

# INNOVATION IN BITS AND BYTES

Informatics Inside  
Frühjahr 2024

# Tagungsband

Herausgeber\*innen:

Uwe Kloos, Hochschule Reutlingen

Natividad Martínez, Hochschule Reutlingen

Gabriela Tullius, Hochschule Reutlingen

# Impressum

## Anschrift:

Hochschule Reutlingen  
Reutlingen University  
Fakultät Informatik  
Human-Centered Computing  
Alteburgstraße 150  
D-72762 Reutlingen

Telefon: +49 7121 / 271-4002  
Telefax: +49 7121 / 271-4032

E-Mail: [informatics.inside@reutlingen-university.de](mailto:informatics.inside@reutlingen-university.de)  
Website: <https://infoinside.reutlingen-university.de>



Hochschule Reutlingen  
Reutlingen University



## Organisationskomitee:

Prof. Dr. rer. nat. Uwe Kloos, Hochschule Reutlingen  
Prof. Dr.-Ing habil. Natividad Martinez, Hochschule Reutlingen  
Prof. Dr. rer. nat. Gabriela Tullius, Hochschule Reutlingen

B. Sc. Manuel Keck  
B. Sc. Felix Meier  
B. Eng. Hewal Altun  
B. Eng. Armin Teubert



©2024 bei den Autoren. Lizenznehmer Hochschule Reutlingen, Deutschland.  
Dieser Artikel ist ein Open-Access-Artikel unter den Bedingungen und Konditionen der  
Creative Commons Attribution (CC BY)-Lizenz. <http://creativecommons.org/licenses/by/4.0>

# Vorwort

Liebe Leserinnen und Leser,

seit über 15 Jahren findet nun die Informatics Inside als Informatikkonferenz an der Hochschule Reutlingen statt, seit einigen Jahren in einem halbjährigen Rhythmus, d.h. im Frühjahr und Herbst. Diese Wissenschaftliche Konferenz des Masterstudiengangs Human-Centered Computing wird von den Studierenden selbst organisiert und durchgeführt. Sie erhalten während ihres Masterstudiums die Gelegenheit sich in einem selbstgewählten Fachgebiet zu vertiefen. Dies kann an der Hochschule, in einem Unternehmen, einem Forschungsinstitut oder im Ausland durchgeführt werden. Gerade diese flexible Ausgestaltung des Moduls „Wissenschaftliche Vertiefung“ führt zu einem sehr breiten Themenspektrum, das von den Studierenden bearbeitet wird. Neben der eigentlichen fachlichen Vertiefung spielt auch die Präsentation und Verteidigung von wissenschaftlichen Ergebnissen eine wichtige Rolle und dies weit über das Studium hinaus. Ein gewähltes Fachgebiet so allgemeinverständlich aufzubereiten und zu vermitteln, dass es auch für Nicht-Spezialisten verständlich wird, stellt immer wieder eine besondere Herausforderung dar. Dieser Herausforderung stellen sich die Studierenden im Rahmen der Frühjahrskonferenz Informatics Inside am 8. Mai 2024.

Das Themenspektrum der diesjährigen Frühjahrskonferenz ist wieder sehr vielfältig und breit gefächert. So erwarten Sie u.a. Beiträge aus dem Gesundheitssektor, der Industriellen Fertigung sowie zu verschiedenen XR-Anwendungen. Das Motto der Veranstaltung „Innovation in Bits and Bytes“ zeigt die Hauptrichtung in die die wissenschaftlichen Projekte gehen: Innovationen. Ohne innovative, neue Ideen, gerade in der Informatik, kann keine Entwicklung von Zukunftslösungen erfolgen. Das gilt gleichermaßen für den privaten als auch wirtschaftlichen Sektor. Die Studierenden haben sich über mehrere Wochen intensiv mit ihren Fachthemen beschäftigt und unterschiedlichste neue Ideen entwickelt. Die Ergebnisse der Projekte sind in den Fachartikeln dieses Tagungsbandes zusammengefasst. Bei der Konferenz werden diese dann präsentiert und diskutiert.

Mit diesem Tagungsband halten Sie diese Fachbeiträge der Studierenden in der Hand und können sich selber ein Bild von ihren Projekten und Arbeiten machen. Ich wünsche Ihnen eine spannende Lektüre, anregende Diskussionen und interessante Gespräche bei der Konferenz.

Prof. Dr. Uwe Kloos  
Human-Centered Computing

# Inhaltsverzeichnis

<b>Vorwort</b> . . . . .	III
<b>Kurze Einführung in die Grundlagen der Zeitreihenanalyse</b>	
Lindeboom, H. . . . .	1
<b>Umgang und Auswirkungen von Datenlücken bei medizinischer Zeitreihenanalyse</b>	
Schrüfer, K. . . . .	11
<b>Analyse der Kernaspekte von Recommender Systems</b>	
Garkusha, B. . . . .	21
<b>Classification of media for adaptive music recommendations</b>	
Keck, M. . . . .	31
<b>Einfluss äußerlicher Faktoren auf die Qualität eines mündlichen Vortrags in VR</b>	
Egler, M. . . . .	41
<b>Handtracking-Genauigkeit für die Erfassung von Musiknoten in der AR-Anwendung</b>	
<b>PianoVision</b>	
Stych, M. . . . .	51
<b>Eignung von Spiel-Design-Elementen in der industriellen Fertigung</b>	
Müller, R. . . . .	61
<b>Visuelle Kommunikation von Verkehrselementen in einer 3D-Fahrumgebung</b>	
Brendli, E. . . . .	71
<b>Autorenverzeichnis</b> . . . . .	81

# Kurze Einführung in die Grundlagen der Zeitreihenanalyse

Hjordis Lindeboom  
Hochschule Reutlingen  
Reutlingen, Deutschland

**Hjordis.Lindeboom@Student.Reutlingen-University.de**

## Abstract

Die Arbeit gibt eine kurze Einführung in die Thematik der klassischen Zeitreihenanalyse. In einer zunehmend digitalisierten Welt mit enorm schnell anwachsenden Datenmengen haben Zeitreihenanalysen eine sehr große Bedeutung. Nach einer kurzen, allgemeinen Definition wird auf mathematisch-statistische Verfahren und Modelle sowie programmtechnische Umsetzungen, z.B. mit Bibliotheken und Funktionen in R und Python, eingegangen.

Dazu wird zunächst ein Überblick über elementare Grundlagen der Zeitreihenanalyse gegeben. Hierbei geht es sowohl um die mathematische Basis von Zeitreihen als auch zentrale Forderungen bei deren Analyse (Stationarität und Ergodizität), das klassische Komponentenmodell, das Box-Jenkins-Verfahren, die Autokorrelation sowie die deskriptive Beschreibung und Modellbildung von univariaten Zeitreihen.

Im Anschluss werden kurz einige univariate Zeitreihenmodelle vorgestellt sowie die mögliche Erweiterung auf mehrdimensionale Zeitreihenanalysen.

---

Betreuer Hochschule: Prof. Dr. Thies  
Hochschule Reutlingen  
[Christian.Thies@Reutlingen-University.de](mailto:Christian.Thies@Reutlingen-University.de)

Informatics Inside Frühjahr 2024  
28. April 2024, Hochschule Reutlingen  
Copyright 2024 Hjordis Lindeboom

## Keywords

Zeitreihenanalyse – Komponentenmodell – Stationarität – Box-Jenkins-Verfahren – Trend – Saisonalität – Autokorrelation.

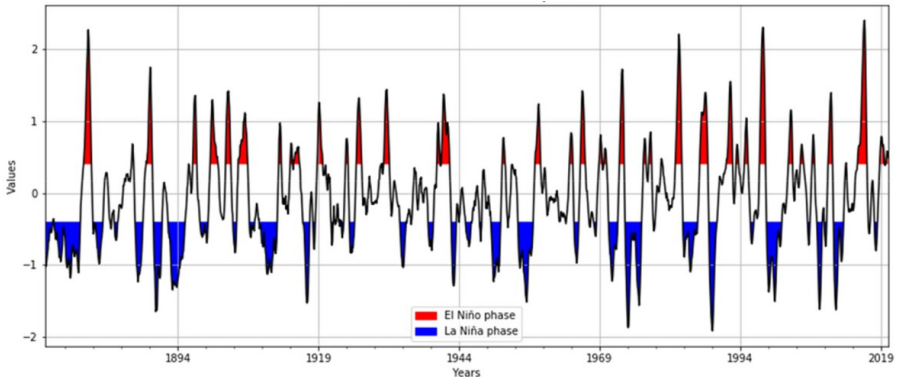
## 1 Einleitung

Eine Zeitreihe (engl. time series) kann als eine zeitlich geordnete Folge von Beobachtungswerten  $y_t$  definiert werden, mit  $t = 1, 2 \dots T$ , d.h. eine Zeitreihe ist eine zeitlich geordnete Folge von Merkmalsausprägungen eines statistischen Merkmals [2, 6], mathematisch betrachtet:

$$\{y_1, \dots, y_T\} = \{y_t\}_{t=1}^T \text{ mit } y_t \in \mathbb{R}$$

Bei der Zeitvariable  $t$  handelt es sich meist um ein diskretes Merkmal mit gleichlangen Abständen zwischen  $t$  und  $t+1$ ; sie kann aber auch stetig sein. Bei den Beobachtungswerten  $y_t$  handelt es sich i.d.R. um eine metrische Variable, alternativ kann sie auch dichotom sein (0-1-Merkmal). [6]

Für Zeitreihen und entsprechende Analysen, die über Regressionsanalysen hinausgehen, existieren zahlreiche Anwendungsbeispiele aus verschiedenen wissenschaftlichen Fachgebieten, zum Beispiel in der Medizin (EKG- und EEG-Diagnostik, biomechanische Bewegungsabläufe, ...), der Meteorologie (Temperaturverläufe, Klimageschichte, Klimawandelmodelle, ...), der Wirtschaft (Verkaufszahlen, Börsenkurse, ...), der Ökologie (Vegetationsentwicklung, Populationsbiologie, ...) und vielen andere Disziplinen. [4, 5, 7, 14]



**Abb. 1: Grafische Darstellung einer Zeitreihe am Beispiel von El-Niño- und La-Niña-Episoden (Oceanic Niño Index:  $ONI \geq +0,5 \text{ } ^\circ\text{C}$  bzw.  $ONI \leq -0,5 \text{ } ^\circ\text{C}$ ); (Abb. 2 in [8]).**

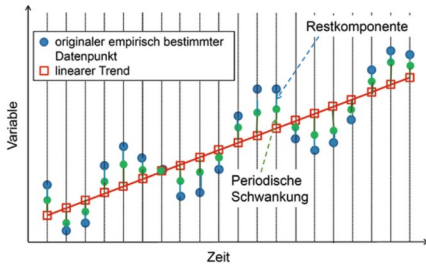
Zur Visualisierung und grafischen Analyse werden die Datenpunkte meist zunächst im y-t-Koordinatensystem dargestellt. Abbildung 1 veranschaulicht dies am Beispiel einer Zeitreihe zu den ozeanischen Klimaphänomenen El-Niño- und La-Niña (außergewöhnlich warme bzw. kalte Meeresströmungen). Die damit in Verbindung stehenden Klimavariablen, Schwankungen der Meeresoberflächentemperatur und des Luftdrucks wurden in Zeitreihenanalysen von Norel et al. untersucht (2021) [8].

Teile von Zeitreihen, die nicht zu stark voneinander abweichen, können mit den bekannten Kenngrößen der deskriptiven Statistik charakterisiert werden (Mittelwert, Varianz, Standardabweichung). Für tiefergehende Analysen von Zeitreihen ist man auf die Bildung von Modellen angewiesen [5, 6, 14]. Anhand vorhandener Daten validierte, statistische Zeitreihenmodelle bilden die Basis für das Erkennen von Gesetzmäßigkeiten, Prognosen, Risikoanalysen oder die Steuerung von Prozessen [6, 14]. Fragestellungen, die im Rahmen von Zeitreihenanalysen behandelt werden, sind u.a. Fragen nach systematischen Veränderungen (langfristige Trends), nach regelmäßigen Mustern, wie z.B. zyklischen Schwankungen, nach möglichen Ausreißern, der Frage, wie

sich die Werte zukünftig weiterentwickeln werden oder auch, ob Zusammenhänge zwischen verschiedenen Zeitreihen existieren [12]. Bei den Methoden der Zeitreihenanalyse unterscheidet man uni- und multivariate Verfahren, je nachdem ob eine oder mehrere Variablen betrachtet werden. Innere Methoden erklären eine Zeitreihe  $y_t$  allein als Funktion der Zeit, äußere Methoden berücksichtigen auch andere Variablen [6: Seite 396]. Bei den Zufallsvariablen  $Y_t$  wird angenommen, dass es sich um eine Folge von unabhängigen, identisch verteilten Zufallsvariablen handelt. Desweiteren spielt im Zusammenhang mit der Bildung von Modellen für Zeitreihen die Annahme bzw. Forderung von Stationarität und Ergodizität eine wichtige Rolle [11, 12], siehe hierzu 2.2 und 2.5 im vorliegenden Text.

Verschiedene Modelle für stationäre Reihen bilden die Basis der Zeitreihenanalyse. Die Visualisierung der Datenpunkte ist ein erster wichtiger Schritt zur Erkennung von Mustern, Trends, gesetzmäßigen Schwankungen und die Auswahl von Modellen. Die zentralen Parameter eines geeignet erscheinenden Modells versucht man dann, u.a. mit Hilfe der empirischen Autokorrelationsfunktion  $r_\tau$  (ACF) hinreichend gut abzuschätzen [12].

## Kurze Einführung in die Zeitreihenanalyse



**Abb. 2: Trend, periodische Schwankung und Restkomponente bei einer Zeitreihe aus (Abb. 11.3 aus [14]).**

## 2 Das Komponentenmodell

Beim klassischen Verfahren werden die Zeitreihen, die durch systematische und zufällige Einflüsse entstehen, in Komponenten zerlegt. Man unterscheidet hierbei drei Elemente als Funktion der Zeit: Trend, Saisonalität und Restkomponente. Beim Trend handelt es sich um eine systematische Veränderung des mittleren Niveaus der Zeitreihe. Bei der Saisonkomponente handelt es sich um periodische Schwankungen und bei der Restkomponente um nicht zu erklärende Einflüsse oder Störungen [11].

Abb. 2 veranschaulicht den Zusammenhang zwischen den Elementen des Komponentenmodells; Abb. 3 ist ein Beispiel für die Zerlegung einer Zeitreihe in die drei Komponenten.

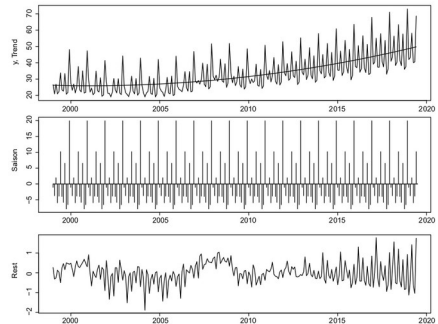
Wenn die Komponenten unabhängig voneinander wirken, addieren sich die Einflüsse auf die Zeitreihe [6, 14]:

$$y_t = T_t + S_t + R_t \text{ (additive Zerlegung)}$$

Wenn die Komponenten abhängig voneinander auf die Zeitreihe wirken, werden die Komponenten miteinander multipliziert [6, 14]:

$$y_t = T_t \cdot S_t \cdot R_t \text{ (multiplikative Zerlegung)}$$

Gemischte Modelle sind möglich. Durch Logarithmieren entsteht aus einer multiplikativen Verknüpfung eine additive. Dies ermöglicht in der Folge die Nutzung von Verfahren bei additiver Zerlegung.



**Abb. 3: Zerlegung der Zeitreihe in Trend (hier: polynomialer Trend 2. Ordnung), periodische Schwankungen (Saison) und Rest (Abb. 2.17 aus [12]).**

### 2.1 Box-Jenkins-Verfahren

Bei diesem weit verbreiteten Ansatz wird zunächst der Trend und anschließend die Saisonalität modelliert und von den Daten abgezogen. Dadurch erhält man eine trend- und saisonbereinigte Datenreihe

$$\hat{y}_t = y_t - T_t - S_t$$

Das Ziel ist eine stationäre Zeitreihe. Damit wird erreicht, dass sich die stochastischen Charakteristika (Mittelwert, Varianz, Kovarianz) über die Zeit nicht ändern und über die empirischen Werte abgeschätzt werden können [11, 12].

### 2.2 Stationarität

Für jeden Zeitpunkt  $t$  einer Zeitreihe liegt nur eine Beobachtung für die Zufallsvariable  $Y_t$  vor. Die Werte für unterschiedliche Zeitpunkte einer Zeitreihe repräsentieren, wenn systematische Einflussfaktoren vorhanden sind, unterschiedliche Zufallsvariablen. Die Forderung der Stationarität stellt sicher, dass sich die statistischen Kenngrößen über die Zeit nicht ändern. Eine Zeitreihe wird allgemein als (schwach) stationär bezeichnet, wenn er mittelwert-, varianz- und kovarianzstationär ist. Diese Annahme bedeutet, dass alle  $y_t$  den gleichen Erwartungswert  $\mu$  sowie die gleiche Varianz  $\sigma^2$  haben [11, 12].

Die empirischen Werte einer Zeitreihe liefern ggf. geeignete Schätzer für die theoretischen Momente [11]. Ein stochastischer Prozess heißt

mittelwertstationär, wenn  $E[Y_t] = \mu$   
für alle  $t \in T$ ,

varianzstationär, wenn  $\text{Var}[Y_t] = \sigma^2$ ,  
für alle  $t \in T$  und

kovarianzstationär, wenn  $\text{Cov}[Y_{t+\tau}, Y_t]$  nur von der Zeitdifferenz  $\tau$  (englisch »lag«) abhängt:

$\text{Cov}[Y_{t+\tau}, Y_t] = \gamma_\tau$ , für alle  $t \in T, \tau \geq 0$

### 2.3 Kovarianz und Autokovarianz

Die empirische Kovarianz einer zweidimensionalen Zufallsgröße  $(X, Y)$  ist als Maß für den linearen Zusammenhang definiert als

$$c_{XY} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad [11, 12]$$

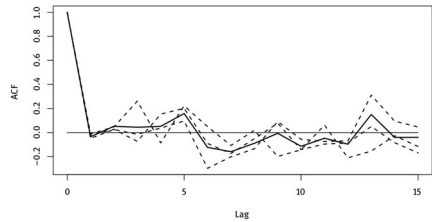
Durch Standardisierung erhält man den empirischen Korrelationskoeffizienten

$$r_{XY} = \frac{c_{XY}}{\sqrt{s_x^2} \sqrt{s_y^2}} \quad [11, 12]$$

Bei einer Zeitreihe wird aus den Beobachtungswerten  $y_1, y_2, \dots, y_N$  für jede Zeitdifferenz  $\tau$ , ein künstlicher zweidimensionaler Datensatz gebildet [12]:

$$(y_{1+\tau}, y_1), (y_{2+\tau}, y_2), \dots, (y_N, y_{N-\tau}).$$

Mit diesen Daten lässt sich die Kovarianz bzw. der Korrelationskoeffizient einer Zeitreihe berechnen – für jede Zeitdifferenz  $\tau$ . Sofern sich Mittelwert und Varianz in den verschiedenen Bereichen der Zeitreihe nicht deutlich unterscheiden, können der globale Mittelwert und die globale Varianz für die Berechnung genutzt werden. Entsprechend [12] ist die empirische Autokovarianzfunktion ( $c_\tau$ ) einer Zeitreihe ( $y_t$ ) bzw. ihre empirische Autokorrelationsfunktion  $r_\tau$  (ACF) wie folgt definiert:



**Abb. 4: ACF einer Zeitreihe; Vergleich der ganzen Reihe mit drei Segmenten (gestrichelte Linien), (Abb. 2.5 in [12]).**

$$c_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} (y_{t+\tau} - \bar{y})(y_t - \bar{y}) \quad \tau \geq 0$$

$$r_\tau = \frac{c_\tau}{c_0} \quad \tau \geq 0$$

(Bei  $\tau < 0$  gilt:  $c_\tau = c_{-\tau}$  und  $r_\tau = r_{-\tau}$ )

Mit Autokorrelation bezeichnet man die Korrelation der Zeitreihen mit sich selbst, d.h. verschiedene Werte der Zeitreihe untereinander. Der ACF-Koeffizient  $r_\tau$  ist ein Maß für die lineare Abhängigkeit für entsprechend weit auseinanderliegende Zeitpunkte, vergleichbar mit dem empirischen Korrelationskoeffizienten [12]. Der Graph der empirischen ACF wird als Korrelogramm bezeichnet [11].

### 2.4 Prüfung der Stationarität

Zur Prüfung der Stationarität bietet es sich an, die Zeitreihe zunächst in mehrere Segmente zu teilen. Anschließend vergleicht man die statistischen Kennzahlen. Die Mittelwerte und Streuungen der verschiedenen Segmente der Zeitreihe sollten nur geringe Unterschiede aufweisen. Letztendlich bleibt eine derartige Bewertung subjektiv, ermöglicht aber eine erste Einschätzung. Grafisch lässt sich die Annahme der Stationarität auch durch die Autokorrelationsfunktion prüfen [11, 12]. In Abb. 4 ist erkennbar, dass die Autokorrelationen nur wenig von der Null abweichen, d.h. die ACF-Werte liegen meist unter ca. 0.3, was darauf hindeutet, dass es keine Korrelation gibt.

## Kurze Einführung in die Zeitreihenanalyse

Der Autokorrelationskoeffizient ACF geht bei stationären Zeitreihen mit wachsendem lag  $\tau$  gegen Null. Für Bildung vieler Modelle ist die ACF eine notwendige, basale Größe.

### 2.5 Ergodizität

Bei der für die Zeitreihenanalyse wichtigen Annahme der Ergodizität wird davon ausgegangen, dass die Stichprobenwerte einer stationären Zeitreihe bei  $t \rightarrow \infty$  gegen die theoretischen Momente der Grundgesamtheit konvergieren, d.h.

$$E[y_t] = \mu \quad \text{und} \quad \text{Var}[y_t] = \sigma^2 \quad [11]$$

Eine stationäre Zeitreihe heißt mittelwertergodisch, wenn

$$\lim_{N \rightarrow \infty} E[(\bar{Y}_N - \mu)^2] = 0 \quad [11]$$

und kovarianzergodisch, wenn für alle Zeitdifferenzen  $\tau$  gilt:

$$\lim_{N \rightarrow \infty} E\left[\frac{1}{N} \sum_{t=1}^N (y_{t+\tau} - \mu)(y_t - \mu)\right] = 0 \quad [11]$$

Ergodizität bedeutet dementsprechend, dass sich die theoretischen Kenngrößen einer Verteilung ( $\mu, \sigma^2, \text{Cov}$ ) aus den empirischen Beobachtungswerten einer Zeitreihe widerspruchsfrei schätzen lassen [12].

### 2.6 Die Trendkomponente

Der Trend einer Zeitreihe lässt sich mit unterschiedlichen Verfahren bestimmen bzw. beschreiben. Eine Möglichkeit ist die lineare Regression:  $T_t = \beta_0 + \beta_1 t$

Allgemein formuliert ergibt sich ein polynomer Trend der Ordnung  $m$  [11]:

$$T_t = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_m t_m$$

Die Modellparameter können über die Methode der kleinsten Quadrate bestimmt werden. Abb. 5 ist ein Beispiel für einen Trend 3. Ordnung. Bei Exponential- oder Potenzmodellen können durch logarithmierte Werte  $\ln(y_t)$  lineare Regressionsmodelle genutzt werden [12].

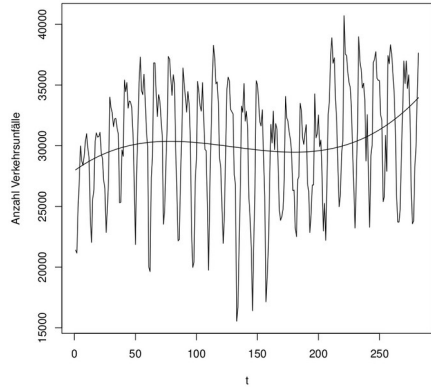


Abb. 5: Zeitreihe mit polynomialem Trend 3. Ordnung ([11]: Abb. 2.6 auf S. 22).

Mitunter ergeben sich Probleme bei der Ermittlung der Parameter nichtlinearer, spezieller Trendmodelle, z.B. bei nichtlinearen Trends mit Sättigungsgrenze. Dann kommen ggf. kompliziertere mathematische Verfahren zum Einsatz [12].

Falls bei der Visualisierung deutlich wird, dass sich eine Trenderfassung durch ein Polynom für den Zeitraum der Reihe nicht in Frage kommt, kann man den Trend auch mit dem Verfahren der gleitenden Durchschnitte erfassen. Hierzu wird die Zeitreihe in gleich lange Segmente unterteilt, für jedes Segment der arithmetische Mittelwert bestimmt und dieser dem zeitlichen Mittelpunkt seines Segmentes zugeordnet. Das Ergebnis ist eine Zeitreihe aus arithmetischen, lokalen Mittelwerten (»glatte Komponente« der Ausgangsreihe), siehe Abb. 6.

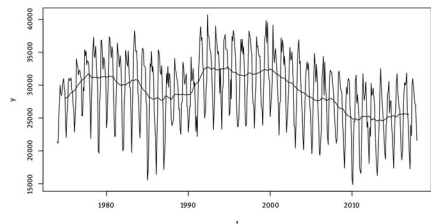


Abb. 6: Zeitreihe aus gleitenden Durchschnittswerten ([11]: Abb. 2.7 auf S. 23).

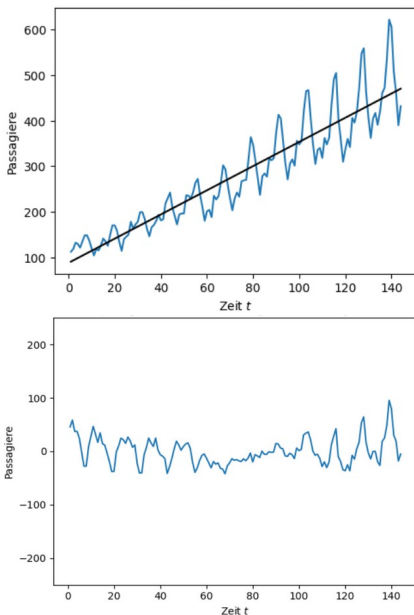
## 2.7 Die Saisonkomponente

Nach einer erfolgreichen Trendbereinigung (oder wenn kein Trend vorhanden ist) schwanken die Werte einer Zeitreihe um die Null. Bei vielen Zeitreihen sind periodische Schwankungen zu beobachten. Zur Saisonbereinigung stellt man zunächst die i.d.R. bekannte Länge der Periode fest. Bei einem jährlichen Muster kann die Saison durch monatlichen Mittel mit 12 Zeitschritten modelliert werden (Saisonperiode  $s = 12$ ), indem man die Differenz der monatlichen Mittelwerte vom Mittelwert der gesamten, trendbereinigten Reihe bildet [3].

Das Ergebnis ist eine trend- und saisonbereinigte Zeitreihe (Abb. 7)  $\hat{y}_t$ :

$$\hat{y}_t = \hat{y}_t - \left( \text{mean}[\hat{y}_{t'}] - \text{mean}[\hat{y}_t] \right)$$

mit  $t' \in 1, \dots, T : \text{mod}(t', s) = \text{mod}(t, s)$



**Abb. 7:** Zeitreihe mit linearem Trend und jährlichem Muster (oben) sowie nach Trend- und Saisonbereinigung (unten), Zeit  $t$  in Monaten. (Abbildungen aus [3]).

Ein Überblick über vorhandene Trend- und Saisonbereinigungsverfahren findet sich bei Schlittgen und Sattarhoff (2020: S. 27-52).

## 3 Modellierung von univariaten Zeitreihen

Für die Modellierung der Autokorrelation bzw. der Zeitreihen existieren zahlreiche Verfahren [3, 5, 12]. Ausgangspunkt hierbei ist der Versuch, die Reihen aus sich selbst zu erklären. Schlittgen und Sattarhoff (2020: S. 69 [12]) formulieren dieses Ziel wie folgt: Wert im Zeitpunkt  $t = f(\text{Werte in vorhergehenden Zeitpunkten}) + \text{Rest}$

$$\text{Formal: } Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t$$

$\varepsilon_t$  ist »weißes Rauschen« (ein reiner Zufallsprozess) mit  $E(\varepsilon_t) = 0$  und  $\text{Var}(\varepsilon_t) = \sigma^2$ , d.h.  $\varepsilon_t \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$ .

Bekannte Zeitreihenmodelle sind das AR-, das MA-, das ARMA-, das ARIMA- und das SARIMA-Modell [3, 5, 12, 14]. Beim AR-Modell (Autoregressiver Prozess) versucht man den aktuellen Wert über eine Linearkombination von  $p$  vorangegangenen Werten (+ weißes Rauschen) vorherzusagen. Beim MA-Modell (Moving Average-Prozess) wird der aktuelle Wert durch eine Linearkombination der zufälligen Fehler von  $q$  vorangegangener Werte vorhergesagt. Durch eine Kombination dieser beiden Modelle entsteht ein ARMA-Modell zur Beschreibung der Autokorrelation, bei welchem der Einfluss der vorangegangenen Werte mit den letzten zufälligen Schwankungen kombiniert wird [3]. Bei den oft verwendeten ARIMA-Modellen (Autoregressive Integrated Moving Average) handelt es sich um eine Erweiterung des ARMA-Modells für nicht-stationäre Zeitreihen [14]. Die Trendbereinigung (durch Differencing) über den Parameter  $d$  ist Teil des Modells [3]. SARIMA (Seasonal Autoregressive Integrated Moving Average) wiederum ist eine spezielle Weiterentwicklung des ARIMA-Modells für die Berücksichtigung saisonaler Effekte [1, 12].

Hinsichtlich der vielen Verfahren zur Parameterschätzung bei Zeitreihenmodellen sind die entsprechenden Ausführungen in Kreiß und Neuhaus (2006) sowie Schlittgen und Satarhoff (2020) zu empfehlen.

### 3.1 Prognosen mit univariaten Zeitreihen

Die Zuverlässigkeit von Prognosen, die auf der Basis von Zeitreihenmodellen formuliert werden, ist mit einigen Unsicherheitsfaktoren verbunden [12: S. 119]:

- Unsicherheiten bei der Modellauswahl, d.h. es stellt sich grundsätzlich die Frage, ob ein anderes Modell zuverlässigere Prognosen liefern könnte.
- Nicht vermeidbare Unsicherheiten durch zukünftige stochastische Einflüsse.
- Zufälligen Fehlern bei der Schätzung der Modellparameter.

Die Einflüsse durch den 1. und den 3. Punkt lassen sich im konkreten Fall nur sehr schwer abschätzen. Beim zweiten Punkt kann man, unter der Annahme normalverteilter Zufallsprozesse Prognoseintervalle (Konfidenzintervalle) bestimmen. Hierfür werden Grenzen für zukünftige Prognosezeitpunkte berechnet, bei denen die zukünftige Zeitreihenvariable  $Y_{N+h}$  mit der Wahrscheinlichkeit  $1-\alpha$  einen Wert aus dem Intervall annimmt [12]. In Abbildungen werden diese Intervalle oft als Bänder dargestellt (vgl. hierzu Abb. 10). Bei der Berechnung der Prognoseintervalle geht man von Gauß-Prozessen aus, d.h. die Gültigkeit der Intervalle hängt davon ab, dass normalverteilte Zufallseinflüsse vorliegen. Die Prognoseintervalle haben dann die Form [12: S. 123]:

$$P(\hat{Y}_{N,h} - z_{1-\alpha/2}\sqrt{V} \leq Y_{N+h} \leq \hat{Y}_{N,h} + z_{1-\alpha/2}\sqrt{V}) = 1 - \alpha$$

$V$  ist die Varianz des Prognosefehlers und  $z_{1-\alpha/2}$  ist das entsprechende Quantil der Standardnormalverteilung.

## 4 Multivariate Zeitreihen

Wenn Zeitreihen mehrerer Variablen vorliegen können diese unabhängig voneinander analysiert werden. Mitunter stellen sich aber Fragen nach Abhängigkeiten oder Interaktionen. Mittels entsprechender Modelle lassen sich auch für mehrdimensionale Zeitreihen Prognosen entwickeln [12]. Das VAR-MA Modell ist z.B. eine vektorbasierte Erweiterung des ARMA Modells, welches mehrere Variablen berücksichtigt.

Statt zwei Zeitreihen ( $X_t$ ), ( $Y_t$ ) als getrennte Größen zu betrachten, können sie auch als Komponenten einer zweidimensionalen Zeitreihe betrachtet werden. Wichtig hierbei sind die auf die einzelnen Zeitreihen bezogenen Kenngrößen  $\mu_x$ ,  $\mu_y$ , die ACFs  $r_x(\tau)$  und  $r_y(\tau)$  sowie die Beschreibung des Zusammenhangs zwischen den beiden Größen. Hierfür lassen sich bei gemeinsam stationären Zeitreihen, analog zu den ACFs der univariaten Zeitreihen, Kreuzkovarianz- und Kreuzkorrelationsfunktion definieren [12].

Die empirische Kreuzkovarianzfunktion zweier Zeitreihen ( $x_t$ ), ( $y_t$ ) lautet:

$$c_{XY}(\tau) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-\tau} (x_{t+\tau} - \bar{x})(y_t - \bar{y}) & \text{für } \tau \geq 0 \\ \frac{1}{N} \sum_{t=1-\tau}^N (x_{t+\tau} - \bar{x})(y_t - \bar{y}) & \text{für } \tau < 0 \end{cases}$$

Die empirische Kreuzkorrelationsfunktion:

$$r_{XY}(\tau) = \frac{c_{XY}(\tau)}{\sqrt{c_X(0)}\sqrt{c_Y(0)}}$$

Die Kreuzkorrelation beim Lag 0 beschreibt den (linearen) Zusammenhang zwischen den beiden Zeitreihen  $X_t$  und  $Y_t$  [12].

Für weitergehende Informationen zu mehrdimensionalen Zeitreihenanalysen, siehe Schlittgen und Satarhoff (2020: S. 187-223).

## 5 Programmiertechnische Umsetzung von Zeitreihenanalysen

Für die Zeitreihenanalyse bietet sich aufgrund der meist großen Datenmengen und des Rechenaufwandes eine programmtechnische Umsetzung der Modellbildung an. Hierfür werden vor allem R oder Python verwendet. Alternativ besteht die Möglichkeit eine Kombination von Python und R zu verwenden.

### 5.1 Python

Python bietet eine große Auswahl an Bibliotheken, unter anderem für die Zeitreihenanalyse und deren Visualisierung. Die Pandas-Bibliothek zum Beispiel enthält neben vielen anderen Funktionen, die Möglichkeit, sogenannte Pandas-Series zu erstellen und zu bearbeiten (Pandas: `pan(ell)da(ta)-s`). Hierzu gibt es Einführungen im Internet, z.B. das Kapitel „Python, Pandas und Zeitserien“ auf der Webseite Python-Kurs [9], das Pandas und numpy nutzt.

Tayeb et al. [13] verwenden ebenfalls die Bibliotheken Pandas und numpy, sowie matplotlib, statsmodel und sklearn. Anstelle von Pandas-Series werden die Daten hier als Pandas-Dataframes eingelesen. Anschließend werden die Daten mit Hilfe der matplotlib über die Funktionen `plot()` und `hist()` visualisiert.

Für die Darstellung der Autokorrelation bzw. der partiellen Autokorrelation wird auf die statsmodel-Bibliothek zurückgegriffen. Die Funktionen `plot_acf()` und `plot_pacf()` ermöglichen dies. Darüber hinaus enthält die statsmodel-Bibliothek die Funktion `seasonal_decompose()`, diese zerlegt Zeitreihen in den Trend, saisonale Informationen und einen Rest mit zusätzlichen Einflüssen und Störgrößen.

### 5.2 R

Für diejenigen, die Python nicht verwenden möchten, ist R eine Alternative.

In der Standardversion von R sind einige Pakete automatisch enthalten, so dass ohne zusätzliche Installation von Paketen bereits Funktionen zur Zeitreihenanalyse zur Verfügung stehen, wie z.B. `arima()` zur Schätzung eines ARIMA-Modells oder auch `filter()` zur Filterung einer Zeitreihe. Einen guten Überblick und Einstieg in die Zeitreihenanalyse mit R bieten Schlittgen und Sattarhoff [12]. Um das volle Potential auszuschöpfen, können zusätzliche Pakete installiert werden. Dazu gehört z.B. `tsapp`, das vom Erstautor erstellt und parallel zur Veröffentlichung entwickelt wurde. Darüber hinaus verweisen die Autoren auf die Pakete `np`, `mFilter`, `vars`, `lmtest`, `dlm`, `car`, `TSA`, `forecast`, `fracdiff`, `tseries`, `leaps`, `MswM`, `urca`, `MASS`, `itsa` und `nlme`. Diese Pakete sollten die meisten Anforderungen der Zeitreihenanalyse abdecken. R bietet auch ohne zusätzliche Pakete eine große Auswahl an relevanten Funktionen, so können Daten eingelesen werden und mittels `ts()` in Zeitreihen umgewandelt werden. Funktionen wie `acf()` zur Autokorrelation, `decompose()` oder auch `arima()` zur Schätzung von ARIMA-Modellen sind standardmäßig vorhanden. Einige dieser Funktionen werden in Paketen wie `tsapp`, erweitert bzw. mögliche Fehlerquellen reduziert.

### 5.3 Python und R

Python und R können auch gleichzeitig verwendet werden. Entweder durch Aufruf und Ausführung vorgefertigter Skripte oder durch direkte Ausführung von Code der jeweils anderen Sprache [10]. Radečić zeigt die Integration von R in Python sowie die Integration von Python in R [10]. So kann ein R-Skript mit der Subprocess-Bibliothek als Subprozess ausgeführt werden oder durch den Import der `rpy2`-Bibliothek direkt auf Inhalte von R zugegriffen werden. In R ermöglicht das Paket `reticulate` sowohl die Ausführung von Skripten als auch die Ausführung von Code. Durch diese Kombination kann auf die Funktionalität beider Sprachen zugegriffen werden.

## 6 Kurzes Fazit

Der vorliegende Text kann nur einen extrem kurzen Einstieg in die elementaren Grundlagen der Zeitreihenanalysen bieten. Bedingt durch die Vielfalt der hierfür zur Verfügung stehenden Verfahren und Modelle, sowie die zum Teil komplizierten mathematischen Hintergründe sind zumindest statistische Grundkenntnisse bei den Nutzer/innen dieser Analysemethoden wichtig. Mittlerweile stehen viele Module und Programmbibliotheken in Python oder R für die Datenanalyse, speziell auch für Zeitreihenanalysen zur Verfügung, aber ohne entsprechende Hintergrundkenntnisse sind Anwendungsfehler vorprogrammiert.

Ein aktuelles Beispiel für die Anwendung von Zeitreihenanalysen sind Prognosen im Zusammenhang mit dem Klimawandel. Zhang, An und Gao (2023) untersuchten diesbezüglich den Verlauf der jährlichen Durchschnittstemperatur in China auf der Basis von Zeitreihenmodellen. In dieser Studie wurde, nach einer Diskussion verschiedener Modellansätze, ein ARIMA-Modell für die Modellierung und Vorhersage ausgewählt. Für Tests auf Stationarität und die Parameterauswahl kamen Python-Bibliotheken wie z.B. statsmodel zum Einsatz. Basierend auf den ACF-Werten und den Testergebnissen wurden die Parameter des ARIMA-Modells geschätzt.

Abb. 8 veranschaulicht den Temperaturverlauf über einen Zeitraum von 184 Jahren von 1838 bis 2021. Die jährliche Durchschnittstemperatur in China zeigt kurzfristige Schwankungen und einen langfristigen Anstieg. Abb. 9 zeigt, zur Evaluierung des Modells mit den geschätzten Parametern, einen Vergleich der Temperaturvorhersagen des Modells mit den historischen Daten aus der Zeitreihe. Die Schattierung entspricht dem berechneten Konfidenzintervall. Abb. 10 stellt die Vorhersage des Modells bei einer Projektion bis 2030 dar (und darüber hinaus). Bei dieser Visualisierung der Vorhersage ist die Skalierung der Y-Achse zu

berücksichtigen. Im Sinne der Vergleichbarkeit wäre eine einheitliche Skalierung der Temperatur zielführend. Unabhängig davon zeigt sich, die auch der für China zu erwartende, langfristige Temperaturanstieg.

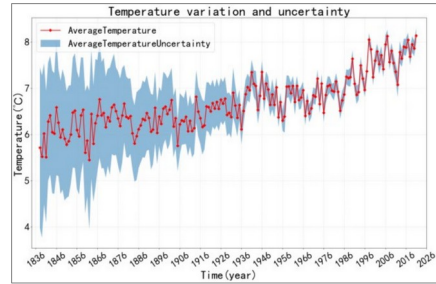


Abb. 8: Zeitreihe zur Durchschnittstemperatur Chinas (Abb. 2 aus [15]).

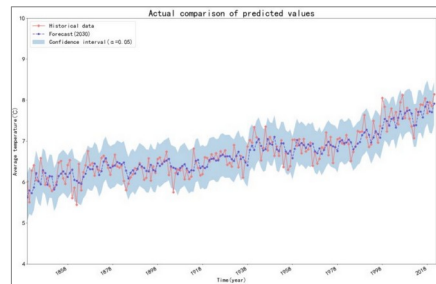


Abb. 9: Vergleich der Temperaturvorhersagen des Modells mit den Werten aus Abb. 8 (Abb. 10 aus [15]).

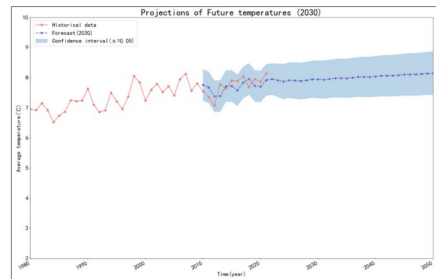


Abb. 10: Prognose der zukünftigen Durchschnittstemperatur in China (Abb. 11 aus [15]).

## Literaturverzeichnis

- [1] Peter J. Brockwell und Richard A. Davis. 2016. Introduction to Time Series and Forecasting (3rd ed.). Springer.
- [2] Peter P. Eckstein. 1999. Zeitreihenanalyse. In: Angewandte Statistik mit SPSS. Gabler Verlag. [https://doi.org/10.1007/978-3-322-92004-1\\_7](https://doi.org/10.1007/978-3-322-92004-1_7)
- [3] Steffen Herbold. 2022. Data-Science-Crashkurs. Kap. 9: Zeitreihenanalyse. dpunkt.verlag. [https://data-science-crashkurs.de/chapters/kapitel\\_09.html](https://data-science-crashkurs.de/chapters/kapitel_09.html)
- [4] Gebhard Kirchgässner und Jürgen Wolters. 2006. Einführung in die moderne Zeitreihenanalyse. Verlag Franz Vahlen München.
- [5] Jens-Peter Kreiß und Georg Neuhaus. 2006. Einführung in die Zeitreihenanalyse. Springer, Berlin, Heidelberg. 388 Seiten.
- [6] Peter von der Lippe. 1993. Deskriptive Statistik. Gustav Fischer Verlag, Stuttgart, Jena.
- [7] David McDowall, Richard McCleary und Bradley J. Bartos. 2019. Interrupted Time Series Analysis. Oxford University Press.
- [8] Mateusz Norel, Michał Kalczyński, Iwona Pińskwar, Krzysztof Krawiec, and Zbigniew W. Kundzewicz. 2021. "Climate Variability Indices—A Guided Tour" Geosciences 11, no. 3: 128. <https://doi.org/10.3390/geosciences11030128>
- [9] Python-Kurs. Python, Pandas und Zeitserien. Abgerufen am 03. April 2024 von [https://www.python-kurs.eu/pandas\\_time\\_series.php](https://www.python-kurs.eu/pandas_time_series.php)
- [10] Dario Radečić. How to Use R and Python Together? Try These 2 Packages. Abgerufen am 03. April 2024 von <https://www.appsilon.com/post/use-r-and-python-together>
- [11] Christina Satarhoff. 2018. Master-Modul 85-011: Zeitreihenanalyse. Skript, 27 Seiten. <https://pdfslide.org/download/link/skript-zeitreihenanalyse-ss2013-bwluni-master-modul85-011-zeitreihenanalyse.html>
- [12] Rainer Schlittgen und Cristina Sattarhoff. 2020. Angewandte Zeitreihenanalyse mit R. De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110694390>
- [13] Hussein Farooq Tayeb, Murat Karabat and Cihan Varol. 2020. Time Series Database Preprocessing for Data Mining Using Python. 2020 8th International Symposium on Digital Forensics and Security (ISDFS), Beirut, Lebanon, 2020, pp. 1-4, doi: 10.1109/ISDFS49300.2020.9116260
- [14] K. Witte. Zeitreihenanalyse. In: Angewandte Statistik in der Bewegungswissenschaft (Band 3). Springer Spektrum, Berlin, Heidelberg, 2019). [https://doi.org/10.1007/978-3-662-58360-9\\_11](https://doi.org/10.1007/978-3-662-58360-9_11)
- [15] Keyu Zhang, Zhulin An und Xuan Gao. 2023. Time Series Temperature Prediction Model Based on Big Data. 2023 IEEE 3rd Internat. Conf. on Electronic Technology, Communication and Information (ICETCI), Changchun, China. S. 1147-1152, <https://ieeexplore.ieee.org/document/10176450>



# Umgang und Auswirkungen von Datenlücken bei medizinischer Zeitreihenanalyse

Katrin Schrüfer  
Hochschule Reutlingen  
Reutlingen, Deutschland

[Katrin.Schruefer@Student.Reutlingen-University.de](mailto:Katrin.Schruefer@Student.Reutlingen-University.de)

## Abstract

Im Allgemeinen gibt es drei Methoden für den Umgang mit fehlenden Daten in Zeitreihen: Löschen, Imputation und Ignorieren. In der Praxis werden noch oft Datensätze gelöscht oder das Thema ignoriert, obwohl die Forschung gerade in den letzten Jahren viele neue Ansätze hervorgebracht hat. Es gibt viele Imputationsmethoden, von klassischen und häufig verwendeten Methoden wie Mittelwert, Median oder Multiple Imputation by Chained Equations bis hin zu Methoden, die auf rekurrenten neuronalen Netzen basieren, wie Long Short Term Memory oder Multi-Directional RNN. Für einen Vitalparameterdatensatz soll eine geeignete Methode für den Umgang mit fehlenden Daten gefunden werden. Aufbauend auf Funde in der Literatur werden die Imputationsmethoden Median, BRITS und SAITS miteinander verglichen. Die Medianimputation erzielte dabei die niedrigsten RMSE und MAE.

## CCS Concepts

Mathematics of computing → Probability and statistics, Computing methodologies → Machine learning

## Keywords

Missing data, Time series analysis, Imputation, bwHealthApp

## 1 Einleitung

Es gibt viele Methoden, um mit fehlenden Daten in Zeitreihen umzugehen. Durch den technischen Fortschritt im Bereich des maschinellen Lernens ist die Zahl der Methoden in den letzten fünf Jahren noch weiter gestiegen. In einigen Veröffentlichungen zu diesen Methoden werden die Auswirkungen auf die nachfolgende Analyse nicht berücksichtigt. Auch wird selten auf die Klassifikation der Datensätze hingewiesen, die für die weitere Methodenwahl wichtig ist. Der Umgang mit fehlenden Daten sollte ein wichtiger Bestandteil jeder Analyse sein, wird aber oft nur am Rande behandelt [22].

### 1.1 Ziel der Arbeit

Die Motivation hinter dieser Arbeit ist, den Umgang mit fehlenden Daten in Zeitreihen im medizinischen Kontext näher zu betrachten. Dabei sollen klassisch und derzeit aktuelle Methoden zum Umgang mit fehlenden Daten anhand der gefundenen Literatur verglichen und nach einer Auswahl auf den Datensatz angewendet werden.

---

Betreuer Hochschule: Prof. Dr. C. Thies  
Hochschule Reutlingen  
Christian.Thies@Reutlingen-  
University.de

Informatics Inside Frühjahr 2024  
08. Mai 2024, Hochschule Reutlingen  
Copyright 2024 Katrin Schrüfer

## 1.2 Weiterer Aufbau der Arbeit

Weitergehend ist die Arbeit wie folgt aufgebaut: In Kapitel 2 werden kurz die Grundlagen zu Zeitreihen, die Definition und Klassifizierungen von Missing Data erläutert und der später verwendete Datensatz vorgestellt. Kapitel 3 enthält die gefundene Literatur und fasst diese kurz zusammen. Kapitel 4 enthält die Anwendung und Ergebnisse der gefundenen Methode mit dem in Kapitel 2 zuvor vorgestellten Datensatz. Darauf folgt in Kapitel 5 die Diskussion mit dem Fazit. Am Ende der Arbeit befindet sich die verwendete Literatur.

## 2 Grundlagen

In diesem Kapitel werden kurz die Grundlagen zu Zeitreihen inklusive Interrupted Time Series und Missing Data erklärt. Außerdem wird hier der später in der Anwendung verwendete Datensatz vorgestellt.

### 2.1 Zeitreihen (Time Series, TS)

Zeitreihen sind Datensätze, die in regelmäßigen Abständen aufgezeichnet werden. Dabei ist zu beachten, dass sich Zeitreihen aus zufälligen Werten mit Trendverhalten und wiederkehrender Saisonalität zusammensetzen. Zeitreihendatensätze bestehen aus einer oder mehreren Informationen, die in einem oder unterschiedlich festen oder flexiblen Zeitintervall gemessen werden [23].

Bei Zeitreihen ist es normal, dass Teile der Daten fehlen, für die weitergehende Analyse ist ein vollständiger Datensatz am besten.

### 2.2 Missing Data

Seit 1976 werden fehlende Daten in drei Mechanismen aufgeteilt: 1) Missing At Random (MAR), 2) Missing Completely At Random (MCAR) und 3) Missing Not At Random (MNAR) [6]. Damit die beste Methode zum Umgang der fehlenden Daten gefunden werden kann, sollte die vorherige Struktur der

fehlenden Daten einem der drei Mechanismen zugeordnet werden [7]. Nachfolgend ein paar wichtige Stichpunkte zu den Klassen.

#### MAR

- Daten, die unabhängig von fehlenden Daten fehlen, aber abhängig sind von beobachteten Daten [7]
- Verzerrung des Ergebnisses bei Ignorieren [7]
- Häufige Klassifikation [7, 22]

#### MNAR

- Daten, die von den fehlenden Daten selbst oder von anderen Daten im Datensatz abhängig sind [7]
- Bei Ignorieren verzerrt das Ergebnis und die Beziehung der Variablen [7]

#### MCAR

- Daten, die rein zufällig fehlen und nicht von den fehlenden Daten oder von den beobachteten Daten abhängig sind [7]
- Keine Verzerrung, kann sicher ignoriert werden [7]
- Häufige Klassifikation, erzeugt einen höheren Standardfehler durch die verkleinerte Stichprobe [7, 22]

Die Menge an fehlenden Daten und die Art der Lücken sind kritische Komponenten, die darüber entscheiden, ob eine Imputationsmethode Erfolg oder Misserfolg hat [22].

Klassifiziert werden die Lücken in Tabelle 1 [22].

**Tabelle 1. Klassifizierung der verschiedenen Typen von Lücken in fehlenden Daten [22, Table 2]**

Typ der Lücke	Kontinuierliches Fehlen
Punktücke	1
Kurze Lücke	Bis zu 3%
Hohe Lücke	3% bis 10%
Sehr hohe Lücke	Über 10%

### 2.3 Datensatz

Die in der Anwendung verwendeten Daten wurden im Rahmen einer Studie mit der bwHealthApp<sup>1</sup> aufgenommen. Die Daten wurden für Patienten im Alltag während der Krebsbehandlung aufgenommen. Insgesamt enthält der Datensatz Daten über 2 unterschiedliche Sensoren. Der Armsensor ist der AS99 der Firma Beurer GmbH<sup>2</sup> und der Ohr Sensor ist der CosinussTwo der Firma Cosinuss GmbH<sup>3</sup>.

Mit dem Armsensor wurde bei 16 Patienten und für den Ohrsensor bei 13 Patienten Daten aufgenommen. In den Datensätzen für den Arm wurden die 6 Variablen aus Tabelle 2 erfasst und für den Ohrsensor die 3 Variablen aus Tabelle 3. Die Messwerte wurden in unterschiedlicher Frequenz aufgezeichnet. Die HR, TEMP und SPO2 wurden pro Sekunde erfasst und die anderen Variablen in 10 Minuten abständen. Die Daten wurden im Zeitraum von Juli bis November 2022 aufgenommen.

**Tabelle 2. Messwert und Einheit für den Armsensor**

Messwert	Einheit
Herzfrequenz (HR)	Bpm (beats per minute)
gelaufenen Schritte am Tag (SC)	Schritte
bewegte Distanz am Tag (DIST)	Kilometer
Aktivität am Tag (AD)	Minuten
verbrauchte Kalorien am Tag (CALs)	Kcal, Kilokalorien
Schlaf am Tag/Nacht (SD)	Minuten

<sup>1</sup> <https://bwhealthapp.reutlingen-university.de/page/home>

<sup>2</sup> <https://www.beurer-shop.de/p/beurer-aktivitaetssensor-as-99-pulse-bluetooth/>

**Tabelle 3. Messwert und Einheit für den Ohrsensor**

Messwert	Einheit
Herzfrequenz (HR)	Bpm (beats per minute)
Temperatur (TEMP)	Grad Celsius
Sauerstoffsättigung (SPO2)	Prozent

Von den 29 Datensätzen wurden 12 nicht weiter in Betracht gezogen, da sie aus zu wenig Daten bestanden. Es mussten Daten von mindestens 20 unterschiedlichen Tagen vorliegen. Der insgesamt Messzeitraum der Datensätze konnte dabei die Mindestanzahl von 20 übersteigen, da an manchen Tagen keine Messwerte aufgenommen wurden. Die unterschiedliche Tagesanzahl mit Messwerten von allen Datensätzen ist in Abbildung 1 zu sehen und die Aufteilung der Datensätze pro Sensor in Tabelle 4.

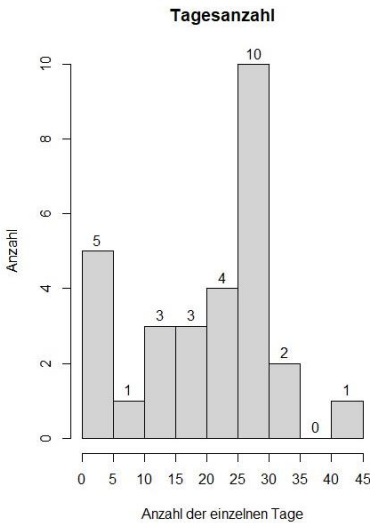
**Tabelle 4. Verwendete Datensätze für Arm- und Ohrsensor**

Sensor	Verwendet	Gesamt
Armsensor	10	16
Ohrsensor	7	13

Es gab eine Vorfilterung der Werte auf physiologisch mögliche Werte. Was eine HR von 35 bis 250, eine Temperatur von 34,5 bis 42 und eine Sauerstoffsättigung von 85 bis 100 bedeutet. Gab es Werte außerhalb dieses Bereiches wurden diese zu fehlenden Werten umgewandelt. Die fehlende Datenrate der Datensätze ist in Tabelle 5 dargestellt.

<sup>3</sup> <https://store.cosinuss.com/products/cosinuss-two?variant=32175832924242>

Durchschnittlich hat bei der Variable HR am wenigsten Daten gefehlt.



**Abbildung 1: Tagesanzahlen der Datensätze von Arm- und Ohrsensor**

**Tabelle 5. Fehlende Datenrate pro Datensatz**

Datensatz Arm = A, Ohr = O		fehlende Datenrate	Datenrate
A1	0.925	A2	0.969
A3	0.949	A4	0.983
A5	0.984	A6	0.967
A7	0.961	A8	0.978
A9	0.982	A10	0.973
O1	0.973	O2	0.737
O3	0.955	O4	0.950
O5	0.704	O6	0.856
O7	0.780		

Die fehlenden Daten in den Datensätzen setzen sich aus kurzen, hohen und sehr hohen Lücken zusammen. Der Datensatz wird als MAR für die vorhandenen Tage und MNAR für die komplett fehlenden Tage

klassifiziert. Im Vergleich fehlen bei den Ohrsensor-Datensätzen weniger Daten als bei den Armsensor-Datensätzen.

### 3 Stand der Forschung

In diesem Abschnitt wird die gefundene Literatur aufgearbeitet.

#### 3.1 Methodik

Es wurde in den Datenbanken PubMed, IEEE und ACM nach Veröffentlichungen gesucht. Die in PubMed gefundene Literatur wurde dabei als relevanter eingestuft als die in anderen Datenbanken gefundene Literatur. Es wurden ausschließlich Publikationen berücksichtigt, die von der Hochschule Reutlingen aus zugänglich sind und in den letzten fünf Jahren veröffentlicht wurden. Die weitere Relevanz der Literatur wurde anhand des Titels und des Abstracts überprüft. Weiter priorisiert wurden neuere Veröffentlichungen.

Es wurde mit dem Term „(“time series“ OR “time series data“) AND (“missing data“ OR “missing value“) AND medical“ gesucht. In PubMed wurden insgesamt 25 Artikel gefunden, weiter in Betracht gezogen wurden davon 19 Artikel. In IEEE wurden insgesamt 51 Artikel zu diesem Term gefunden. Weiter in Betracht gezogen wurden 25 Artikel. In ACM wurden 500 Artikel gefunden. Durch weitere Einschränkung auf die Kategorien „Research Article“ und „Survey“ ergaben sich insgesamt 237 Artikel, von denen 29 weiter in Betracht gezogen wurden.

#### 3.2 Erkenntnisse aus der Literatur

Generell gibt es drei Möglichkeiten, um mit fehlenden Daten umzugehen, Löschung [1, 11], Imputation und Ignoranz [22]. Dabei wird meistens eine Art der Imputation verwendet.

##### 3.2.1 Löschen

Bazo-Alvarez et. al. untersuchte 60 Interrupted Time Series-Studien, die im Jahr 2019 veröffentlicht wurden. Von nur 13 der 60

Studien wurde die Behandlung von fehlenden Daten berücksichtigt, dabei arbeiteten 11 mit der Löschung nicht vollständiger Daten (complete case analysis, CCA) [1]. CCA ist die listweise Löschung von Fällen in denen Variablen, die in der Analyse betrachtet werden, fehlen. Die zwei übrigen Studien verwendeten unterschiedliche Mixed-Effects Modelle [1]. Weiter wurden in den 60 Studien fast nur mit aggregierten Daten gearbeitet, obwohl alle Studien auf feingranulareren Daten Zugriff hatten. Die Auswirkungen von fehlenden Daten bei aggregierten Daten wurden bei allen Studien nicht berücksichtigt [1]. Daten, die auf der Individualebene MAR sind, werden auf der aggregierten Ebene zu MNAR [1]. Nur zwei der Studien führte explizit die vermutete Kategorie ihrer fehlenden Daten auf, einmal MAR und MNAR [1].

### 3.2.2 Erkenntnisse

Durch Verwendung von CCA können Informationen verloren gehen und es führt zu einer möglichen Verzerrung der Analyse [1].

Eine andere Methode, um mit fehlenden Daten umzugehen sind Imputationen. Dabei werden verschiedene Verfahren verwendet, um die fehlenden Daten mit neu erstellten oder sich im Datensatz befindenden Daten zu ersetzen. Klassische Imputationsmethoden, um mit Zeitreihen umzugehen, sind die Mittelwert- [2,5,9] und Median-Imputation [2,12], wobei der Mittelwert oder der Median die fehlenden Werte ersetzt. Multiple Imputation (MI) [16], und ihre weit verbreitete Variante MI by Chained Equations (MICE) [3-5,8,9,14], ist ebenfalls eine klassische Imputationsmethode. Bei MICE werden erst willkürlich die fehlenden Werte initialisiert und dann die Gleichungen durch Regressionsmodelle iterativ aktualisiert. Methoden basierend auf K-Nearest Neighbor (KNN) [2,9,14] und Matrix Factorization (MF) [9,14] werden auch verwendet. In letzter Zeit wurden viele Deep-Learning-Architekturen basierend auf rekurrente neuronale Netze (RNN) [2,3,9,10,14,19], Convolutio-

nal Neural Networks (CNN) [10], Autoencoder (AE) [3, 15,19], Generative Adversarial Networks (GAN) und Transformer [2,19] veröffentlicht.

Imputationsmethoden aufbauend auf RNN sind z.B. Long Short Term-Memory (LSTM) und Methoden basierend auf Gated Recurrent Units (GRU). Diese Methoden sind allerdings auch nicht ohne Nachteil, da sie schlechter mit MNAR klassifizierten fehlenden Daten umgehen können [3]. Weitere Methoden, um mit fehlenden Daten umzugehen, wären die Interpolation. Klassische Interpolationsmethoden wie verschiedene Varianten des Spline-Modells [5,13] sind in der Literatur zu finden.

Für MICE wurde gezeigt, dass die Trennung von Imputationsprozess und Vorhersageprozess zu suboptimalen Ergebnissen führt, da das zweistufige Modelltraining die Korrelation zwischen den fehlenden Daten und den vorhergesagten Ergebnissen ignoriert [4]. Die häufig verwendeten Daten Imputationsmethode Last Observed Carried Forward (LOCF) und das Äquivalent Next Observation Carried Backward (NOCB) sind intuitiv in dem sie die Datenlücken mit dem letzten Wert vor oder dem ersten Wert nach der Datenlücke füllen. Diese Methoden sind aber nicht für längere Datenlücken geeignet, da diese Ansätze zu Verzerrungen führen können und etwaige Ausreißer nicht abgefangen werden [2, 3, 5].

### 3.2.3 Gefundene Literatur zu Imputation und Interpolation

Le, Do und Nguyen verglichen 2023 verschiedene Imputationsmethoden anhand des mittleren absoluten Fehlers (mean absolute error MAE). Sie verglichen Mittelwert-Imputation, Median-Imputation, LOCF und K-Nearest Neighbors Imputation (KNNI), sowie neuere Techniken wie Bidirectional Recurrent Imputation for Time Series (BRITS [21]), Transformer, und Self-attention-based imputation for time series (SAITS [20]) [2].

Um die Methoden vergleichen zu können, fügten sie künstliche, gleichmäßig verteilte fehlende Daten hinzu. Der spätere Vergleich wurde unter der MCAR Annahme durchgeführt [2]. Der Originaldatensatz, ein crowd-sourced Fitbit Datensatz, der in 2016 aufgenommen wurde, wurde in zwei Datensätze aufgeteilt. Ein Datensatz mit Daten zur täglichen Aktivität von 10 Teilnehmern ohne fehlende Daten mit insgesamt 420 Einträgen. Der zweite Datensatz umfasste alle Daten zur täglichen Aktivität des Datensatzes mit einer fehlenden Datenrate von ca. 36,7% mit über 2000 Einträgen [2].

Beide Datensätze wurden mit 10% bis 80% fehlende Datenrate für zufällige Werte aller Features simuliert. Da der zweite Datensatz die Originalrate der fehlenden Daten von ca. 36,7% enthielt summiert sich die Gesamtfehlrate bei diesem Datensatz auf 43% bis 87% [2]. Für den ersten Datensatz bei einer fehlenden Datenrate von unter 30% hatte die KNNI-Methode den niedrigsten MAE, darüber performte SAITS am besten. Bestätigt wurde die SAITS Methode als beste Methode im zweiten Datensatz, da dort die fehlende Datenrate bei 43% beginnt [2].

J. Zhang und P. Yin verknüpften 2019 bidirektionales LSTM mit Denoising Autoencoder für eine Imputation Methode, die mit MCAR und MNAR klassifizierte fehlende Daten umgehen kann [3]. Verglichen wurden die statischen Imputationsmethoden MissForest, Expectation Maximization, LOCF, MICE und eine RNN basierende Methode namens Multi-Directional RNN (MRNN). Die Methoden wurden auf 4 vollständigen Datensätzen aus der UCI Machine Learning Repository (Gesture, Eye State, Occupancy, EEG) und einem klinischen Vitalparameterdatensatz, dem eICU Collaborative Research Database, verglichen. Der eICU Datensatz, wurde auf einen Intervall von einer Stunde modelliert, was die MNAR fehlenden Daten herbeiführt [3]. Bei allen Datensätzen erzielte die von Zhang und Yin vorgeschlagene Methode einen niedrigeren RMSE. Auch bei Veränderung der fehlenden Datenrate zu

MNAR wurde ein niedrigerer RMSE erzielt [3].

Wang et. al. stellen 2023 ein bidirektionales zeit- und aufmerksamkeitsgekoppeltes Netzwerk (BiT-MAC) vor, um sowohl die Intra-Zeitreihen-Kopplung als auch die Inter-Zeitreihen-Kopplung der fehlenden Werte zu berücksichtigen [9]. Getestet wurde BiT-MAC auf dem PhysioNet/CiC 2012 und dem COVID-19 Datensatz. Es wurde der MAE und der mittlere relative Fehler (MRE) verglichen [9]. Die nicht RNN Methoden Mittelwert-Imputation, KNN, MF und MICE, sowie die RNN basierenden Methoden GRU-D MRNN, BRITS und BiT-MAC wurden verglichen. Für alle Methoden erzielte BRITS die besten MAE und MRE Ergebnisse auf beiden Datensätzen [9].

Li et. al. verglichen 2023 GRU-D, MRNN, KNN, MF, MICE und ihre Methode Multivariate Time Series for Survival Prediction (MTSSP) [10]. MTSSP kombiniert Imputation und Klassifikation. Sie verglichen den RMSE auf den MIMIC-III und MIMIC-IV Datensätzen mit einer zusätzlichen 10% fehlenden Datenrate. Dabei schnitt ihr Modell MTSSP mit einem RMSE von 0,0899 für MIMIC-III Datensatz und einen RMSE von 0,0586 für den MIMIC-IV Datensatz deutlich am besten ab [10]. Am zweitbesten war für den MIMIC-III Datensatz GRU-D mit einem RMSE von 0,1083 und für MIMIC-IV Datensatz war MRNN mit 0,0032 besser als GRU-D [10].

Zhang et. al. stellte 2023 Medical ConvTransformer-BERT Modell (MedCT-BERT) aufbauend auf der GAN-Variante GAIN und ConvTransformer vor [17]. Damit die GAIN-Variante auch unterschiedlichen Sequenzlängen durch fehlende Daten und Mustern parallel verarbeiten kann [17]. Sie verglichen GAIN mit SAITS und BRITS auf ihren RMSE bei unterschiedlichen Fehlmengen, dabei erhielt GAIN auch bei 60% Fehlmenge den besten RMSE [17].

## Umgang und Auswirkungen von Datenlücken bei medizinischer Zeitreihenanalyse

M und Narayan verglichen 2023 typisch verwendete Imputationsmethoden auf dem MIMIC-III Datensatz auf den Umgang mit MNAR [18]. Auswahlmodelle oder Mischungsmodelle (pattern mixture models, PMM) sind typische Techniken für MNAR Längsschnittdaten, da sie die Verteilung auf Grundlage des Musters der fehlenden Daten, sowie die bedingte Wahrscheinlichkeitsverteilung der Messwerte der fehlenden Daten berücksichtigen [18]. M und Narayan verglichen KNN, MICE und PMM anhand der normalisierten quadratischen Abweichung (normalized root-mean-square deviation, NRMSD), dabei zeigte sich, dass PMM mit einem NRSMD von 0,1782 deutlich besser als die anderen Methoden war. Zu beachten ist aber, dass für bestimmte Experimente MICE den NRMSD von PMM übertraf [18]. Am Schluss warnen M und Narayan vor übermäßig komplexen Modellen [18].

### 3.3 *Verwendete Methoden*

Die auf den Datensatz in Kapitel 2 angewendeten Methoden sind Medianimputation, BRITS [21] und SAITS [20]. Im nachfolgenden werden die verwendeten Methoden kurz erklärt.

Es wurde sich für den Vergleich mit diesen Methoden entschieden, da sowohl BRITS [21] als auch SAITS [20] auf einem medizinischen Vitalparameterdatensatz mit guten Ergebnissen getestet wurden. Der Median sollte hier eine unkomplizierte und klassische Vergleichsbasis liefern.

#### 3.3.1 *Median*

Der Median wird für jede Variable pro Aufnahme Tag berechnet. Anschließend wurden alle fehlenden Werte mit dem Median der Variable ersetzt.

#### 3.3.2 *BRITS [21]*

BRITS wurde von Cao et. al. 2018 entwickelt. Die von BRITS erzeugten Werte werden als Teil des RNN-Grafen behandelt und können dadurch während der Backpropagation aktualisiert werden. BRITS stellt keine

spezifischen Annahmen an den Datensatz voraus, da BRITS über ein bidirektionales rekurrent dynamisches System lernt. Dadurch kann BRITS mit mehreren korrelierten fehlenden Werten und nichtlinearer Dynamik umgehen. Für weitere Informationen über BRITS lesen sie [21].

#### 3.3.3 *SAITS [20]*

SAITS wurde von Du, Cote, und Liu 2023 entwickelt. SAITS besteht aus zwei gewichteten Blöcken mit diagonal maskierter Selbstaufmerksamkeit (DMSA, diagonally-masked self-attention block). Die DMSA erfassen sowohl die zeitlichen Abhängigkeiten als auch die Merkmalskorrelationen zwischen den Zeitschritten. Durch die Möglichkeit den zwei DSMA-Blöcken dynamisch Gewichte zuzuweisen, kann die Aufmerksamkeitskarte und die Informationen über die fehlenden Daten genauer repräsentiert werden. Für weitere Informationen zu SAITS lesen sie [20].

## 4 *Anwendung*

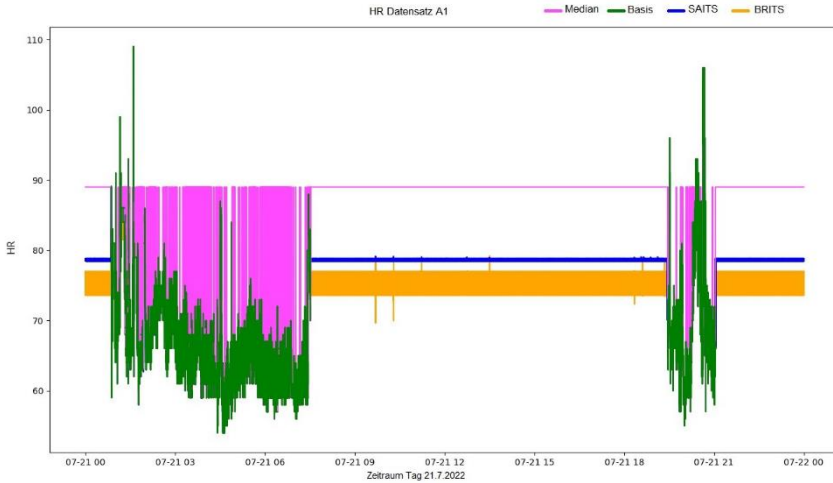
In diesem Abschnitt werden die gefundenen Ansätze an dem in Kapitel 2 beschriebenen Datensatz getestet.

### 4.1 *Ergebnisse und Vergleich des RMSE*

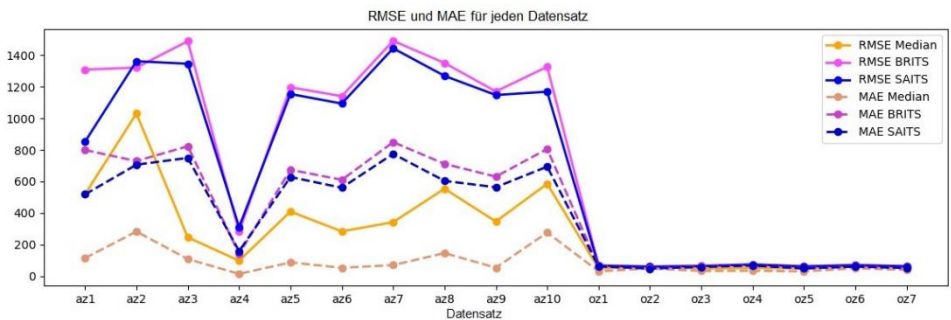
Für die Imputation von SAITS [20] und BRITS [21] wurden die Implementationen aus dem GitHub-Repository des SAITS [20] Artikel verwendet. Der Code wurde auf die Struktur der Daten angepasst. Weiter wurde eine Sequenzlänge von 6 und eine Trainingsdauer von 100 Epochen für das Training der Daten gewählt. Ein Ausschnitt der Grunddaten aus Datensatz A1 für den HR ist in Abbildung 2 zu sehen. Zu sehen sind die Basisdaten in grün, der Median in rosa, SAITS in blau und BRITS in orange. Wie hier gut zu erkennen ist, gibt es für die grünen Linien mehrere größere und kleinere Lücken in den aufgenommenen HR-Werten. Auch gut erkenntlich ist der in rosa imputierten Medianwert für den Tag, der knapp unter 90bpm

liegt. Wie in Abbildung 2 zu sehen, lassen sich aus der Medianimputation keine weiteren Aussagen zur HR-Struktur über den Tag sagen. Bei SAITS und BRITS war dies ebenfalls nicht möglich, beide pendelten sich auf

einen Wertebereich ein der selten überstiegen wurde. Dieses Verhalten ist auch bei den anderen Variablen und Datensätzen zu finden.



**Abbildung 2. HR von A1 für den Tag 20.7.2022, Basisdaten in grün, Median in rosa, SAITS in blau und BRITS in orange**



**Abbildung 3. RMSE und MAE für alle Datensätze**

Die Imputationsmethoden werden anhand des RMSE und des MAE verglichen, dieser wird für den gesamten Datensatz errechnet. Für die Berechnung wurden die fehlenden Werte der Basisdaten mit Null aufgefüllt. Die Ergebnisse sind in Abbildung 3 grafisch

dargestellt. Auf der X-Achse sind alle Datensätze aufgereiht. Der errechnete RMSE oder MAE für die Imputation ist auf der Y-Achse festgehalten.

Für Abbildung 3 ergibt sich daraus ein Liniendiagramm, das sowohl die Unterschiede

zwischen den Datensätzen als auch die Unterschiede der Imputationen innerhalb eines Datensatzes zeigt. Wie in der Abbildung zu sehen, erreicht die Medianimputation den besten RMSE und MAE für jeden Datensatz.

## 5 Diskussion und Fazit

Es wurden drei Imputationsmethoden für den Datensatz verglichen, SAITS, BRITS und Median pro Tag. Dabei waren die Werte für den RMSE und MAE für den Median am besten. Es ist aber nicht auszuschließen, dass eine hier nicht verwendete Methode ein besseres Ergebnis geben kann. Gerade da es viele Methoden gibt, um mit fehlenden Daten umzugehen und es in Zukunft nur noch mehr geben wird. Wie die Literatur aber zeigt, werden nur selten die Klassifikation der fehlenden Daten, sowie die verschiedenen großen Datenlücken berücksichtigt. Es muss weitere Forschung in diesem Bereich getätigt werden, auch für mehr Übertragbarkeit der verschiedenen Methoden auf andere Datensätze. Der Umgang mit fehlenden Daten ist ein wichtiger Teil jeder Analyse von Daten und die möglichen Auswirkungen eines gewählten Ansatzes dürfen nicht übersehen werden. Dabei sollten nicht nur immer kompliziertere Methoden in Betracht gezogen werden.

## Literaturverzeichnis

- [1] J. Bazo-Alvarez, T. P. Morris, J. R. Carpenter and I. Petersen. Current Practices in Missing Data Handling for Interrupted Time Series Studies Performed on Individual-Level Data: A Scoping Review in Health Research. *Clinical Epidemiology*, 2021, vol. 13, pp. 603–613.
- [2] L. P. Le, T. T. Do, T. Nguyen. Data Imputation for Multivariate Time-series Data. *International Conference on Knowledge and Systems Engineering (KSE)*, 2023, vol. 15, pp. 1-6
- [3] J. Zhang, P. Yin. Multivariate Time Series Missing Data Imputation Using Recurrent Denoising Autoencoder. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019, pp. 760-764
- [4] H. Gao, R. Li, J. Wang, H. Zhao, S. Yan, L. Ma. Research on Bidirectional Recurrent Imputation of Multivariate Time Series for Clinical Outcomes Prediction. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022, pp. 954-960
- [5] Y. Yin, C. -A. Chou. An Enhanced Imputation Approach for Spatio-Temporal Clinical Data. *IEEE International Conference on Automation Science and Engineering (CASE)*, 2022, pp. 813-818
- [6] D. B. Rubin. Inference and missing data. *Biometrika*, 1976, vol. 63, no. 3, pp. 581-592
- [7] H. M. Ahmed, B. Abdulrazak, F. G. Blanchet, H. Aloulou, M. Mokhtari. Long Gaps Missing IoT Sensors Time Series Data Imputation: A Bayesian Gaussian Approach. *IEEE Access*, 2022, vol. 10, pp. 116107-116119
- [8] J. Fonseca, X. Liu, H.P. Oliveira, T. Pereira. Mortality prediction using medical time series on TBI patients. *Computer methods and programs in biomedicine*, 2023, vol. 242
- [9] Q. Wang, G. Chen, X. Jin, S. Ren, G. Wang, L. Cao, Y. Xia. BiT-MAC: Mortality prediction by bidirectional time and multi-feature attention coupled network on multivariate irregular time series. *Computers in biology and medicine*, 2023, vol. 155
- [10] B. Li, Y. Shi, L. Cheng, Z. Yan, X. Wang und H. Li. MTSSP: Missing Value Imputation in Multivariate Time Series for Survival Prediction, *International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1-8
- [11] W. Cha, I. Yun, C.M. Nam, J.Y. Nam und E.C. Park. Evaluation of Assisted

- Reproductive Technology Health Insurance Coverage for Multiple Pregnancies and Births in Korea. *JAMA network open*, 2023, vol. 6, no. 6
- [12] W. Huang, A.Y. Meir, B. Olapeju, G. Wang, X. Hong, M. Venkataramani, T.L. Cheng, T. Igusa, L. Liang und X. Wang. Defining longitudinal trajectory of body mass index percentile and predicting childhood obesity: methodologies and findings in the Boston Birth Cohort. *Precision nutrition*, 2023, vol. 2, no. 2
- [13] O. Eales, K.E.C. Ainslie, C.E. Walters, H. Wang, C. Atchison, D. Ashby, C.A. Donnelly, G. Cooke, W. Barclay, H. Ward, A. Darzi, P. Elliott, S. Riley. Appropriately smoothing prevalence data to inform estimates of growth rate and reproduction number. *Epidemics*, 2022, vol. 40
- [14] Y. Luo. Evaluating the state of the art in missing data imputation for clinical data. *Briefings in bioinformatics*, 2022, vol. 23, no. 1
- [15] G.H. Kwak, L. Ling, P. Hui. Predicting the Need For Vasopressors in the Intensive Care Unit Using an Attention Based Deep Learning Model. *Shock (Augusta, Ga.)*, 2021, vol. 56, no. 1, pp. 73-79
- [16] S. Wambua, L. Malla, G. Mbevi, A.P. Nwosu, T. Tuti, C. Paton, S. Cheburet, A. Many, M. English, E.A. Okiro. The indirect impact of COVID-19 pandemic on inpatient admissions in 204 Kenyan hospitals: An interrupted time series analysis. *PLOS global public health* vol. 1, no. 11
- [17] K. Zhang, K. Niu, Y. Zhou, W. Tai and G. Lu. MedCT-BERT: Multimodal Mortality Prediction using Medical ConvTransformer-BERT Model. *IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2023, pp. 700-707
- [18] F. B. M und S. Narayan. Analysis of Multiple Imputation Techniques in Healthcare Data. *1st International Conference on Circuits, Power and Intelligent Systems (CCPIS)*, 2023, pp. 1-6
- [19] M. Isgut, L. Gloster, K. Choi, J. Venugopalan and M. D. Wang. Systematic Review of Advanced AI Methods for Improving Healthcare Data Quality in Post COVID-19 Era. *IEEE Reviews in Biomedical Engineering*, 2023, vol. 16, pp. 53-69
- [20] W. Du, D. Cote, Y. Liu. SAITS: Self-Attention-based Imputation for Time Series. *Expert Systems with Applications (ESWA)*, 2023, vol. 219, issue C
- [21] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, Y. Li. BRITS: Bidirectional Recurrent Imputation for Time Series. *Advances in Neural Information Processing Systems*, 2018, vol. 31 (NeurIPS 2018)
- [22] D. Adhikari, W. Jiang, J. Zhan, Z. He, D.B. Rawat, U. Aickelin, H.A. Khorshidi. A Comprehensive Survey on Imputation of Missing Data in Internet of Things. *ACM Computing Surveys*, 2022, vol. 55 issue. 7 no. 133, pp 1–38
- [23] J. Sanchez. *Time Series for Data Scientists. Data Management, Description, Modeling and Forecasting*. Cambridge University Press, 2023



©2024 Katrin Schrüfer. Lizenznehmer Hochschule Reutlingen, Deutschland. Dieser Artikel ist ein Open-Access-Artikel unter den Bedingungen und Konditionen der Creative Commons Attribution (CC BY)-Lizenz. <http://creativecommons.org/licenses/by/4.0/>

# Analyse der Kernaspekte von Recommender Systems

Boris Garkusha

Hochschule Reutlingen  
Reutlingen, Deutschland

Boris.Garkusha@Student.Reutlingen-University.de

## Abstract

Empfehlungsdienste, zu Englisch Recommender Systems, werden sowohl im professionellen als auch im Haushalt heutzutage täglich verwendet. Es handelt sich dabei meistens um Algorithmen, die Nutzern automatisch Objekte aus einem Katalog empfehlen sollen. Aufgrund des Umfangs an Faktoren, welche bei der Entwicklung eines Empfehlungsdienstes zu beachten sind, kann es für Entwickler schwierig sein, passende Empfehlungen zu generieren. Diese Arbeit erläutert, welche Faktoren für erfolgreiche Empfehlungsdienste notwendig sind. Diese Kernaspekte werden konkret spezifiziert und ihrer Bedeutsamkeit nach priorisiert. Diese Faktoren sollen Entwicklern bei der Erstellung eines Empfehlungsdienstes helfen. Anhand einer Literaturrecherche können erfolgreich Kernaspekte von diversen Recommender Systems beschrieben werden. Eine weitere Literaturrecherche zu den erforschten Aspekten wird zur Verifizierung und Evaluierung der Faktoren verwendet. Diese Arbeit ermöglicht es Entwicklern, besser zu verstehen, welche Faktoren für sie wichtig sind, falls sie einen Empfehlungsdienst planen.

---

Betreuer Hochschule: Prof. Dr. Uwe Kloos  
Hochschule Reutlingen  
Uwe.Kloos@Reutlingen-  
University.de

Informatics Inside Frühling 2024  
08. Mai 2024, Hochschule Reutlingen  
Copyright 2024 Boris Garkusha

## CCS Concepts

• **Information systems~Information retrieval** → **Users and interactive retrieval** → **Personalization** • Software and its engineering → Software creation and management → Designing software → Requirements analysis

## Keywords

recommendation systems, hybrid recommenders, critiquing-based recommenders, user evaluations, software planning

## 1 Einleitung

Empfehlungsdienste sind für viele Webseiten und Dienste unvermeidbar. Wie der Name impliziert, handelt es sich dabei um Algorithmen oder Werkzeuge, die Nutzern Empfehlungen generieren sollen. [1] Im Gegensatz zum gewöhnlichen Handel kann nicht für jeden Kunden eine 100 % persönliche Empfehlung geboten werden, und oftmals wollen das die Nutzer von diversen Plattformen auch nicht. Viele Nutzer wollen sich überraschen lassen und Empfehlungen bekommen, die sie nicht haben würden, falls sie direkt nach einem Objekt gesucht hätten. [1] Umso wichtiger ist es also, dass das System für den entsprechenden Fall geeignet ist. Daher ergibt sich die Forschungsfrage dieses Papers: Welche Faktoren sind bei der Erstellung eines Empfehlungsdienstes relevant und wie sollten diese priorisiert werden?

Um diese Frage beantworten zu können, werden in diesem Paper erfolgreiche Empfehlungsdienste analysiert und wichtige

Kernaspekte identifiziert. Diese Kernaspekte werden evaluiert und ihrer Relevanz nach sortiert. Die Verwendung dieser Aspekte soll Entwicklern und Forschern eine Basis bieten, die verwendet werden kann, um Empfehlungsdienste gestalten zu können.

## 2 Definition Empfehlungsdienst

F. Ricci et al. definieren Recommender Systems (RS), auch Empfehlungsdienste genannt, als Software-Werkzeuge und Techniken, die Nutzern für sie wahrscheinlich interessante Artikel empfehlen sollen. [1] Dabei ist der „Artikel“ ein Objekt in einem Katalog, unabhängig von der Kategorie; es kann sich also um Musik, Videos, Nachrichten etc. handeln. Jedes RS spezifiziert sich gewöhnlicherweise auf eine einzelne Art von Artikel und ist entsprechend in Nutzeroberflächen für den gewählten Typ integriert, wie beispielsweise die Startseite von Netflix, die den Nutzern durch ein RS ausgewählte Videos anzeigt. [2]

## 3 Auswahl der Kategorien

Recommender Systems haben alle dasselbe Ziel: Die Empfehlung eines Objektes aus einem Katalog von Objekten für den Nutzer. Wie diese Empfehlung getroffen wird, unterscheidet sich allerdings stark von Dienst zu Dienst: Es gibt mehrere Arten von Empfehlungsdiensten, wie Dienste auf Basis der Geschmäcker von ähnlichen Nutzern oder Dienste auf Basis vorheriger Auswahlen der Nutzer oder solche, die beide diese Aspekte kombinieren. [1] Diese Systeme alle als gleich zu betrachten und direkt zu vergleichen führt daher nicht zu schlüssigen Ergebnissen. Für diese Vertiefung werden deshalb Kategorien definiert, in die sich die einzelnen Dienste einteilen lassen können. Diese Kategorien sind angelehnt an die Empfehlungsstrategien, die F. Ricci et al. beschrieben haben.

**Basierend auf vorherigem Nutzerverhalten**

Das System speichert vorherige Auswahlen des Nutzers. Die ausgewählten Items haben alle mehrere verschiedene Aspekte, wie beispielsweise das Thema im Fall eines Nachrichtendienstes. Das System empfiehlt dem Nutzer dann Objekte, welche ähnliche Charakteristiken zu den vorherigen haben.

**Basierend auf ähnlichem Nutzerverhalten**

Auch als kollaborative Auswahl bekannt. Diese Art speichert das Verhalten aller Nutzer und findet anhand dieser Daten ähnliche Nutzer zu dem Aktuellen. Wenn eine bestimmte Anzahl an ähnlichen Nutzern ein Objekt mögen, von welchem der aktuelle Nutzer nichts weiß, wird dem aktuellen Nutzer dann dieses Item empfohlen.

**Kategorie 1 – Basierend auf Nutzerkriterien**

Das System kommuniziert direkt mit dem Nutzer und versucht, aus seinen/ihren Angaben Empfehlungen zu erstellen. Li Chen et al. beschreiben den standardmäßigen Ablauf eines solchen Systems: Der Nutzer wird direkt nach den Hauptsuchkriterien gefragt und basierend darauf kriegt er/sie eine oder mehrere initiale Empfehlungen. [3] Jetzt kann sich der Nutzer entweder direkt für ein Objekt entscheiden oder er/sie kann weitere Kriterien bezüglich des gesuchten Produkts definieren. Dieser Schritt wird dann so oft ausgeführt, bis der Nutzer zu einer Entscheidung kommt.

**Kategorie 2 – Hybrid**

Anstatt Kategorien 1.1 und 1.2 getrennt zu beobachten, wird eine Kombination verwendet. Dieser hybride Empfehlungsvorgang betrachtet sowohl das eigene Nutzungsverfahren als auch das von ähnlichen Nutzern derselben Plattform. Es wird dieses Verfahren betrachtet, da es deutlich häufiger auftaucht als die getrennten Methoden und da die getrennten Methoden auf ähnlichen Daten und Aspekten der zu empfehlenden Objekte beruhen. [1]

Oftmals gibt es Systeme, welche hauptsächlich auf Nutzerdaten beruhen, diese aber

## Analyse der Kernaspekte von Recommender Systems

durch Kriteklemente ergänzen. Ein einfaches Beispiel wäre YouTube: Während die Empfehlungen auf dem hybriden Vorgang aufgebaut sind, erlaubt die Plattform den Nutzern auch, Videos gut oder schlecht zu bewerten. Diese Bewertungen werden dann für zukünftige Empfehlungen beachtet. Wichtig ist es, hier eine Unterscheidung zu ziehen: Obwohl diese Systeme Kriteklemente besitzen, basieren sie nicht exklusiv auf dieser Kritik, wie Kategorie 1 es beschreibt. Dementsprechend dient die Kritik als zusätzliche Daten bezüglich des Nutzerverhaltens. Als Folge wird für diese Vertiefung beispielweise YouTube nicht in Kategorie 1 präsentiert, obwohl es Kritiken erlaubt. Im Vergleich dazu basiert beispielweise MovieLens primär auf Nutzerfeedback und benutzt dann zusätzlich dazu die Kriterien anderer, ähnlicher Nutzer, um zu einem Entschluss zu kommen.

### 4 Auswahl Empfehlungsdienste

Drei Dienste wurden auf Basis ihrer Relevanz, öffentlicher Dokumentation und ihres Erfolgs ausgewählt.

#### Kategorie 1 – Basierend auf Nutzerkriterien

1. MovieLens

#### Kategorie 2 – Hybrid

1. Netflix
2. Amazon

### 5 Analyse Empfehlungsdienste

In diesem Kapitel werden die ausgewählten Empfehlungsdienste analysiert.

#### 5.1 MovieLens

MovieLens<sup>1</sup> ist ein kritikbasierter Empfehlungsdienst in Form einer Webseite. Er soll

in der Lage sein, dem Nutzer Filme zu empfehlen, indem er/sie andere bisher gesehene Filme mit einer Sternbewertung bewertet. Zusätzlich dazu kann der Nutzer sogenannte Tags zu seinem Profil hinzufügen. Jeder Film besitzt mehrere dieser Tags, was bei der Empfehlung beurteilt wird. Zusätzlich verfügt die Webseite über vier unterschiedliche Empfehlungsalgorithmen, die von dem/der Nutzer\*in auswählbar sind:

1. Eine nicht-personalisierte Empfehlung, basierend komplett auf dem Verhalten anderer Nutzer und deren Empfehlungen.
2. Eine Empfehlung, die komplett auf den ausgewählten Tags basiert.
3. Eine Empfehlung, die die Bewertungen des Nutzers mit denen anderer Nutzer vergleicht (dafür müssen mindestens 15 Bewertungen gegeben werden). [4]
4. Eine Empfehlung, die ebenfalls auf den Bewertungen des Nutzers und denen anderer Nutzer beruht, allerdings einen anderen Algorithmus namens *Funk Singular Value Decomposition* dafür verwendet. [4] Eine *Singular Value Decomposition* (SVD) ist eine auf der Fourier-Transformation basierende Methode in der linearen Algebra, mit der ein Datensatz reduziert und gleichzeitig nutzbarer gemacht wird, indem Korrelationen vorhergesagt werden können. Bei Empfehlungsdiensten eignet sie sich gut, um Korrelationen zwischen Nutzern und Katalogobjekten voraussagen zu können. [5] Daten, die mathematisch durch Matrizen dargestellt werden können, können mittels der SVD fallgerecht behandelt werden. Die *Funk SVD* ist eine Variante der SVD, die spezifisch für den Fall verwendet wird, wenn Datenmatrizen, wie es häufig mit echten Daten der Fall ist, Lücken haben. Die *Funk SVD* ignoriert die Lücken und rechnet mit den gegebenen Daten, um Voraussagen zu berechnen. [6]

---

<sup>1</sup> <https://movielens.org/>

MovieLens sticht durch die Auswahl eines Empfehlungsalgorithmus heraus. Die Ersteller\*innen der Webseite haben sich dafür entschieden, da verschiedene Algorithmen unterschiedliche Stärken bezüglich der Auswahl besitzen und sie mit den daraus erstandenen Nutzerdaten evaluieren wollten, wie akkurat die Empfehlungen dem Nutzerverhalten entsprechen. [4]

MovieLens ist ein interessanter Fall, da die Webseite eine komplett persönliche Empfehlung ohne Einwirkung anderer Nutzer\*innen generieren kann. Andere Empfehlungsdienste basieren in den meisten Fällen auf dem Vergleich des eigenen Nutzerverhaltens mit den Vorlieben anderer Nutzer\*innen, die ähnliches Verhalten gezeigt haben. Die Abhängigkeit von Daten anderer Nutzer\*innen kann, abhängig vom System, problematisch sein. Besonders neue Empfehlungsdienste haben oftmals Probleme mit der Generierung von Empfehlungen aufgrund eines Datenmangels. [9] Das wird die Cold-Start-Problematik genannt. MovieLens bietet hier, durch komplette Unabhängigkeit anderer Nutzerdaten, eine potenzielle Lösung des Problems.

## 5.2 Netflix

Netflix<sup>2</sup> ist, nach Angaben auf ihrer Unternehmensprofilseite, eines der größten Unterhaltungsdienste der Welt mit mehr als 270 Millionen bezahlten Abonnements. [15] Hauptsächlich dient Netflix als ein Streaming-Service – eine Applikation, mit der sich Videos über das Internet schauen lassen. Bei diesen Videos handelt es sich größtenteils um originale oder lizenzierte Filme und Fernsehsendungen.

Netflixs Empfehlungen basieren auf proprietären Algorithmen. [2] Für verschiedene Arten von Empfehlungen wird jeweils ein anderer Algorithmus verwendet. Laut C. Gomez-Urbe et al. verlieren Nutzer der Plattform Interesse und wechseln auf eine

andere Plattform, wenn sie nicht in der Lage sind, innerhalb von 60 bis 90 Sekunden einen interessanten Film/eine interessante Serie zu finden. [2] Entsprechend spielt die Geschwindigkeit, mit der gute Empfehlungen erstellt werden können, eine wichtige Rolle. Netflix sortiert Videos auf der Homepage der Applikation in verschiedene Kategorien ein, welche einem relativ vage definierten Thema entsprechen. Der dafür verwendete Algorithmus beruht auf den Daten des Nutzers – dessen Zu- und Abneigungen – muss aber weit verwendbar sein und kann als Resultat nicht so persönlich sein, wie man annehmen könnte.

Zusätzlich zu diesen Genre-basierten Reihen hat Netflix auch noch Genre-unabhängige Empfehlungen, die vollends auf Nutzerverhalten basieren und den ganzen Katalog in Betracht ziehen.

Eine weitere Reihe auf der Homepage basiert komplett auf Ähnlichkeit zwischen verschiedenen Videos. Netflix empfiehlt, Film A zu gucken, weil er viele Ähnlichkeiten mit Film B aufweist, den der Nutzer vor Kurzem gesehen hat. Diese Empfehlungen sind verhaltensunabhängig in dem Sinne, dass sie nur die Charakteristiken der Filme vergleichen und nicht in Betracht ziehen, was der Nutzer oder andere Nutzer des Services gesehen haben.

Damit der/die Nutzer\*in möglichst schnell zu einer Entscheidung kommt, basiert Netflixs Empfehlungsdienst darauf, dem Nutzer verschiedene Arten von Empfehlungen, basierend auf verschiedenen Faktoren, zu präsentieren. Die Genres, die auf der Homepage gezeigt werden, sind abhängig von den Vorlieben des Nutzers und anderer, ähnlicher Nutzer, sind aber an sich nicht sehr personalisiert. Eine andere Rubrik ist komplett personalisiert und wieder eine andere basiert allein auf Ähnlichkeiten verschiedener Videos. Es ist daher sehr schwer für Nutzer zu

<sup>2</sup> <https://www.netflix.com/>

verstehen, welche Empfehlungen aus welchem Grund gewählt wurden. Relevanz und Personalisierbarkeit werden für Geschwindigkeit getauscht.

### 5.3 Amazon

Amazon ist eine der weltweit größten E-Commerce-Plattformen weltweit. [7] Als eine E-Commerce-Plattform ist der Nutzen eines Empfehlungsdienstes selbstverständlich: Amazon gewinnt direkt davon, dass Nutzer\*innen Käufe durch ihren Dienst durchführen, und der Empfehlungsdienst ist dafür da, potenziellen Kund\*innen für sie interessante Artikel anzuzeigen.

Schon in 1998 verwendete Amazon einen Empfehlungsdienst, der auf artikelbasierten, kollaborativen Filtern basiert. [8] B. Smith et al. beschreiben, wie sich der Algorithmus von standardmäßigen, nutzerbasierten, kollaborativen Filtern abwandelt. Statt nach ähnlichen Nutzern zu suchen, werden bei Amazon erstmal ähnliche Artikel gesucht, zu denen, für die der/die Nutzer\*in Interesse gezeigt hat. Ähnliche Artikel sind in dem Fall diejenigen, die ähnliche Nutzer\*innen wahrscheinlich sind, zusätzlich zu dem ausgewählten Artikel kaufen. Durch diese Ketten an zusammenhängenden Artikeln lassen sich dann Empfehlungen erstellen. Bereits gesehene oder bestellte Artikel werden davor noch aussortiert.

Zeit spielt bei der Berechnung dieser Empfehlungen eine große Rolle. B. Smith et al. geben ein Beispiel, bei dem zwei Bücher, die fünf Monate nacheinander gekauft wurden, weniger Relevanz aufweisen als zwei Bücher, die am selben Tag gekauft wurden. [8] Diese Metrik kann auch zur Spezifikation verwendet werden. Speicherkarten werden oft zur selben Zeit wie Kameras gekauft, aber nicht andersrum, weshalb beim Kauf einer Kamera zwar Speicherkarten, beim Kauf einer Speicherkarte aber keine Kameras empfohlen werden.

Ein weiterer Teil von Amazons Empfehlungsdienst sind Nutzerbewertungen und

Rezensionen. [9] Amazon verfügt über eine Sternwertung pro Artikel – ein Stern für schlecht, fünf Sterne für gut. Es ist unklar, in wie weit Amazon bei der Empfehlungsgenerierung Nutzerbewertungen beachtet. Allerdings zeigt sich, dass Artikel mit höheren Bewertungen generell mehr Käufe vorweisen können und entsprechend häufiger Kunden empfohlen werden.

## 6 Auswahl Kernaspekte

Aus der Analyse der gegebenen Literatur konnten Merkmale festgestellt werden, die für das Erstellen eines Empfehlungsdienstes beachtet werden sollten. Diese Aspekte repräsentieren Ziele oder Ansprüche, die wichtig sind, bei der Erstellung eines Empfehlungsdienstes zu beachten.

### 1: Relevanz

Der/die Nutzer\*in muss das Gefühl haben, die empfohlenen Objekte entsprechen seinem/ihrem Geschmack.

### 2: Personalisierbarkeit

Dem/der Nutzer\*in müssen genügend Optionen gegeben werden, damit er/sie die Empfehlung mit Zuversicht akzeptieren kann.

### 3: Diversität

Ein besonderer Nutzen von Recommender Systems ist die Empfehlung von Optionen, von denen der Nutzer nichts wusste. Es ist wichtig für Empfehlungsdienste, diese Funktion zu erfüllen, um sich von Suchmaschinen abzusetzen.

### 4: Cold-Start Behandlung

Recommender Systems, die auf Nutzerdaten basieren, haben ein großes Problem zum Anfang des Betriebes, da die Anzahl der Nutzerdaten nicht ausreichend ist. [10] Ein gutes RS muss in der Lage sein, eine Lösung für dieses Problem zu bieten.

### 5: Geschwindigkeit

Abhängig vom System kann es wichtig sein, dass der Nutzer sehr schnell zu einer Empfehlung kommt.

## 7 Evaluierung Kernaspekte

In diesem Kapitel werden die ausgewählten Kernaspekte mittels weiterer Literatur kritisch analysiert.

### 7.1 Relevanz

Relevanz ist eines der wichtigsten Merkmale eines guten Empfehlungsdienstes, ist aber schwer, konsequent zu erzielen. Abhängig vom System müssen verschiedene Methoden zur Auswahl von Artikeln verwendet werden.

H. Jiang et al. beschreiben einen Ansatz, der auf dem Discrete-Choice-Modell basiert. [11] Hierbei werden die Wahrscheinlichkeiten berechnet, mit denen Nutzer einen Artikel aus mehreren ähnlichen Artikeln auswählen. Das Ziel ist dann, die Wahrscheinlichkeit, dass ein Artikel A ausgewählt wird, zu maximieren. Dieser Ansatz verfügt über mehrere Vorteile, mit denen die Relevanz besser berechenbar ist und somit Empfehlungen besser erstellt werden können. Es kann eine *Conversion Rate* berechnet werden, die klar mathematisch darstellt, welcher Teil einer Nutzergruppe sich für einen Artikel entschieden hat. Gleichzeitig muss bessere Relevanz nicht mehr für bessere Diversität getauscht werden und es kann, abhängig vom Fall, die relevantere Metrik benutzt werden.

Ein anderer Ansatz wäre der von S. Raza et al. Sie beschreiben eine zweischichtige Architektur für die Verwendung von Empfehlungsdiensten bei Nachrichtenseiten. [12] Die erste Schicht enthält ähnliche Nachrichten zu dem momentanen Artikel. Die zweite Schicht verfügt dann über Nutzerinformationen wie vorherige Aufrufe und Bewertungen. Eine Verlustfunktion berechnet den Effekt von anderen, nicht-empfohlenen Artikeln, um Diversität einzuführen.

### 7.2 Personalisierbarkeit

Personalisierbarkeit ist unterschiedlich anzuwenden für jedes System. Manche Systeme basieren komplett darauf, anpassbar zu sein,

während es für andere wichtig ist, Empfehlungen möglichst unkompliziert für den Nutzer zu erstellen. Die Personalisierbarkeit eines Empfehlungsdienstes steht entsprechend im Kontrast zu der Geschwindigkeit. Sollte das Ziel des Nutzers sein, möglichst schnell eine Empfehlung zu kriegen, sollte Personalisierung außenvorgelesen werden. Im Gegensatz allerdings bevorzugen viele erfahrene Nutzer\*innen, Empfehlungen für seinen/ihren ganz spezifischen Fall zu erhalten. In diesem Fall spielt es eine größere Rolle, dass der/die Nutzer\*in so viele Angaben zu seinem/ihrer Fall geben kann, wie möglich. Natürlich ist es möglich, beides zu haben: viele Dienste haben eine Startseite, die schnelle, weniger personalisierte Ergebnisse liefert und zusätzlich eine detailliertere Suchfunktion, mit der nach spezifischen Artikeln und dazugehörigen Empfehlungen gesucht werden können.

Einen Ansatz zur Personalisierbarkeit bietet, wie in Kapitel 5.1 beschrieben, MovieLens. Die Personalisierung passiert in einem Kreislauf: die Webseite empfiehlt einen Film, der Nutzer bewertet diesen und die Webseite verwendet diese Bewertungen, um besser angepasste Empfehlungen zu erstellen. MovieLens zeigt sich daher als eine Webseite, welche für erfahrene Filmliebhaber viel Nutzen zeigt, allerdings im Vergleich zu den Durchschnittsfilmfans weniger interessant wäre.

Hohe Personalisierbarkeit ist ein absolutes Muss für Empfehlungsdienste, welche für wichtige Entscheidungen verwendet werden. Ein Beispiel wäre die Wahl von Software, welche für eine längere Zeit verwendet werden soll: selbst wenn die Software kostenlos ist, wird dennoch Zeit und Aufwand und im Fall eines Unternehmens daher auch Geld verwendet, um die Verwendung der Software zu meistern.

### 7.3 Diversität

Durch Recherche im Feld der Empfehlungsdienste hat sich gezeigt, dass Diversität ein

wichtiger Aspekt ist, um bessere Empfehlungen zu generieren. [13] S. Vargas et al. beschrieben, wie erhöhte Diversität zu erhöhter Zufriedenheit von Nutzern und Attraktivität gegenüber der Plattform führen.

Die Unabhängigkeit zweier Artikel kann mathematisch bestimmt werden. Die meisten Algorithmen, die zu erhöhter Diversität führen sollen, sortieren Empfehlungslisten neu, um Artikel mit erhöhter Diversität weiter nach vorne zu schieben. [13] Dafür können verschiedene Algorithmen verwendet werden, um eine gute Balance aus Alt und Neu zu erstellen, die dem gegebenen Fall entspricht.

Diversität kann auch mittels Genres realisiert werden. Wenn es beispielsweise um Filmempfehlungen geht, können zu den empfohlenen Filmen auch welche hinzugefügt werden, deren Genre nicht dem vorherigen Nutzerverhalten entspricht. Das hat auch den Vorteil, dass die meisten Nutzer\*innen bereits das Konzept von Genres verstehen und leicht zwischen empfohlenen Filmen unterscheiden können, indem sie sich die angegebenen Stichworte anschauen. Es gibt ein paar Faktoren, die verwendet werden können, um Genre-Diversität zu messen, wie die generelle Ausbreitung verschiedener Genres in einer Empfehlungsliste oder die Redundanz von Artikeln desselben Genres.

### 7.4 Cold-Start Behandlung

Es ist zu beachten, dass es nicht nur eine einzelne Art von Cold-Start gibt. [10] Ein Beispiel: Bei Amazon gibt es zwei Faktoren: Nutzer und Artikel. Bei Verwendung des Dienstes durch eine\*n neuen Nutzer\*in gibt es einen Nutzer-Cold-Start, da das System noch keine Information über ähnliche Nutzer besitzt. Zusätzlich gibt es allerdings einen Produkt-Cold-Start, da neue Produkte noch kein assoziiertes Nutzerverhalten haben. Bestimmte Produkte werden bestimmten Nutzergruppen empfohlen, und wenn das System nicht weiß, für welche Nutzer das neue Produkt interessant sein könnte, kann es auch nicht personalisiert empfohlen werden.

In einem größeren Rahmen leiden neue Systeme ganz vom Cold-Start. Es ist schwer, Empfehlungen für irgendeinen Fall zu generieren, wenn keine vorherigen Fälle bekannt sind.

Einen Ansatz zur Lösung dieser Problematik stellen Stichwörter oder Tags dar. Ein neues Buch kann verschiedene Tags, basierend auf verschiedenen Elementen erhalten: dem/der Autor\*in, dem Genre, der Perspektive etc. Besonders bei aussagestarken Tags wie dem/der Autor\*in ist das Cold-Start-Problem gut zu behandeln, da Fans von dem/der Autor\*in automatisch das neue Buch empfohlen werden können.

Anders können einem neuen System oder einem Artikel Verhaltensdaten gleich zu Beginn gefüttert werden. Wenn beispielsweise davon ausgegangen werden kann, dass ein neuer Film dem Verlauf eines älteren Films, was Empfehlungen angeht, ähnelt, können diese Daten dem neuen Film als Startpunkt dienen. Im Laufe der Zeit, durch weiteres Nutzerverhalten, präzisieren sich Empfehlungen für den neuen Film dann, und er wird Nutzer\*innen empfohlen, für die er potenziell relevanter wäre.

Die Wichtigkeit der Behandlung des Cold-Start-Problems ist also stark vom Fall abhängig. Sollte es sich um ein kritikbasiertes System halten, ist die Problematik direkt nicht so schädlich wie bei hybriden oder gar ganz nutzerdatenbasierten Systemen.

### 7.5 Geschwindigkeit

Wie sich im Fall von Netflix gezeigt hat, kann die Geschwindigkeit eines Empfehlungsdienstes abhängig vom Fall das wichtigste, aber auch das irrelevanteste Kriterium sein. Wenn es darum geht, dem Nutzer eine so präzise und relevante Empfehlung wie möglich zu geben, spielt Geschwindigkeit so gut wie keine Rolle. Auf der Startseite eines Streaming-Dienstes gibt es im Gegensatz nichts Wichtigeres, als dass der/die Nutzer\*in schnell etwas Passendes findet. Um ein anderes Beispiel zu verwenden: Bei der

Verwendung von Suchmaschinen wie Google ist es häufig der Fall, dass Nutzer\*innen so schnell wie möglich das finden, wonach sie suchen. Es ist bekannt, dass eine sehr geringe Anzahl von Nutzer\*innen jemals auf die zweite Seite der Ergebnisse navigiert. Wahrscheinlicher ist es, dass stattdessen eine neue Suche mit einem ähnlichen, angepassten Suchterm stattfindet.

Es gibt verschiedene Methoden, um die Geschwindigkeit, mit der relevante Empfehlungen generiert werden, erhöht werden können. E. Kannout et al. verwenden einen Ansatz, basierend auf Assoziierungsregeln, welche die Wahrscheinlichkeit, mit der zwei oder mehr Artikel miteinander verknüpft sind, darstellen. [14] Die Idee ist, dass, wenn weniger Objekte betrachtet werden, die Berechnung der Empfehlungen schneller wird. Mittels der Assoziierungen von Objekten können schnell Objekte aus den potenziellen Empfehlungen entfernt werden. Darunter leidet allerdings die Diversität, was, abhängig vom Fall, wichtig ist zu beachten.

## 8 Priorisierung Kernaspekte

Anhand der untersuchten Beispiele von Empfehlungsdiensten und der Literaturrecherche ergibt sich die folgende Priorisierung der untersuchten Kernaspekte.

### 1: Relevanz

In so gut wie jedem Fall ist die Relevanz von höchster Priorität. Empfehlungen haben keinen Nutzen, sollte der/die Nutzer\*in das Gefühl haben, die Empfehlungen seien komplett irrelevant. Eine Ausnahme stellt ein Fall dar, bei dem ein\*e Nutzer\*in ausschließlich komplett neue Artikel ohne irgendeine Relevanz zum eigenen Geschmack empfohlen haben möchte. Eine zu hohe Relevanz hat die Gefahr, dass der Empfehlungsdienst praktisch zu einer Suchmaschine wird, die dem Nutzer nur Artikel bezüglich der einen Sache empfiehlt, von der der Dienst dankt, dass das ist, was der Nutzer sucht. Dennoch ist die Relevanz von extrem hoher Bedeutung für alle anderen Aspekte.

MovieLens versucht, erhöhte Relevanz durch die Verwendung verschiedener Algorithmen zu erreichen, während Amazon und Netflix die Relevanz der Empfehlungen an andere Nutzer\*innen als Basis zur Berechnung neuer Empfehlungen verwenden.

### 2: Cold-Start Behandlung

Spontan betrachtet hört sich Cold-Start nicht nach einem riesigen Problem an, doch falls man einen Empfehlungsdienst in seinen Einzelteilen betrachtet, wird schnell klar, dass die Problematik in vielen Aspekten auftritt und ein System so gut wie unbrauchbar macht, sollte es nicht proaktiv behandelt werden. Die Verwendung von externen Datenbanken kann hier stark von Nutzen sein. Ein Beispiel wäre ein Filmempfehlungsdienst, der auf einer externen Filmdatenbank basiert. Diesen Filmen wurden bereits Stichwörter zugeteilt, wodurch ein guter Startpunkt für Empfehlungen gegeben ist. Ein Startpunkt muss auf jeden Fall für alle relevanten Punkte gegeben sein.

MovieLens umgeht das Problem nahezu komplett, indem das System die direkte Kommunikation mit dem/der Nutzer\*in antritt. Alle Empfehlungen basieren direkt auf den eigenen Wahlen, wodurch die Datensammlung zwar mehr Zeit kostet und aufwändiger wird, der/die Nutzer\*in aber genauere Empfehlungen kriegt. Eine solche Methode ist sehr vom Einsatzfall abhängig: MovieLens ist ein Dienst für Enthusiasten, die gerne etwas Zeit benutzen würden, um bessere Empfehlungen zu erhalten. Dasselbe lässt sich nicht über Amazon oder Netflix sagen, welche lieber schnellere, simplere Empfehlungen für ihre Kunden generieren. Entsprechend ist Cold-Start bei diesen Diensten ein größeres Problem, für welches beide Dienste ihre eigenen Lösungsmethoden verwenden, aber dennoch darunter leiden.

### 3: Diversität

Gute Diversität führt bewiesenermaßen zu besseren Ergebnissen. [12] Das lässt sich gut

## Analyse der Kernaspekte von Recommender Systems

vorstellen, sollte man sich einen Empfehlungsdienst ausdenken, der gar keine Diversität aufweist: es würde immer nur dieselbe Art von Artikel wieder und wieder empfohlen werden. Empfehlungsdienste sind, der Definition nach, nicht dasselbe wie Suchmaschinen und sollten sich nicht ganz so robotisch anfühlen. Dennoch muss gut darauf geachtet werden, dass der Grad an Diversität dem Produkt entspricht.

Netflix verwendet verschiedene Empfehlungsalgorithmen für verschiedene Kategorien, welche alle nebeneinander präsentiert werden. Somit ist hohe Diversität geboten. MovieLens bietet auch einen Grad an Diversität durch die Verwendung mehrerer, auswählbarer Algorithmen. Bei Amazon lässt sich Diversität als ein größeres Problem betrachten, da das Gefühl, dass nur populäre Produkte empfohlen werden, merkbar ist.

### 4: Personalisierbarkeit & Geschwindigkeit

Diesen beiden Aspekten wurde dieselbe Priorität zugewiesen, da sie von sich aus abgeheigt sind. Höhere Personalisierbarkeit kommt automatisch mit reduzierter Geschwindigkeit und umgekehrt. Es ist daher umso wichtiger, die Prioritäten schon vor der Implementierung eines Empfehlungsdienstes klarzustellen.

MovieLens und Netflix stehen sich in diesen Rubriken direkt gegenüber. Dies lässt sich anhand der jeweiligen Zielnutzer\*innen erklären: MovieLens ist für Filmliebhaber, während Netflix für den Durchschnittsmenschen ist. Amazon trifft eine gute Mitte. Die E-Commerce-Seite bietet eine Anzahl an Filtern, welche auf verschiedenen Produktkategorien basieren, mit denen sich Empfehlungen besser personalisieren lassen. Doch auch ohne diese Filter lassen sich schnell passende Produkte durch einen starken Suchalgorithmus finden.

## Diskussion

Die endgültige Priorisierung der gewählten Aspekte ist aufgrund mangelnder Quellen

nicht immer zutreffend, doch die Ergebnisse sollten dennoch Entwicklern helfen, einen Leitfaden für die Gestaltung ihres Empfehlungsdienstes zu erstellen. Optimalerweise sollten auch weitere Faktoren betrachtet werden, die wegen des Umfangs der Arbeit ausgelassen wurden.

## Fazit und Ausblick

In diesem Paper wurden fünf Kernaspekte spezifiziert und deren Bedeutung und Implementierung beispielhaft beschrieben. Wie sich gezeigt hat, sind Empfehlungsdienste ein sehr weiträumiges Forschungsgebiet mit viel Spielraum – jedes System kann andere Empfehlungen verwenden.

## Literaturverzeichnis

- [1] Ricci, F., Rokach, L., & Shapira, B. (2021). Recommender Systems: techniques, applications, and challenges. In Springer eBooks (pp. 1–35). [https://doi.org/10.1007/978-1-0716-2197-4\\_1](https://doi.org/10.1007/978-1-0716-2197-4_1)
- [2] Gomez-Uribe, C. A., & Hunt, N. T. (n.d.). The Netflix Recommender System. *ACM Transactions on Management Information Systems*, 6(4), 1–19. <https://doi.org/10.1145/2843948>
- [3] Chen, L., & Pu, P. (2011). Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction*, 22(1–2), 125–150. <https://doi.org/10.1007/s11257-011-9108-6>
- [4] Ekstrand, M. D., Kluver, D., Harper, F. M., & Konstan, J. A. (2015). Letting Users Choose Recommender Algorithms. *Proceedings of the Ninth ACM Conference on Recommender Systems*. <https://doi.org/10.1145/2792838.2800195>
- [5] Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A Constant Time Collaborative Filtering Al-

- gorithm. *Information Retrieval Journal*, 4(2), 133–151.  
<https://doi.org/10.1023/a:1011419012209>
- [6] Vineslas, M. (2021, December 7). How Does the Funk Singular Value Decomposition Algorithm work in Recommendation Engines? *Medium*. Retrieved April 2, 2024, from <https://medium.datadriveninvestor.com/how-funk-singular-value-decomposition-algorithm-work-in-recommendation-engines-36f2fbf62cac>
- [7] Shvartsman, D. (2024, 6. März). Amazon: The World's Most Powerful Economic and Cultural Force. Investing.com. Abgerufen am 17. März 2024, von <https://www.investing.com/academy/statistics/amazon-facts/#:~:text=The%20latest%20statistics%20presented%20below.has%20300%20million%20active%20users>
- [8] Smith, B. & Linden, G. (2017b). Two Decades of Recommender Systems at Amazon.com. *IEEE Internet Computing*, 21(3), 12–18.  
<https://doi.org/10.1109/mic.2017.72>
- [9] Leino, J. & Räihä, K. (2007). Case Amazon: Ratings and reviews as part of a recommendations. *Proceedings Of The 2007 ACM Conference On Recommender Systems, RecSys 2007*.  
<https://doi.org/10.1145/1297231.1297255>
- [10] Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. *ICUIMC '08: Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication*.  
<https://doi.org/10.1145/1352793.1352837>
- [11] Jiang, H., Qi, X. & Sun, H. (2012). Choice-Based Recommender Systems: A Unified Approach to Achieving Relevancy and Diversity. *Social Science Research Network*.  
<https://doi.org/10.2139/ssrn.1989238>
- [12] Raza, S., Bashir, S. R., Naseem, U., Liu, D. D. & Reji, D. J. (2022). Incorporating Accuracy and Diversity in a News Recommender System. *2022 IEEE 9th International Conference On Data Science And Advanced Analytics (DSAA)*.  
<https://doi.org/10.1109/dsaa54385.2022.10032442>
- [13] Vargas, S., Baltrunas, L., Karatzoglou, A. & Castells, P. (2014). Coverage, redundancy and size-awareness in genre diversity for recommender systems. *RecSys '14: Proceedings Of The 8th ACM Conference On Recommender Systems*.  
<https://doi.org/10.1145/2645710.2645743>
- [14] Kannout, E., Nguyen, H. S. & Grzegorski, M. (2022). Speeding up recommender systems using association rules. In *Lecture Notes in Computer Science* (S. 167–179).  
[https://doi.org/10.1007/978-3-031-21967-2\\_14](https://doi.org/10.1007/978-3-031-21967-2_14)
- [15] *Netflix - overview - profile*. (n.d.). Retrieved April 28, 2024, from <https://ir.netflix.net/ir-overview/profile/default.aspx>



# Classification of media for adaptive music recommendations

Manuel Keck

Reutlingen University, Germany

Manuel.Keck@Student.Reutlingen-University.de

## Abstract

Music is heard in all areas of everyday life. With the diverse range of streaming portals, users have millions of tracks at their fingertips anytime, anywhere. To cope with this variety of music, users create playlists that categorize their favorite music. Creating playlists demands time, and numerous users haven't designated a specific playlist for particular situations or moods in their daily routines. If users wish to listen to a particular music genre in a given situation, they must manually select it to transition from a negative mood to a positive one. When selecting music, it must fulfill certain criteria so that it helps the user in the situation. This selection is only partially taken over by streaming providers through predefined playlists and is not individually tailored to the user. This study examines the potential for classifying music and proposes a system for mood-based recommendations of music songs. The system is developed using the Spotify developer API and the examined classifiers are validated with this system.

## CCS Concepts

• **General and reference** → Surveys and overviews; • **Human-centered computing** → **Web-based interaction**.

## Keywords

Mood, Mood Transformation, Classification, Recommender Systems. Spotify

## 1 Introduction and Motivation

In recent years, the amount of music data available has grown significantly due to the

introduction of streaming platforms. As reported by the Swedish music streaming provider, Spotify<sup>1</sup>, data availability has increased five-fold by 2023. According to a statistics from Spotify [14], the number of available songs is projected to increase from 20 million in 2014 to 100 million in 2023.

With such an abundance of tracks at their disposal, users often find it challenging to pinpoint the most fitting one for various situations. Moreover, they are left to navigate this selection process largely on their own, with minimal support from playlists tailored to specific situations or moods.

This work addresses a subfield of Schulz's work [13], which delineates the concept of adaptive music for mood regulation as a plausible means of mood modulation.

The current task involves developing a system to generate fitting song recommendations based on both the current mood and a desired mood, which users manually select within the system's current version. The system aims to effect mood changes through the suggested song choices.

Since there is no universal taxonomy for emotions where the numerous researchers in the domain of cognitive psychology agreed on,

<sup>1</sup><https://www.spotify.com> Accessed: March 23, 2024

---

Supervisor University

Dr. Gabriela Tullius  
Reutlingen University  
Gabriela.Tullius@Reutlingen-  
University.de

Informatics Inside Spring 2024  
08th May 2024, Reutlingen University  
Copyright 2024 Manuel Keck

it is hard to determine music which can compensate negative moods [6]. However, in recent years, there has been a significant advancement in the field of research investigating the correlation between music and mood. There are several publications about the impact of music on the mood of the listeners but they are based on assumptions or evaluated questionnaires like the work of Ahmad and Rana [1].

A study conducted by Spotify in February 2023 as part of the Sonic Science 2.0<sup>2</sup> research project provides evidence that sound can effectively engage all crucial areas of the brain, influencing emotions, cognition, and interaction.

The initial Sonic Science study was conducted in 2021 by Spotify to investigate the effects of digital audio on the brain. With the latest research in Sonic Science 2.0 they have proven the impact of music while measuring the electrical attributes of the human skin changed while listening to music. With electrodermal activities the emotional arousal of listened audio content could be measured. This is based on the medical principle that the excitation triggered by the sympathetic nervous system makes our body sweat.

## 2 Aim of the Work

This work explores the classification of digital available music for possible use in a mood-based recommender system under the premise of the following research question: *RQ*: Which criteria and characteristics could be taken into consideration when classifying auditory media content for personalized recommender systems that aim to influence mood?

To answer this research question, a general overview of recommender systems and the need of music classifiers is discussed. After a short insight of personality, a system will be introduced, that allows for user interaction

<sup>2</sup>[https://spotify\\_presse.prowly.com/231726-sonic-science-20](https://spotify_presse.prowly.com/231726-sonic-science-20) Accessed: March 23, 2024

to validate specified classifiers to transform predefined moods.

The interactive website will be provided to a selected group of users for validation. The validation results can then be used to adjust the values and further refine the selection.

## 3 Related Work

This work follows previous findings described by Schulz [13] in a past conference contribution at *Informatic Inside* from Reutlingen University in Spring 2021. The findings have been refined and integrated into a proprietary validation system, which offers mood-based song recommendations to its user base.

One research area that addresses this topic is that of recommender systems. Recommender systems are developed to present suitable content to a user, as described by Deldjoo et al. [4]. Nowadays there are several conferences dedicated to this research area. One of these conferences is *ACM Recommender Systems Conference (RecSys)*<sup>3</sup>. The research topics from this conference are closely linked to conferences that are dedicated to information retrieval like the *International Society for Music Information Retrieval (ISMIR)*<sup>4</sup> or *Special Interest Group on Information Retrieval (SIGIR)*<sup>5</sup>.

In Schulz's work on emotion-based in-car interaction patterns, she explains how the driver's emotional state can have a negative affect on road safety [13]. Further, the author describes how to detect the emotional state in cars and how to monitor them.

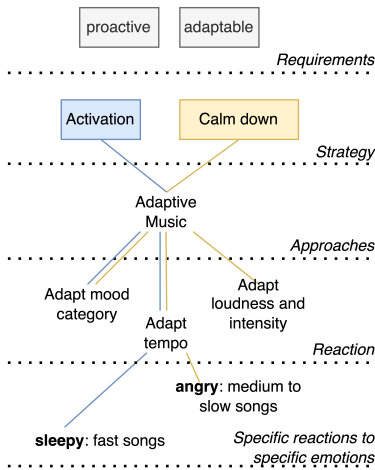
With this results, a low-level and a high-level pattern is introduced, where in particular parts from the low-level pattern are covered in this work. The low-level pattern describes an approach to offer decision support for the use of different approaches to regulate and

<sup>3</sup><https://recsys.acm.org> Accessed: March 23, 2024

<sup>4</sup><https://ismir.net> Accessed: March 23, 2024

<sup>5</sup><https://dl.acm.org/sig/sigir> Accessed: March 23, 2024

maintain emotions of drivers in cars. In particular, there are five layers in the low-level pattern and two areas. The first area is designed for mood regulation, that will be covered in this work. The main objective of this area is to regulate a negative mood which can be dangerous in terms of road safety, into a positive mood. The second area is mood maintenance where an existing mood will be supported. Figure 1 illustrates the pattern described by Schulz in a simplified format, which serves to highlight its significance and relevance for the present study.



**Figure 1: Low-level pattern for mood regulation, based on Schulz [13]**

The initial layer in the mood regulation area outlines *requirements* for managing emotion regulation processes. In the next layer, there is the *strategy* to regulate the mood, whereby in particular „activation“ and „calm down“ will be considered. In the third layer, there are *approaches* to pursue the strategy from layer two. The fourth layer is named *reaction*, where the attributes of the music are listed and the last layer is an extension of the previous layer, named *specific reactions to specific emotions*.

## 4 Method

First, a general overview about how recommendations are created is needed. To get the state of the art regarding recommender systems, a literature research is performed with Google Scholar, to get results from novel publications.

To get insights about the state of the art regarding information retrieval and music recommender systems, ScienceDirect<sup>6</sup> is considered to identify publications from 2023 and 2024 with the following search term: „*music recommendations AND classification AND mood*“.

The comprehensive state of the art overview from Deldjoo et al. could be identified [4]. This overview was used to get further literature. With the technique of *citation pearl growing* by Markey and Cochrane [8], 30 publications were identified and are partially used to analyze the current state of the art regarding recommender systems, classification of media and mood-based recommendations. In addition, media content provider were identified and their research parts were evaluated. Here, Spotify Research<sup>7</sup> and Netflix<sup>8</sup> are promising but since music is the main focus of the work, Spotify will be used.

In Section 6, it is demonstrated that media classification is not only feasible but also effective, and this approach will be evaluated using the system described therein.

## 5 Classification of Media

Section 3 describes Spotify’s Sonic Science, which proves that music can affect our mood. The subsequent phase involves delineating a classification scheme for music, which will be utilized to present recommendations to users with the aim of changing their mood. Classifying media remains a significant obstacle in the field of data science. However,

<sup>6</sup><https://www.sciencedirect.com> Accessed: March 23, 2024

<sup>7</sup><https://research.atspotify.com> Accessed: March 23, 2024

<sup>8</sup><https://research.netflix.com> Accessed: March 23, 2024

there has been significant progress in this research topic over the last five years, which is summarized in the following sections.

## 5.1 Recommendations

Recommender system research topics, driven by commercial interests, has had a significant increase over the past due to competitions like *Netflix Prize* for movie recommendations described by Bennett and Lanning from Netflix [3] or the already mentioned yearly *ACM Recommender Systems Challenge*, co-organized by Spotify.

In essence, a recommender system suggests media content to a user that aligns with their individual preferences or past consumption patterns. Deldjoo et al. [4] provide an insightful examination of the current state of the art in content-driven music recommendations, which further underscores the significance of such systems. Recommender systems are designed to suggest media content that a user might like. They are developed to increase the user experience while consuming media content.

According to Musto et al. [10], a potential approach for recommendations is to use content-based methods that suggest items with similar characteristics to those that the user has previously shown interest in. This can be accomplished by creating a profile of the user and the items based on descriptive features derived from analyzing textual content or low-level attributes of the items. Using these low-level attributes, Knees and Schedl [7] describe this as contextual metadata including genre, instrument, mood or listening situation.

Deldjoo et al. [4] categorize those attributes into high-level attributes and low-level attributes. High-level attributes contain metadata like genre, artist, tags or Wikidata<sup>9</sup>, whereas low-level attributes are features, extracted from the core audio signal like pitch,

<sup>9</sup><https://www.wikidata.org> Accessed: March 23, 2024

timbre or tempo, using audio signal processing and machine learning techniques.

There is a significant research field dedicated to extracting information from music or audio. Deldjoo et al. compiled all relevant papers from the inaugural RecSys conference and ISMIR, as well as those published between 2010 and 2020 in other venues. They consolidated these findings into a model, elaborated upon in the subsequent sections.

## 5.2 Music classifiers

With a better comprehension of the requirements for generating recommendations in recommender systems, let us explore how to collect and categorize the relevant data. Deldjoo et al. present an „onion model“ which is shown in figure 2 below and described in the following subsections. The model is read from the inside out and the associated attributes become more subjective on the outside.

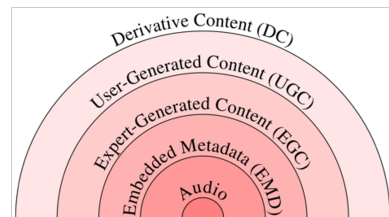


Figure 2: „Onion model“, based on Deldjoo et al. [4]

**5.2.1 Audio** Audio is at the center of the model. This layer describes the digital representation of the content. It outlines this digital signal as physical signal, without any semantic. The semantic dimension would be created within the listeners when perceiving the audio.

Different algorithms try to detect and experience those features that listeners do. Deldjoo et al. describe time-domain descriptors like loudness or zero-crossing-rate, spectral descriptors like roll-off or mel frequency

cepstral coefficients, tonal descriptors like melody or chroma and last, rhythm descriptors like onsets or beats.

**5.2.2 Embedded Metadata** After analyzing the physical audio signal we can move one step further and have a look on embedded metadata (EMD). EMD refers to descriptive and technical metadata such as record identifiers and accompanying data, comprising information like artist name, track name, album name, lyrics, label, year of recording, producer, production settings and even different multimedia data such as video clips of a song or album cover artworks.

This data is very closely connected to the music item and often directly embedded in the digital container of the audio (e.g. ID3 tags in MP3) or externally recorded in separate files for external databases like MusicBrainz<sup>10</sup>.

**5.2.3 Expert-Generated Content** Expert-generated content (EGC) is music metadata, assigned from a trained listener from musico-logical perspective. Comprised attributes are genre, style or mood. In addition, the music gets contextualized in terms of era, origin and trends, where trends need to be set in retrospect. EGC provides more detailed descriptors of the content compared to the previous layers. This includes attributes like instrumentation, sound quality, vocal properties, structural properties, lyrical content or song message.

In addition, information from the performing artists can be collected, like for example biographies. Examples, where such data can be found is Music Genome Project, AllMusic<sup>11</sup> or Wikipedia<sup>12</sup>. The Music Genome Project is a project, where these expert-labelled data is compared with audio features, as described by Prockup et al. in their work [11].

To sum up, obtaining this data is less scalable due to the need of experts.

<sup>10</sup><https://musicbrainz.org> Accessed: March 23, 2024

<sup>11</sup><https://www.allmusic.com> Accessed: March 23, 2024

<sup>12</sup><https://www.wikipedia.org> Accessed: March 23, 2024

**5.2.4 User-Generated Content** According to Deldjoo et al., user-generated content (UGC) describes published data on public websites like blogs, wikis, discussion forums, posts, chats, tweets, podcasts, pins, digital images, videos, audio files and other forms of media that users had created.

This kind of data is referred as community metadata by Whitman et al. [15] in their corresponding work. The research area around music information retrieval (MIR) deals with getting this data. Murthy et al. [9] provide a detailed review of eight different MIR-related tasks, including emotion recognition. This is one example of how to obtain data in the area of content-based MIR.

One UGC platforms is Last.fm<sup>13</sup>. Additional UGC solutions include user-generated playlists, which combine suitable songs, and song explanations, such as those provided by Genius<sup>14</sup>.

**5.2.5 Derivative Content** Derivative content refers to new creations, based on the original content. This comprises re-use of original content such as in remixes, mash-ups, cover versions or parodies. This kind of content is often performed or produced by amateurs on platforms like YouTube<sup>15</sup>.

## 5.3 Meta-Classifiers

Spotify Sonic Science 2.0 provides an advanced research for the described classifiers. In this research, Spotify uses those values in combined acoustic attributes such as *acousticness*, *speechiness*, *bounciness*, *instrumentalness* or *danceability*. A brief overview about all available attributes can be found in the documentation of Spotify Web API<sup>16</sup>.

According to the results of the study by Spotify Sonic Science 2.0, users tend to listen to

<sup>13</sup><https://www.last.fm> Accessed: March 23, 2024

<sup>14</sup><https://genius.com> Accessed: March 23, 2024

<sup>15</sup><https://www.youtube.com> Accessed: March 23, 2024

<sup>16</sup><https://developer.spotify.com/documentation/web-api/reference/get-recommendations> Accessed: March 23, 2024

songs with the corresponding attributes in special situations. For example, *acousticness* is an attribute for songs, that will be listened in moments of concentration or relaxation, whereas *danceability* will be found in songs listened to in social activities like sports or partying, but also in daily tasks like cooking or eating.

This discovery holds significant practical implications for the development of interactive websites. Rather than relying solely on classifier attribute combinations, Spotify employs meta-classifiers to define combinations of classifier attributes, which are then utilized to generate mood-based recommendations.

## 5.4 Personality

Regarding to Spotify Research there is a correlation between personality traits and musical preferences. Anderson et al. [2] investigate the lack of research in the fields of combining personality and music preferences. Based on their research, it is feasible to infer the Big Five personality traits from the *Big Five Inventory*, a 44-item questionnaire devised by John et al. in 1991 [5], which assesses openness, conscientiousness, extraversion, agreeableness, and emotional stability. In addition to a user study, the results were combined within a machine learning model to predict someone's personality using their preferences and demographic information. The findings reveal a pronounced correlation between personality traits and music listening characteristics, as extensively detailed in the study by Anderson et al. [2].

According to the interactive website developed in this work, there are two main options to integrate the user's personality to fetch suitable recommendations. The first one is the earlier mentioned 44-item questionnaire to determine the personality. In the envisioned website, an alternative strategy could involve leveraging the individual's Spotify account instead of independently determining personality traits. This integration

eliminates the need for separate personality assessments, as the user's listening history within their Spotify profile already encapsulates their preferences.

## 6 Evaluation

To address the research question, the meta-classifiers identified in section 5.3 were implemented on an interactive website. The purpose of this implementation was to validate the feasibility of a system capable of influencing moods in a personalized recommender system. This section offers details on the design and evaluation of the interactive website developed for recommending adaptive music.

According to the *strategy*-layer by Schulz, described in section 3, „activation“ and „calm down“ are used and renamed to „energized“ and „relaxed“ to get more general moods. The list of potential positive moods has been expanded by „happy“ and „focused“.

The moods „angry“, „tired“, „stressed“ and „shocked“ are implemented as negative moods in the current version. These values were chosen as examples for the first version of the website.

In the work of Schulz, the *reaction*-layer of the low-level pattern describes attributes that music should have in order to be able to regulate a certain mood. The last layer extends the *reaction*-layer and describes attributes to maintain in particular „sleepy“ or regulate moods like „angry“.

In this work, this is realized with the Spotify meta-classifiers. To be more precise, for each mood transformation there are individual values for the meta-classifiers to fetch suitable recommendations. With the integration of the meta-classifiers, no additional database for expert-generated content, described in section 5.2.3, is needed.

### 6.1 Concept

To generate suitable recommendations, the needed data should be identified and the

availability checked. On the interactive website, Spotify Premium users will be able to login and select a *current mood* and a *target mood*. With this manual selection, it will be tried to transform a negative mood into a positive mood. Therefore, information about classifying content is needed to select suitable content for each mood transformation. Due to the individual Spotify logins, the personalization is guaranteed and the recommendations are based on the users preferences. Once the recommendation makes the transformation feasible, users can respond with either „yes“ or „no“. These responses will be evaluated to validate the chosen values, once a substantial amount of data has been collected. The values can then be adjusted to fetch better recommendations.

## 6.2 Requirements

To formulate the functional- and non-functional requirements (FR, NFR), the set of rules of Rupp et al. [12] was used. In the following tables 1 and 2, the most important FR and NFR are documented in a superficial approach.

**Table 1: Functional Requirements**

ID	Description
FR1	The system should have a Spotify login method.
FR2	The system has to provide a possibility to select the current mood and target mood.
FR3	The system must be able to request mood-based recommendations from Spotify Web API.
FR4	The system should have buttons to submit a response regarding the suggested recommendations.
FR5	The system requires database integration to store feedback.

**Table 2: Non-Functional Requirements**

ID	Description
NFR1	The system should be accessible as a public website 24h per day.
NFR2	The system must handle the Spotify access token for the estimated lifetime of the token of 3600s (1h).
NFR3	The system must be able to onboard up to 25 users.
NFR4	The system must be usable on mobile and desktop platforms.

## 6.3 Values

To get suitable recommendations for each mood transformation, separate request URLs will be created to request the Spotify Web API. In addition to the top three tracks and top two artists over the last six weeks, the values shown in table 3 are used to request Spotify’s recommendations endpoint. There are way more available attributes from Spotify but with the attributes shown in table 3 the best results were achieved. The values are initially based on a clustering publication from Aldrete on RPubS named *Spotify Music Moods*<sup>17</sup> and continued with meaningful values following the same principle.

## 6.4 Implementation

The website is developed in React<sup>18</sup> with Next.js<sup>19</sup>. The website can be found under the following URL: <https://changeyourmood.vercel.app>. The user’s feedback is collected within a Postgres database. The landing page of the website offers the Spotify login method. After logging in, users are able to see their top ten tracks and top ten artists from the past

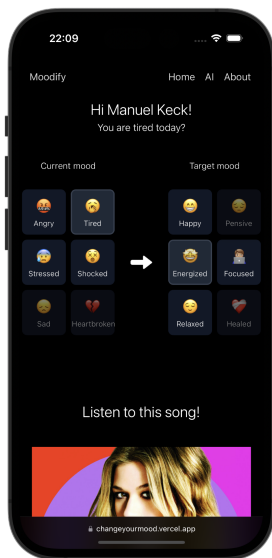
<sup>17</sup><https://rpubs.com/mary18/860196> Accessed: March 23, 2024

<sup>18</sup><https://react.dev> Accessed: March 23, 2024

<sup>19</sup><https://nextjs.org> Accessed: March 23, 2024

**Table 3: Meta-Classifier values, own representation, based on Aldrete**

	target_energy	target_danceability	target_tempo
angry - happy	0.8	0.8	120
angry - energized	0.8	0.8	140
angry - relaxed	0.4	0.4	80
angry - focused	0.2	0.2	120
tired - happy	0.5	0.7	120
tired - energized	0.9	0.9	140
tired - relaxed	0.2	0.2	60
tired - focused	0.4	0.4	100
stressed - happy	0.4	0.5	120
stressed - energized	0.9	0.9	100
stressed - relaxed	0.1	0.1	100
stressed - focused	0.2	0.2	100
shocked - happy	0.4	0.5	90
shocked - energized	0.9	0.9	160
shocked - relaxed	0.2	0.3	80
shocked - focused	0.5	0.2	120

**Figure 3: Website: Mood selection. Own representation.**

six weeks. This feature has been designed to enhance users' awareness of their personal preferences. After that, a mood transformation matrix is presented, as shown in figure 3. At the moment, there are implemented four different *current moods* and also four *target moods*. The user is permitted to select any combination of the options, with the exception of those that have been greyed out. After selecting a *current mood* and a *target mood*, the Spotify Web API gets called to fetch recommendations. The attributes described previously and shown in table 3 are included in this request URL. After fetching a recommendation, the user can play the song in the built-in Spotify Web Player and skip through 20 recommendations for each transformation. Finally, the user can indicate whether the selected song has the ability to change their

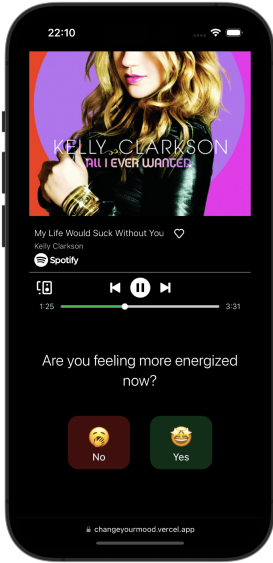


Figure 4: Website: Submit feedback. *Own representation.*

current mood to the *target mood*, by submitting „yes“ or „no“, as shown in figure 4.

### 6.5 Validation

After approximately six weeks of collecting data, there are some first recognizable patterns in the stored dataset.

A total of 60 feedbacks were received, with five of these being negative. The initial moods of „angry“ and „shocked“ were selected less frequently than „tired“ and

„stressed“, as shown in figure 5. The most frequent negative feedback was for the mood „energized“, indicating that this needs to be adjusted again.

To answer the research question, it is sufficient to use meta-classifiers such as *target\_energy*, *target\_danceability* or *target tempo*. It should be noted that multiple classifiers may be combined in this meta-classifiers.

### 6.6 Limitations

There are some limitations in the current implementation of the proposed system. First of all, the usage of the website is restricted to Spotify Premium users only. In addition, there is no dynamic adjustment of the values implemented. If a user submits „negative“ feedback, no new recommendations are provided. This goes hand in hand with the limitation, that the same recommendations will always be provided, if there is no change in top three tracks or top two artists. This is caused by the deterministic characteristic of the Spotify Web API recommendations endpoint. Another limitation is, that some users may connect personal, historical events with tracks. In that particular case, the recommendation might not be suitable. Finally, achieving the *target mood* comprehensively may require additional tracks beyond the provided recommendations. Therefore, a graph should be implemented to recommend new tracks with dynamically adjusted values to reach a *target mood*.

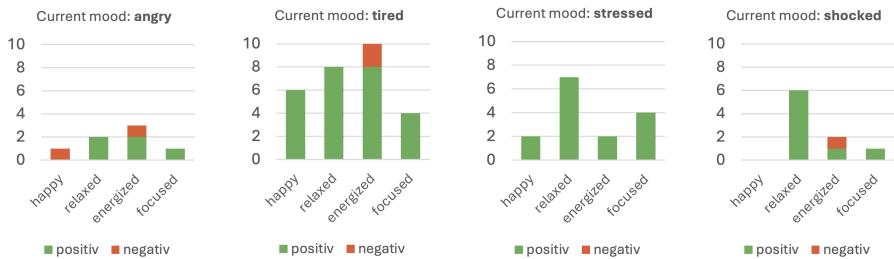


Figure 5: Evaluation of mood transformations. *Own representation.*

## References

- [1] Nawaz Ahmad and Afsheen Rana. “Impact of Music on Mood: Empirical Investigation”. In: *Research on Humanities and Social Sciences* (2015). ISSN: Online) 2225-0484. URL: <https://ssrn.com/abstract=2696883>.
- [2] Ian Anderson et al. ““Just the Way You Are”: Linking Music Listening on Spotify and Personality”. In: *Social Psychological and Personality Science* 12.4 (2021), pp. 561–572. DOI: 10.1177/1948550620923228. URL: <https://doi.org/10.1177/1948550620923228>.
- [3] James Bennett, Stan Lanning, et al. “The netflix prize”. In: *Proceedings of KDD cup and workshop*. Vol. 2007. New York. 2007, p. 35.
- [4] Yashar Deldjoo, Markus Schedl, and Peter Knees. “Content-driven music recommendation: Evolution, state of the art, and challenges”. In: *Computer Science Review* 51 (2024), p. 100618. ISSN: 1574-0137. DOI: <https://doi.org/10.1016/j.cosrev.2024.100618>. URL: <https://www.sciencedirect.com/science/article/pii/S1574013724000029>.
- [5] Oliver P John, Eileen M Donahue, and Robert L Kentle. “Big five inventory”. In: *Journal of personality and social psychology* (1991).
- [6] Marius Kaminskis and Francesco Ricci. “Contextual music information retrieval and recommendation: State of the art and challenges”. In: *Computer Science Review* 6.2-3 (2012), p. 106.
- [7] Peter Knees and Markus Schedl. “A survey of music similarity and recommendation from music context data”. In: *ACM Trans. Multimedia Comput. Commun. Appl.* 10.1 (Dec. 2013). ISSN: 1551-6857. DOI: 10.1145/2542205.2542206. URL: <https://doi.org/10.1145/2542205.2542206>.
- [8] Karen Markey and Pauline A Cochrane. “Online training and practice manual for ERIC data base searchers”. In: (1981), p. 19.
- [9] YV Srinivasa Murthy and Shashidhar G Koolagudi. “Content-based music information retrieval (cb-mir) and its applications toward the music industry: A review”. In: *ACM Computing Surveys (CSUR)* 51.3 (2018), pp. 1–46.
- [10] Cataldo Musto et al. “Semantics and content-based recommendations”. In: *Recommender systems handbook*. Springer, 2012, pp. 251–298.
- [11] Matthew Prockup et al. “Modeling Genre with the Music Genome Project: Comparing Human-Labeled Attributes and Audio Features.” In: *ISMIR*. 2015, pp. 31–37.
- [12] SOPHISTen Rupp C. “Requirements-Engineering und -Management”. In: *Hanser Verlag* 7. Auflage (2021), pp. 357–367.
- [13] Franziska Katharina Schulz. “Patterns of in-car interaction based on emotion recognition”. In: *Informatics Inside Frühling 2021* (2021).
- [14] Spotify. “Anzahl der verfügbaren Tracks auf Spotify in ausgewählten Monaten von August 2014 bis September 2023 (in Millionen) [Graph]”. In: *Statista* (2023). URL: <https://de.statista.com/statistik/daten/studie/378806/umfrage/anzahl-der-verfuegbaren-songs-auf-spotify/>.
- [15] Brian Whitman and Steve Lawrence. “Inferring descriptions and similarity for music from community metadata”. In: *ICMC*. 2002.



# Einfluss äußerlicher Faktoren auf die Qualität eines mündlichen Vortrags in Virtual Reality

Markus Egler  
Hochschule Reutlingen  
Reutlingen, Deutschland  
[Markus.Egler@Student.Reutlingen-University.de](mailto:Markus.Egler@Student.Reutlingen-University.de)

## Abstract

Das Präsentieren vor einer größeren Anzahl an Menschen ist in jedem Lebensabschnitt eines Menschen relevant. In der Schule werden Vorträge gehalten, bei der Arbeit werden Ergebnisse und Produkte vorgestellt und Freunden erzählt man Geschichten. Für jede dieser Situationen ist es wichtig, qualitativ gut präsentieren zu können. In dieser Arbeit soll recherchiert werden, welchen Einfluss äußerliche Faktoren im Allgemeinen auf einen Vortrag haben können, sowie existierende Techniken untersucht werden, die einen Vortrag qualitativ verbessern. Die Ergebnisse der Recherche werden im zweiten Teil der Arbeit mit einer Studie ergänzt. Dabei wird die Effektivität des Trainings mit einem Simulator in VR erfasst, welcher einen Vortrag vor einem Vorlesungssaal nachahmt. Zusätzlich wird darauf geachtet, ob sich die Qualität des Vortrags über mehrere Iterationen der Benutzung des Simulators

verbessert, obwohl der Vortragende währenddessen über negative Gesichts- und Körperemotionen der Zuschauer abgelenkt wird. Das Ergebnis ist, dass die Anzahl an negativen Stimuli auch einen negativen Einfluss auf die Qualität des Vortrages hat, diese Qualität jedoch mit einer höheren Anzahl an Vorträgen steigt.

## CCS Concepts

- Human-centered computing → Human computer interaction (HCI) → Interaction paradigms → **Virtual reality**
- Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods → **User Studies**

## Keywords

Virtual Reality, Presentation, External Factors, Public Speaking, Speech evaluation

---

Betreuer Hochschule: Prof. Dr. Curio  
Hochschule Reutlingen  
[crisobal.curio@reutlingen-university.de](mailto:crisobal.curio@reutlingen-university.de)

Informatics Inside Frühling 2024  
08. Mai 2024, Hochschule Reutlingen  
Copyright 2024 Markus Egler

# 1 Einleitung

Die Einleitung sieht zunächst in Kapitel 1.1 vor, die Motivation und die Ziele zu erläutern. Kapitel 1.2 gibt einen Überblick zur Richtung, in welche die Forschung geht. Das Kapitel 1.3 geht genauer auf die Methodik bei der Erstellung der Arbeit ein, während das Kapitel 1.4 die Struktur der Arbeit erläutert.

## 1.1 Motivation und Ziele

Das Thema des öffentlichen Sprechens und des Vortragens vor einer gewissen Anzahl an Menschen ist heutzutage noch immer sehr relevant. Ist die Anzahl der Zuschauer klein, mit ein bis drei Personen, wie das bei einem Bewerbungsgespräch üblich ist, oder groß, mit 50-100 Personen, wie bei einer Präsentation in einem Vorlesungssaal, die Qualität eines Vortrags darf sich nicht verringern. Die Kunst des Präsentierens ist dabei genauso wichtig im Alltag von Studenten, wie im späteren Berufsleben, wo zum Beispiel ein Produkt einem Kunden vorgestellt werden muss.

Leider ist nicht jede Person in gleicher Weise kompetent darin, einen Punkt herüberzubringen oder ein sachliches Thema auf eine interessante Weise vorzustellen. Wie jede andere Fähigkeit muss auch diese geübt und trainiert werden, um darin besser zu werden. Jedoch hat man nicht oft die Möglichkeit, vor einer größeren Menschenmenge, wie das in einem Vorlesungssaal der Fall ist, seinen Vortrag zu üben. Hier kommt Virtual Reality (VR) ins Spiel, denn in einer VR-Umgebung kann man nicht nur einen Vorlesungssaal simulieren; es ist außerdem möglich, ein animiertes Publikum bereitzustellen. So kann ein Vortrag vor einem Publikum realistisch simuliert und mehrere Male in VR durchlaufen werden. In dieser Arbeit wird auf einen im Masterprojekt „VR-Lab“ entwickelten Trainingssimulator für das öffentliche Sprechen in VR aufgebaut. Dabei sollen die Zuschauer und Animationen des Trainingssimulators für diese Arbeit speziell im Fokus stehen, da

diese die Qualität des Vortrags beeinflussen können.

Die qualitative Bewertung der Vorträge wird möglich gemacht über die verwendeten Sensoren der VR-Brille (HP Reverb G2 Omnicapt) und verschiedenste Skripte und Pakete in Unity. Die Aspekte, die dadurch untersucht werden können, sind unter anderem: Puls, Füllwörter, Sprechtempo, Augenkontakt und räumliche Blickverfolgung. Schließlich soll herausgefunden werden, wie sich äußerliche Faktoren, also zum Beispiel negative oder positive Stimuli über die Zuschauer und deren Animationen, auf die Qualität eines Vortrags auswirken, gemessen an den oben genannten Aspekten und dem Stress, welcher der Vortragende empfindet.

## 1.2 Forschungsfragen

RQ1: Welche bereits existierenden Systeme und Techniken gibt es, VR oder anderweitig, mit denen die Qualität eines Vortrags verbessert werden kann?

Personen, die Schwierigkeiten bei Vorträgen haben, sprechen oft von Nervosität, wenn sie vor einer größeren Menschenmenge vortragen müssen. Wie kann trotzdem eine hohe Qualität des Vortrags aufrechterhalten oder die Qualität eines Vortrags verbessert werden? Diese Frage soll mithilfe einer gezielten Literaturrecherche geklärt werden.

RQ2: Welche Auswirkungen haben äußerliche Faktoren und deren Anzahl, wie zahlreiche negative Stimuli von den Zuschauern, auf einen Vortrag, welcher in VR durchgeführt wird?

Diese Forschungsfrage zielt darauf ab, welchen Einfluss äußerliche Faktoren in VR genau auf den Vortragenden haben, wenn die Zuschauer und deren Animationen nur simuliert sind. Außerdem soll überprüft werden, welchen Einfluss die Anzahl der negativen Stimuli hat. Diese Frage soll über eine Studie mit dem Simulator in VR beantwortet werden.

Einfluss äußerlicher Faktoren auf die Qualität eines mündlichen Vortrags in Virtual Reality

### 1.3 Methodik

Die Methoden und das Vorgehen während der Forschung, ist in den Vorgang einer Literaturrecherche einzuordnen. Die Literaturrecherche beginnt mit den Grundlagen des Themas via Artikel und der Aufstellung von Forschungsfragen. Anhand der Forschungsfragen werden dann Keywords aufgestellt. Danach erfolgt die gezielte Suche über wissenschaftliche Datenbanken. Verwendete Datenbanken umfassten EDDI, IEEE, Pubmed und ACM. Gefundene Quellen müssen schließlich noch auf Qualität geprüft werden, anhand von dem Erscheinungsdatum der Quelle, sowie dem Verlag, in dem die Quelle publiziert wurde.

Zusätzlich zu der Literaturrecherche, welche Einblicke in die Auswirkungen von äußerlichen Faktoren auf Präsentationen gibt, soll eine Studie in dem Trainingssimulator in VR durchgeführt werden. Die Studie soll verschiedenen äußerliche Faktoren (störende Gesichts- und Körperanimationen der Zuschauer) und deren Auswirkungen in einer VR-Umgebung untersuchen. Dafür sollen die Versuchspersonen einen Vortrag unter verschiedenen äußerlichen Faktoren halten, um deren Einfluss herauszufinden.

### 1.4 Struktur der Arbeit

In dieser Arbeit werden unterschiedliche Systeme dargestellt, welche sich mit der Bewertung der Qualität eines Vortrags beschäftigen, sei das in VR oder der echten Welt. Kapitel 2 beschreibt bereits existierende Systeme in VR auf dem heutigen Stand, welche sich mit äußerlichen Faktoren und deren Einfluss auf einen Vortrag sowie die Qualität dessen beschäftigen. In Kapitel 3 werden die Ergebnisse der Recherche nach der ersten Forschungsfrage erläutert, ähnliche Systeme, welche bereits existieren, und die Qualität eines Vortrags bewerten. Zusätzlich werden die Ergebnisse der Studie in der zweiten Forschungsfrage aufgezeigt, in Tabellenform. Kapitel 4 beschreibt eine Diskussion der Ergebnisse. Kapitel 5 ist das abschließende Fazit.

## 2 State of the Art

In diesem Kapitel werden einige Systeme und Methodiken vorgestellt, mit welchen man die Qualität eines Vortrags messen und verbessern kann. Außerdem wird darüber informiert, wie Stress und Nervosität eine Präsentation positiv oder negativ beeinflussen. Es werden Projekte und Studien betrachtet, welche im VR-Bereich arbeiten und versuchen, die Qualität eines Vortrags zu fördern. Dies soll erläutern, inwiefern sich das durchgeführte Projekt und die Ergebnisse, welche über Tests entstanden sind, sich mit anderen State of the Art – Projekten unterscheiden.

In Referenz [5] wird ein System beschrieben, welches in VR verankert ist und Live-Feedback von den Zuschauern sowie Analysen zur Qualität des Vortrags liefert. Dabei wird auch von Glossophobia gesprochen, einer Angst vor dem Sprechen vor einem größeren Publikum, und der Notwendigkeit von einer privaten und sicheren Umgebung, welche VR über Simulation bereitstellen kann. Dies soll eine größere Stressreaktion vorbeugen, vor allem da der Vortrag mehrmals geübt werden kann. Ein Vorteil von VR ist außerdem, dass es eine Möglichkeit bietet, einen Vortrag vor einem großen Publikum, so oft man will, durchzuführen. Aufgezeichnete Daten sind unter anderem Sprechrate, Füllwörter und Grammatik. Aufgrund der aufgezeichneten Daten sollen die Zuschauer in der entsprechenden Weise in Echtzeit reagieren.



Abbildung 1: Benutzeroberfläche des Sprech-Simulators in Virtual Reality. Quelle: Referenz [5].

Die Analyse der Sprache funktioniert über die „Windows Speech Library“, welche es ermöglicht, eine Log-Datei, also ein gesamtes Transkript des Vortrags zu erstellen. Aus der Datei werden verschiedene Wörter herausgefiltert. Zudem gibt es eine Liste an Wörtern, um herauszufinden, ob spezielle Wörter öfter verwendet werden als andere. Auf dieser Liste können auch Füllwörter vorhanden sein, welche nicht benutzt werden sollen, da man in seinem Wortlaut möglichst präzise sein will.

Für das Live-Feedback werden die Avatare der Zuschauer leicht animiert. Jedoch haben die Zuschauer nur drei Gesichtszustände. Ebenso fällt auf, dass die Emotion auf den Gesichtern der Zuschauer nur durch Augenkontakt verändert wird. Schaut der Vortragende dem Zuschauer in die Augen, so zeigt dieser einen positiveren Gesichtsausdruck. Zu diesem Projekt wurde auch eine kleine Studie durchgeführt, bei der zehn Teilnehmer einen kleinen Vortrag über ein zufälliges Thema halten sollten. Als Ergebnis gab es ein kleines Feedback von den einzelnen Teilnehmern zu ihrem Level von Komfort bei der Durchführung ihres Vortrags, leider keine spezifischen Daten zu Füllwörtern etc. Außerdem sah der Vorlesungssaal, welcher benutzt wurde, nicht sehr realitätsnah aus und auch die Zuschauer, welche in Abbildung 1 zu sehen sind, wurden recht simpel gehalten. Dadurch können komplexere Emotionen über die Animationen der Zuschauer nicht sehr präzise übermittelt werden.

In Referenz [6] wird ein „Serious Game“ in VR beschrieben, mit einem entwickelten Stressanalyse-Modell, welches an der Stimme erkennen soll, unter wie viel Stress ein Vortragender steht, während er einen Vortrag hält. Außerdem liefert das entwickelte System Feedback über animierte Zuschauer in Echtzeit.

Es werden verschiedene Algorithmen des maschinellen Lernens benutzt, um den Stress in einem Vortrag erkennen zu können. Ein akustisches Signal soll von verschiedenen

Algorithmen des maschinellen Lernens analysiert werden, auf unterliegende Emotionen in Merkmalen der Stimme. Die Umgebung ist in Unity programmiert, wobei ein Publikum mit mehreren Zuschauern simuliert wird. Aufgrund des detektierten Stress-Levels im Vortragenden soll der Zuschauer dann auf eine bestimmte Weise reagieren. Die Reaktionen können als positive Emotionen in Zuschauern auftreten, bei Vortragenden mit hohen Stress-Levels, oder als negative Emotionen im wettkampforientierten Modus. Ist man auf den Wettkampf aus, bekommt man auch einen High-Score für seinen Vortrag. Genauer sind die verschiedenen negativen Emotionen der Zuschauer in Abbildung 2 zu sehen.



**Abbildung 2: Screenshot des Publikums mit verschiedenen positiven beziehungsweise negativen Emotionen von links nach rechts.**

**Quelle: Referenz [6].**

Für die Algorithmen zur Analyse der Stimme fiel der „Random Forest Algorithmus“ mit 82 % Genauigkeit als am präzisesten auf. Negativ fällt die schlechtere Qualität der Avatare für die Zuschauer aus. Die Avatare wirken eher wie ein Cartoon als reale Menschen.

Einfluss äußerlicher Faktoren auf die Qualität eines mündlichen Vortrags in Virtual Reality

### 3 Ergebnisse

Die in dieser Arbeit erzielten Ergebnisse beziehen sich auf die untersuchten Methoden aus der Literaturrecherche und wie diese genutzt werden, um Vorträge zu bewerten. Außerdem werden die wichtigsten Aspekte herausgearbeitet, um einen Vortrag in der echten Welt zu bewerten und zu beeinflussen. Ein anderes wichtiges Ziel ist es, herauszufinden, inwiefern die Qualität eines Vortrags verbessert werden kann.

Die weiteren Ergebnisse beziehen sich auf die zweite Forschungsfrage, also welche Auswirkungen äußerliche Faktoren in VR und die Anzahl der negativen Stimuli über die simulierten Zuschauer auf die Qualität eines Vortrags haben. Dazu wurde eine Studie durchgeführt, die in Kapitel 3.2 beschrieben wird. Es soll evaluiert werden, wie bedeutend die Verbesserungen in der Qualität des Vortrags wirklich sind.

#### 3.1 Methoden zur Qualitätsverbesserung eines Vortrags

In diesem Kapitel wird sich mit den Systemen und Methoden in der realen Welt beschäftigt, welche die Qualität eines Vortrags messen und äußerliche Faktoren und deren Auswirkung untersuchen. Dabei wird auch darauf geachtet, wie genau den Personen geholfen werden soll.

##### 3.1.1 Awareness- and Competing Response Training

Eine Technik, die Nervosität und Angst während eines Vortrags zu kontrollieren und gezielt zu bekämpfen versucht, wird Awareness Training genannt. Diese Methodik, welche in Referenz [1] und [3] beschrieben wird, fokussiert sich dabei auf die Verbesserung der Qualität eines Vortrags. Es ist eine Technik im psychologischen Raum, bei der dem Vortragenden bewusst gemacht wird, wann er Fehler begeht, welche sich wiederum negativ auf die Qualität des Vortrags auswirken.

Awareness Training fokussiert sich darauf, die Flüssigkeit eines Vortrags aufrechtzuerhalten, indem man störende auditive und körperliche Antworten auf den Stress während eines Vortrags abgewöhnt. Eine körperliche Antwort kann zum Beispiel das Fahren mit der Hand durch die Haare sein. [3]

Eine in Referenz [1] beschriebene auditive Antwort, welche abtrainiert werden soll, sind die sogenannten Füllwörter, welche Sätze unpräzise und stockend machen. Wichtig bei dem Durchführen des Awareness Trainings ist, dass andere Faktoren, welche den Vortrag beeinflussen, überwacht werden, sodass die gesamte Qualität der Präsentation nicht sinkt. [1]

Der Beginn des Awareness Trainings, beschrieben in Referenzen [1,2,3], kommt einher mit einem drei- bis fünfminütigen Vortrag zu einem zufälligen Thema, ohne vorheriges Training. Dieser Vortrag soll die Baseline für eine Studie sein, mit der spätere Vorträge verglichen werden können. Teilnehmer bekamen fünf bis zehn Minuten Vorbereitungszeit, um sich mit dem Thema vertraut zu machen.

Daraufhin beginnt der Trainingsteil der Studie, aufgeteilt in Videotraining und In-Vivo Training. [1,3]

Bei dem Videotraining wird dem Teilnehmer ein Video gezeigt, bei dem verschiedene negative Einflüsse, wie Füllwörter, hervorgehoben und verständlich für den Teilnehmer erklärt werden. Daraufhin soll der Teilnehmer dann selbst die negativen Einflüsse in einem weiteren Video feststellen und zu dem genauen Zeitpunkt, an dem diese stattfinden, die Hand anheben. Dies soll zeigen, dass der Teilnehmer weiß, welche Aktionen und Gewohnheiten in einem Vortrag überhaupt die Qualität dessen verringern. Zudem soll der Teilnehmer Unflüssigkeiten während des Vortrags sofort bemerken und auf diese reagieren können. Das Training wurde erst fortgeführt, wenn bei einem Video 85-100 % aller Fehler erkannt worden waren. [1]

Im In-Vivo Training wird das Videotraining bei dem eigenen Vortrag gleich in Aktion gesetzt. Der Teilnehmer soll seine rechte Hand heben, wenn er selbst eine Sprechunflüssigkeit durchführt, und seine linke Hand, wenn er den Reiz verspürt, eine Sprechunflüssigkeit auszuführen. Dieses Vorgehen wird so oft wiederholt, bis eine 75 % Reduktion an Störungen im Redefluss gegenüber der Baseline festgestellt wird. [1]

Eine weitere Technik, um gezielt Gewohnheiten abzutrainieren, nennt sich das „Competing Response Training“ (CRT). Bei dem CRT versucht man dem Vortragenden spezielle Gewohnheiten, welche dieser während eines Vortrags durchführt, abzugewöhnen. Dies soll geschafft werden, indem eine gegenspielende neue Gewohnheit antrainiert wird. Ein Beispiel, welches erwähnt wird, ist reguliertes Atmen, um Stottern zu bekämpfen. Für das CRT, welches zusätzlich zum Awareness Training verwendet werden kann, wird deswegen etwas länger trainiert, etwa ein- bis dreimal die Woche. Zusätzlich wird die Technik der „Response Description“ angewandt, das heißt, es wird für den Teilnehmer genau der angewandte Prozess beschrieben, sowie die spezifische nervöse Gewohnheit, welche behandelt werden soll. [3]

Durch eine häufigere Wiederholung des Trainings und eine Kombination von Awareness Training und CRT wird eine höhere Reduktion der negativen Gewohnheiten während eines Vortrags angemerkt, als das in Referenz [1] der Fall ist, wo nur Awareness Training angewandt wird.

Für die Überprüfung und Validierung der Ergebnisse werden die Teilnehmer einzeln über mehrere Faktoren auch von externen Prüfern bewertet. Zu den Faktoren gehören unter anderem Augenkontakt oder die Benutzung von Füllwörtern.

Referenz [1] zeigt eine Bewertung der Fähigkeiten des öffentlichen Sprechens vonseiten externer Prüfer als Ergebnis in der folgenden Tabelle.

**Tabelle 1: Bewertung von Fähigkeiten des öffentlichen Sprechens. Links, mit Werten der Baseline, rechts mit Werten des Awareness-Trainings.**

**Quelle: Referenz [1].**

Public speaking abilities (external rating)		
Appeared comfortable	3,5 (2-4)	3,25 (1-5)
Acceptable voice projection	4 (2-5)	4,25 (2-5)
Appropriate speech rate	4,25 (3-5)	3,75 (3-5)
Eye contact with audience	2 (1-3)	1,75 (1-3)
Fluent speech	3,5 (2-4)	3,75 (2-5)
Appeared nervous	3,5 (3-4)	3 (1-4)
Appropriate movements	3,75 (3-5)	3,25 (2-4)
Sounded out of breath	4,75 (4-5)	3,75 (2-5)
Appropriate gestures	3,5 (3-4)	3,75 (2-5)
Used fillers	2,5 (1-5)	3,5 (2-5)
Appeared confident	3,5 (3-4)	3,5 (2-5)
Overall rating of ability	3,5 (2-4)	3,5 (2-4)

Als Ergebnis in Referenz [3] konnten Unflüssigkeiten in einem Vortrag bis hin zu 89% verringert werden. Eine nervöse Angewohnheit, welche unter anderem mit dem CRT angegangen wurde, ist das Schnalzen mit der Zunge.

### 3.2 Ergebnisse der Studie

In diesem Kapitel wird erläutert, wie die Studie konzipiert und aufgebaut wurde, wie die verschiedenen Tests im Detail durchgeführt wurden, die Ergebnisse, und wie diese Daten/Ergebnisse evaluiert wurden.

#### 3.2.1 Aufbau der Studie

Die Studie ergänzt das Masterprojekt „Game of Talk“ aus dem VR-Lab. Es wurde sich für die Studie jedoch zusätzlich ein Konzept der Schwierigkeitsgrade überlegt, also eine Erhöhung der negativen Stimuli über die Zuschauer während des Vortrags. Diese negativen Stimuli bestehen aus Animationen der Zuschauer im Vorlesungssaal. So werden Gesichtsanimationen manipuliert, um negative Emotion herüberzubringen, sowie Körperanimationen manipuliert, um den Vortragenden abzulenken. In dem Trainingssimulator, welcher einen Vorlesungssaal an der Hochschule Reutlingen im Gebäude 9 simuliert, können die eigenen Vorträge auf verschiedene Aspekte, wie die Anzahl an Füllwörtern, das Sprechtempo, Augenkontakt, Blickverteilung im Raum, etc. untersucht werden.

Einfluss äußerlicher Faktoren auf die Qualität eines mündlichen Vortrags in Virtual Reality

Ähnlich konzipiert, wie das in den Referenzen [1] und [3] geschildert ist, wird den Studenten am Anfang des Experiments zehn Minuten Zeit gegeben, um einen groben Rahmen zu ihrem zufälligen Vortrag zu gestalten. Danach wird den Probanden etwas Zeit gegeben, sich mit der VR-Brille und der virtuellen Umgebung vertraut zu machen.

Daraufhin sollen die Teilnehmer der Studie einen zweiminütigen Vortrag über das gegebene Thema halten. Die Vorträge sollen für Einzelpersonen wiederholt werden, mit insgesamt drei Vorträgen von jedem Teilnehmer. Dies ist wichtig, um den Fortschritt der Teilnehmer über einen längeren Zeitraum und mehrmalige Übung mit dem Trainingssimulator zu überprüfen. Dadurch soll auch die Effektivität des Trainingssimulators überprüft werden.

Der erste Vortrag dient als Baseline ohne Hilfestellung und Feedback und wird auf der ersten Schwierigkeitsstufe durchgeführt. Die Schwierigkeitsstufe 1 bedeutet, dass Zuschauer im Publikum neutral wirken und bei Augenkontakt positiv reagieren. Nach dem ersten Vortrag wird dem Teilnehmer Feedback gegeben über die Oberfläche des Trainingssimulators und gezieltes Feedback zu einzelnen Auffälligkeiten extern. Dies soll Aspekte des in Referenz [1] beschriebenen Awareness Trainings miteinbeziehen.

Der zweite Vortrag hat einen höheren Schwierigkeitsgrad (Stufe 2). Dabei vermitteln die Zuschauer dem Vortragenden negative Stimuli über Manipulation der Gesichtsanimationen und einer einhergehenden Wiedergabe von negativen Emotionen. Wird ein Zuschauer angeschaut, so zeigt dieser eine negative Emotion. Nach dem Vortrag wird auch wieder Feedback gegeben.

Der dritte Vortrag hat den höchsten Schwierigkeitsgrad (Stufe 3). Zuschauer zeigen negative Animationen und zusätzlich ist ein Störfaktor am äußeren Blickfeldrand eingebaut.

Die Tests sollen auch die Möglichkeit der weiteren Erforschung anbieten, indem man Techniken des Awareness Trainings mit VR kombiniert. Es lohnt sich deshalb, da Awareness Training gutes Feedback gibt, aber Vorträge in VR so oft man will wiederholt werden können.

### 3.2.2 Ergebnisse der Studie

Die Ergebnisse der Studie sind unterteilt in die drei Präsentationen pro Teilnehmer (fünf insgesamt), in einer großen Tabelle. Angegeben wird der prozentuale Anteil an Füllwörtern im Vergleich zu den Gesamtwörtern, welche verwendet wurden. Weiterhin wird der Blickkontakt mit dem Publikum an sich, sowie der Augenkontakt mit einzelnen Zuschauern in Prozent angegeben. Dabei bezieht sich die Prozentzahl auf die gesamte Blickverteilung im Vorlesungssaal. Schließlich werden der Puls in „Beats per Minute“ (BPM), die Lautstärke in Dezibel (dB) und das Sprechtempo in „Words per Minute“ (WPM) angegeben. Für das Sprechtempo sind Werte zwischen 100 und 120 WPM bis zu höchstens 150 WPM anzustreben, wobei ein Wert von 150 WPM als recht hoch angesehen und als Zeichen von Nervosität gedeutet wird. Die Ergebnisse werden im Folgenden in Tabelle 2 explizit präsentiert.

**Tabelle 2: Ergebnisse der drei verschiedenen Präsentationen der fünf verschiedenen Teilnehmer (T1, ..., T5). Die grüne Farbe markiert eine Verbesserung der vorherigen**

**Präsentation, gelb einen ähnlichen Wert und rot eine Verschlechterung.**  
Quelle: Eigene Darstellung.

<b>Präsentation 1</b>	T1	T2	T3	T4	T5
Füllwörter	3.47 %	4.66 %	4.48 %	4.15 %	5.22 %
Blickkontakt	38 %	28 %	32 %	51 %	39 %
Augenkontakt	33 %	36 %	39 %	33 %	37 %
Lautstärke	74 dB	77 dB	73 dB	75 dB	73 dB
Sprechtempo	143 WPM	141 WPM	131 WPM	148 WPM	148 WPM
Puls	113.3 BPM	96.83 BPM	85.04 BPM	98.33 BPM	88.12 BPM
<b>Präsentation 2</b>					
Füllwörter	2.42 %	6.04 %	2.04 %	2.08 %	1.87 %
Blickkontakt	25 %	30 %	49 %	45 %	42 %
Augenkontakt	27 %	36 %	37 %	30 %	41 %
Lautstärke	72 dB	78 dB	74 dB	75 dB	73 dB
Sprechtempo	129 WPM	142 WPM	122 WPM	147 WPM	119 WPM
Puls	101.28 BPM	95.33 BPM	86.25 BPM	83.75 BPM	82.26 BPM
<b>Präsentation 3</b>					
Füllwörter	2.87 %	5.52 %	3.41 %	2.83 %	4.35 %
Blickkontakt	35 %	33 %	45 %	50 %	47 %
Augenkontakt	40 %	25 %	29 %	33 %	36 %
Lautstärke	73 dB	75 dB	73 dB	73 dB	75 dB
Sprechtempo	153 WPM	161 WPM	138 WPM	128 WPM	160 WPM
Puls	104.08 BPM	104.11 BPM	86.42 BPM	87.32 BPM	85.71 BPM

### 3.2.3 Evaluation der Studie

Untersucht wird in diesem Kapitel, inwiefern die Ergebnisse, welche im vorherigen Kapitel über verschiedene Tabellen dargestellt wurden, aussagekräftig sind.

Auffallend ist, dass die prozentuale Anzahl an Füllwörtern, zu den normalen Wörtern,

sich generell für die zweite Präsentation verringert wurde, jedoch für die dritte Präsentation oft wieder etwas erhöht hat. Dies zeigt, dass das Feedback und Üben mit dem Trainingssimulator grundsätzlich effektiv ist, jedoch stark ablenkende negative Stimuli in Präsentation drei sehr störend sind. Für vier von fünf Teilnehmern steigt aber trotzdem

Einfluss äußerlicher Faktoren auf die Qualität eines mündlichen Vortrags in Virtual Reality der prozentuale Anteil nicht über den der ersten Präsentation hinaus.

Der Blick- und Augenkontakt mit dem Publikum und den Zuschauern hat sich über den Verlauf der Vorträge für alle Teilnehmer leicht verbessert. Dies ist passiert, obwohl mehrere Störfaktoren zwischen Präsentation eins und drei eingebaut wurden. Dies lässt darauf schließen, dass das Feedback und das wiederholte Training sehr hilfreich waren.

Der Puls hat sich für alle Teilnehmer zwischen Präsentation zwei und drei etwas erhöht. Dies kann an den zusätzlichen Störfaktoren liegen. Außerdem fällt auf, dass der Puls in der ersten Präsentation meist etwas höher war als in den darauffolgenden Präsentationen. Nach Befragung der Teilnehmer wird klar, dass man ohne Feedback und Hilfestellung etwas nervöser ist.

Das Sprechtempo hat sich für alle außer einem Teilnehmer zwischen Präsentation zwei und drei erhöht. Zusammen mit einem generell höheren Puls lässt das darauf deuten, dass Teilnehmer etwas nervöser waren und deswegen schneller redeten.

## 4 Diskussion

In diesem Kapitel sollen die Erkenntnisse aus der Arbeit validiert werden. Wichtige Faktoren, welche ein VR-System mit einer simulierten Umgebung und Zuschauern, mit der realen Welt und einem Vortrag vor echten Menschen unterscheiden, sollen hier weiter erläutert werden.

Die Ergebnisse aus der Analyse der Studien, welche mit Awareness Training arbeiten, zeigen, dass einzelne Gewohnheiten, welche die Qualität eines Vortrags verringern, abgewöhnt werden können. Was sich aus diesen Studien aber meist nicht herauslesen lässt, ist, welche Auswirkungen die Reduzierung auf die Qualität des gesamten Vortrags hat. Es stellt sich die Frage, ob trotz der Reduzierung einer Gewohnheit nicht vielleicht eine andere ebenso negative Gewohnheit auftritt. Ein Beispiel dafür ist die Bekämpfung der Gewohnheit des Schnalzens mit der Zunge

mit einem längeren Durchatmen. Hier ist es möglich, dass längere Pausen entstehen, welche den Redefluss unterbrechen.

Äußere Faktoren, welche einen Einfluss auf die Qualität eines Vortrags haben, können vielseitig sein. Diese, welche über den Trainingssimulator in VR untersucht wurden, beziehen sich hauptsächlich auf Stress, welcher durch den Sensor für den Herzschlag in der VR-Brille gemessen wird und auf negative Stimuli der Zuschauer. Dazu gehört die Anzahl der Zuschauer, die Animationen der Zuschauer, welche verschiedene Emotionen übermitteln sollten, sowie zusätzliche Störfaktoren. Zu beobachten war, dass negative Stimuli der Zuschauer auch einen negativen Einfluss auf den Stress des Vortragenden hatten. Leider konnten im Rahmen der Studie nur drei verschiedene Vorträge durchgeführt werden; es wäre interessant zu wissen, inwieweit sich der Stress der Teilnehmer insgesamt senken lassen könnte.

Wenn man die Ergebnisse von Systemen, welche Awareness Training benutzen, und denen, welche den Trainingssimulator in VR benutzen, miteinander vergleicht, fällt auf, dass die Ergebnisse aus Systemen mit VR-Komponente, genauere Aussagen zur objektiven Qualität des Vortrags bieten, wobei zusätzlich ein größeres Publikum eingesetzt werden kann. Deswegen wäre ein Training in VR vorzuziehen. Eine mögliche Überlegung wäre noch, das Awareness Training vor der Simulation in VR durchzuführen und so beide Techniken zu kombinieren. Einzelne ungewollte Gewohnheiten könnten reduziert werden und man hätte zusätzlich alle Vorteile des VR-Trainingssimulators. Das Training in VR bietet außerdem den Vorteil, dass auch als Einzelperson zu Hause geübt werden kann, solange diese Person die notwendige VR-Brille und das Programm dafür hat. Zusätzlich soll das Training mit dem VR-Trainingssimulator über einen längeren Zeitraum, die Angst des Vortragens vor einem größeren Publikum bekämpfen, dadurch, dass der Vortrag oft wiederholt werden kann

und die negativen Stimuli über einen längeren Zeitraum hochskaliert und variiert werden können.

## 5 Fazit

Das Training mithilfe des Trainingssimulators in VR ist hilfreich darin, aufkommende Defizite während einer Präsentation zu reduzieren. Im Vergleich zu den durchgeführten Tests, welche Techniken des Awareness Training verwenden, eine psychologische Technik, sind ähnliche Resultate erreicht worden. In der Zukunft könnte es hilfreich sein, zusätzlich das Training in VR mit Techniken des Awareness Trainings zu kombinieren, um deren Effektivität zu studieren. Jedoch sind beide Möglichkeiten des Trainings anwendbar und größtenteils effektiv an sich, für spontanes Vortragen vor mehreren Menschen. Der Vorteil, welcher durch das Trainieren in VR entsteht, ist, dass das Vortragen vor einem großen Publikum geübt werden kann. Dies ist hilfreich, da viele Personen nervös vor dem Sprechen vor größeren Menschenmengen sind und dies in VR in einer sicheren Umgebung geübt werden kann, so oft man will. Somit kann auch die Wurzel, die das eigentliche Problem meist auslöst, angegangen werden.

## Literaturverzeichnis

[1] C. Montes, M. Heinicke, and D. Geierman. Awareness training reduces college students' speech disfluencies in

public speaking. *Journal of applied behavior analysis*. DOI: 10.1002/jaba.569, 2019.

- [2] B. Pawlik and C. Perrin. Reducing speech disfluencies during publik speaking using brief habit reversal. *Journal of applied behavior analysis*. DOI: 10.1002/jaba627, 2019.
- [3] S. Ortiz, M. Deshais, R. Miltenberger et al. Decreasing nervous habits during public speaking: A component analysis of awareness training. *Journal of applied behavior analysis*. DOI: 10.1002/jaba882, 2021.
- [4] S. Zhang, W. Guo and G. Hu. Exploring universal speech attributes for speaker verification. *IEEE*. DOI: 10.1109/ICASSP.2017.7953179, 2017.
- [5] B. Truong, T. Le, K. Le et al. Public Speaking Simulator with Speech and Audience Feedback. *IEEE*. DOI: 10.1109/ISMAR-Adjunct57072.2022.00184, 2022.
- [6] Arushi, R. Dillon and A. Teoh. Real-time Stress Detection Model and Voice Analysis: An Integral VR-based Game for Training Public Speaking Skills. *IEEE*. DOI: 10.1109/CoG52621.2021.9618989, 2021.



©2024 Markus Egler. Lizenznehmer Hochschule Reutlingen, Deutschland. Dieser Artikel ist ein Open-Access-Artikel unter den Bedingungen und Konditionen der Creative Commons Attribution (CC BY-NC) -Lizenz. <http://creativecommons.org/licenses/by/4.0/>

# Evaluation der Hand-Tracking-Genauigkeit für die Erfassung von Musiknoten in der Augmented Reality-App PianoVision

Michael Stych

Hochschule Reutlingen

Reutlingen, Deutschland

Michael.Stych@student.Reutlingen-University.de

## Zusammenfassung

Die innovative Anwendung PianoVision<sup>1</sup> verbindet das Klavierspiel mit Augmented Reality und ermöglicht es dem Benutzer, auf einer virtuellen Klaviatur zu spielen. Die gespielten Klaviertasten werden durch das Hand-Tracking der Meta Quest 2 Brille erfasst. Diese Arbeit evaluiert die Genauigkeit der Notenerfassung durch dieses Hand-Tracking. Es werden zwei Klavierstücke unterschiedlicher Komplexität gespielt und analysiert: Klavierstück 1 mit einzelnen, nacheinander gespielten Noten und Klavierstück 2 mit gleichzeitig gespielten Noten. Die von der PianoVision Anwendung erfassten Noten wurden mit den tatsächlich gespielten Noten des MIDI-Keyboards verglichen und die Genauigkeit und Verzögerung für beide Klavierstücke berechnet. Die Ergebnisse zeigen, dass die Genauigkeit der Handregistrierung für einfache Klavierstücke ausreichend ist. Bei komplexeren Stücken werden viele Noten durch das Verdecken der Mittel- Ring- und kleinen Finger spät oder garnicht registriert, dabei nimmt die Genauigkeit ab und die Verzögerung zu. Für ein optimales Spielerlebnis sollte die Latenz minimiert und die Erkennung verdeckter Finger verbessert werden.

## CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; • **Applied computing** → *Interactive learning environments*.

<sup>1</sup><https://www.pianovision.com/>

## Keywords

Augmented Reality, Hand-Tracking, Virtual Piano, learning

## 1 Einleitung

Ein Musikinstrument zu erlernen, bleibt für viele Menschen ein unerfüllter Traum. Viele geben den Musikunterricht schnell wieder auf [3]. Die Gründe dafür sind der hohe Zeitaufwand, die Schwierigkeit Noten zu lesen, die Kosten für den Unterricht und die fehlenden Erfolgserlebnisse [5, 7]. Mixed Reality (MR) Anwendungen sollen das Erlernen eines Instrumentes erleichtern. Diese versuchen Schwierigkeiten wie das Notenlesen zu überbrücken, indem Klaviernoten visuell dargestellt werden. Hand-Tracking und Musical Instrument Digital Interface (MIDI) Keyboards ermöglichen es, die gespielten Noten zu erfassen und kontinuierlich Feedback zu geben. Dieses Feedback soll den Musikunterricht ersetzen [9].

Molero et al. zeigen, dass MR-Anwendungen mit Gamification-Komponenten einen positiven Effekt auf Motivation und Lernerfolg

---

Betreuer/-in Hochschule

Prof. Cristobal Curio  
Hochschule Reutlingen  
Cristobal.Curio@Reutlingen-  
University.De

Informatics Inside Frühjahr 2024  
08. Mai 2024, Hochschule Reutlingen  
Copyright 2024 Michael Stych

beim Erlernen eines Instrumentes haben können [7]. Zudem zeigt M. J. Cook in seiner Arbeit, dass Augmented Reality (AR) im Musikunterricht eingesetzt werden kann, um das Lernen und das Verständnis zu verbessern. Die Integration von AR bietet die Möglichkeit, interaktiv mit der Technologie zu interagieren und dadurch ein tieferes Verständnis für die Musiktheorie zu entwickeln [2].

Mit Hilfe dieser Lernanwendungen ist es möglich das Verständnis komplexer Musiktheorie und -techniken zu erlernen. Nach der Studie von Kangxu Cui ist es möglich eine signifikante Verbesserung der musikalischen Fähigkeiten aufzubauen [4].

Besonders gut funktionieren Klavierlern-Anwendungen mit Hilfe eines MIDI-Keyboards, um Noten aufzunehmen und Feedback zu erhalten. So kann das Klavierspielen ohne Notenkenntnisse gelernt und verbessert werden [9]. Anwendungen wie PianoVision nutzen ein virtuelles Keyboards anstelle von MIDI-Keyboards. Das Hand-Tracking der Brille registriert die gespielten Noten, sobald die Hand die virtuelle Tastatur berührt. Diese Anwendung ermöglicht das Erlernen des Klavierspiels ohne physisches Klavier, ohne Lehrer und ohne Notenkenntnisse. Das Ziel dieser Arbeit ist die Genauigkeit der Notenerkennung durch das virtuelle Klavier der Anwendung PianoVision und die Hand-Tracking-Funktion der Oculus Meta Quest 2<sup>2</sup> zu evaluieren.

### 1.1 Ziel der Arbeit

Das Ziel dieser Arbeit ist die Evaluierung der AR Anwendung PianoVision hinsichtlich der Genauigkeit der Interaktion zwischen dem Hand-Tracking und des virtuellen Keyboards zur Notenaufnahme. Daraus leiten sich folgende Forschungsfragen ab, die in Tabelle 1 dargestellt werden.

**Tabelle 1: Forschungsfragen**

<b>RQ1</b>	Wie hoch ist die Genauigkeit der Notenaufnahme durch das virtuelle Klavier der PianoVision Anwendung und der Hand-Tracking-Funktion der Oculus Meta Quest 2?
<b>RQ2</b>	Welche Auswirkungen hat die Komplexität des Musikstücks auf die Genauigkeit der Notenerkennung des virtuellen Klaviers?

## 1.2 Aufbau der Arbeit

In Kapitel 2 werden die grundlegenden Prinzipien zum Verständnis der Arbeit erläutert. Dabei wird auf das Klavierspielen und die Grundlagen der MR eingegangen. In Kapitel 3 wird der Stand der Technik präsentiert. In Kapitel 4 wird die Methodik erläutert, mit der das Experiment zur Validierung der Genauigkeit des Hand-Tracking in der PianoVision Anwendung durchgeführt wird. In Kapitel 5 werden die Ergebnisse aus dem Experiment präsentiert. Aufbauend auf den Ergebnissen werden diese in Kapitel 6 interpretiert. Abschließend folgt das Fazit dieser Arbeit.

## 2 Grundlagen

Dieser Abschnitt erläutert die Grundlagen des Klavierspielens, der Musiknotation von MR-Anwendungen, der Oculus Meta Quest 2 und der Passthrough-Funktion. Diese Grundlagen bilden die Basis für die Untersuchung der PianoVision Anwendung.

### 2.1 Klavier spielen

Das Spielen des Klaviers ist eines der beliebtesten musikalischen Aktivitäten und begleitet die Menschheit seit dem 15. Jahrhundert. Die Tasten des Klaviers repräsentieren verschiedene Noten. Durch das Spielen bestimmter Tasten auf dem Klavier entstehen Melodien, Akkorde und Harmonien, die schließlich für das menschliche Ohr ein Lied ergeben.

<sup>2</sup><https://www.meta.com/de/en/quest/products/quest-2/>

Um das Klavier spielen zu können sind Fähigkeiten erforderlich wie das Verständnis der Musiktheorie, gute Handkoordination und Disziplin [6].

## 2.2 Musiknoten

Musiknoten werden durch Symbole auf Notenblättern grafisch dargestellt. Durch die Notation können die Tonhöhe, die Tondauer und die Tonlautstärke von Noten festgehalten werden. Das Ergebnis ist ein standardisiertes Medium, das von Menschen interpretiert und gelesen werden kann [8].

## 2.3 Digitale Musiknotation

MIDI ist eine digitale Notation der Musiknoten. Diese ist von Maschinen sowie von Menschen lesbar und lässt sich leicht bearbeiten. Im Gegensatz zu herkömmlichen Audioformaten wie *MPEG-1 Audio Layer 3* (MP3) werden keine Audiodateien, mit ihren Audiosignalen gespeichert, sondern eine digitale Musiknotation äquivalent zur Musiknotation auf dem Blatt erstellt. Ein Vorteil der MIDI Dateien ist, dass sie in Notenblätter und umgekehrt umgewandelt werden können. Das ist mit Audiodateien wie MP3 nur mit großem Aufwand möglich [8].

## 2.4 PianoVision

Die AR-Anwendung PianoVision kombiniert die reale Welt mit virtuellen Elementen, um ein interaktives und immersives Erlebnis zu schaffen. In PianoVision wird die AR genutzt, um Benutzern ein virtuelles Klavier zur Verfügung zu stellen. Das Spielen auf diesem wird ermöglicht, indem Handbewegungen verwendet werden, die durch Hand-Tracking erfasst werden. Durch diese Technologie können Benutzer Klavier spielen lernen, üben und verbessern, ohne ein physisches Klavier zu benötigen. Die Genauigkeit des Hand-Trackings ist entscheidend für die Präzision und das Benutzererlebnis in MR-Klavieranwendungen.

## 2.5 Oculus Meta Quest 2

Die Oculus Meta Quest 2 ist eine Virtual Reality (VR) Brille, die eine immersive VR-Erfahrung für Spiele, Unterhaltung, Simulationen, Lernanwendungen und weiteren Szenarien bietet. Die Brille besitzt einen Qualcomm Snapdragon XR2 Prozessor. Das Display hat eine Auflösung von 3.664 x 1.920 Pixel und hat eine Bildwiederholfrequenz von bis zu 90 Hz. Die kabellose Verbindung ermöglicht Bewegung ohne externe Kabel. Die VR-Brille ist geeignet für Räume von bis zu 15 Quadratmetern. Integrierte Tracking-Kameras sorgen für präzises Hand-Tracking [1]. Für das Experiment dieser Arbeit wird die Meta Quest 2 VR-Brille benutzt.

## 2.6 Hand-Tracking

Das Hand-Tracking ist eine Funktionalität vieler VR-Brillen und ermöglicht es die Hände in der virtuellen Welt wiederzuspiegeln. Dabei werden diese durch äußere Sensoren der VR-Brille erkannt. Durch Sensoren und Algorithmen wird die Position der Hände auf die virtuelle Welt projiziert. Eine Herausforderung stellt die Genauigkeit und die Registrierungsgeschwindigkeit dar [9].

## 2.7 Passthrough

Die Passthrough-Funktion der Meta Quest 2 nutzt vier externe Kameras, die in das VR-Headset integriert sind. Die Kameras übertragen die physische Umgebung des Benutzers in Echtzeit auf das Display und ermöglichen dem Benutzer eine transparente Sicht auf die Umgebung durch Schwarz-Weiß-Bilder. Die Passthrough-Funktion von der Meta Quest 2 ermöglicht eine nahtlose Integration der realen und virtuellen Welt. Der Benutzer kann dadurch zwischen beiden Welten wechseln und eine bessere räumliche Orientierung erhalten. Diese Funktion bildet die Grundlage für eine breite Palette von Anwendungen, wie beispielsweise die Unterstützung der Interaktion mit realen Objekten in der

VR-Umgebung oder die Schaffung immersiver Lernerfahrungen [1].

### 3 Stand der Technik

In der Arbeit von Abdllkarim et al. [1] wird ein methodischer Rahmen zur Bewertung der Genauigkeit von Hand-Tracking-Systemen in der virtuellen Realität vorgestellt. Mithilfe einer Fallstudie mit dem Meta Quest 2 System wurden präzise Tests zur Positionsgenauigkeit und Gestenerkennung durchgeführt. Dabei wurden verschiedene Handbewegungen und -gesten untersucht. Die Ergebnisse zeigen, dass das Meta Quest 2 Hand-Tracking-System eine durchschnittliche Positionsgenauigkeit von 1,5 Zentimetern aufweist. Jedoch wurden Herausforderungen bei der Gestenerkennung festgestellt. Die Fehlerquote beim Erkennen der richtigen Gesten liegt bei 12 %.

In der Arbeit von Labrou et al. [5] wird eine Methode zur Klavierperformance-Erfassung für MR-Klavierlernanwendungen präsentiert. Das System erreicht eine Genauigkeit von durchschnittlich 1,8 Zentimeter bei der Handbewegungserfassung. Jedoch sind bei schnellen oder komplexen Bewegungen, wie Trillern und großen Sprüngen, Fehlerraten von bis zu 5 % zu verzeichnen. Die verzeichnete Latenzzeit von durchschnittlich 20 ms kann bei einigen Szenarien die Immersion beeinträchtigen.

Die Arbeit von Rigby et al. [9] stellt piARno - An Augmented Reality Piano Tutor vor. Das System kombiniert AR mit Klavierunterricht, um Schülern eine immersive Lernerfahrung zu bieten. Dabei können sie in Echtzeit die Hände des Klavierlehrers beobachten. Der Fokus der Arbeit lag auf der Evaluierung der Genauigkeit und Wirksamkeit von piARno. Die Ergebnisse zeigen, dass das System eine durchschnittliche Positionsgenauigkeit von 1,2 Zentimetern bei der Verfolgung der Handbewegungen erreicht. Trotzdem wurden während der Studie Herausforderungen

bei der Gestenerkennung festgestellt, insbesondere bei komplexen Handbewegungen. Die Erkennung dieser Bewegungen führte zu einer Fehlerquote von 8 %. Diese Arbeit zeigt das Potenzial von piARno als effektives Werkzeug für das Klavierlernen auf, hebt jedoch auch die Notwendigkeit hervor, die Gestenerkennung weiter zu verfeinern, um ein noch genaueres und effizienteres Lernerlebnis zu gewährleisten.

### 4 Methodik

In diesem Kapitel wird die Methodik des Experiments zur Evaluierung der PianoVision App definiert. Zunächst wird die Hardware beschrieben, die für das Experiment verwendet wird, einschließlich der Oculus Meta Quest 2, eines MIDI-Keyboards und eines Computers.

Dann wird auf den Versuchsaufbau eingegangen. Dieser beinhaltet die Eigenschaften der Versuchsumgebung und der Positionierung der Hardware, sowie den Ablauf des Experiments. Darüber hinaus wird die Auswahl der Musikstücke beschrieben, die für das Experiment verwendet werden. Es werden zwei Musikstücke verwendet: "Twinkle Twinkle Little Star" und "Ode to Joy". Die Musikstücke wurden nach verschiedenen Kriterien ausgewählt, darunter die Anzahl die Geschwindigkeit der gespielten Noten und ob das Lied mit einer Hand oder beiden Händen gespielt wird.

Folgend wird die Datenerfassung erläutert, die die Audio-, Video- und MIDI-Daten umfasst. Die Video- und Audiodaten werden mit der Screen-Capturing-Funktion der Meta Quest 2 aufgenommen. Die gespielten Noten werden simultan als MIDI-Daten mit einem MIDI-Keyboard aufgenommen und stellen die Grundwahrheit dar.

Schließlich wird die Methodik zur Analyse der Daten vorgestellt. Es werden verschiedene Metriken verwendet, um die Genauigkeit der Anwendung zu bewerten, darunter die

Fehlerrate, die Verzögerung und die Gesamtgenauigkeit.

Abschließend wird das Ziel des Experiments erläutert, nämlich die Genauigkeit der PianoVision Anwendung bei der Erfassung von Handbewegungen und gespielten Noten zu bewerten. Die Ergebnisse des Experiments werden in Kapitel 5 der Arbeit vorgestellt.

## 4.1 Hardware & Tools

In diesem Kapitel werden die Geräte und die Software beschrieben, die für die Untersuchung der Genauigkeit der PianoVision Anwendung verwendet werden.

### Oculus Meta Quest 2

Die Oculus Meta Quest 2 wird als AR-Brille während des Versuchs verwendet, um mit der PianoVision Applikation zu interagieren und die Handbewegungen beim spielen des Klaviers zu erfassen.

### Screen-Capture

Die Screen-Capture-Funktion von Meta Quest 2 wird genutzt, um die Interaktion mit der PianoVision-Anwendung aufzuzeichnen. Im durchzuführenden Experiment sollen die Bild-Frames, auf denen die gespielten Noten aufgezeichnet werden, mit den tatsächlich gespielten Noten verglichen werden.

### MIDI-Keyboard

Als MIDI-Keyboard wird das M-Audio Keystation 61 MK3<sup>3</sup> eingesetzt, um die Interaktionen mit den anschlagdynamischen Tasten während des Klavierspiels aufzuzeichnen.

### Ableton

Mit dem DAW (Digital Audi Workstation) Ableton<sup>4</sup> werden die gespielten Noten des MIDI-Keyboards in Form von MIDI-Dateien aufgenommen.

### PianoVision

Mit der AR Applikation PianoVision wird das Experiment durchgeführt. Es wird die Genauigkeit der registrierten Klaviertasten durch das Hand-Tracking erfasst.

<sup>3</sup><https://www.m-audio.com/keystation-61-mk3>

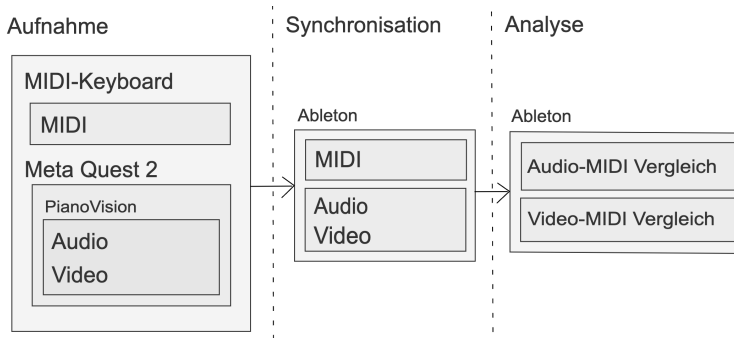
<sup>4</sup><https://www.ableton.com/>

## 4.2 Versuchsaufbau

In Abbildung 1 wird der Ablauf des Experiments dargestellt. Die Meta Quest 2 und das MIDI-Keyboard werden in einem Abstand von 50 cm voneinander positioniert. Der Betrachtungswinkel auf das Keyboard beträgt 35 Grad. Es wurde darauf geachtet, Bewegungen auf ein Minimum zu reduzieren. Der Versuch findet unter guten Lichtverhältnissen statt. Das virtuelle Klavier der PianoVision-Anwendung wird über dem MIDI-Keyboard positioniert und präzise kalibriert, um sicherzustellen, dass jede Taste optimal übereinstimmt. Die Interaktionen mit den anschlagdynamischen Tasten des MIDI-Keyboards werden über MIDI in der DAW Ableton aufgezeichnet. Die vorliegenden MIDI-Daten bilden die Grundlage für den Vergleich mit den von der PianoVision Anwendung erfassten Daten. Diese erfassten Daten werden in Form von Audio- und Videodaten aufgezeichnet und im Anschluss synchronisiert, um eine Vergleichbarkeit herzustellen. Es wurde festgestellt, dass die mit der Screen-Capture-Funktion der Meta Quest 2 aufgenommenen Videodaten nicht synchron mit den entsprechenden Audiodaten verlaufen. Die Audiospur war um 200 ms verzögert. Das Problem wurde behoben und die Audio- und Videodaten wurden synchronisiert. Anschließend wurden die Audio- und Videodaten mit den gleichzeitig extern aufgenommenen MIDI-Daten des MIDI-Keyboards verglichen. Um die Genauigkeit des Vergleichs sicherzustellen, wurden die Videodaten Bild-Frame für Bild-Frame verglichen. Auf diese Weise wurde die Übereinstimmung der aufgezeichneten MIDI-Keyboard-Daten mit den PianoVision-Daten überprüft.

## 4.3 Musikstückauswahl

Für die Evaluierung der PianoVision-Anwendung werden Musikstücke anhand verschiedener Kriterien ausgewählt, um die Genauigkeit und Effektivität der Anwendung bei verschiedenen Arten des Klavierspiels



**Abbildung 1:** Diese Abbildung bildet den Versuchsablauf ab. Während des Versuchs werden MIDI-, Audio- und Videodaten gleichzeitig aufgezeichnet. Anschließend werden die verschiedenen Daten synchronisiert, um sie für die Analyse zusammenzuführen.

zu beurteilen. Die ausgewählten Kriterien decken ein breites Spektrum von Spielsituationen ab, von der Notenzahl bis hin zur Spielgeschwindigkeit. Dieser Ansatz soll eine umfassende Bewertung der PianoVision Anwendung Präzision und Zuverlässigkeit über verschiedene musikalische Anforderungen und Fähigkeiten hinweg ermöglichen. Die Kriterien umfassen:

#### **Einzelne Noten**

Dieses Kriterium bezieht sich auf die Leistung der Anwendung, die Genauigkeit beim Spielen von einzelnen, nacheinander folgenden Noten zu messen.

#### **Mehrere Noten**

Hier wird die Genauigkeit beim gleichzeitigen Spielen mehrerer Noten bewertet, was für Akkorde und komplexe Harmonien relevant ist.

#### **Langsame Geschwindigkeit**

Musikstücke, die wenige Noten pro Minute enthalten, werden ausgewählt, um die Leistung der Anwendung bei langsameren Stücken zu testen.

#### **Schnelle Geschwindigkeit**

Im Gegensatz dazu umfasst dieses Kriterium Musikstücke mit vielen Noten pro Minute,

was die Fähigkeit der Anwendung testet, bei schnellen Stücken präzise zu sein.

- Lied 1: Twinkle Twinkle Little Star
- Lied 2: Ode to Joy

**Twinkle Twinkle Little Star** Das Lied besteht aus 42 Noten und ist 26 Sekunden lang. Das sind 1,6 Noten pro Sekunde oder 97,2 Noten pro Minute.

**Ode to Joy** Das Lied besteht aus 100 Noten und ist 40 Sekunden lang. Das sind 1,6 Noten pro Sekunde oder 150 Noten pro Minute.

## **4.4 Datenerfassung**

Es werden Audio-, Video- und MIDI-Datensätze erfasst und synchronisiert die beispielsweise ebenfalls in 2 dargestellt werden. Die Interaktion wird daraufhin mit der PianoVision Anwendung und der Meta Quest 2 analysiert.

**4.4.1 Audio- & Videoaufnahme** Die Anwendung PianoVision wird über die Screen-Capturing-Funktion der Meta Quest 2 aufgezeichnet. So können die vom System registrierten, gespielten Klaviertasten visualisiert und im Nachhinein ausgewertet werden. Die Größe des Videos beträgt 1024x1024 Pixel und ist MPEG4 codiert. Die Bildrate beträgt

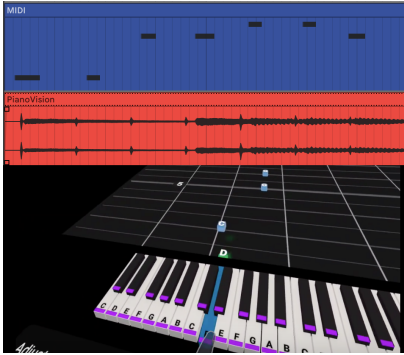


Abbildung 2: Die Abbildung zeigt die Spuren der Grundwahrheit (MIDI-Datei in blau), die Audiospur der aufgenommenen PianoVision-Datei (rot) und das dazugehörige Video in der Software Ableton.

24 fps (Bilder pro Sekunde). Die Videodatei beinhaltet einen Stereo-Audiokanal, der auch zur Analyse benutzt wird.

4.4.2 *MIDI-Aufnahme* Zusätzlich zu den Audio- und Videodaten werden gleichzeitig MIDI-Daten über das MIDI-Keyboard aufgezeichnet. Diese Daten enthalten Informationen über die gespielten Noten, die Anschlagsdynamik und die Länge der Töne. Die MIDI-Datensätze dienen als Referenzwert und Grundwahrheit und ermöglichen die genaue Verifizierung der erkannten Noten und ihrer Eigenschaften. Der Eingangs-Latenz der MIDI-Aufnahmen liegt bei 5ms. Diese werden zu den Verzögerungswerten des Versuchs addiert, um Absolutwerte zu erhalten.

### 4.5 Analyse

Eine detaillierte Analyse wird vorgenommen, um die Genauigkeit der PianoVision-Anwendung bei der Erfassung der aufgenommenen Noten zu evaluieren. Dabei werden verschiedene Metriken berücksichtigt:

- **Fehlerrate:** Die Anzahl der falsch erkannten oder nicht erkannten Noten im Vergleich zu den tatsächlich gespielten Noten. Dabei werden die Fehlerraten aller Finger analysiert.
- **Verzögerung:** Die Zeitdifferenz zwischen dem Spielen einer Note auf dem MIDI-Keyboard und ihrer Erkennung durch die Anwendung.
- **Gesamtgenauigkeit:** Die prozentuale Genauigkeit der PianoVision Anwendung über zwei Musikstücke mit unterschiedlicher Komplexität hinweg.

Diese Analyse ermöglicht es, die Leistungsfähigkeit der Anwendung in Bezug auf das Erfassen von registrierten Noten zu beurteilen.

## 5 Ergebnisse

In diesem Kapitel werden die Ergebnisse dieser Untersuchung beschrieben. Es werden die in Kapitel 4.5 aufgeführten Analysekriterien ausgewertet.

### 5.1 Fehlerrate

Tabelle 2 zeigt die jeweilige absolute Fehler rate in den Durchgängen der Musikstücke. Dabei handelt es sich um nicht registrierte oder falsch registrierte Noten.

Tabelle 2: Absolute Fehlerrate

	Lied 1	Lied 2
Durchgang 1	6	24
Durchgang 2	2	31
Durchgang 3	6	28
<b>Gesamt</b>	14	83

Die prozentualen Fehlerraten der jeweiligen Musikstücke werden mit folgender Formel berechnet:

$$Fehler(\%) = \frac{Gesamtnoten}{Gesamt\ fehler} \times 100$$

Daraus ergeben sich folgende prozentuale Fehler der Musikstücke:

- Lied 1: 11,11 %
- Lied 2: 26,66 %

**5.1.1 Lied 1** Die absolute Noten- und Fehlerrate der einzelnen Finger in Lied 1 kann man der Tabelle 3 entnehmen. Basierend auf den Daten hat der Mittelfinger eine Fehlerrate von 12,5%, der Ringfinger eine Fehlerrate von 66,67% und der kleine Finger eine Fehlerrate von 50%.

**Tabelle 3: Absolute Fehler- und Notenrate je Finger der linken Hand in Lied 1**

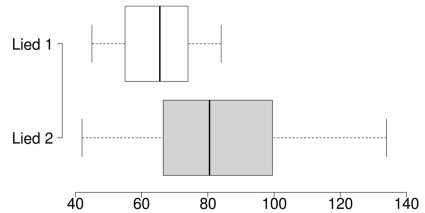
Finger	Fehler #	Noten #
Daumen	3	8
Zeigefinger	4	14
Mittelfinger	1	8
Ringfinger	4	6
kleiner Finger	2	4

**5.1.2 Lied 2** Die absolute Noten- und Fehlerrate der einzelnen Finger in Lied 2 kann man in Tabelle 4 einsehen. Die Gesamtfehler der linken Hand liegen bei 31,94 % und bei der rechten Hand bei 25,88 %. Aus den Daten kann man ableiten, dass die Ringfinger mit etwa 40,68 % für den höchsten Prozentsatz der Fehler verantwortlich sind, gefolgt von den Mittelfingern mit etwa 30,51 % und den kleinen Fingern mit 25,42 %.

## 5.2 Verzögerung

Aus Abbildung 3 können die Verzögerungsdaten der jeweiligen Lieder entnommen werden. Die Anzahl der Datenpunkte (registrierte Noten) beträgt bei Lied 1: 106 und bei Lied 2: 217. Der Median liegt bei Lied 1 bei 65,5 ms und bei Lied 2 bei 80,5 ms. Demnach ist der Median in Lied 2 um 15 ms höher als in Lied 1. Der Interquartilsabstand (IQR) beträgt 33

ms für Lied 1 und 44,5 ms für Lied 2. Hier zeigt sich, dass die Verzögerungen in Lied 2 ungleichmäßiger verteilt sind, der IQR ist um 11,5 ms höher. Während Lied 1 nur einen Ausreißerwert bei 87 ms aufweist, zeigt Lied 2 mehrere Ausreißer mit einem Maximum von 136 ms.



**Abbildung 3: Boxplot der aufgezeichneten Verzögerungen von Lied 1 & Lied 2**

## 5.3 Gesamtgenauigkeit

Die Gesamtgenauigkeit setzt sich, wie aus Formel 1 zu entnehmen ist, aus den zuvor erhobenen Daten zusammen:

Die Gesamtgenauigkeit (Genauigkeit) kann berechnet werden mit der Formel:

$$\text{Genauigkeit} = 100 - \frac{\text{Fehleranzahl}}{\text{Notenanzahl}} \times 100 \quad (1)$$

Hierbei sind:

Genauigkeit : Gesamtgenauigkeit in Prozent

Fehleranzahl : Gesamte Anzahl der Fehler

Notenanzahl : Gesamte Anzahl der Noten

Hieraus folgen die in Tabelle 1 dargestellten Genauigkeiten zu den Musikstücken und die Gesamtgenauigkeit.

Insgesamt ergibt sich für diese Untersuchung eine Gesamtgenauigkeit aller Musikstücke von 77,23 % wie man Tabelle 5 entnehmen kann.

## 6 Diskussion

In diesem Kapitel werden die Ergebnisse der analysierten Daten interpretiert. Es wird auf

**Tabelle 4: Anzahl der Noten pro Finger und die absolute Fehlerrate in Lied 2**

Finger	Noten pro Finger			Absolute Fehlerrate	
	Bezeichnung	Rechte Hand	Linke Hand	R. Hand	L. Hand
Daumen	1. Finger	60	12	7	5
Zeigefinger	2. Finger	84	4	9	0
Mittelfinger	3. Finger	48	9	13	5
Ringfinger	4. Finger	24	18	15	9
Kleiner Finger	5. Finger	18	6	11	4
<b>Gesamt</b>		228	72	59	23

**Tabelle 5: Genauigkeit einzelner Lieder und die Gesamtgenauigkeit**

	Genauigkeit
Lied 1	88,89 %
Lied 2	73,33 %
Gesamt	77,23 %

die Fehlerrate eingegangen und in Anschluss auf die Verzögerung, sowie die Gesamtgenauigkeit. Die eingangs aufgeführten Fragestellungen werden beantwortet.

### 6.1 Fehlerrate

Es zeigt sich, dass die Fehlerrate der Erkennung der Noten mittels der PianoVision-Anwendung mit der Komplexität der Musikstücke steigt. Ebenso verschlechtert sich die Erkennung des Tastenanschlags mit steigender Komplexität. Der Hauptgrund dafür ist insbesondere die Verdeckung der durch die Mittel-Ring- und Zeigefinger gespielten Noten, da diese durch die davor positionierten Finger verdeckt werden. Diese Finger sind für 69,51 % aller Fehler verantwortlich.

### 6.2 Verzögerung

Die Untersuchung der Videodaten ergab: Die gleichzeitig oder in schneller Abfolge gespielten Noten werden erst später erfasst, aufgrund der Finger, die andere verdeckten. Dies erklärt die verzögerte Registrierung dieser Töne und die Häufung von Ausreißern in Lied 2. Die Verzögerungsdaten von Lied 1 und Lied 2 weisen sowohl Gemeinsamkeiten als auch Unterschiede auf. Beide Datensätze zeigen eine positive Schiefe: die Mehrheit der Verzögerungswerte liegt unterhalb des Medians. In beiden Liedern sind Ausreißer vorhanden, die auf vereinzelte, sehr lange Verzögerungen hinweisen. Lied 2 hat einen höheren Median und einen größeren IQR, was auf eine stärkere Variabilität der Verzögerungen hindeutet. Lied 2 weist eine Häufung von Ausreißern auf, die durch verdeckte Noten und die damit verbundene verzögerte Erfassung erklärt werden kann.

### 6.3 Gesamtgenauigkeit

Die Ergebnisse spiegeln sich auch in den Gesamtgenauigkeiten der ausgewählten Musikstücke wider. Je komplexer ein Lied ist, desto geringer ist die Gesamtgenauigkeit der Erkennungsleistung mittels Hand-Tracking. Die eingangs gestellte Forschungsfrage 1:

"Wie hoch ist die Genauigkeit der Notenerkennung durch die Anwendung PianoVision?" kann mit 77,28 % beantwortet werden. Forschungsfrage 2: "Welche Auswirkungen hat die Komplexität des Musikstücks auf die Genauigkeit der Notenerkennung des virtuellen Klaviers?" kann folgend beantwortet werden: Die prozentuale Genauigkeit für das Lied 1 liegt bei 88,88 % und für das komplexere Lied 2 bei 73,33 %. Das ist ein Unterschied von 15,55 %.

## 7 Fazit

Die AR-Anwendung PianoVision ermöglicht das Klavierspielen mit der VR-Brille Oculus Meta Quest 2 ohne physisches Klavier, ohne Lehrer und ohne Notenkenntnisse. Die Genauigkeit der Notenerkennung wurde getestet und war für einfache Klavierstücke ausreichend. Es stellte sich jedoch heraus, dass das Hauptproblem nicht das gleichzeitige Spielen und Registrieren der Noten ist, sondern beim Hand-Tracking die Verdeckung von Mittelfinger, Ringfinger und kleinem Finger durch davon positionierte Finger. Dies wirkt sich stark auf die Genauigkeit und die Verzögerung aus. Um ein immersives und optimales Lernerlebnis zu ermöglichen, muss die Hand-Tracking-Technologie weiterentwickelt werden. Eine Lösung könnte darin bestehen, maschinelle Lernverfahren zur Vorhersage der Fingerpositionen einzusetzen.

## Literatur

- [1] Diar Abdulkarim u. a. „A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the Meta Quest 2“. In: *Behavior Research Methods* 56.2 (Feb. 2024), S. 1052–1063. DOI: 10.3758/s13428-022-02051-8.
- [2] Mervyn John Cook. „Augmented Reality: Examining its value in a music

technology classroom. Practice and potential.“ In: *Waikato Journal of Education* 24.2 (Nov. 2019), S. 23–38. DOI: 10.15663/wje.v24i2.687.

- [3] Alejandro Cremaschi u. a. „Students who quit music lessons: recent research and recommendations for teachers“. In: *MTNA e-Journal* 6 (Feb. 2015), S. 15.
- [4] Kangxu Cui. „Artificial intelligence and creativity: piano teaching with augmented reality applications“. In: *Interactive Learning Environments* 31.10 (2023), S. 7017–7028. DOI: 10.1080/10494820.2022.2059520.
- [5] Katerina Labrou u. a. „Following the Master’s Hands: Capturing Piano Performances for Mixed Reality Piano Learning Applications“. In: CHI EA ’23. , Hamburg, Germany, Association for Computing Machinery, 2023. ISBN: 9781450394222. DOI: 10.1145/3544549.3585838.
- [6] Cheryl D. Metcalf u. a. „Complex hand dexterity: a review of biomechanical methods for measuring musical performance“. In: *Frontiers in Psychology* 5 (2014). DOI: 10.3389/fpsyg.2014.00414.
- [7] D. Molero u. a. „A novel approach to learning music and piano based on mixed reality and gamification“. In: *Multimedia Tools and Applications* 80.1 (2021), S. 165–186. DOI: 10.1007/s11042-020-09678-9.
- [8] Meinard Müller. *Fundamentals of Music Processing*. Springer, 2021. DOI: 10.1007/978-3-030-69808-9.
- [9] Liam Rigby, Burkhard C. Wünsche und Alex Shaw. „piARno - An Augmented Reality Piano Tutor“. In: OzCHI ’20. Sydney, NSW, Australia: Association for Computing Machinery, 2021, S. 481–491. ISBN: 9781450389754. DOI: 10.1145/3441000.3441039.



©2024 Michael Stych. Lizenznehmer Hochschule Reutlingen, Deutschland.

Dieser Artikel ist ein Open-Access-Artikel unter den Bedingungen und Konditionen der Creative Commons Attribution (CC BY)-Lizenz. <http://creativecommons.org/licenses/by/4.0>

# Eignung von Spiel-Design-Elementen in der industriellen Fertigung

Rebecca Müller

Hochschule Reutlingen  
Reutlingen, Deutschland

Rebecca.Mueller@Student.Reutlingen-University.de

## Abstract

Bei der Gamification werden Spiel-Design-Elemente in einem spielfremden Kontext eingesetzt. Der bislang wenig erforschte Einsatz in der industriellen Fertigung zeigt auf, dass dieser die Motivation und die Leistung verbessern kann. In dieser Arbeit wird analysiert, welche spezifischen Spiel-Design-Elemente für eine Gamification-Anwendung in der industriellen Fertigung geeignet sind. Hierfür wird mittels einer systematischen Literaturrecherche erörtert, welche Gamification-Konzepte in diesem Kontext bislang erfolgreich getestet wurden. Es wurden dreizehn Kriterien identifiziert, die durch insgesamt acht Elemente erfüllt werden können. Es wird ein interaktives User-Interface vorgestellt, das eine spezifische Auswahl dieser Kriterien für den Einsatz in der industriellen Fertigung ermöglicht. Anhand dieser Auswahl wird angezeigt, welche Spiel-Design-Elemente geeignet sind.

## CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)** → **HCI theory, concepts and models** • **Human-centered computing** → **Interaction design** →

---

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

Informatics Inside Frühjahr 2024  
08. Mai 2024, Hochschule Reutlingen  
Copyright 2024 Rebecca Müller

Interaction design process and methods →  
User centered design • **Applied computing**  
→ **Operations research** → **Industry and  
manufacturing**

## Keywords

Gamification, Spiel-Design-Elemente, industrielle Fertigung

## 1 Einleitung

Gamification ist in den Bereichen der Bildung und des Gesundheitswesens wissenschaftlich fundiert. In diesen Bereichen konnten bereits einige Vorteile der Anwendung von Gamification belegt werden. Im Kontext der industriellen Fertigung ist diese Anwendung vergleichsweise wenig erforscht, obwohl Vorteile versprechen, ebenfalls mit der Anwendung im industriellen Kontext zu wirken. So konnten schon einige positive Wirkungen, wie beispielsweise die Erhöhung von Motivation und Leistung gezeigt werden [13]. Ein generelles Interesse an dem Einsatz von Gamification in diesem Umfeld ist bekannt [2, 9]. Dieses Kapitel umfasst das Ziel, die Methodik und die weitere Strukturierung dieser Arbeit.

### 1.1 Ziel der Arbeit

Dieser Artikel widmet sich der Fragestellung: Welche Spiel-Design-Elemente eignen sich für den Einsatz in der industriellen Fertigung? Die Beantwortung dieser Frage soll in einer Übersicht resultieren, die zeigt, wel-

che Spiel-Design-Elemente bereits erfolgreich im Kontext der industriellen Fertigung getestet wurden. Die Darstellung dieser Übersicht erfolgt in einem interaktiven User-Interface. Dieses zeigt welche Spiel-Design-Elemente unter bestimmten Voraussetzungen und Zielen geeignet sind.

## 1.2 Methodik

Für die Beantwortung der Forschungsfrage wurde eine systematische Literaturrecherche durchgeführt. Zur Durchführung der Recherche wurden die Suchterme (gamification) AND (production) sowie, das deutsche Pendant, (gamification) AND (produktion) verwendet. Diese Suchterme zielen darauf ab Literaturen zu finden, in denen Gamification-Anwendungen bereits im industriellen Umfeld eingesetzt und getestet wurden. Die in diesem Artikel aufgeführten Literaturquellen wurden in den Fachdatenbanken Springer Link<sup>1</sup>, ACM Digital Library<sup>2</sup> und ResearchGate<sup>3</sup> gefunden. Einschlusskriterien für die ausgewählte Literatur waren der Einsatz von Gamification-Anwendungen im Bereich der Produktion und zugehörigen Logistik. Außerdem sollten die, in den Artikeln, eingesetzten Gamification-Anwendungen positive Ergebnisse erzielt haben. Zusätzlich wurden weitere Literaturquellen aufgenommen, die in den gefundenen Arbeiten verwendet wurden. Zuletzt wurden grundlegende Arbeiten zur Gamification und Produktion aufgenommen. Die Erarbeitung des interaktiven User-Interfaces basiert ebenfalls auf der Literaturrecherche sowie der Zusammenfassung und Strukturierung aller daraus gewonnenen Erkenntnisse. Diese beziehen sich auf die Voraussetzungen und Ziele für den Einsatz von Gamification-Anwendungen im industriellen Umfeld sowie auf die Erkenntnisse zu den erfolgreich eingesetzten Spiel-Design-Elementen.

## 1.3 Aufbau der Arbeit

In Kapitel 2 werden zunächst die Grundlagen für diese Arbeit beschrieben. Diese umfassen die Definitionen der Gamification und der Spiel-Design-Elemente sowie die Beschreibung der für diese Arbeit relevanten Spiel-Design-Elemente. Außerdem wird der Kontext der industriellen Fertigung definiert. In Kapitel 3 werden bisherige Forschungen zur Gamification im industriellen Arbeitsumfeld zusammengefasst. In Kapitel 4 werden Spiel-Design-Elemente bestimmten Voraussetzungen und Zielen zugeordnet. Dadurch wird eine Übersicht gegeben, unter welchen dieser Bedingungen welche Spiel-Design-Elemente besonders für eine Gamification-Anwendung in der industriellen Fertigung in Betracht gezogen werden können. In Kapitel 5 wird das in dieser Arbeit erstellte interaktive User-Interface und seine Funktionen beschrieben. In Kapitel 6 wird eine Zusammenfassung dieser Arbeit gegeben. Abschließend wird in Kapitel 7 eine Einordnung der Arbeit vorgenommen sowie ein Fazit gezogen.

## 2 Grundlagen

Folgend wird zunächst die industrielle Fertigung definiert, in deren Kontext die Gamification in diesem Artikel behandelt wird. Anschließend wird auf die Gamification eingegangen, indem diese definiert wird. Anschließend werden alle relevanten Spiel-Design-Elemente, die im Verlauf des Artikels relevant sind, beschrieben. Zuletzt werden allgemeine, zu berücksichtigende, Aspekte behandelt.

### 2.1 Industrielle Fertigung

Die Produktion wird von Kellner et al. [2020] als die „Erzeugung von Gütern durch die Kombination von Produktionsfaktoren“ [5] definiert. Allgemein liegt es im Sinne des

<sup>1</sup> <https://link.springer.com/>

<sup>2</sup> <https://www.acm.org/>

<sup>3</sup> <https://www.researchgate.net/>

Produktionsmanagements die Arbeitsleistung zu erhöhen. Maßgeblich hierfür ist unter anderem die Erhöhung der Leistungsbereitschaft der Mitarbeiter. Diese setzt sich aus der physischen und der psychischen Leistungsbereitschaft zusammen. Die psychische Leistungsbereitschaft kann laut Kellner et al. [2020] mit der Motivation eines Mitarbeitenden gleichgesetzt werden [5]. Die Gamification zielt darauf ab, diese Motivation zu erhöhen und mittels dieser Motivationssteigerung die Leistungen der Nutzer zu verbessern [13].

### 2.2 Gamification

Bislang gibt es noch keine allgemeingültige Definition der Gamification. Für diese Arbeit wird die Definition von [13] angenommen, da diese keine spezifischen Ziele beinhaltet und in dieser Arbeit möglichst umfassende Ziele und Voraussetzungen aufgenommen werden. Somit ergibt sich folgende Definition nach Sailer [2016]: „Gamification ist ein Prozess der spielerischen Gestaltung von Aktivitäten in einem spielfremden Kontext durch die Verwendung von Spiel-Design-Elementen.“ [13]. Spiel-Design-Elemente sind die grundlegenden Bausteine einer Gamification-Anwendung. Sie können in Abhängigkeit ihrer Darstellung und Form unterschiedlich wirken [13]. Die folgenden Spiel-Design-Elemente finden Anwendung im weiteren Verlauf dieses Artikels. Es ist nicht die Absicht eine vollständige Aufzählung aller Spiel-Design-Elemente darzustellen.

#### Punkte

Punkte sind eines der am häufigsten eingesetzten Spiel-Design-Elemente in Gamification-Anwendungen. Sie können für das Abschließen verschiedener Aufgaben vergeben werden und visualisieren den aktuellen Spielstand. Dadurch kann dem Nutzer, je nach Anwendung der Punkte, transparent vermittelt werden, wann dieser ein höheres Level erreicht oder wie die Leistung im Vergleich zu anderen Nutzern ist [13]. Bei der

Gestaltung einer Punktevergabe sollte bedacht werden, inwiefern diese die Nutzer beeinflussen. Es kann sein, dass das Vergeben von Punkten das Stresslevel der Nutzer erhöht, beispielsweise durch einen erhöhten Leistungsdruck durch den Punkttestandvergleich mit anderen Nutzern. Dieses Stresslevel könnte reduziert werden, indem Aufgaben mehrmals bearbeitet werden können, deren Abschluss mit Punkten belohnt wird [10].

#### Abzeichen

Werden vordefinierte Leistungen eines Nutzers erbracht können diese mit Abzeichen belohnt werden [10]. Sie können dabei verschieden definierte Leistungen repräsentieren, wie z.B.:

- Die erstmalige Ausführung einer Aktivität.
- Das Erreichen eines bestimmten Punkttestands.
- Das Abschließen gewisser Aufgaben innerhalb einer vorgegebenen Zeitspanne.

Das Erreichen dieser Abzeichen macht für die Nutzer zu erreichende Ziele transparent. Zusätzlich können Abzeichen zur Orientierung für Nutzer beitragen, indem sie aufzeigen, welche Aktionen möglich sind. Vor allem für neue Nutzer einer Gamification-Anwendung ist dies besonders hilfreich. Beispielsweise kann eine Liste aller Abzeichen, mit den Bedingungen zum Erhalt dieser, hilfreich sein. Zusätzlich können Abzeichen als Statussymbol dienen und die Motivation der Nutzer durch die Fortschrittsanzeige der Abzeichen stärken. In einem Wettbewerb können sie zum Vergleich verschiedener Nutzer dienen [13].

#### Fortschrittsanzeige

Individuelle Leistungen eines Nutzers können innerhalb eines definierten Zeitraums mit Fortschrittsanzeigen abgebildet werden. Sie zeigen dem Nutzer an, welche Aufgaben oder Teilziele sie bisher erreicht haben. Außerdem machen sie für den Nutzer sichtbar

wie viele Teilziele bereits erreicht wurden und noch nötig sind zur Vervollständigung eines Gesamtziels. Fortschrittsanzeigen sind für alle Leistungsstärken unterschiedlicher Nutzer motivierend [13].

### **Bestenlisten**

Bestenlisten stellen Nutzer in einer Auflistung dar, die nach einer bestimmten Variable sortiert ist. Sie können den Nutzern Zielsetzungen aufzeigen. Diese sollten klar definiert, zeitgebunden, messbar und vor allem erreichbar sein. Bestenlisten bringen die Nutzer in eine, für alle einsehbare, Reihenfolge ihrer erreichten Leistungen. Somit ist dieses Spiel-Design-Element von einem hohen kompetitiven Charakter geprägt. Es sollte beachtet werden, dass dies bei Nutzern sowohl zu einer Motivationssteigerung als auch zu einer Demotivation, je nach erbrachter Eigenleistung, führen und den Leistungsdruck dadurch erhöhen kann [13]. Es sollten Überlegungen angestrebt werden, inwieweit diese Listen eingesehen werden können und welche Ziele damit erreicht werden sollen. So kann es beispielsweise hilfreich sein einen Nutzer nur in dem Kontext des nächstbesseren und -schlechteren Nutzer abzubilden [11].

### **Level**

Die Gestaltung von Level kann von verschiedenen Faktoren abhängig gemacht werden, beispielsweise durch das Erreichen eines gewissen Punktestandes. Je nach Zielsetzung kann der Aufstieg in ein höheres Level an das Erreichen von Aufgaben in einem bestimmten Zeitraum oder an eine bestimmte Anzahl an Ausführungen einer Aktivität geknüpft sein [10]. Das Erreichen eines Aufstiegs in ein höheres Level kann mit neuen Fähigkeiten oder der Verbesserung definierter Attribute verbunden sein. Dadurch werden Nutzer dazu animiert, sich länger mit den Inhalten der Gamification-Anwendung zu beschäftigen [14].

### **Narrativ**

Mit einem Narrativ ist das Unterlegen einer

Gamification-Anwendung mit einer Geschichte gemeint. Die Geschichte kann sich an der realen Situation orientieren, an der die Anwendung anknüpft. Ein Narrativ kann die Immersion erhöhen und Handlungen eine Bedeutung zukommen lassen. Außerdem kann es genutzt werden, um ein Ziel vorzugeben. Es sollte beachtet werden, dass Nutzer stets das Gefühl haben sollten, einen Einfluss auf die Geschichte zu haben. Andernfalls sollten sie den Handlungsstrang des Narrativs als spannend und lohnend empfinden [13].

### **Avatar**

Mit Avataren kann eine Repräsentation eines Nutzers geschaffen werden, mit der sich dieser identifizieren kann. Avatare können ebenso zur Immersion in eine Gamification-Anwendung beitragen. Bei der Gestaltung von Avataren ist zu beachten, dass diese gut durchdacht und ansprechend sein sollten, um eine spielerische Erfahrung zu erzeugen [13].

### **Nicht-Spieler-Charaktere**

Nicht-Spieler-Charaktere sind Charaktere, die nicht aktiv von den Nutzern gesteuert werden können. Sie können hilfreich in einer Gamification-Anwendung mit Narrativ sein, indem sie beispielsweise Aufgaben an die Nutzer übergeben und ihr Gefühl der Bedeutsamkeit erhöhen [13].

Im Allgemeinen sollten neben einer besonderen Eignung der Spiel-Design-Elemente, je nach Voraussetzungen und Zielen, weitere Aspekte beachtet werden. Die Art der Implementierung sowie die Gestaltung der Spiel-Design-Elemente spielen eine entscheidende Rolle bei ihrer Wirkung [13, 14]. Ebenso wurde festgestellt, dass das Kombinieren der Spiel-Design-Elemente und deren Interaktion miteinander wirkungsvoller sind als isolierte Elemente [10, 13]. Um die Wirkung einer Gamification-Anwendung zu maximieren bietet es sich weiter an, allgemeine psychologische Grundlagen zur Motivation heranzuziehen [11, 13]. Bei der Erwägung von

## Eignung von Spiel-Design-Elementen in der industriellen Fertigung

Teambildungen sollte besonders darauf geachtet werden, welche Ziele damit erreicht werden sollen. Teams können einerseits das Gefühl des Zusammenhalts und dadurch eine Motivationssteigerung, aber auch Konflikte durch einen Wettbewerbscharakter schaffen [14]. Zuletzt sollten immer ethische Richtlinien, wie beispielsweise die freiwillige Teilnahme, sowie die allgemeinen Arbeitsschutzbedingungen Berücksichtigung finden [11, 13].

### 3 Stand der Forschung

Korn et al. entwickelten verschiedene Gamification-Anwendungen, die sie mit beeinträchtigten Menschen evaluierten. Zunächst stellten sie einen Entwurf vor, den sie „(Production) Tetris“ [7, 8] betitelten. In diesem wurde der Fortschritt auf einem Bildschirm mit an Tetris angelehnte Blöcke visualisiert. Die Farben der Blöcke wechselten von Grün zu Rot. Diese Farbcodierung spiegelte die Bearbeitungszeit eines Arbeitsschrittes gemessen an der Durchschnittszeit des Nutzers wider. Diese wurde ebenfalls mit einem grauen Block visualisiert, der sich mit dieser Durchschnittszeit bewegte [6]. Getestet wurde diese Anwendung mit dem Zusammenbauen eines Auto-Fahrgestells mit Lego-Bausteinen [7].

Die zweite Gamification-Anwendung war „Circles & Bars“ [7]. Hier wurde das Zusammenbauen von Metallscheren als Arbeitsgang ausgewählt. Es wurde eine Fortschrittsanzeige mit einem Kreis und zwei Balken auf die Arbeitsfläche projiziert. Der Kreis änderte seine Farbe ebenfalls nach der Bearbeitungsgeschwindigkeit von Grün zu Rot. Zusätzlich wurde der Radius des Kreises reduziert und mit einem Würfel in der Mitte des Kreises die Prozessnummer abgebildet. Nach Abschluss eines Arbeitsschrittes wurde der erste Fortschrittsbalken gefüllt. Das Beenden einer Sequenz führte zur Verschiebung des Füllungsstandes des eben genannten Balkens in den zweiten Balken, zur Visualisierung des Gesamtfortschritts [7].

Der dritte Ansatz war eine Fortschrittsanzeige in Form einer Pyramide, wie diese Anwendung ebenfalls betitelt wurde. In dieser stellten die einzelnen Stufen Arbeitsschritte dar. Diese Stufen zeigten mittels Farbcodierung von Grün zu Rot die Arbeitsgeschwindigkeit an. Wurde ein Fehler im Arbeitsschritt gemacht färbte sich die Stufe sofort rot. Auf den Stufen der Pyramide befand sich ein Avatar des Nutzers, der diese erklimmt. Wenn alle Stufen der Pyramide grün waren und somit alle Arbeitsschritte fehlerfrei bearbeitet wurden, erhielt der Avatar einen Pokal [10].

Die vorangegangenen Gamification-Anwendungen wurden in einem Interview mit zehn Instruktoren aus zwei Manufakturen der Automobilbranche diskutiert. Es zeigte sich, dass ein großes Interesse an den Gamification-Anwendungen besteht. Bevorzugt wurde hierbei das Design der Pyramide. Es wurde eine sehr hohe generelle Akzeptanz einer Gamification-Anwendung in der Automobilbranche festgestellt, sofern die Einhaltung der spezifischen Anforderungen der Branche gewährleistet wird [8].

Sailer untersuchte in [13] die Wirkung von Gamification auf die Motivation und Leistung im industriellen Kontext. Hierfür wurden zwei Studien durchgeführt. Zunächst eine Feldstudie im Bereich der Intralogistik. 103 Teilnehmer, gleichmäßig aufgeteilt in eine Experimental- und eine Kontrollgruppe, testeten den Vorgang der Kommissionierung. Der Arbeitsvorgang der Experimentalgruppe war angereichert mit diversen Spiel-Design-Elementen. Es zeigte sich, dass durch diese Elemente sowohl Aspekte der Motivation als auch der Leistung gefördert wurden. In der weiterführenden Simulationsstudie wurde die Kommissionierung online mit einzelnen Spiel-Design-Element-Gruppen angereichert, um die Wirkung der einzelnen Elemente untersuchen zu können [13].

Caserman et al. [2023] untersuchten Generationsunterschiede in der Akzeptanz einer

Gamification-Anwendung, wie auch die Akzeptanz der dafür verwendeten Spiel-Design-Elemente. Hierfür wurde eine App entwickelt, die mit 367 Angestellten getestet wurde. Es wurde festgestellt, dass eine allgemeine Akzeptanz von Gamification-Anwendungen im Arbeitsbereich von den Generationen abhängig ist. So akzeptierte Generation X und Y die Anwendung am ehesten und konnte sich eine tatsächliche Einbindung in den Arbeitsalltag vorstellen. Generation Z hingegen konnte sich eine Einbindung in den Arbeitsalltag nur bedingt vorstellen [1].

## 4 Anwendung

Im folgenden Kapitel wird dargelegt welche Spiel-Design-Elemente für den Einsatz in der industriellen Fertigung geeignet sind. Die folgenden Unterkapitel entsprechen den Hauptkriterien, die sich aus einzelnen Kriterien zusammensetzen. Die Hauptkriterien fassen unterschiedlichste Voraussetzungen und Ziele zusammen, die für eine Umsetzung von Gamification-Anwendungen in der industriellen Fertigung berücksichtigt werden können. Aus einem oder mehreren Kriterien leiten sich geeignete Spiel-Design-Elemente ab. So ergibt sich am Ende dieses Kapitels eine Übersicht der Spiel-Design-Elemente zugeordnet zu den Hauptkriterien.

### 4.1 Sicherheit

Das Hauptkriterium *Sicherheit* deckt Prozessabläufe ab, die fehleranfällig und risikobehaftet sind. In [10] wird aufgezeigt, dass in diesem kritischen Arbeitsumfeld die Ablenkung durch Gamification-Anwendungen so gering wie möglich gehalten werden sollte. Dazu gehört ebenfalls, dass eine geringe Interaktion oder, wenn möglich, keine Interaktion mit der Anwendung stattfindet. Somit lassen sich zu dem Hauptkriterium *Sicherheit* zwei Kriterien zuordnen:

- Geringe bis keine Interaktion mit den Spiel-Design-Elementen
- Ablenkung durch die Gamification möglichst gering halten

Nach [10] können diese Kriterien mit dem Spiel-Design-Element **Fortschrittsanzeige** umgesetzt werden.

### 4.2 Persönlichkeit

In [4] wird festgestellt, dass die Persönlichkeitsmerkmale der Geduld und Kompetitivität einen Einfluss auf die wahrgenommene Arbeitslast haben. Dies sollte in der Wahl der Spiel-Design-Elemente berücksichtigt werden. Weiter können Persönlichkeitsmerkmale anhand von Spielertypen, wie in [12] vorgenommen werden. Achiever bereichern sich vorwiegend damit Punkte zu sammeln und einen Stufenaufstieg zu erreichen. Explorer finden einen Anreiz darin nach interessanten oder raren Elementen Ausschau zu halten. Die Interaktion mit der Spielwelt ist für sie relevant. Socialiser haben Interesse an der Interaktion mit anderen Nutzern. Killer legen den Fokus darauf auf andere Spieler einwirken zu können [12].

Zu dem Hauptkriterium *Persönlichkeit* lassen sich folglich folgende Kriterien zuordnen:

- Hohe Geduld und hohe Kompetitivität
- Spielertyp: Achiever
- Spielertyp: Explorer
- Spielertyp: Socialiser
- Spielertyp: Killer

Schätzen sich die Nutzer als geduldig und wettbewerbsorientiert ein, kann der Einsatz von **Bestenlisten** in Betracht gezogen werden [14]. Falls diese Eigenschaften der Nutzer überwiegend gering sind, sollte auf den Einsatz dieses Spiel-Design-Elements verzichtet werden. Überwiegt der Spielertyp Achiever empfiehlt sich der Einsatz von **Abzeichen, Level oder Punkten** [10, 12]. Für den Spielertyp Explorer ist es empfehlenswert versteckte oder gar geheime Spiel-Design-Elemente einzubinden. **Punkte** können notwendig sein, um weiteres Explorieren zu ermöglichen. Der Spielertyp Socialiser profitiert von Kooperationen mit anderen Spielern, was mittels Teams bewirkt werden

Eignung von Spiel-Design-Elementen in der industriellen Fertigung

kann [14]. Ein Motivationsanreiz können **Punkte oder Abzeichen** darstellen, durch deren Erreichen das Freischalten von Kommunikationswegen erfolgt. **Punkte** eignen sich, um dem Spielertypen Killer das Freischalten von Fähigkeiten zu ermöglichen, die das Einwirken auf andere Spieler gestatten [12]. **Abzeichen** können ebenfalls ansprechend für diesen Spielertyp sein [10].

### 4.3 Motivation

Eine Steigerung der Motivation, begründet auf der Selbstbestimmungstheorie nach [3], wurde von Sailer in [13] untersucht. Nach der Subtheorie der psychologischen Grundbedürfnisse werden drei dieser Grundbedürfnisse definiert. Das Bedürfnis nach Kompetenzerleben beschreibt das Gefühl sich selbst als effektiv im Umgang mit seinem Umfeld zu erfahren. Das Erleben von Autonomie zeigt sich mit psychologischer Freiheit, bewusstem und zielgerichteten Handeln und sich selbst als Verursacher dieser Handlungen anzuerkennen. Soziale Eingebundenheit wird mittels einer positiven Wechselwirkung mit anderen Menschen erlebt [13]. Dem Hauptkriterium *Motivation* ordnen sich folgende Kriterien unter:

- Förderung des Kompetenzerlebens
- Förderung der Aufgabenbedeutsamkeit
- Erleben sozialer Eingebundenheit

**Abzeichen, Bestenlisten und Fortschrittsanzeigen** erwiesen sich als fördernd für das Kompetenzerleben und die Aufgabenbedeutsamkeit. Das Erleben sozialer Eingebundenheit kann durch die Elemente **Avatar, Narrativ** und **Nicht-Spieler-Charaktere** gefördert werden [13].

### 4.4 Generation

In [1] wird die Akzeptanz der Spiel-Design-Elemente bezüglich der Generationsunterschiede untersucht. Hierfür teilen sie die Generationen folgendermaßen auf:

- Generation X: 16 – 24 Jahre
- Generation Y: 25 – 40 Jahre

- Generation Z: über 40 Jahre

Diese Unterteilung stellt ebenfalls die Kriterien des Hauptkriteriums *Generation* dar. Das Spiel-Design-Element **Avatar** wurde von allen Generationen favorisiert. **Abzeichen** wurden von den Generation Y und Z befürwortet. **Bestenlisten** und **Punkten** standen alle Generationen neutral gegenüber [1].

### 4.5 Übersicht der Hauptkriterien

In Tabelle 1 ist eine Übersicht aller zuvor beschriebener Hauptkriterien mit ihren geeigneten Spiel-Design-Elementen. Bei Mehrfachnennung der Spiel-Design-Elemente stehen solche an erster Stelle, die am häufigsten dem entsprechenden Hauptkriterium zugeordnet werden können. Die Reihenfolge der Elemente ist von am häufigsten zu am wenigsten häufig zu den Hauptkriterien passend sortiert.

**Tabelle 1: Übersicht der Hauptkriterien und ihren geeigneten Spiel-Design-Elementen**

Hauptkriterien	Spiel-Design-Element/e
Sicherheit	Fortschrittsanzeige
Persönlichkeit	Punkte, Abzeichen, Bestenlisten, Level
Motivation	Abzeichen, Bestenlisten, Fortschrittsanzeigen, Avatar, Narrativ, Nicht-Spieler-Charaktere
Generation	Avatar, Abzeichen, Bestenlisten, Punkte

## 5 Interaktives User-Interface

Nicht alle industriellen Arbeitsschritte umfassen dieselben Voraussetzungen und Ziele. Deshalb wurde ein interaktives User-Inter-

face entwickelt, das eine spezifische Auswahl der, in dieser Arbeit erarbeiteten, Hauptkriterien und Kriterien ermöglicht. Dadurch kann eingesehen werden, welche Spiel-Design-Elemente für einen spezifischen Arbeitsprozess und seine individuellen Bedingungen am geeignetsten sind. Abbildung 1 zeigt das interaktive User-Interface<sup>4</sup>.

Mit der Checkbox der Hauptkriterien können alle zugehörigen Kriterien ausgewählt werden, wie in Abbildung 1 bei dem Hauptkriterium *Sicherheit* zu sehen ist. Einzelne Kriterien können ebenso separat über die jeweiligen Checkboxes an- und abgewählt werden.

Zu jeder Auswahl werden die Prozentzahlen der in dieser Arbeit als passend ermittelten Spiel-Design-Elemente angezeigt. Die Prozentzahl der Elemente ergibt sich aus der Anzahl passender Spiel-Design-Elemente im Verhältnis zu den ausgewählten Kriterien. Sie wird zusätzlich in einem runden Fortschrittsbalken visualisiert. Außerdem wird in einem Balkendiagramm angezeigt wie viele Kriterien die jeweiligen Elemente erfüllen. Bei Auswahl eines Spiel-Design-Elements wird unter Beispiele ein entsprechender erklärender Kurztext und ein grafisches Beispiel dazu angezeigt.

## Gamification in der Fertigung

### Eignung von Spiel-Design-Elementen anhand von Kriterien

Durch das Auswählen von Kriterien wird angezeigt, welche Spiel-Design-Elemente geeignet sind. Durch das Auswählen eines Spiel-Design-Elements wird ein Beispiel angezeigt.

**Sicherheit**

- Geringe bis keine Interaktion
- Ablenkung so gering wie möglich

**Persönlichkeit**

- Hohe Geduld und hohe Kompetenzität
- Spielertyp: Achiever
- Spielertyp: Explorer
- Spielertyp: Socialiser
- Spielertyp: Killer

**Motivation**

- Förderung des Kompetenzerlebens
- Förderung der Aufgabenbedeutsamkeit
- Erleben sozialer Eingebundenheit

**Generation**

- Generation X: 16 - 24 Jahre
- Generation Y: 25 - 40 Jahre
- Generation Z: über 40 Jahre

50%	67%	50%	50%
Punkte	Abzeichen	Fortschrittsanzeige	Bestenlisten
0%	34%	0%	0%
Level	Narrativ	Avatar	Nicht-Spieler-Charaktere

Anzahl der erfüllten Kriterien der jeweiligen Spiel-Design-Elemente

Spiel-Design-Elemente	Anzahl Kriterien
Punkte	3
Abzeichen	4
Fortschrittsanzeige	3
Bestenlisten	3
Level	0
Avatar	2
Narrativ	0
Nicht-Spieler-Charaktere	0

Beispiele

Abzeichen repräsentieren verschiedene definierte Leistungen, wie bspw.:

- Die erstmalige Ausführung einer Aktivität.
- Das Erreichen eines bestimmten Punktestands.
- Das Abschließen gewisser Aufgaben innerhalb einer vorgegebenen Zeitspanne.

Sie tragen zur Transparenz bei, indem sie aufzeigen, welche Aktionen möglich sind und welche Bedingungen an den Erhalt dieser geknüpft ist. In einem Wettbewerb können sie zum Vergleich verschiedener Nutzer dienen.

Michael Sailer. 2016. Die Wirkung von Gamification auf Motivation und Leistung. Springer Fachmedien Wiesbaden, Wiesbaden.

Abbildung 1: Beispielauswahl des interaktiven User-Interfaces

<sup>4</sup> <https://gitlab.reutlingen-universi-ty.de/muellerr/vv>

## 6 Zusammenfassung

In diesem Artikel werden Spiel-Design-Elemente der Gamification im Kontext der industriellen Fertigung betrachtet. Es wird erörtert welche Elemente sich unter spezifischen Voraussetzungen und Zielsetzungen eignen. Hierfür wurden Kriterien, die für eine Umsetzung relevant sein können zu thematisch zusammengehörigen Hauptkriterien zusammengefasst. Anhand der Kriterien lassen sich verschiedene Spiel-Design-Elemente festmachen, die sich besonders eignen. So lässt sich aus Tabelle 1 die Zuordnung der Spiel-Design-Elemente zu den Hauptkriterien entnehmen. Für eine individuelle Auswahl von Hauptkriterien und Kriterien wurde ein interaktives User-Interface erstellt. Dieses ermöglicht eine Empfehlung von Spiel-Design-Elementen zu spezifischen Anwendungsfällen.

## 7 Diskussion und Fazit

Es zeigt sich, dass die Wahl der Spiel-Design-Elemente einer Gamification-Anwendung in der industriellen Fertigung von verschiedenen Voraussetzungen und den angestrebten Zielen anhängig ist. Trotzdem sollte beachtet werden, dass eine Auswahl von Spiel-Design-Elementen für eine Gamification-Anwendung von weiteren Faktoren, als der bisher belegten Wirksamkeit der dargelegten Forschung abhängig ist. So ist diese ebenfalls stark von der Darstellung und Implementierung abhängig, wie in Kapitel 2 beschrieben ist. Weiter sollten allgemeine Bedingungen des Arbeitskontextes Beachtung finden und eine Anwendung an diese angepasst werden.

Bei der Nutzung des interaktiven User-Interfaces sollte beachtet werden, dass dieses sich nur auf die genannten Forschungsergebnisse bezieht. Es kann sinnvoll sein andere Spiel-Design-Elemente für eine Anwendung in Betracht zu ziehen, deren Vorteile mit der bisherigen Forschung noch nicht belegt wurden.

## Literaturverzeichnis

- [1] Polona Caserman, Kim A. Baumgartner, Stefan Göbel, and Oliver Korn. 2023. A best practice for gamification in large companies: An extensive study focusing inter-generational acceptance. *Multimed Tools Appl.* DOI: <https://doi.org/10.1007/s11042-023-16877-7>.
- [2] WonJoon Chung and Cliff S. Shin, Eds. 2017. *Advances in Affective and Pleasurable Design*. Advances in Intelligent Systems and Computing. Springer International Publishing, Cham. DOI: <https://doi.org/10.1007/978-3-319-41661-8>.
- [3] Edward L. Deci and Richard M. Ryan. 1993. Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik* 39 (1993) 2, S. 223-238. *Zeitschrift für Pädagogik* 39. DOI: <https://doi.org/10.25656/01:11173>.
- [4] Makenzie Dolly, Ashish Nimbarte, and Thorsten Wuest. 2024. The effects of gamification for manufacturing (GfM) on workers and production in industrial assembly. *Robotics and Computer-Integrated Manufacturing* 88, 102722. DOI: <https://doi.org/10.1016/j.rcim.2024.102722>.
- [5] Florian Kellner, Bernhard Lienland, and Maximilian Lukesch. 2020. *Produktionswirtschaft*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [6] Oliver Korn. 2012. Industrial playgrounds. In *Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, New York, NY, USA, 313–316. DOI: <https://doi.org/10.1145/2305484.2305539>.

- [7] Oliver Korn, Markus Funk, and Albrecht Schmidt. 2015. Towards a gamification of industrial production. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. ACM, New York, NY, USA, 84–93. DOI: <https://doi.org/10.1145/2774225.2774834>.
- [8] Oliver Korn, Peter Muschick, and Albrecht Schmidt. 2017. Gamification of Production? A Study on the Acceptance of Gamified Work Processes in the Automotive Industry. In *Advances in Affective and Pleasurable Design*, WonJoon Chung and Cliff S. Shin, Eds. Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, 433–445. DOI: [https://doi.org/10.1007/978-3-319-41661-8\\_42](https://doi.org/10.1007/978-3-319-41661-8_42).
- [9] Oliver Korn, Albrecht Schmidt, and Thomas Hörz. 2013. Augmented manufacturing. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, New York, NY, USA, 1–8. DOI: <https://doi.org/10.1145/2504335.2504356>.
- [10] Oliver Korn, Annika S. Schulz, and Belinda J. Hagley. 2022. Gamification: Grundlagen, Methoden und Anwendungsbeispiele. In *Digitale Lernwelten – Serious Games und Gamification*, Wolfgang Becker and Maren Metz, Eds. Springer Fachmedien Wiesbaden, Wiesbaden, 43–63. DOI: [https://doi.org/10.1007/978-3-658-35059-8\\_4](https://doi.org/10.1007/978-3-658-35059-8_4).
- [11] Janaki Kumar and Mario Herger. 2013. *Gamification at work. Designing engaging business software* (1. ed.). Interaction Design Foundation, Århus.
- [12] Richard Bartle. 1996. *Hearts, clubs, diamonds, spades: Players who suit MUDs*. Journal of MUD research.
- [13] Michael Sailer. 2016. *Die Wirkung von Gamification auf Motivation und Leistung*. Springer Fachmedien Wiesbaden, Wiesbaden.
- [14] Daniel Tolks and Michael Sailer. 2021. Gamification als didaktisches Mittel in der Hochschulbildung. In *Digitalisierung in Studium und Lehre gemeinsam gestalten*. Springer Fachmedien Wiesbaden, Wiesbaden, 515–532. DOI: [https://doi.org/10.1007/978-3-658-32849-8\\_29](https://doi.org/10.1007/978-3-658-32849-8_29).



©2024 Rebecca Müller. Lizenznehmer Hochschule Reutlingen, Deutschland. Dieser Artikel ist ein Open-Access-Artikel unter den Bedingungen und Konditionen der Creative Commons Attribution (CC BY)-Lizenz. <http://creativecommons.org/licenses/by/4.0/>

# Visuelle Kommunikation von Verkehrselementen in einer 3D-Fahrumgebung

Eduard Brendli

Hochschule Reutlingen  
Reutlingen, Deutschland

Eduard.Brendli@Student.Reutlingen-University.de

## Abstract

Die moderne Fahrzeugtechnologie hat in den letzten Jahren eine rasante Entwicklung erlebt, insbesondere durch die zunehmende Verbreitung von Fahrassistenzsystemen. Diese Systeme verbessern nicht nur die Sicherheit im Straßenverkehr, sondern entlasten auch den Fahrer bei der Navigation und Überwachung der Verkehrssituation. Anders als in einer realen Umgebung können Assistenzsystemen auch in einer Fahrsimulation von großem Gebrauch sein.

Im Vergleich zu den Assistenzsystemen in der realen Welt, treten in der virtuellen Welt andere Probleme auf. So stellt sich die Frage, wie passive Verkehrselemente in einer Fahrumgebung erkannt werden, nach welchen Regeln dies geschehen soll und wie die visuelle Kommunikation aussehen kann. Hierbei spielen Faktoren wie Entfernung, visuelle Darstellung sowie aktive und passive Eingriffe in das Fahrgeschehen eine Rolle. Das Ziel dieser Arbeit besteht darin, ein tieferes Verständnis für die Funktionsweise der

Erkennung von Verkehrszeichen in der realen Welt zu erlangen, den bisherigen Einsatz zu untersuchen und eine mögliche Implementierung eines visuellen Systems für Verkehrselemente in einer 3D-Fahrumgebung aufzuzeigen. Das Ergebnis soll abbilden, wie ein realitätsnahe 3D-Fahrsimulation das Gefühl erweckt, als wäre man selbst am Steuer eines Fahrzeuges.

## CCS Concepts

Computing methodologies -> Computer graphics / Modeling and simulation

## Keywords

Verkehrsschildererkennung; Verkehrsschilder; Visuell; Fahrsimulation; 3D; Straßenverkehr; Fahrzeug; Kamera; Assistenzsystem; Autonomes Fahren; Verkehrssituation

## 1 Einleitung

Die Verkehrszeichenerkennung ist ein bedeutender Bestandteil von modernen Fahrzeugen, der in den letzten Jahren aufgrund der zunehmenden Verbreitung von Fahrassistenzsystemen und der Entwicklung autonomer Fahrzeuge erheblich an Bedeutung gewonnen hat. Diese Systeme spielen eine entscheidende Rolle bei der Verbesserung der Sicherheit im Straßenverkehr und tragen dazu bei, die Fahrer bei der Navigation und der Überwachung der Verkehrssituation zu entlasten. Im Vergleich zu den Assistenzsystemen in der realen Welt treten in der virtuellen Welt jedoch unterschiedliche Probleme auf. So stellt sich die Frage, ob passive

---

Betreuer Hochschule: Prof. Dr. Uwe Kloos  
Hochschule Reutlingen  
Uwe.Kloos@Reutlingen-  
University.de

Informatics Inside Frühling 2024  
08. Mai 2024, Hochschule Reutlingen  
Copyright 2024 Eduard Brendli

Verkehrselemente durch das Analysieren des Bildausschnittes der Fahrumgebung erkannt werden sollen oder Verkehrselemente durch Lokalisieren des Fahrers und des Objektes zu einer Interaktion führen. Dabei ist dies eine zentrale Herausforderung im Zusammenhang mit der Erkennung und visuellen Darstellung von Verkehrsschildern in einer 3D-Fahrumgebung wie der von OpenDS. Die Integration dieser passiver Verkehrselemente in einer virtuellen Fahrsimulation eröffnet dabei neue Möglichkeiten für die Fahrzeugsteuerung und die Interaktion mit der Umgebung. Dieses Paper beleuchtet diese Bedeutung und die Herausforderungen der Schilderkennung in der realen als auch in einer digitalen Fahrumgebung und zeigt dessen visuelle Darstellung auf.

## 2 Forschungsfragen

Dieses Kapitel zeigt die Fragen, welche mit Hilfe der Arbeit beantwortet werden sollen. Die folgende Tabelle 1 zeigt die herausgearbeiteten Forschungsfragen auf.

**Tabelle 1. Forschungsfragen**

ID	Frage
RQ1	Wie werden Verkehrselemente in der realen Welt in einem Fahrzeug erkannt und dargestellt?
RQ2	Welche Herausforderungen muss eine 3D-Fahrsimulation zur Erkennung von Verkehrszeichen bewältigen?

## 3 Motivation

Die Verkehrszeichenerkennung ist ein wichtiger Bestandteil des zukünftigen autonomen Fahrens und hat das Potenzial, die Straßensicherheit zu verbessern und den Fahrer zu entlasten. Es ist daher wichtig, die Funktionsweise, die Herausforderungen und das Potential der Verkehrszeichenerkennung zu verstehen. Daraus können Schlüsse für die Integration in einer Fahrumgebung ausfindig

gemacht werden. Die Beantwortung der Forschungsfragen RQ1 und RQ2 sollen dabei helfen, ein besseres Verständnis für das System zu gewinnen, Herausforderungen zu identifizieren sowie eine mögliche Integration in einer Fahrumgebung zu ermöglichen.

## 4 Methodik

In diesem Kapitel wird das Konzept und die Struktur der Arbeit erläutert. Der Schwerpunkt liegt auf der Literaturrecherche, die anhand der Forschungsfragen durchgeführt wird. Hierbei werden unterschiedliche Suchbegriffe verwendet, die aus den Forschungsfragen abgeleitet sind. Tabelle 2 zeigt eine Verknüpfung zwischen den Suchbegriffen und den Forschungsfragen.

**Tabelle 2. Suchterme**

Forschungsfrage	Suchterm
RQ1	Traffic sign recognition AND Car AND function AND Visualization
RQ2	Traffic sign recognition AND 3D Driving Simulation AND Visualization AND Challenges

Für die Literaturrecherche werden die aufgelisteten Suchterme mithilfe von folgenden Suchquellen gesucht:

**Tabelle 3. Suchquellen**

Quelle	Link
ACM	<a href="https://dl.acm.org/">https://dl.acm.org/</a>
IEEE	<a href="https://ieeexplore.ieee.org/Xplore/home.jsp">https://ieeexplore.ieee.org/Xplore/home.jsp</a>
Google Scholar	<a href="https://scholar.google.com/">https://scholar.google.com/</a>
Science Direct	<a href="https://www.sciencedirect.com/">https://www.sciencedirect.com/</a>
Springer	<a href="https://link.springer.com/">https://link.springer.com/</a>

## Visuelle Kommunikation von Verkehrselementen in einer 3D-Fahrumgebung

Nachfolgend wurden zudem gewisse Kriterien zur besseren Einordnung definiert.

- Ausschließlich Veröffentlichungen zu Verkehrszeichenerkennung in Bezug auf Fahrzeuge und Fahrsimulationen
- Veröffentlichungsdatum: 2012-heute

Weiterführend beschäftigt sich Kapitel 5 mit dem Stand der Wissenschaft, indem Erkenntnisse aus der Literaturrecherche beschrieben werden. In Kapitel 6 werden Grundlagen der Technik sowie die Funktionsweise erläutert. Im siebten Kapitel werden die existierenden Herausforderungen und Hürden des Systems nähergebracht und die Integration in einer Fahrumgebung begutachtet. Anschließend wird im Kapitel 8 das Fazit schlussfolgernd zusammengefasst und die möglichen Innovationen des Systems unter Kapitel 9 niedergeschrieben.

### 5 Stand der Wissenschaft

In den vergangenen Jahren hat sich die Technologie der Verkehrszeichenerkennung erheblich weiterentwickelt und ist zu einem dynamischen Forschungsfeld geworden. Die Anwendung kamerabasierter Verkehrszeichenerkennungssysteme in Fahrzeugen wird in zahlreichen wissenschaftlichen Publikationen untersucht. Diese Technologie beruht auf der Bilderkennung, wobei Verkehrsschilder mithilfe von Kamerabildern und fortschrittlichen Ansätzen des maschinellen Lernens und Deep Learning erfasst und klassifiziert werden können [1]. Studien wie die von Yu, Dingjun et al. [4] erläutern das Vorgehen einer solchen Bilderkennung und Kategorisierung, insbesondere im Zusammenhang mit neuronalen Netzen.

In früheren Studien wurden auch die Erfassung von optionalen Schilderergänzungen beziehungsweise Zusatzzeichen wie zeitbezogen, nass oder richtungsbestimmend sowie deren Syntax untersucht, um zukünftige Missverständnisse bei den Systemen zu ver-

meiden [2]. Weiterhin wurden Untersuchungen durchgeführt, um zu verstehen, wie verschiedene Witterungsbedingungen, wie Verdeckung und Dunkelheit, die Erkennung von Verkehrsschildern beeinflussen können [3]. Eine weitere Forschungsrichtung beschäftigt sich mit der Verkehrszeichenerkennung mittels dynamischer CNNs (Convolutional Neural Networks) und vergleicht deren Leistungsfähigkeit anhand unterschiedlicher Methodiken, insbesondere im Kontext von Schildern mit hoher Geschwindigkeitsbeschränkung [5]. Die Studie der Universität Zagreb in Kroatien untersuchte zudem die Verkehrszeichenerkennung von Fahrzeugen verschiedener Hersteller [7].

Die kontinuierliche Forschung an der Lernfähigkeit neuronaler Netze ist ebenfalls von Bedeutung. Unterschiedliche Herangehensweisen und Trainingsmethoden werden erforscht, um wiederkehrende Muster und Veränderungen im Straßenverkehr zu erkennen und zu kategorisieren [6]. Trotz einiger Herausforderungen zeigen diese Forschungsergebnisse, dass eine immer präzisere Verkehrszeichenerkennung in naher Zukunft möglich ist. Dies legt nahe, dass weitere Fortschritte bei der Entwicklung autonomer Fahrzeuge zu erwarten sind.

### 6 Grundlagen

In diesem Kapitel wird grundlegendes zur Verkehrszeichenerkennung, dessen benötigte Komponente und der Funktionsweise erläutert. Anschließend wird die 3D Fahrumgebung genauer beschrieben. Dabei handelt es sich um eine Software, welche an der Hochschule Reutlingen verwendet und weiterentwickelt wird. Sie dient zur Simulation einer Fahrumgebung.

#### 6.1 3D-Fahrumgebung

Im Rahmen der Forschung umfasst die visuelle Implementierung von Verkehrsschildern einen 3D-Fahrsimulator. Für diesen Zweck wird die Simulationssoftware OpenDS verwendet, die an der Hochschule Reutlingen für Forschungszwecke zur Verfügung steht.

OpenDS ist ein fortschrittlicher Fahr Simulator, dessen Kernkomponenten vom deutschen Forschungszentrum für künstliche Intelligenz (DFKI) entwickelt wurden. Der Quellcode dieser Software ist unter der Open-Source-Lizenz GNU GPL frei verfügbar und in der Programmiersprache Java geschrieben. Diese Offenheit ermöglicht es, den Fahr Simulator nach Bedarf zu erweitern, indem zusätzliche Verkehrselemente in die vorhandene Umgebung integriert werden können.

OpenDS bietet eine Vielzahl von Funktionen und unterstützt verschiedene Simulationsszenarien, darunter Verkehrssimulation, Ampelschaltungen, Wettersimulationen, Eye-Tracking sowie Analysetools für das Fahrverhalten. Darüber hinaus umfasst die Software verschiedene vorgefertigte Auto- und Straßenmodelle sowie Fahraufgaben, die es den Forschenden ermöglichen, realistische Fahrsituationen zu simulieren und zu untersuchen.

## 6.2 Verkehrszeichenerkennung

Die Nutzung eines in der Frontscheibe integrierten Kamerasystems bietet die Möglichkeit, Verkehrszeichen während der Fahrt zu erkennen und dem Fahrer in Echtzeit anzuzeigen. Dadurch kann der Fahrer entlastet werden und potenzielle Gefahren durch unangepasste Geschwindigkeiten vermieden

werden. Die Verkehrszeichenerkennung basiert dabei auf dem Einsatz von maschinellem Lernen, einem Zweig der künstlichen Intelligenz. Je nach den Anforderungen und dem Verwendungszweck können unterschiedliche Ansätze verwendet werden.

Grundlegend funktioniert die Verkehrszeichenerkennung durch die kontinuierliche Erfassung von Bildern mit einer Rate von mehreren Frames pro Sekunde. Für jedes Einzelbild werden mithilfe von Algorithmen des maschinellen Lernens auffällige Merkmale identifiziert, die auf das Vorhandensein von Verkehrsschildern hinweisen. Die Grundlage dieser Merkmalsidentifikation bildet ein Datensatz, der die Verkehrsschilder enthält und mit den entsprechenden Kategorien versehen ist. Ein vereinfachtes Beispiel dieses Vorgehens zeigt Abbildung 1.

Dieser Prozess umfasst mehrere Schritte, darunter die Vorverarbeitung der Bilddaten, die Merkmalsextraktion und schließlich die Klassifizierung der erkannten Merkmale anhand des vorliegenden Datensatzes. Durch kontinuierliches Training und Optimierung des Modells kann die Genauigkeit und Zuverlässigkeit der Verkehrszeichenerkennung verbessert werden, um eine präzise und zuverlässige Erfassung der Verkehrsschilder in verschiedenen Verkehrssituationen zu gewährleisten.

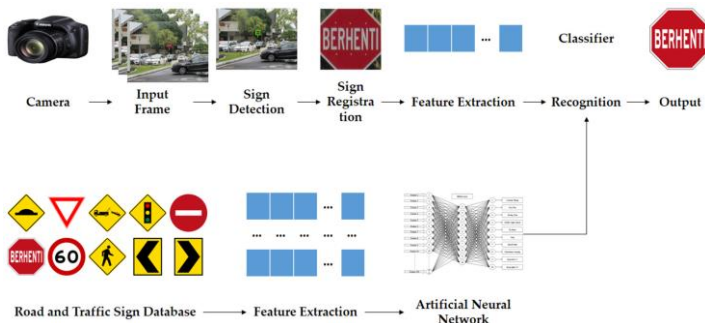


Abbildung 1: Funktionsweise einer Verkehrszeichenerkennung

Quelle: [18]

## 7 Integration der visuellen Darstellung

Die Integration von passiven Verkehrselementen beziehungsweise Straßenschildern kann in mehreren Schritten gegliedert werden. Als erstes wird auf die allgemeine Visuelle Kommunikation in der realen Welt eingegangen. Anschließend wird die Fahrsimulation aufgegriffen und thematisiert. Bestandteil dessen ist die Integration mit Berücksichtigung der Verkehrszeichen sowie den Rahmenbedingungen in der Virtuellen Umgebung. Im Anschluss folgt die Integration beziehungsweise Implementierung der Visualisierung von Straßenschildern in der Fahrsimulation OpenDS.

### 7.1 Visuelle Kommunikation

Die visuellen Kommunikationsmöglichkeiten von Verkehrsschildern in Fahrzeugen spielen eine entscheidende Rolle für die Sicherheit und Effizienz des Straßenverkehrs. Diese Kommunikation erfolgt hauptsächlich über das Anzeigen und Interpretieren von Verkehrszeichen auf dem Armaturenbrett, dem Infotainmentdisplay oder dem Head-Up-Display des Fahrzeugs. Die dargestellten Informationen können verschiedene Arten von Verkehrszeichen umfassen, darunter Geschwindigkeitsbegrenzungen, Verkehrshinweise, Warnungen vor Gefahren, Vorfahrtsregelungen und andere verkehrsrelevante Informationen. Diese werden in der Regel durch Symbole oder Textanzeigen visualisiert, die dem Fahrer sofort erkennbare und verständliche Informationen liefern. Die visuelle Darstellung von Verkehrsschildern zielt dabei darauf ab, dem Fahrer eine klare Orientierung über die Verkehrssituation zu geben und ihn bei der sicheren und effizienten Navigation zu unterstützen.

### 7.2 Fahrsimulation

Um die Integration von Verkehrselementen in den OpenDS Fahrsimulator zu ermöglichen, war es erforderlich, eine umfassende Recherche durchzuführen. Dabei mussten

verschiedene Aspekte berücksichtigt und analysiert werden, darunter mögliche Visualisierungen von Verkehrselementen sowie die Struktur und der Aufbau der Fahrsimulation selbst. Zunächst galt es, potenzielle Visualisierungen von Verkehrselementen zu untersuchen und zu bewerten [9].

Dazu gehörte die Analyse verschiedener Ansätze zur Darstellung von Verkehrsschildern, Markierungen auf der Straße und anderer relevanter Elemente. Es wurden verschiedene Optionen in Betracht gezogen, um sicherzustellen, dass die visuelle Darstellung der Verkehrselemente sowohl realistisch als auch benutzerfreundlich ist. Parallel dazu wurde die Fahrsimulation von OpenDS genauer betrachtet, insbesondere deren Struktur und Aufbau. Dabei wurde untersucht, wie Szenen und Umgebungen innerhalb des Simulators organisiert sind und wie die Integration neuer Elemente erfolgen kann. Durch diese Analyse konnte festgestellt werden, wie Verkehrselemente in die bestehende Infrastruktur des Fahrsimulators integriert werden können.

### 7.3 DrivingTask

Eine wichtige Komponente im Zusammenhang mit den Schildern ist die Datei "scene.xml", die sich innerhalb des jeweiligen DrivingTask-Ordners befindet. Eine DrivingTask bildet dabei eine vollständige Szene ab. Verglichen werden kann dies mit unterschiedlichen Spielwelten, wobei jede Spielwelt eine unterschiedliche DrivingTask widerspiegelt. Vor jedem Start muss hierfür eine DrivingTask ausgewählt werden, welche im Anschluss geladen wird. Geladen werden die Daten aus der „scene.xml“ wodurch diese Datei die zentrale Schnittstelle zur Konfiguration und Definition der einzelnen Objekte und Einstellungen, die in einem bestimmten Szenario vorkommen, darstellt. Sie fungiert dabei als eine Art Verzeichnis, alle erforderlichen Informationen für die Erstellung und Darstellung eines Szenarios enthält. Dabei werden sowohl die vi-

suellen als auch die funktionalen Aspekte berücksichtigt. Die einzelnen Objekte, wie Straßen, Gebäude, Fahrzeuge, Verkehrsschilder und andere Elemente, werden hier definiert und in der Szene positioniert. Ein Beispiel von Objekten in der Fahrumgebung ist in Abbildung 2 einzusehen.



**Abbildung 2: Schilderobjekte in der 3D-Fahrumgebung OpenDS des Future Mobility Lab**

Quelle: Eigene Aufnahme

Die Konfiguration in der "scene.xml" umfasst eine Vielzahl von Parametern und Einstellungen, die die Eigenschaften und das Verhalten der Szenenobjekte bestimmen. Dazu gehören beispielsweise Position, Größe, Rotation, Farbe, Textur, Interaktion mit anderen Objekten und vieles mehr. Diese Einstellungen sind entscheidend für die korrekte Darstellung und das realistische Verhalten der Elemente während der Simulation.

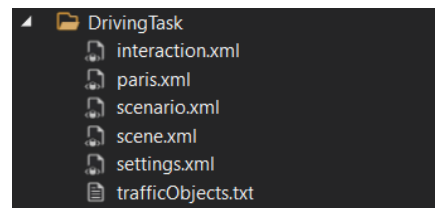
## 7.4 Rahmenbedingung

Um die Darstellungen dem Fahrer zu übermitteln, sollen die geltenden Verkehrselemente im Sichtfeld des Fahrers sein. Dazu muss eine Regel integriert werden, die aussagt, in welchem Radius das aktuell gültige Verkehrselement visuell angezeigt werden soll. Im Realen Umfeld werden dabei Schilder erkannt, die ausschließlich im Sichtfeld des Fahrers beziehungsweise der integrierten Kamera liegen. Im Virtuellen Umfeld ist dies schwieriger zu Implementieren. Schilder, welche in einer 3D-Fahrumgebung als Objekte angesehen werden, verfügen meistens

über Koordinaten, welche ihre Position bestimmen. Auch das fahrende Fahrzeug in der Simulation befindet sich auf gewissen Koordinaten. So kann die Rahmenbedingung mithilfe der Werte des Schildes und des Fahrzeuges definiert werden.

## 7.5 Vorkehrungen mit Skript

Da die Architektur von OpenDS sich primär auf die DrivingTasks bezieht, wird eine dynamische Implementierung der vorhandenen Schilder in der Fahrumgebung benötigt. Hierfür wurde ein Skript erstellt, welches die benötigte DrivingTask anschaut, es die enthaltenen Verkehrselemente identifiziert und in eine Datei namens „trafficObjects“ schreibt. Dies ist notwendig, da anderweitig keine Kommunikation zwischen der DrivingTask, den Objekten und die visuelle Anzeige der geltenden Verkehrsschilder entstehen kann. Eine vereinfachte Darstellung des Aufbaues einer DrivingTask mit integrierten trafficObjects ist in der nachfolgenden Abbildung 3 ersichtlich.



**Abbildung 3: Aufbau einer DrivingTask von OpenDS mit trafficObjects**

Quelle: Eigene Aufnahme

## 7.6 Integration

Die Integration folgte mit dem Einbetten von Verkehrsschildern als Objekte, sodass diese im Blickfeld zu sehen sind, jedoch die Sicht auf die Straße nicht behindern. Dabei wurde die Verwendung eines TrafficSignLocator verwendet, in welchem die vorhandenen und integrierten Verkehrsschilder integriert sind. In der Folgenden Abbildung 4 ist diese Visuelle Darstellung eines Geschwindigkeitsbegrenzungsschildes zu sehen.



**Abbildung 4: Eigene Umsetzung von Verkehrszeichen in der 3D-Fahrumgebung OpenDS des Future Mobility Lab**

**Quelle: Eigene Aufnahme**

Dieses stimmt mit den aktuellen Begebenheiten der Simulation beziehungsweise der befahrenden Straße überein. Die Auswahl und Platzierung der Umgebungsinformationen erfolgte strategisch an der rechten unteren Seite, um eine effektive Kommunikation mit dem Fahrer zu ermöglichen. Dieser kann dadurch zum Beispiel die Geschwindigkeit mit dem Tachowert gegenüberstellen und bei Überhöhung beispielweise drosseln.

## 8 Fazit

In den letzten Jahren hat das Thema der Verkehrszeichenerkennung erhebliche Fortschritte gemacht und ist zu einem aktiven Forschungsfeld gewachsen. Diese Fortschritte basieren auf fortschrittlichen Methoden der Bilderkennung, insbesondere auf dem Einsatz von maschinellem Lernen und Deep Learning. Die kontinuierliche Weiterentwicklung von Verkehrszeichenerkennungssystemen ist von entscheidender Bedeutung für die Entwicklung intelligenter Verkehrssysteme und autonomer Fahrzeuge. Sie trägt dazu bei, die Zuverlässigkeit und Genauigkeit dieser Systeme signifikant zu verbessern, was wiederum die Sicherheit im Straßenverkehr erhöht.

Ein weiterer wichtiger Aspekt der Forschung auf diesem Gebiet ist die Integration von Verkehrsschildererkennung in realistischen Fahrumgebungen. Dies umfasst sowohl die

physische reale Welt als auch virtuelle Simulationen. Durch die Schaffung realitätsnaher Fahrumgebungen können Forscher unter kontrollierten Bedingungen verschiedene Situationen und Szenarien testen und evaluieren. Dies ermöglicht es, die Leistung von Verkehrszeichenerkennungen unter realen Bedingungen zu überprüfen, ohne die Sicherheit oder Integrität realer Verkehrssituationen zu gefährden.

Darüber hinaus bieten virtuelle Simulationen den Vorteil, dass sie flexibel anpassbar sind und verschiedene Szenarien reproduzierbar machen, was wiederum eine präzise Analyse und Optimierung der Systeme ermöglicht. Somit spielen realistische Fahrumgebungen eine entscheidende Rolle bei der Weiterentwicklung und Validierung von Verkehrszeichenerkennungssystemen und tragen maßgeblich zur Sicherheit und Effizienz des Straßenverkehrs bei. Auch dieses Paper, und die dafür unternommenen Forschung und Entwicklung, trägt zu der kontinuierlichen Weiterentwicklung von Fahrsimulationen bei.

## 9 Innovationspotential

Die Verkehrszeichenerkennung hat in den letzten Jahren einen bedeutenden Fortschritt erfahren. Die Kamerabasierte Verkehrszeichenerkennung wird aktiv in der Forschung untersucht und ist ein bedeutendes Anliegen bei der Entwicklung autonomer Fahrzeuge. Der neueste Ansatz im Bereich der Verkehrszeichenerkennung in Fahrzeugen beinhaltet die Erweiterte Erkennung von einer Vielzahl von Schildern, einschließlich Geschwindigkeitsschildern, Stoppschildern, Warn- und Hinweisschildern sowie Ampeln und deren Grün- und Rotphasen. Durch die Einführung dieser Technologie kann das autonome Fahren verbessert werden und das autonome Befahren von Städten in Zukunft ermöglicht werden.

Darüber hinaus ermöglichen realitätsnahe Fahrumgebungen Entwicklern und Forschern, neue Fahrzeugtechnologien zu testen

und zu validieren, indem sie realistische Testumgebungen für autonome Fahrzeuge und Fahrerassistenzsysteme bereitstellen [10]. Durch die Simulation von realen Straßenbedingungen, Verkehrsdichte und Wetterverhältnissen können Entwickler die Leistung ihrer Technologien bewerten und verbessern. Durch die Fortschritte in der KI können Fahrsimulationen in naher Zukunft zudem auch intelligentes Verhalten von virtuellen Verkehrsteilnehmern wie Fußgängern, Radfahrern und anderen Fahrzeugen simulieren. Dies ermöglicht es den Benutzern, reale Verkehrssituationen zu erleben und ihre Reaktionsfähigkeiten zu verbessern [11]. Insgesamt tragen diese Fortschritte dazu bei, Fahrsimulationen zu einem wichtigen Werkzeug für die Ausbildung von Fahrern, die Entwicklung autonomer Fahrzeugtechnologien und die Erforschung von Verkehrsicherheitsmaßnahmen zu machen.

## Literaturverzeichnis

- [1] Yihan Lai. 2022. Traffic Sign Recognition Based on Deep Learning Technique. In Proceedings of the 6th International Conference on Information System and Data Mining (ICISDM '22). Association for Computing Machinery, New York, NY, USA, 62–68. <https://doi.org/10.1145/3546157.3546167>
- [2] T. Wenzel, S. Brueggert and J. Denzler, "Additional Traffic Sign Detection -- A Comparative Study," 2015 IEEE 18<sup>th</sup> International Conference on Intelligent Transportation Systems, 2015, pp. 794-799, doi: 10.1109/ITSC.2015.134.
- [3] Chuwen Song. 2020. Assessing the Impact of Various Image Corruptions on the Recognition of Traffic Signs by Machine Learning. In Proceedings of the 12th International Conference on Computer Modeling and Simulation (ICCMS '20). Association for Computing Machinery, New York, NY, USA, 47–50. <https://doi.org/10.1145/3408066.3408071>
- [4] Yu, Dingjun & Wang, Hanli & Chen, Peiqiu & Wei, Zhihua. Mixed Pooling for Convolutional Neural Networks, 2014, Page 364-375. doi: 10.1007/978-3-319-11740-9\_34.
- [5] Babić, Darko, Dario Babić, Mario Fiolić, and Željko Šarić. 2021. "Analysis of Market-Ready Traffic Sign Recognition Systems in Cars: A Test Field Study" *Energies* 14, no. 12: 3697. <https://doi.org/10.3390/en14123697>
- [6] Xiucui Guo, Changxuan Zhao, and Yaodong Wang. 2019. Traffic Sign Recognition Based on Joint Convolutional Neural Network Model. In Proceedings of the 2<sup>nd</sup> International Conference on Big Data Technologies (ICBDT '19). Association for Computing Machinery, New York, NY, USA, 200–203. <https://doi.org/10.1145/3358528.3358555>
- [7] Reif, K. (2014). Fahrerassistenzsysteme. In: *Automobilelektronik*. ATZ/MTZ-Fachbuch. Springer Vieweg, Wiesbaden. [https://doi.org/10.1007/978-3-658-05048-1\\_12](https://doi.org/10.1007/978-3-658-05048-1_12)
- [8] Islam, Kh Tohidul, and Ram Gopal Raj. 2017. "Real-Time (Vision-Based) Road Sign Recognition Using an Artificial Neural Network" *Sensors* 17, no. 4: 853. <https://doi.org/10.3390/s17040853>
- [9] Peter Foytik, Craig Jordan, and R. Michael Robinson. 2017. Exploring simulation based dynamic traffic assignment with a large-scale microscopic traffic simulation model. In Proceedings of the 50th Annual Simulation Symposium (ANSS '17). Society for Computer Simulation International, San Diego, CA, USA, Article 11, 1–12.
- [10] Andreas Riegler, Andreas Riener, and Gerald Ostermayer. 2023. From Me to We: Combining Driving Simulation

## Visuelle Kommunikation von Verkehrselementen in einer 3D-Fahrumsgebung

and Traffic Simulation for Holistic Usability and Safety Research. In Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '23 Adjunct). Association for Computing Machinery, New York, NY, USA, 294–296. <https://doi.org/10.1145/3581961.3609840>

[11] Bihao Wang, Marwa Mahmoud, Javier Echevarría Cuesta, Hannah Close,

Quentin Stafford-Fraser, and Peter Robinson. 2018. Enhanced Traffic Simulation for Improved Realism in Driving Simulators. In Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18). Association for Computing Machinery, New York, NY, USA, 170–174.

<https://doi.org/10.1145/3239092.3265962>



©2024 Eduard Brendli. Lizenznehmer Hochschule Reutlingen, Deutschland. Dieser Artikel ist ein Open-Access-Artikel unter den Bedingungen und Konditionen der Creative Commons Attribution (CC BY)-Lizenz. <http://creativecommons.org/licenses/by/4.0/>



# Autorenverzeichnis

Brendli, E., 71

Egler, M., 41

Garkusha, B., 21

Keck, M., 31

Lindeboom, H., 1

Müller, R., 61

Schrüfer, K., 11

Stych, M., 51

---

Hochschule Reutlingen  
Reutlingen University  
Fakultät Informatik  
Human-Centered Computing  
Alteburgstraße 150  
D-72762 Reutlingen

Telefon: +49 7121 / 271-4002  
Telefax: +49 7121 / 271-4032

E-Mail: [informatics.inside@reutlingen-university.de](mailto:informatics.inside@reutlingen-university.de)  
Website: <https://infoinside.reutlingen-university.de>