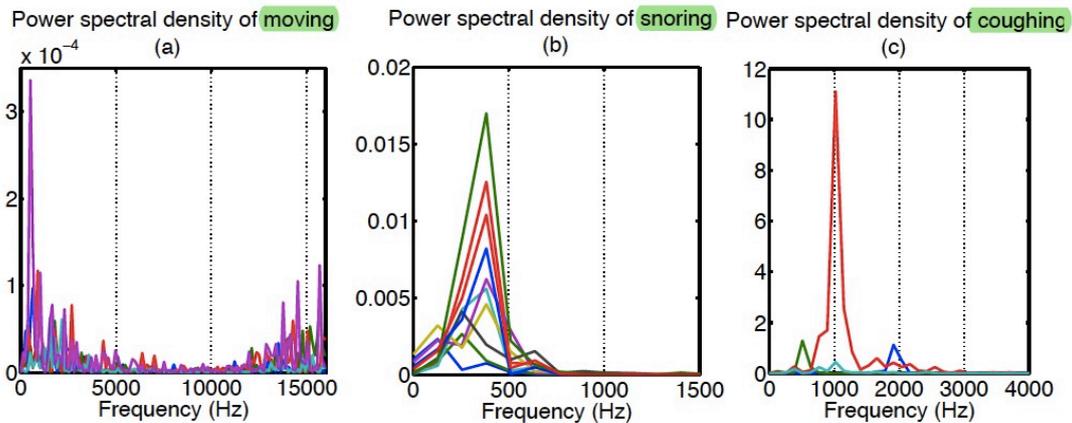
iSleep allows the user to track the sleep efficiency over time and relate irregular sleep patterns to possible causes.

### 3.2 Principles behind audio-signal-based system



**Fig.13** Architecture of iSleep System

From Fig.13, the iSleep can collect sound from microphone in smart phones, and through filter the left sounds will be detected into different behaviors during sleeping.



**Fig.14** (a), (b) and (c) show the power spectral density of typical moving, snoring and coughing events, respectively

In Fig.14, the different behaviours, have been separated according to different sound frequency, thus it is possible to monitor sleep with microphone in a smartphone.

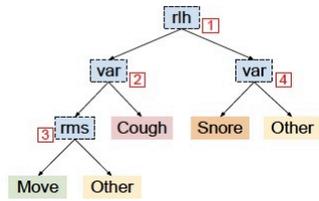


Figure 11: The decision tree for event detection. The splitting conditions are shown in Fig. 14(a), 14(b) and 14(c).

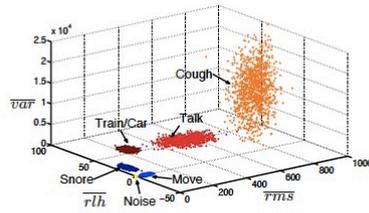


Figure 12: The sound feature vectors of different events in feature space.

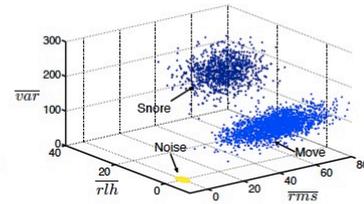


Figure 13: The sound feature vectors of snore, move and noise in the feature space.

### Fig.15

In Fig.15, it shows how to detect the behaviour according to the sound. They design a decision-tree based classifier to detect sleep-related events. Decision tree is robust to errors and widely used for classification problems with attribute-value pairs and discrete output values [6]. The middle one also shows the splitting conditions of the decision tree. The dotted rectangles indicate the splitting features, and the leaf nodes denote the classification results. The splitting features and thresholds are determined based on the information gain calculated using entropy.

Here are some selected benchmarks:

The ratio of low-band to high-band energies ( $r/h$ ).

The root mean square ( $rms$ ), which captures the loudness of sound.

The third feature is variance ( $var$ ), which reflects how far the amplitudes of acoustic signals within the frame are spread out. For example, the  $var$  of a frame associated with body movement is typically much lower than that of snoring or coughing.

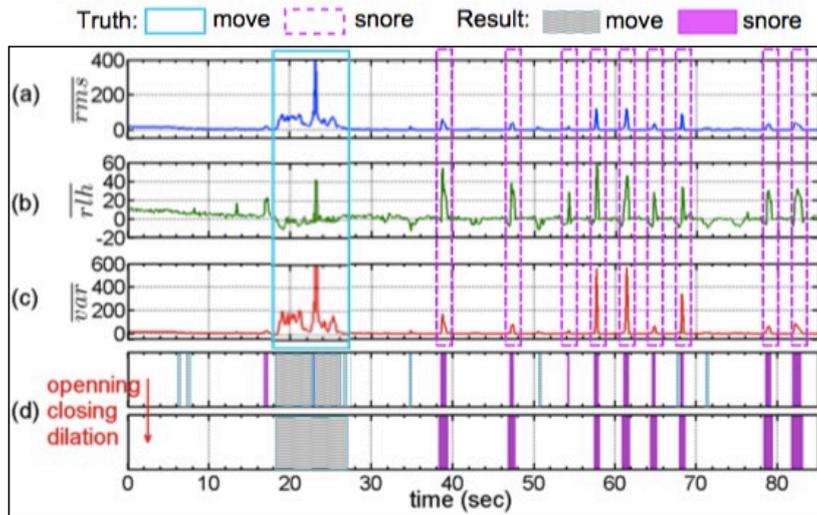
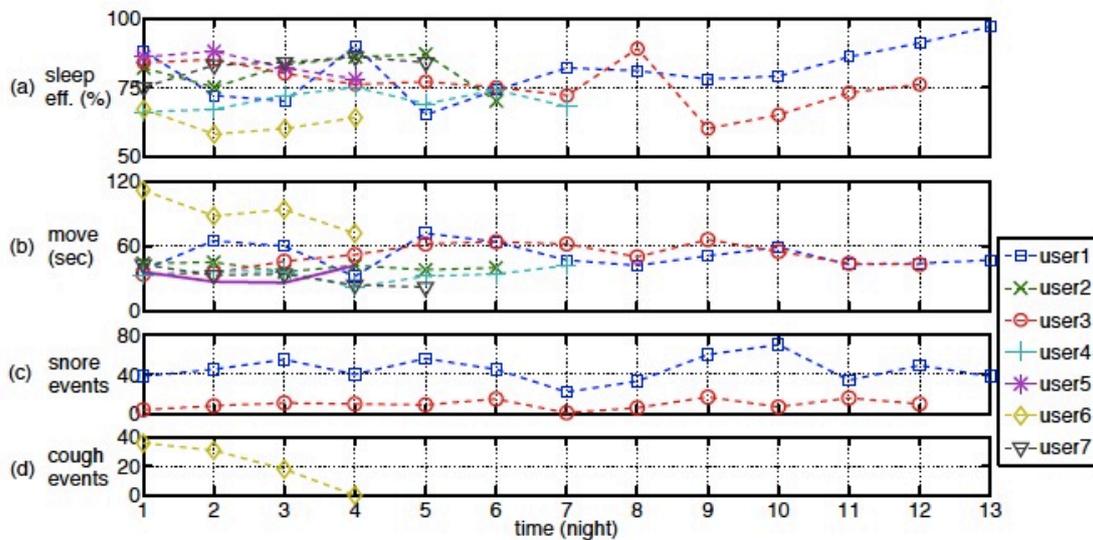


Figure 16: The event detection process of a 85-second acoustic signal that contains sound of body movement and snoring. (a), (b) and (c) show

the features for each frame, respectively. (d) shows the event detection results before and after opening, closing and dilation operations.

In Fig.16 the examples of different sounds have been showed. Sound of bodymovements or snoring have clear differences.



**Fig.17** The sleep efficiency and sleep events of different users during the long-term experiment. (a) The sleep efficiency at each night. (b) The total time (in seconds) of body movement at each night. (c) The number of snoring events at each night. (d) The number of coughing events at each night.

In Fig.17, the user results during a long term experiment is showed.

But after the sounds data are collected and detected into different behaviors, how to evaluate the sleep quality of users?

A1	Time to go bed at night
A2	Minutes taken to fall asleep
A3	Get-up time in the morning
A4	Hours of actual sleep per night
B1	Cannot sleep within 30 minutes
B2	Wake up in the middle of the night or early morning
B3	Cannot breath comfortably
B4	Cough or snore loudly

**Table 1:** Metrics from PSQI that iSleep uses to estimate the sleep quality.

Duration of Sleep	A4
Sleep Disturbance	B1, B2, B3, B4
Sleep Latency	A2, B1
Sleep Efficiency	A1, A3, A4

**Table 2:** The left column is the components in PSQI that can be derived from detected events. The right column is the metrics that are used to calculate the corresponding component score.

In iSleep implementation of actigraphy-based estimation, iSleep adopts similar method proposed in [7], where the sleep/wake state of a minute is determined by taking 4 previous minutes and 2 following minutes into account. The model takes the form:  $D = 0.125(0.15A_{-4} + 0.15A_{-3} + 0.15A_{-2} + 0.08A_{-1} + 0.21A_0 + 0.12A_{+1} + 0.13A_{+2})$

And the duration of each wake state lasts around 10 minutes. The sleep efficiency is defined as the ratio of actual sleep time to total in-bed time:

$$\text{Sleep Efficiency} = \frac{T_{\text{sleep}}}{T_{\text{sleep}} + T_{\text{wake}}}$$

For the long-term sleep quality estimation, iSleep calculates the scores of 4 components listed in Table 2 from PSQI. In order to calculate the component scores, iSleep measures the metrics listed in Table. 1 based on the detected events. However, some of the metrics used to calculate the score of Sleep Disturbance can not be measured by iSleep. For example, some of them are related to the bedroom's temperature, whether having dream, or feeling pain during sleep.



**Fig.18** The user interface of iSleep. (a) The screen showing sleep efficiency and sleep states over night. (b) The screen showing the sleep events detected overnight and the normalized loudness of each event. (c) The screen showing the history of sleep efficiencies and events.

From Fig.18, the user interface is showing the recent user sleep quality results. It is easy to read and the results are also obvious for users.

### 3.3 Discussion of audio-signal-based system

iSleep is designed for operating over night while the user is asleep, it is a unobtrusive APP, which has minimized the burden on the user.

iSleep provides fine-grained measurement, such as overall sleep efficiency and the occurrences of events that may interrupt sleep, such as cough and snore. Such fine-grained sleep profiling helps the user understand what factors affect their sleep quality.

iSleep has robust monitoring accuracy across different users, smartphones and sleep environments.

The users' privacy has been strictly protected. Due to the inherently private nature of sleep, any concern (or even suspension) of privacy breach may prevent the adoption of sleep monitoring technology. So the system processed the sensor samples on the fly and only keep sleep-related data (e.g., the number/loudness of snores), instead of sending any raw sensor samples to a remote server, because they may capture sensitive information such as audio of sleep talks, conversations before/after sleep and etc.

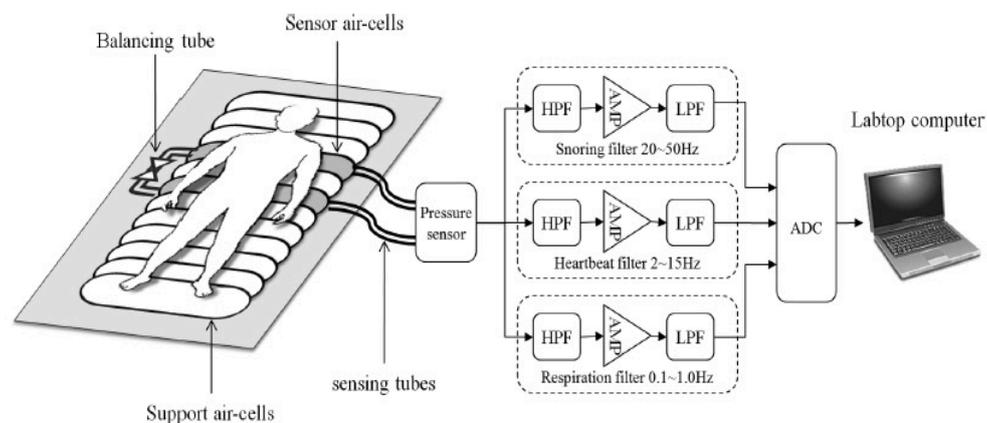
On the other hand, it requires that user has to have a smart phone, and needs to leave the phone somewhere close to the bed (up to several meters). Because the built-in microphone of smartphone is designed for capturing close vocals, and usually has low sensitivity. Moreover, many sleep-related events only

generate low-intensity sound. For example, the intensity of the sound from a roll-over movement is typically only several dB higher than that of ambient noise.

## 4 Sleep-Monitoring Based on Pressure

### 4.1 Idea of pressure-based system

Nonconstrained sleep monitoring system and algorithms using Air-Matress with balancing tube method[8], is used as the model in this part.

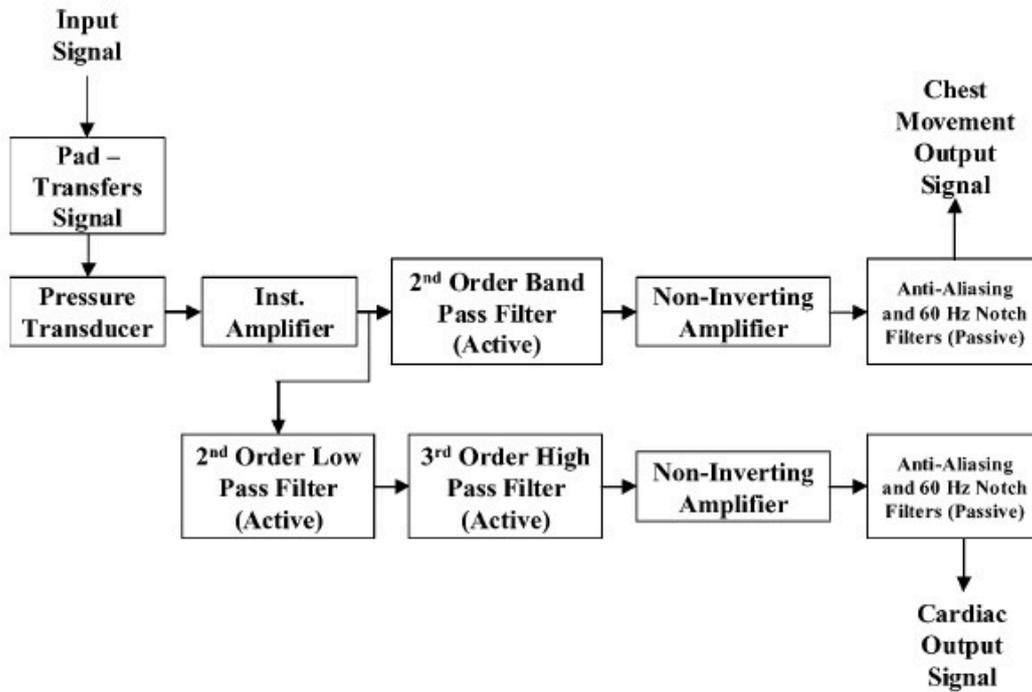


Overall structure of the measurement system using AMBT method.

**Fig.19 The structure of air mattress**

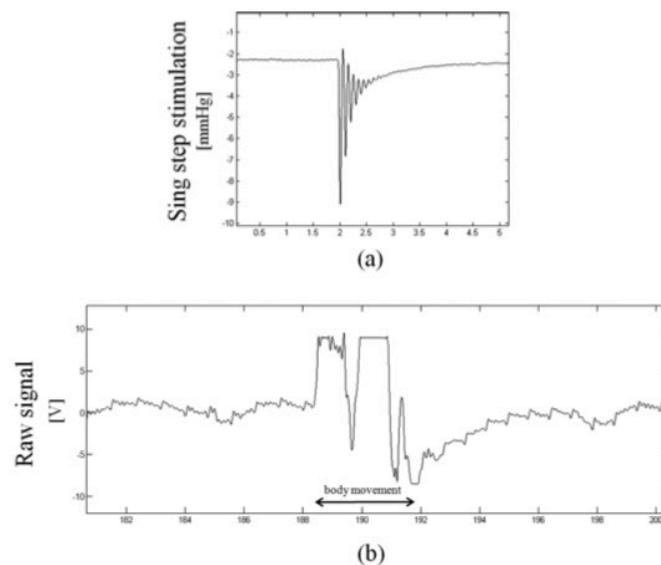
The bed-type sensor system using the air-matress with balancing tube (AMBT) method to noninvasively monitor the signals of heartbeat, respiration, and events of snoring, sleep apnea and body movement of subject on the system. The proposed system consists of multiple cylindrical air cells, two sensor cells and 18 support cells, and the small physiological signals were measured by the changes in pressure difference between the sensor cells, and the dc component was removed by balancing tube that is connecting the sensor cells. Using newly developed AMBT method, heartbeat, respiration, snoring, and body movement signals were clearly measured. For the concept of a home healthcare system, two automatic processing algorithms were developed: one is to estimate the mean heart and respiration rates for every 30 s, and another one is to detect the snoring, sleep apnea, and body movement events from the measured signals. In the beat- to-beat heart rate and breath-by-breath respiration rate analyses, the correlation coefficients of the heart and respiration rates from the proposed AMBT method compared with reference methods, electrocardiogram, and respiration effort signal from piezoelectric belt, were 0.98 ( $p < 0.01$ ) and 0.96 ( $p < 0.01$ ), respectively. Sensitivity and positive predictive value (PPV) of the detection algorithm for snoring event were 93%, 96%, for sleep apnea event were 93%, 88%, and for body movement event were 86%, 100%, respectively. These findings support that ABMT method provides an accurate and reliable means to monitor heartbeat, respiration activities and the sleep events during sleep[9].

### 4.2 Principles behind pressure-based system



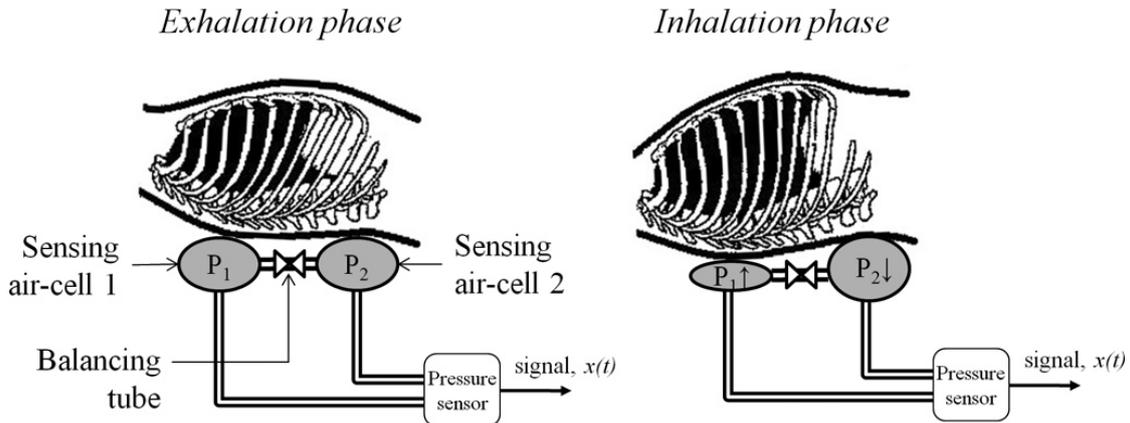
**Fig.20 Structure of pressure-based system**

The idea of the system is relative easy, and practical. When the signal is simple, the system is more robust and result is more exact.



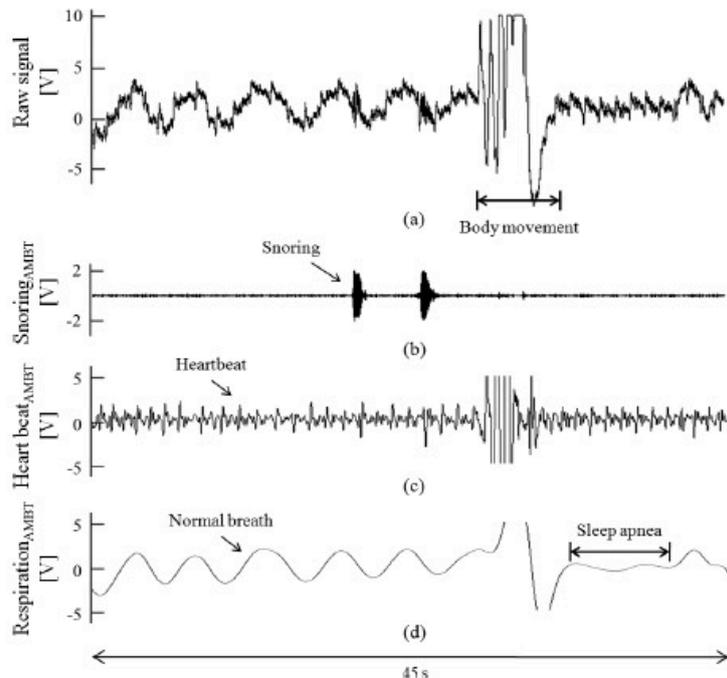
**Fig.21 Equalization of differential pressure by the balancing tube. (a) Step response by dropping 1 kg weight at t = 2. (b) Time trace example of removal of the dc component occurred by changing the body position.**

In Fig.21, the signal of pressure changed obviously with the changing of weight on the tube. So it is possible to use pressure signal to monitor sleeping.



**Fig.22 Example of the located position of the sensor air-cells and an illustration of pressure changes during the inhale and exhale movements**

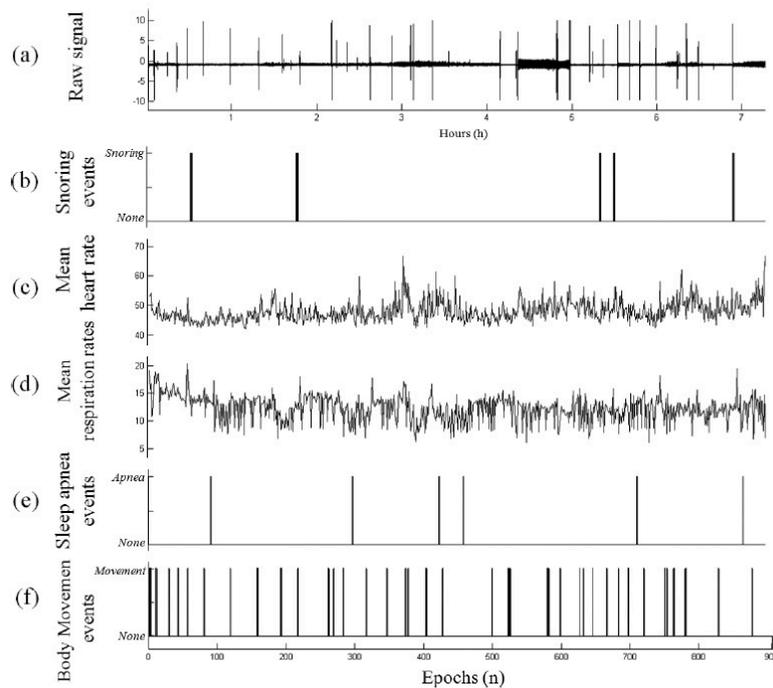
From Fig.22, the behaviour while breathing on bed can cause pressure change on sensor, in this way, the breath rate can be monitored during sleeping in bed.



**Fig.23 Time trace example of the output signals from the proposed system. (a) Raw signal with several signal components, e.g., heartbeat, respiration, snoring, body movement, and sleep apnea. (b) Filtered snoring AMBT signal with two times of snoring events. (c) Filtered heartbeat AMBT**

signal with body movement event. (d) Filtered respiration AMBT signal with sleep apnea event.

From Fig.23, the different behaviour during sleeping, such as heart beating or snoring will cause different pressure changing signals. That will lead to the monitoring detected results.



**Fig.24 Result of home sleep analysis using the AMBT system. (a) Raw signal. (b) Detected snoring events. (c) Estimated mean heart rates. (d) Estimated mean respiration rates. (e) Detected sleep apnea events. (f) Detected body movement events.**

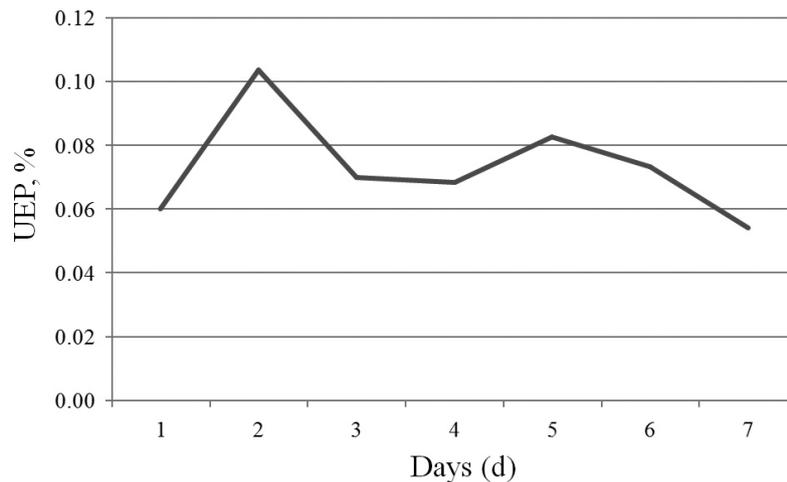
Fig. 24 shows one day results of home sleep monitoring using the AMBT system. The mean heart rate and respiration rate were estimated using the described auto-correlation method, and the snoring, apnea, and body movement were detected by the event detection algorithm. Data of seven and half hours were analysed, and the results show several body movements, snoring, and sleep apnea events during sleep.

After the signal can be detected, a way to evaluate sleeping quality need to be defined by these events and monitored.

In Fig.22, the heart rates shows four to five times of very low frequency cycles during the entire sleep, and each cycle has a period about one and half hours. Several times of snoring, sleep apnea, and body movement events were also detected. When body movements occurred, the AMBT signals were saturated and could not be processed to detect the heart rate or event information. The percentage of undetectable epochs due to the body movement was defined as:

Undetectable epoch percentage (UEP)

$$= \frac{\text{a number of body movement occurred epochs}}{\text{a total number of epochs}} \times 100(\%)$$



**Fig.25 UEP scores for 7 days**

UEP scores were calculated by (2) for 7 days in Fig. 23. The average UEP score was 7.33% with a standard deviation +1.62%.

### 4.3 Discussion of pressure-based system

For this pressure-based air-mattress, the idea is simple and easy to be accepted by normal users, and it is also contactless.

The signal is also robust and more exact.

But in the other way, it needs some extra cost for the sensors, and the results are not so obvious for users to see.

## 5 Sleep-Monitoring Based on Accelerometer

The accelerometer-based system is also designed with mattress, but the signal to monitor is totally different.

### 5.1 Idea of accelerometer-based system

In Monitoring body positions and movements during sleep using WISPs[10], it presents a sleep monitoring system based on the WISP platform - active RFID-based sensors equipped with accelerometers. And how the system accurately infers fine-grained body positions from accelerometer data collected from the WISPs attached to the bed mattress. Movements and their duration are also detected by the system.

## 5.2 Principles accelerometer-based system

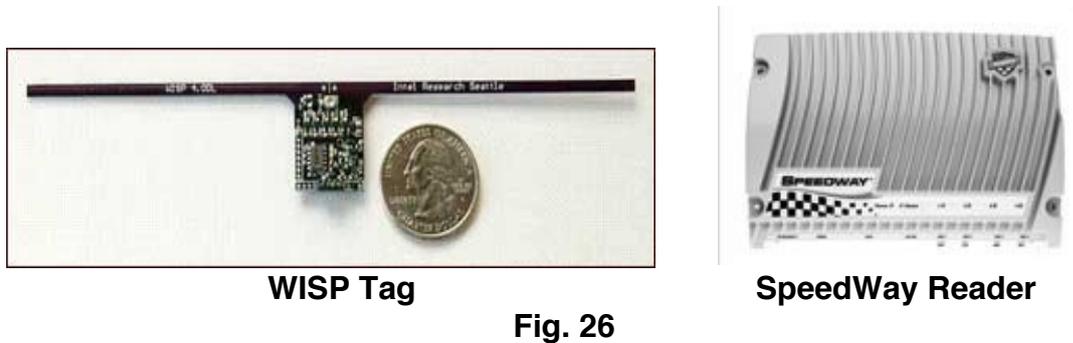
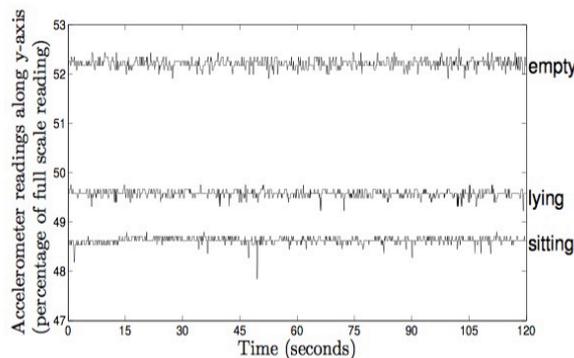
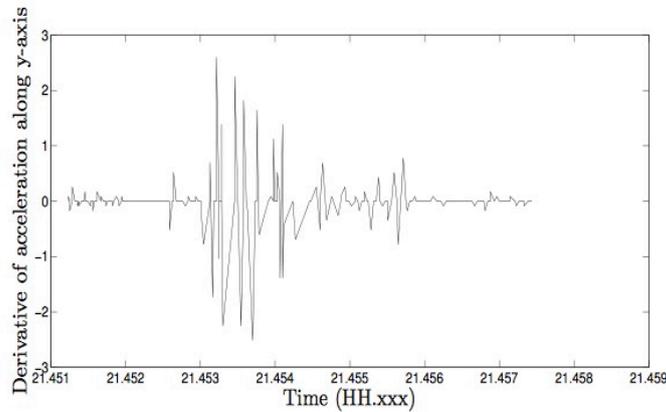


Fig.26 showed the sensors.

Recently WISP tags have been used for recognizing daily activities. In [11], the authors instrument everyday objects (e.g., glass, plate, books) with WISP tags that have accelerometers. Later, daily activities are inferred from the traces of objects that are moved. In [12], WISPs are used for sensing and monitoring of exercises involving free weights. The authors instrument free weights (e.g., dumbbells) and body parts (e.g., wrists) with WISPs. Then the accelerometer sensor readings from the tags are used to infer the exercise being done and the association between the user and the particular weight(s) being used. [13] presents a wireless neural interface that uses WISPs. It provides the neuroscientists a wireless, battery-free method of monitoring neural signals.

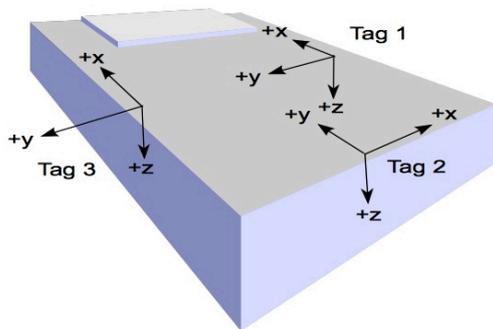


**Fig.27: Accelerometer Reading Variation for Empty/Lying/Sitting**



**Fig. 28 Accelerometer Reading along y-axis during a Movement**

Fig.27 shows that empty, lying and sitting situation of user can cause different signal for accelerometer. Fig.28 shows a movement in bed can cause a certain change of accelerometer signal. With this, it is possible to monitor the position changing during sleeping.

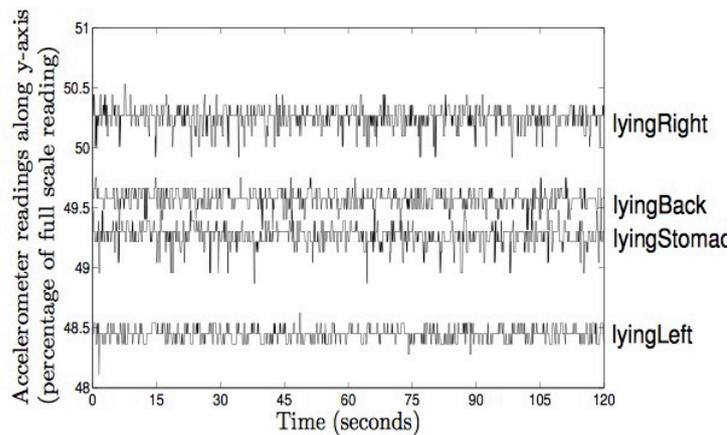


**Fig.29 Accelerometers setup**



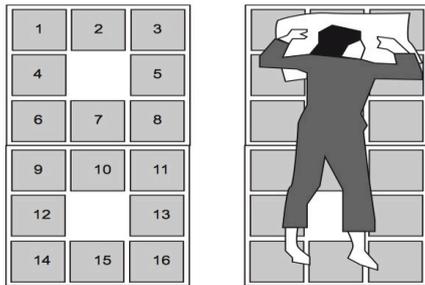
**Fig.30 Usual sleep position**

According to Fig.29 and 30, to monitor the usual sleep position require three accelerometers, each one at each side of the user, and one beside the legs.

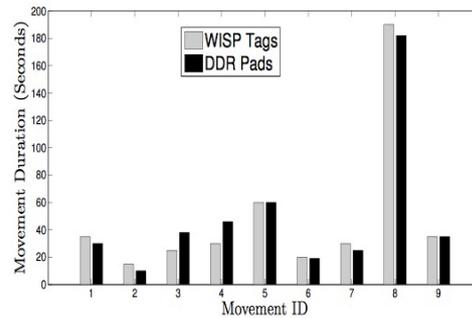


**Fig.31 Accelerometer Reading Variation for Different Lying Positions**

From Fig.31, the position during sleeping is also able to give different accelerometer signals. And in this way, the sleep position is also able to be monitored.



**Fig.32 The DDR Pad has 8 binary contact buttons around the side, but the middle portion of the pad does not have a button.**



**Fig.33 Durations of the Movements Detected by System and DDR Pads**

To test the system, 16 DDR pads are used to measure the results, like Fig.32, and are compared to the accelerometer-based system. According to Fig.33, the results of the system is quite exact, almost as same as DDR pads, but much easier to use.

### 5.3 Discussion accelerometer-based system

Accelerometer-based system is easy to set up and it is also battery free. But it costs extra for sensors and can only monitor limited signs. Only sleeping position is not totally enough for evaluating sleep quality.

## 6 CONCLUSION

Comparing to clinical sleep monitoring, home sleep monitoring is lower cost, more flexible and better for long term monitoring. In this paper, we have a general view of recent different technologies about home sleeping monitoring. Different technologies have their own advantages and disadvantages. By choosing the right one for users, their characters should be considered. For instance, some users do not want to keep smartphones close to them while sleeping, then iSleep is not suitable for them.

And for all of the technologies, there are still no medical standards for sleeping quality, thus they cannot analyse sleeping according to the feedback signs they have collected. Thus, the sleeping monitoring results are not always suitable for all situation or users.

In the future, home sleep monitoring will require more combination between technologies and medical science. In this way, the devices can serve people more correctly.

## 7 REFERENCES

- [1] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng, "Tracking Vital Signs During Sleep Leveraging Off-the-shelf WiFi," *MobiHoc*, pp.267-276, 2015
- [2] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures," in *MobiCom*, 2014, pp. 617–628.
- [3] P. Sebel, M. Stoddart, R. Waldhorn, C. Waldman, and P. Whitfield, *Respiration, the breath of life*. Torstar Books New York, 1985.
- [4] J. F. Murray, "The normal lung: the basis for diagnosis and treatment of pulmonary disease," Saunders, 1986.
- [5] Tian Hao, Guoliang Xing, and Gang Zhou, "iSleep: Unobtrusive Sleep Quality Monitoring using Smartphones," *SenSys*, 2013
- [6] T. Mitchell. *Machine Learning*. McGraw-Hill.
- [7] J. Webster et al. An activity-based sleep monitor system for ambulatory use. *Sleep*, 5(4):389–399, 1982.
- [8] Jae Hyuk Shin, Young Joon Chee, Do-Un Jeong, and Kwang Suk Park, "Nonconstrained sleep monitoring system and algorithms using Air Mattress with balancing tube method," *IEEE*, 2009
- [9] David C. Mack, James T. Patrie, Paul M. Suratt, Robin A. Felder, and Majd Alwan, "Development and Preliminary Validation of Heart Rate and Breathing Rate Detection Using a Passive, Ballistocardiography-Based Sleep Monitoring System," *IEEE*, 2009
- [10] Enamul Hoque, Robert F. Dickerson, John A. Stankovic, "Monitoring Body Positions and Movements During Sleep using WISPs," *Wireless Health*, 2010
- [11] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall. Recognizing daily activities with rfid-based sensors. In *UbiComp*, 2009.
- [12] R. Chaudhri, J. Lester, and G. Borriello. An rfid based system for monitoring free weight exercises. In *SenSys*, 2008.
- [13] D. Yeager, J. Holleman, R. Prasad, J. Smith, and B. Otis. Neuralwisp: A wirelessly powered neural interface with 1-m range. *IEEE Transactions on Biomedical Circuits and Systems*, To Appear.
- [14] Feng Lin, Yan Zhuang, Chen Song, Aosen Wan, Yiran Li, Changzhan Gu, Changzhi Li, and Wenyao Xu, "SleepSense: A Noncontact and Cost-Effective Sleep Monitoring System," *IEEE*, 2016