



Entwicklung einer Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion

Von der Fakultät für Informatik, Wirtschafts- und Rechtswissenschaften
der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades
und Titels

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

angenommene Dissertation

von Herrn Marvin Leon Norda,
geboren am 05.02.1994 in Norden

Tag der Disputation: 23.09.2025

Erstgutachter: Prof. Dr.-Ing. Axel Hahn

Zweitgutachter: Prof. Dr.-Ing. Sven Carsten Lange

Abstract

As the most natural form of communication, speech offers the potential to make collaboration between humans and machines in production processes more efficient. Although the control of technical systems in the consumer sector such as vehicles, smartphones and home automation systems has been available as a standard HMI modality for years, the control of machines by voice is currently still the exception. For the targeted use of voice control as an industrial HMI, objective influencing factors and quantitative parameters derived from them are required in order to predict the efficiency potential. In this thesis, a generally valid method for evaluating the efficiency of voice control as an industrial HMI in production is developed based on the three influencing factors of menu complexity, type and number of walking paths and the number of machines to be operated simultaneously. This enables manufacturers of production machines and end users to evaluate their existing or newly introduced production processes with regard to the efficiency of voice control as a supplement or replacement for a conventional HMI. The systematic identification of application scenarios that benefit from voice control in terms of efficiency enables production companies to make a confident investment decision.

Kurzfassung

Sprache bietet als natürlichste Form der Kommunikation das Potenzial, die Kollaboration zwischen Mensch und Maschine in Produktionsabläufen effizienter zu gestalten. Obwohl die Steuerung von technischen Systemen im Consumer-Bereich wie bei Fahrzeugen, Smartphones und Hausautomatisierungen bereits seit Jahren standardmäßig als HMI-Modalität zur Verfügung steht, ist die Steuerung von Maschinen per Sprache aktuell noch die Ausnahme. Für den zielgerichteten Einsatz einer Sprachsteuerung als industrielles HMI werden objektive Einflussfaktoren und davon abgeleitete quantitative Parameter benötigt, um das Effizienzpotenzial zu prognostizieren. In dieser Arbeit wird auf Basis der drei Einflussfaktoren Menükomplexität, Art und Anzahl von Laufwegen sowie der Anzahl simultan zu bedienender Maschinen eine allgemeingültige Methode zur Evaluation der Effizienz einer Sprachsteuerung als industrielles HMI in der Produktion entwickelt. Diese ermöglicht es Herstellern von Produktionsmaschinen und Endanwendern, ihre vorhandenen oder neu einzuführenden Produktionsabläufe hinsichtlich der Effizienz einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen HMI zu bewerten. Durch die systematische Identifikation von Anwendungsszenarien, die von einer Sprachsteuerung hinsichtlich ihrer Effizienz profitieren, werden Produktionsunternehmen in die Lage versetzt, sicher eine Investitionsentscheidung treffen zu können.

Inhalt

Abkürzungen	IX
1 Einleitung	1
1.1 Motivation	2
1.2 Forschungslücke	3
1.3 Ziele und Forschungsfragen.....	4
1.4 Anforderungsliste (Lastenheft).....	6
1.5 Aufbau der Arbeit	8
2 Stand der Forschung	10
2.1 Zusammenhang zwischen Haupt-, Neben-, Rüst- und Brachzeiten sowie Produktionsprozessen und Produktionsabläufen	11
2.2 Konventionelle haptisch-visuelle Human-Machine Interfaces.....	13
2.3 Neuartige nutzerzentrierte Human-Machine Interfaces.....	17
2.3.1 Gestensteuerung / Gesture-based Interaction.....	18
2.3.2 Eye Tracking / Gaze Control / Eye Movement Control	19
2.3.3 Brain-Computer Interface	21
2.3.4 Sprachsteuerung / Voice Control.....	22
2.4 Normen zur Ergonomie der Mensch-System-Interaktion, Ergonomischen Gestaltungsgrundsätzen und Evaluationsprinzipien	26
2.4.1 Definition von Gebrauchstauglichkeit, Effektivität, Effizienz und Zufriedenstellung im Kontext von Human-Machine Interfaces.....	28
2.4.2 Gestaltung von konventionelle haptisch-visuellen Human-Machine Interfaces als Ausgangsbasis	29
2.4.3 Gestaltung von Spracheingaben eines nutzerzentrierten Human-Machine Interface	30
2.4.4 Gestaltung von nutzerzentrierten Human-Machine Interfaces zur Steigerung der Gebrauchstauglichkeit.....	31
2.4.5 Prozesse zur Gestaltung von nutzerzentrierten Human-Machine Interfaces.....	32
2.4.6 Ergonomische Gestaltungsgrundsätze auf Basis der Sicherheit von Produktionsmaschinen	33
2.4.7 Evaluation von Human-Machine Interfaces in industriellen Produktionsabläufen.....	33

2.5	Einflussfaktoren auf die Gebrauchstauglichkeit industrieller Human-Machine Interfaces	34
2.6	Effizienzpotenzial einer industriellen Sprachsteuerung in der Produktion	39
2.7	Etablierte Konzepte, Methoden und Werkzeuge zur Evaluation der Effizienz von Produktionsabläufen	41
2.7.1	Lean Production	42
2.7.2	Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung e. V. (REFA)	43
2.7.3	Root Cause Analysis (RCA)	44
2.7.4	Methods-Time Measurement (MTM)	45
2.7.5	Think Aloud Method & NASA Task Load Index (NASA-TLX)	46
3	Entwicklung einer Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion.....	49
3.1	Vorgehensweise bei der Entwicklung	49
3.2	Die entwickelte Methode	50
3.2.1	Anzahl der simultan zu bedienenden Maschinen (Mehrmaschinenbedienung)	52
3.2.2	Komplexität der Menüstruktur im konventionellen Graphical User Interface	54
3.2.3	Eingabezeiten im konventionellen Graphical User Interface	54
3.2.4	Messung/Berechnung von Laufwegzeiten im Produktionsablauf	55
3.2.5	Berechnung der Eingabedauer einer Sprachsteuerung über Zeitäquivalente	56
3.2.6	Kategorisierung von Laufwegen	57
3.2.7	Berechnung des Effizienzpotenzials	57
3.2.8	Template zur Anwendung in der industriellen Praxis	58
4	Anwendungsszenarien, Interaktionsprinzipien und Einflussfaktoren für eine Sprachsteuerung	61
4.1	Steigerung der Produktivität von Produktionsabläufen durch effiziente Sprachsteuerung	62
4.1.1	Produktionsautomaten	63
4.1.2	Roboter	64
4.1.3	Werkzeugmaschinen	65
4.1.4	Produktionsanlagen	67

4.2	Einflussfaktoren auf die Effizienz einer Sprachsteuerung in industriellen Produktionsabläufen	69
5	Quantifizierung des Einflussfaktors »Komplexität von Menüstrukturen«	71
5.1	Die Wizard-of-Oz-Methode im Vergleich zur Implementierung einer Sprachsteuerung	72
5.2	Implementierung einer Sprach- und Touchsteuerung in eine Produktionsmaschine.	74
5.2.1	Sprachsteuerung	76
5.2.2	Touchsteuerung	77
5.3	Design of Experiment.....	79
5.4	Ergebnisse.....	87
5.4.1	Statistik.....	87
5.4.2	Einfluss der Menüebenen auf die zeitliche Effizienz.....	88
5.4.3	Einfluss der Anzahl an Interaktionen auf die zeitliche Effizienz.....	90
5.4.4	Lerneffekte.....	92
5.4.5	Subjektive Einschätzung der Probanden (Fragebogen)	93
5.4.6	Zeitäquivalente als quantifizierte Effizienzfaktoren	94
6	Quantifizierung der Einflussfaktoren »Laufwege« und »Mehrmaschinenbedienung«	97
6.1	Eigenschaften von Laufwegen	97
6.2	Eigenschaften des Produktionsablaufs.....	99
6.3	Teilautomatisierte Produktionsabläufe mit Laufwegen in der Produktion	100
7	Validierung der entwickelten Methode.....	110
7.1	Design of Experiment Probandenstudie Validierung.....	111
7.2	Komplexität von Menüstrukturen	114
7.3	Laufwege und Mehrmaschinenbedienung	118
7.4	Lerneffekte.....	121
7.5	Abweichungen zwischen der entwickelten Methode und den real gemessenen Interaktionszeiten (Eingabe-/Laufwegzeiten)	125
8	Diskussion der Ergebnisse im Kontext der Forschungsfragen	127
8.1	Methode zur Evaluation der Effizienz einer Sprachsteuerung	127
8.2	Industrielle Produktionsabläufe mit Effizienzpotenzial durch Sprachsteuerung	128

8.3	Einflussfaktoren auf die Effizienz einer Sprachsteuerung	129
8.4	Parameter zur Quantifizierung des Effizienzpotenzials.....	131
8.5	Entwicklung von Produktionsabläufen zur Nutzung des Effizienzpotenzials von Sprachsteuerung.....	132
9	Fazit und Ausblick	134
	Literatur	XI
	Abschließende Erklärung	XX

Abkürzungen

ALS	Amyotrophe Lateralsklerose
ANOVA	Analysis of Variance
API	Application Programming Interface
ASR	Automatic Speech Recognition
BCI	Brain-Computer Interaction
BDE	Betriebsdatenerfassung
BEP	Break-Even-Point
CNC	Computerized Numerical Control
CPPS	cyberphysikalische Produktionssysteme
DoE	Design of Experiment
EEG	Elektroenzephalografie
ERP	Enterprise Resource Planning
GBI	Gesture-based Interaction
GGs-AR	Gaze-Gesture-Speech Augmented Reality
HCI	Human-Computer Interface
HRC	Human-Robot Collaboration
HMI	Human-Machine Interface
IPC	Industrie-PC
KI	Künstliche Intelligenz
LSTM	Long Short-Term Memory
MES	Manufacturing Execution System
MTI	Mensch Technik Interaktion
MMI	Multi Media Interface
MMS	Mensch-Maschine-System
MOST	Maynard Operation Sequence Technique

MRI	Mensch-Roboter-Interaktion
MTM	Methods-Time Measurement
MTM-UAS	Methods-Time Measurement Universal Analysis System
NUI	Natural User Interface
REFA	Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung e. V. (vormals: Reichsausschuß für Arbeitszeitermittlung)
SCADA	Supervisory Control and Data Acquisition
SCI	Spinal Cord Injury
SDK	Software Development Kit
SOTA	state-of-the-art
OEE	Overall Equipment Effectiveness
TDNN	Time Delay Neural Network
UCD	User Centered Design

1 Einleitung

Die Anforderungen des globalen Marktes entwickeln sich in Richtung zunehmend individualisierter Produkte mit gleichzeitig immer kürzeren Lebenszyklen. Parallel dazu verstärkt die beschleunigte Entwicklung neuer Technologien diesen Trend. Abnehmende Losgrößen, verkürzte Produktlebenszyklen und eine sogenannte Hyper-Individualisierung von Produkten erschweren die wirtschaftliche Automatisierung von Produktionsprozessen. In Zukunft wird die Flexibilität, mit der sich Produktionssysteme an die Bedürfnisse des Marktes anpassen können, ein entscheidender Wettbewerbsfaktor sein. [BNP21, KHJ23, TDK21]

Im Kontext der Industrie 4.0 wurden autonome cyber-physische Produktionssysteme (CPPS) implementiert, die eine hohe Produktivität, Prozessqualität und Prozesssicherheit in zunehmend komplexen Produktionsprozessen gewährleisten. Die Automatisierung hat dadurch ein derart hohes Niveau erreicht, dass der traditionelle Maschinenbediener zu einem Supervisor geworden ist, der mit einer zunehmenden Menge an Informationen umgehen muss. [Fri20, GSL14, LFK14, LoDr20, Lor20, MLA22].

Das Konzept der Industrie 5.0 erweitert die ursprünglichen Ansätze der Industrie 4.0, welche sich auf Digitalisierung und Künstliche Intelligenz zur Steigerung der Effizienz und Flexibilität von Produktionsprozessen konzentrierten, um einen nutzerzentrierten Ansatz (engl.: human-centric approach) bei dem neben der Technologie auch der Mensch bei der Entwicklung von Produktionsabläufen mit einbezogen wird. [BNP21, TDK21] Ziel ist es, die Intelligenz, das übergeordnete Prozess-Verständnis und die Wahrnehmung des Menschen mit all seinen Sinnen in seiner neuen Rolle als Supervisor im Produktionsablauf mit der Produktivität, Geschwindigkeit und Präzision moderner Produktionsmaschinen zu kombinieren. Dafür bedarf es nutzerzentrierter und gebrauchstauglicher Human-Machine Interfaces (HMI), um Produktionsabläufe mit hohen Flexibilitätserfordernissen, die sich wirtschaftlich nicht vollständig automatisieren lassen, durch reduzierte Maschinenbedienzeiten effizienter zu gestalten. [EKM20, KHJ23, Lot12a] Aufgrund dieser neuen Rolle des Bedieners müssen die Anforderungen an HMIs neu evaluiert und sowohl neuartige organisatorische als auch technologische Methoden zur Anpassung der HMI erforscht werden [GSL14, JHR21].

Sprache bietet als natürlichste Form der Kommunikation das Potenzial, komplexe Menüstrukturen durch direkte Befehle zu durchbrechen, Laufwege zu vermeiden und eine ortsunabhängige Bedienung mehrerer Maschinen zu ermöglichen [Koe23, Lot12b, RBR20]. Die Eingabe von Befehlen per Sprachsteuerung bietet nicht nur das Potenzial, in gewissen Anwendungsszenarien

und Bediensituationen schneller als andere HMI-Modalitäten zu sein, sondern bietet auch die Möglichkeit, Tätigkeiten zu parallelisieren. So können simultan zu einer wertschöpfenden Tätigkeit nicht wertschöpfende Tätigkeiten wie Dokumentationstätigkeiten, ohne den Einsatz von Händen oder Augen per Sprache durchgeführt werden. [VaBo16] Eine Gestensteuerung, wie sie bereits in Consumer Applikationen wie Smartphones oder MMIs moderner Fahrzeuge integriert ist, bietet diese Möglichkeit nicht, da eine Haupttätigkeit bei der Interaktion mit dieser HMI-Modalität unterbrochen werden muss [MIK17].

Die Lücke zwischen vollautomatisierter Serienfertigung und manueller Einzelfertigung kann ein flexibles, nutzerzentriertes und damit besonders gebrauchstaugliches, respektive effizientes HMI wie eine Sprachsteuerung schließen. Insbesondere aufgrund der jüngsten Entwicklungen im Bereich Künstlicher Intelligenz (KI) sind Sprachsteuerungen mittlerweile robust genug gegenüber Störgeräuschen in industriellen Umgebungen.

Produktionsmaschinen sind jedoch komplexe, kostspielige und teilweise gefährliche technische Systeme, die von geschulten Fachkräften oft in lauter Umgebung bedient werden. Diese Herausforderungen, zusammen mit negativen Erfahrungen mit Sprachsteuerungen in Consumer-Anwendungen, erhöhen die Hemmschwelle für HMI-, Steuerungs- und Maschinenhersteller sowie Betreiber von Produktionsmaschinen, Sprachsteuerung als HMI-Modalität zu implementieren. Erst wenn die Effizienzvorteile einer Sprachsteuerung ausreichend quantifiziert wurden, wird diese durch Unternehmen als praktikable Lösungen für die aktuellen Herausforderungen der industriellen Digitalisierung in Betracht gezogen.

1.1 Motivation

Da im Durchschnitt 85 % der Unternehmenskosten bei der Teilefertigung und Montage anfallen, wirkt sich eine Reduzierung von nicht wertschöpfenden Tätigkeiten positiv auf die Produktivität aus [VaBo16]. Dies ist besonders im Hochlohnland Deutschland von entscheidender Bedeutung, auch wenn Produktionsunternehmen bei der Erprobung oder gar flächendeckenden Nutzung von Sprachsteuerung noch zögerlich sind [EKM20, Koe23, Lot12a, Lot12b]. Neben Bedenken hinsichtlich der Robustheit in lauten Störgeräuschumgebungen liegt dies vor allem daran, dass das Effizienzpotenzial einer Sprachsteuerung in der Produktion noch nicht systematisch untersucht wurde [FjKa22].

Eine Untersuchung der Effizienz von Sprachsteuerung als Vertreter von nutzerzentrierten HMIs in der Industrie 5.0 generiert Erkenntnisse, die zur Ableitung von übergeordneten Methoden zur

Bewertung und Auslegung von Sprachsteuerung benötigt werden. Aktuell fehlen diese Gestaltungsrichtlinien, werden aber gleichzeitig dringend benötigt, um nutzerzentrierte HMIs zur Steigerung der Wertschöpfung und Mitarbeiterzufriedenheit in der industriellen Produktion zu etablieren. [PLK24]

Die Motivation für diese Arbeit ist es, das bisher ungenutzte Effizienzpotenzial einer Sprachsteuerung systematisch zu bewerten und für industrielle Anwendungen nutzbar zu machen. Dafür soll eine »*Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion*« entwickelt werden, mit der identifizierte Einflussfaktoren auf die Effizienz einer Sprachsteuerung anhand von Parametern bewertet werden. Dadurch sollen Unternehmen in die Lage versetzt werden, ihre bestehenden, aber auch neu entwickelte Produktionsabläufe zunächst unter Verwendung eines konventionellen HMIs zu bewerten und anschließend das Effizienzpotenzial einer Sprachsteuerung als Ergänzung oder Austausch zu diesem zu prognostizieren. Damit soll effektiv der Weg für die Sprachsteuerung hin zu einem industriellen Standard-HMI, wie es konventionelle HMIs bereits sind, geebnet und übergeordnet die nachhaltige Sicherung des Produktionsstandort Deutschlands unterstützt werden.

1.2 Forschungslücke

Haptisch-visuelle HMIs wie Taster, Schalter, Drehregler, Touchscreens, Bildschirme, Tastaturen und Mäusen werden als konventionelle HMIs bezeichnet. Im Consumer-Bereich ist die Steuerung von Funktionen technischer Systeme per Sprache bereits als Standard etabliert (vgl. **Kapitel 2**). Die Gründe zur Nutzung einer Sprachsteuerung anstelle konventioneller HMI-Modalitäten im Consumer-Bereich sind dabei unterschiedlich (vgl. **Kapitel 2**). Die Möglichkeit zur technischen Integration einer automatischen Spracherkennung (engl.: automatic speech recognition (ASR)) in Form einer Sprachsteuerung in eine Produktionsmaschine wurde bereits in zahlreichen Veröffentlichungen untersucht (vgl. **Kapitel 2.3.4**). Dabei wurde die technische Machbarkeit bereits an Roboter(-armen), Computerized Numerical Control (CNC)-Dreh-/Fräsmaschinen, Produktionsautomaten oder Enterprise Resource Planning (ERP)-Systemen nachgewiesen (vgl. **Kapitel 2.3.4**).

Seit vielen Jahren wird bereits zum Thema Sprachsteuerung von Produktionsmaschinen wie auch zu vielen anderen nutzerzentrierten HMIs wie Eye Movement Control, Gestensteuerung oder Brain-Computer Interfaces (BCI) geforscht (vgl. **Kapitel 2.3**). Bisher sind jedoch nur die Sprachsteuerung und zum Teil die Gestensteuerung zur Steuerung technischer Systeme im Con-

sumer-Bereich etabliert (vgl. **Kapitel 2.3**). Das bedeutet, dass die Sprachsteuerung mit ihren potenziellen Vorteilen gegenüber konventionellen HMIs seit Jahren am nächsten an der Schwelle zu einem industriellen Standard-HMI steht. Aktuelle Produktionsmaschinen wie Produktionsautomaten, Roboter, Werkzeugmaschinen oder Produktionsanlagen werden heutzutage jedoch weiterhin vornehmlich mit konventionellen HMIs ausgestattet (vgl. **Kapitel 2.2 & 4**).

Vereinzelte Veröffentlichungen darauf hingewiesen, dass die Auswahl eines effizienten HMI für den Produktionsablauf an einer Produktionsmaschine nicht trivial ist (vgl. **Kapitel 2.5 & 2.6**). Die Veröffentlichungen, in denen Auswahlkriterien für ein HMI thematisiert werden, behandeln dieses Thema jedoch entweder auf einem sehr hohen Abstraktionslevel oder mit der Effizienz nur als Nebenziel. Stattdessen stehen andere Kriterien wie die übergeordnete Gebrauchstauglichkeit (engl.: usability) oder die Benutzererfahrung im Fokus (vgl. **Kapitel 2.5 & 2.6**). Auch wenn einige Studien das Effizienzpotenzial einer Sprachsteuerung im Vergleich zu konventionellen HMIs qualitativ bewerten, mangelt es an detaillierten quantitativen Aussagen, um das Effizienzpotenzial systematisch nachweisen zu können.

1.3 Ziele und Forschungsfragen

Ziel dieser Arbeit ist es, eine vom wertschöpfenden Produktions- oder Montageprozess in den Hauptzeiten unabhängige und damit allgemeingültige Methode zu entwickeln, um die Effizienz einer Sprachsteuerung als industrielles HMI in Form einer robusten ASR und eines Audio-Feedbacks im Vergleich zu einem konventionellen, haptisch-visuellen HMI in der Produktion zu evaluieren.

Dazu werden unterschiedliche Produktionsabläufe identifiziert, die hinsichtlich ihrer Effizienz von einer Sprachsteuerung profitieren. Anschließend werden übergeordnete Einflussfaktoren auf die Effizienz abgeleitet, die auf Basis von technischen Parametern quantitativ bewertet werden. Alle Einflussfaktoren und zugehörigen Parameter, die einen signifikanten Einfluss auf das Effizienzpotenzial einer Sprachsteuerung haben, werden in einer quantitativen Methode zusammengefasst. Dabei sollen die Parameter möglichst einfach messbar und gleichzeitig ausreichend valide sein, um die Praxistauglichkeit der Methode bei ausreichend hoher Genauigkeit zu gewährleisten. Mithilfe der Methode sollen Unternehmen in die Lage versetzt werden, ohne die Integration einer Sprachsteuerung in ihre Produktionsmaschinen das Effizienzpotenzial in ihren Produktionsabläufen zu bewerten, um so eine objektive Entscheidungsgrundlage für oder gegen eine Sprachsteuerung zu schaffen. Dadurch soll die Hürde für die Industrialisierung einer Sprachsteuerung zur Reduzierung von Nebenzeiten in Produktionsabläufen gesenkt werden, um

die damit verbundene die Produktivität von Produktionsabläufen zu steigern. Die Ziele dieser Arbeit werden im Folgenden aufgelistet.

Ziele

1. Entwicklung einer vom Fertigungs- oder Montageprozess in den Hauptzeiten unabhängigen und damit allgemeingültigen Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles HMI
2. Identifikation von Einflussfaktoren und Quantifizierung dieser anhand von abgeleiteten Parametern, die das Effizienz-Verhältnis zwischen einem konventionellen HMI und einer nutzerzentrierten Sprachsteuerung abbilden
3. Herleitung einer objektiven und quantitativen Entscheidungsbasis für den Einsatz einer nutzerzentrierten Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen HMI
4. Senkung der Investitionsschwelle für Unternehmen zur Investition in eine Sprachsteuerung als HMI durch eine vorherige Prognose der Reduzierung von Nebenzeiten
5. Entwicklung von Handlungsempfehlungen zur systematischen Nutzung des Effizienzpotenzials einer Sprachsteuerung in der Produktion

Die beschriebenen Ziele werden auf Basis einer Hauptforschungsfrage und vier zugehörigen Teilforschungsfragen bearbeitet. Dabei werden mithilfe der Teilforschungsfragen einzelne Aspekte der Hauptforschungsfrage detaillierter betrachtet.

Hauptforschungsfrage

- Wie muss eine Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion gestaltet sein?

Teilforschungsfragen

- Welche Produktionsabläufe lassen sich durch eine Sprachsteuerung im Vergleich zu einem konventionellen HMI effizienter gestalten?
- Welche Einflussfaktoren sind für den effizienten Einsatz einer Sprachsteuerung als industrielles HMI einer Produktionsmaschine relevant?
- Welche Parameter können genutzt werden, um das Effizienzpotenzial einer industriellen Sprachsteuerung im Vergleich zu einem konventionellen HMI zu quantifizieren?

- Lässt sich die Effizienz von bestehenden Produktionsabläufen durch den Einsatz einer industriellen Sprachsteuerung signifikant steigern oder ist eine Anpassung von Produktionsabläufen an diese nutzerzentrierte HMI-Modalität notwendig, um das Effizienzpotenzial zu nutzen?

1.4 Anforderungsliste (Lastenheft)

Aus den in **Kapitel 1.3** formulierten Zielen und Forschungsfragen werden folgende Anforderungen an die zu entwickelnde Methode abgeleitet und in einer Anforderungsliste (im technischen Kontext auch Lastenheft genannt) zusammengefasst.

Anforderung 1

Die zu entwickelnde Methode muss die Möglichkeit bieten, das Effizienzpotenzial einer Sprachsteuerung in einem Produktionsablauf, basierend auf Daten des aktuellen Produktionsablaufs, mit einem konventionellen, haptisch-visuellen HMI quantitativ zu bewerten (vgl. **Anforderung 1**). Die Möglichkeit zur quantitativen Evaluation gewährleistet die Unabhängigkeit vom subjektiven Empfinden der Mitarbeitenden im Produktionsablauf. Da dies der wesentliche Vorteil der zu entwickelnden Methode im Vergleich zu existierenden und etablierten Konzepten, Methoden und Werkzeugen (vgl. **Kapitel 2.7**) ist, bei denen jeweils der Produktionsablauf mit dem konventionellen und dem neuartigen HMI bewertet werden muss, nimmt diese Anforderung den ersten Platz ein. Die Evaluierung von Software oder softwarebasierten Systemen wird in **Kapitel 2.4.7** definiert.

Anforderung 2

Auf Basis der quantitativen Evaluation muss eine Prognose der nicht wertschöpfenden, zu reduzierenden Nebenzeiten des Produktionsablaufs, in denen Mitarbeitende mit der Produktionsmaschine über das konventionelle HMI interagieren, möglich sein (vgl. **Anforderung 2**). Dadurch wird die zeit- und kostenintensive Integration einer Sprachsteuerung in einem Produktionsablauf, der gegebenenfalls vom Einsatz dieser HMI-Modalität gar nicht profitiert, vermieden.

Anforderung 3

Die Evaluation des Effizienzpotenzials einer Sprachsteuerung muss auf Basis der aufgewendeten Interaktionszeit inklusive etwaiger Fehleingaben, so wie es die »DIN EN ISO 9241–11 *Ergonomie der Mensch-System Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte*« von 2018 beschreibt, erfolgen (vgl. **Anforderung 3 & Kapitel 2.4.1**) [DIN18]. Die Interaktionszeit be-

schreibt dabei die Eingabezeit in ein HMI und gegebenenfalls anfallende Laufwege zwischen unterschiedlichen HMIs oder einem Arbeitsplatz an einer Produktionsmaschine und einem HMI (vgl. **Kapitel 5 & 6**).

Anforderung 4

Des Weiteren muss die zu entwickelnde Methode unabhängig vom wertschöpfenden Produktions- oder Montageprozess in den Hauptzeiten sein, um allgemeingültig für die Evaluation des Effizienzpotenzials einer Sprachsteuerung im Vergleich zu einem konventionellen, haptisch-visuellen HMIs zu sein (vgl. **Anforderung 4**).

Anforderung 5

Zuletzt muss die Prognose der potenziell zu reduzierenden Nebenzeiten auf Basis der Evaluation des Effizienzpotenzials einer Sprachsteuerung präziser als die interindividuellen Unterschiede von Mitarbeitenden in der Evaluation sein (vgl. **Anforderung 5**). Nur dadurch wird gewährleistet, dass die Methode Unternehmen überhaupt eine sichere Entscheidungsbasis für oder gegen die Investition in eine Sprachsteuerung bietet.

Die oben beschriebenen Anforderungen werden im Folgenden aufgelistet.

Anforderungsliste (Lastenheft)

1. Quantitative Evaluation des Effizienzpotenzials einer nutzerzentrierten Sprachsteuerung im Vergleich zu einem konventionellen haptisch-visuellen Human-Machine Interface
2. Prognose der potenziell zu reduzierenden Nebenzeiten auf Basis von Einflussfaktoren des konventionellen Produktionsablaufs ohne die Notwendigkeit der Integration einer Sprachsteuerung
3. Evaluation des Effizienzpotenzials auf Basis der aufgewendeten Interaktionszeit inklusive etwaiger Fehleingaben (gemäß DIN EN ISO 9241-11:2018)
4. Unabhängigkeit vom wertschöpfenden Produktions-/Montageprozess in den Hauptzeiten des Produktionsablaufs und daraus resultierende Allgemeingültigkeit der Methode
5. Höhere Präzision der Prognose als interindividuelle Unterschiede in Interaktionszeiten von Bedienern zur Gewährleistung der praktischen Nutzbarkeit

1.5 Aufbau der Arbeit

In **Kapitel 2** wird mit dem Stand der Forschung begonnen. Zuerst wird die Verschiebung des Fokus in Richtung teilautomatisierter Produktionsabläufe vorgestellt, in denen Mensch und Maschine kollaborieren. Dieser wird in der (Weiter-)Entwicklung des Konzeptes der Industrie 4.0 zur Industrie 5.0 beschrieben. Damit verbunden ist eine steigende Bedeutung von nutzerzentrierten HMIs zur Reduzierung von Nebenzeiten und zur Steigerung der Produktivität von Produktionsprozessen. Dies gilt vor allem für Hochlohnländer wie Deutschland.

Zu Beginn des darauffolgenden **Kapitels 2.1** werden die Begriffe »Haupt-«, »Neben-« und »Brach-« sowie »Rüstzeiten« aus der Terminologie des »Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung e. V.« (REFA) vorgestellt. Anschließend werden die Begriffe »Produktionsprozess« und »Produktionsablauf« sowie die der Einsatz und die Abgrenzung beider im Kontext dieser Arbeit definiert.

Danach wird in **Kapitel 2.2** der aktuelle Stand der Technik von konventionellen haptisch-visuellen HMIs als Ausgangsbasis für diese Arbeit beschrieben. Im Kontrast dazu folgen in **Kapitel 2.3** gemäß dem Konzept der Industrie 5.0 die vier im industriellen Kontext relevantesten neuartigen und nutzerzentrierte HMIs Gestensteuerung (vgl. **Kapitel 2.3.1**), Eye Movement Control (vgl. **Kapitel 2.3.2**), Brain-Computer Interfaces (BCI) (vgl. **Kapitel 2.3.3**) und Sprachsteuerung (vgl. **Kapitel 2.3.4**).

In **Kapitel 2.4** und den zugehörigen **Kapiteln 2.4.1 bis 2.4.7** werden aktuelle Normen zur Ergonomie der Mensch-System-Interaktion sowie zu ergonomischen Gestaltungsgrundsätzen und Evaluationsprinzipien vorgestellt. Danach werden die Bestandteile jeder Norm, die in dieser Arbeit zur Entwicklung einer »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« genutzt werden, beschrieben.

Im Anschluss werden in **Kapitel 2.5** Einflussfaktoren auf die Gebrauchstauglichkeit, die der Effektivität, Effizienz und Zufriedenstellung übergeordnet ist (vgl. **Kapitel 2.4.1**), aus der aktuellen Forschung vorgestellt. In **Kapitel 2.6** wird auf konkrete Effizienzpotenziale von Sprachsteuerung eingegangen, die in aktuellen Forschungsarbeiten untersucht werden.

Als letztes Kapitel des Stands der Forschung werden in **Kapitel 2.7** etablierte Konzepte, Methoden und Werkzeuge vorgestellt, die entweder als Vorbild für die in dieser Arbeit entwickelte Methoden dienen oder aus denen Bestandteile in der entwickelten Methode genutzt werden.

Nach dem Stand der Forschung wird in **Kapitel 3** zunächst die übergeordnete Vorgehensweise zur Entwicklung der Methode vorgestellt (vgl. **Kapitel 3.1**), bevor diese im Detail mit ihren einzelnen Schritten in den **Kapiteln 3.2.1 bis 3.2.7** vorgestellt wird. Abschließend wird in **Kapitel 3.2.8** ein Template zur Erfassung aller notwendigen Parameter zur Evaluation der Effizienz einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen HMI eingeführt.

In **Kapitel 4** wird der Weg über Anwendungsszenarien, die jeweiligen Interaktionsprinzipien zwischen Mensch und Maschine sowie die Identifikation von Einflussfaktoren bis hin zur Quantifizierung über abgeleitete Parameter exemplarisch an unterschiedlichen Arten von Produktionsmaschinen erläutert.

Die **Kapitel 5 und 6** beinhalten die Beschreibung einer Probanden- und einer Industriestudie zur Quantifizierung der Einflussfaktoren **Komplexität von Menüstrukturen, Art und Anzahl von Laufwegen** sowie **Anzahl simultan zu bedienender Maschinen**. Die Ergebnisse ermöglichen es Unternehmen mithilfe der Methode, ohne die tatsächliche Integration einer Sprachsteuerung in ihre Produktionsmaschine das Effizienzpotenzial zu bewerten und die Reduzierung von Nebenzeiten zu prognostizieren (vgl. **Kapitel 1.4**).

Danach wird die Methode in **Kapitel 7** gemäß den Anforderungen aus dem Lastenheft validiert (vgl. **Kapitel 1.4**). Dazu werden Interaktionsprinzipien aus realen Anwendungsszenarien der Industriestudie und aktuellen Veröffentlichungen in einer Probandenstudie an einer eigens entwickelten Technologieplattform nachgestellt. Diese verfügt über eine konventionelles, haptisch-visuelles HMI in Form eines Touchscreens und gleichzeitig über eine neuartige, nutzerzentrierte Sprachsteuerung. Zur Validierung bedienen alle Probanden entlang eines festgelegten Versuchsablaufs die Technologieplattform sowohl über den Touchscreen als auch die Sprachsteuerung. Die gemessenen Eingabe- und Laufwegzeiten werden anschließend mit der Prognose zur Reduzierung von Nebenzeiten auf Basis der Evaluation der Effizienz einer Sprachsteuerung mithilfe der entwickelten Methode verglichen. Dadurch lassen sich die interindividuellen Abweichungen der Probanden ins Verhältnis zur Prognose der Reduzierung von Nebenzeiten auf Basis der Evaluation des Effizienzpotenzials einer Sprachsteuerung stellen (vgl. **Kapitel 1.4**).

Abschließend werden in **Kapitel 8** die Ergebnisse der vorangegangenen **Kapitel 3 bis 7** auf Basis der aufgestellten Ziele und Forschungsfragen aus **Kapitel 1.3** diskutiert. Zuletzt wird in **Kapitel 9** ein Fazit gezogen und ein Ausblick auf zukünftige anknüpfende Forschungsarbeiten gegeben.

2 Stand der Forschung

Im 20. Jahrhundert lag der Fokus der Produktion noch auf der Entwicklung effektiver und fehlerfreier Prozesse. Im Rahmen von Industrie 4.0 wurde dieser Ansatz um das Streben nach hoch-effizienten Prozessen erweitert. Dadurch veränderte sich die Zusammenarbeit zwischen Mensch und Maschine. Während diese beiden Parteien in der Vergangenheit klar getrennt waren, wuchsen sie innerhalb der Industrie 4.0 über immer umfangreichere HMIs zu einem symbiotischen Mensch-Maschine-System (MMS) zusammen. Anstelle von einzelnen Interaktionen an einer Maschine steuern und überwachen Maschinenbediener heutzutage mehrere verteilte, zum Teil intelligente und autonome Produktionssysteme. [AhZA17, GSL14, VSL21]

Aufgrund des Wandels in der Interaktion zwischen Mensch und Maschine, von Maschinenbedienern, die einzelne Maschinen bedienen, hin zu Supervisoren, die ein Netz an autonomen und intelligenten Produktionssystemen steuern und überwachen, müssen die Anforderungen an neuartige, nutzerzentrierte HMIs, aber auch konventionelle, haptisch-visuelle HMIs neu bewertet werden. Zusätzlich müssen auch neuartige Organisationsmethoden von Produktionsabläufen wie die Mehrmaschinenbedienung von CPPS hinsichtlich der Bedürfnisse und Fähigkeiten der Maschinenbediener untersucht und angepasst werden. [GSL14, JHR21, MLA22, PLK24, VSL21]

In **Abbildung 1** wird der beschriebene Wandel der Anforderungen an HMIs in der Transition von Industrie 4.0 zu Industrie 5.0 abgebildet.

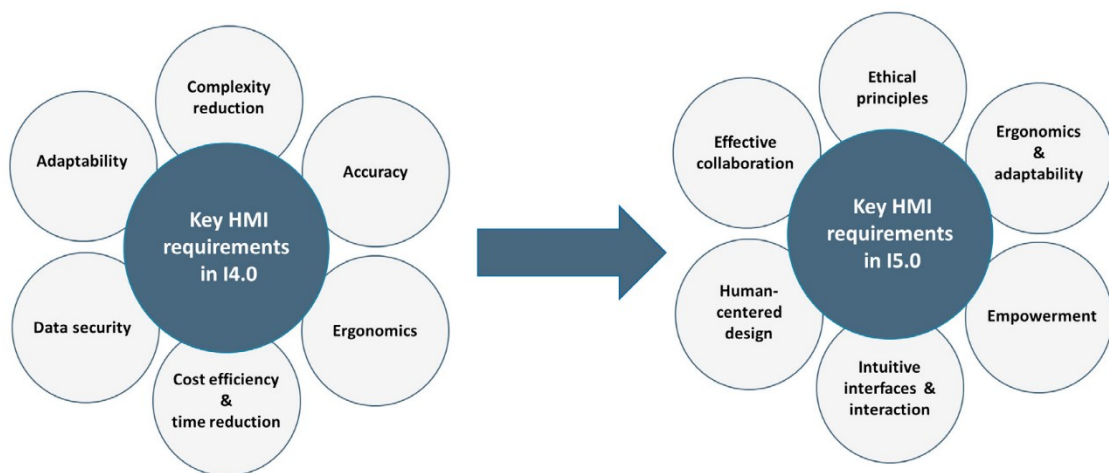


Abbildung 1: Wandel der Anforderungen an HMIs von Industrie 4.0 (I4.0) zu Industrie 5.0 (I 5.0) [PLK24]

Während in der Industrie 4.0 die Anforderungen an HMIs noch durch die Maschine und das HMI selbst definiert waren, werden diese in der Industrie 5.0 durch die Maschinenbediener getrieben. In der Industrie 4.0 mussten zunächst grundsätzliche Hürden hinsichtlich Genauigkeit, Datensicherheit und der Reduzierung der Komplexität von HMI genommen werden, bevor daraus in der Industrie 5.0 übergeordnete Anforderungen wie die effektive Kollaboration von Mensch und Maschine, ethische Prinzipien und intuitive HMIs und Produktionsabläufe abgeleitet werden können. Anforderungen an die Ergonomie und Anpassbarkeit sowie kosten- und zeittechnische Effizienz von HMI waren bereits in der Industrie 4.0 vorhanden (vgl. **Kapitel 2.4**). In der Industrie 5.0 rücken diese jedoch noch mehr in den Fokus und werden um Anforderungen an ein nutzerzentriertes Design (engl.: human-centered design) und Empowerment des Maschinenbedieners zur Bewältigung seiner neuen Rolle als Supervisor erweitert. [BNP21, PLK24]

2.1 Zusammenhang zwischen Haupt-, Neben-, Rüst- und Brachzeiten sowie Produktionsprozessen und Produktionsabläufen

Produktionsprozesse (Fertigungs-/Montageprozesse) sind standardisierter Arbeitsabläufe, in denen durch maschinelle und/oder manuelle Be- und Verarbeitung von Rohstoffen oder Zwischenprodukten ein verkaufsfähiges Produkt produziert wird. Im Unterschied zum wertschöpfenden Produktionsprozess beschreiben Produktionsabläufe die systematische Abfolge aller Arbeitsschritte im Produktionsprozess, die zur Herstellung eines Produkts notwendig sind. Gemäß der REFA-Definition findet die Wertschöpfung im Produktionsprozess in den sogenannten »Hauptzeiten« statt (vgl. **Kapitel 2.7.2**). Damit ist die Zeit gemeint, in der beispielsweise ein Roboter eine Schweißnaht entlang der Verbindungsstelle zweier Karosserieteile zieht. Diese Hauptzeiten bildet zusammen mit den sogenannten »Neben-« und »Brachzeiten« die sogenannte »Betriebsmittelzeit«. Die in dieser Arbeit untersuchte Interaktion mit einem HMI zählt zu den Nebenzeiten, während die Brachzeit das ablaufbedingte oder erholungsbedingte Unterbrechen des Produktionsablaufs beschreibt. Der Produktionsablauf beschreibt die Reihenfolge aller Arbeitsschritte in der übergeordneten Betriebsmittelzeit, bestehend aus Hauptzeiten, Nebenzeiten und Brachzeiten sowie vorgelagerten Rüstzeiten. [Bin10, BuSc21, REF25, ScSc14, Tsc83]

In **Abbildung 2** werden exemplarisch unterschiedliche Produktionsprozesse geordnet nach ihrem Mechanisierungsgrad dargestellt. In der ersten Spalte wird mit der »Manuellen Produktion« begonnen, bei der Mitarbeitende entweder händisch oder mit handgeführten Werkzeugen Pro-

dukte verpacken, Baugruppen zusammensetzen, Schraubverbindungen lösen oder Bauteile entgraten. Die zweite Spalte zeigt Beispiele für Produktionsprozesse der »*Maschinellen Produktion*«, bei der Mitarbeitende zwar mithilfe von angetriebenen Werkzeugen, die aber immer noch händisch geführt werden, Löcher bohren, Schraubverbindungen anziehen, Oberflächen bearbeiten oder Bleche zuschneiden. In der dritten Spalte werden Produktionsprozesse der »*Teilautomatisierten Produktion*« gezeigt, in denen Mitarbeitende mit Maschinen im Rahmen von vor- oder nachbereitenden Tätigkeiten in Rüst- und Nebenzeiten interagieren, die Wertschöpfung in den Hauptzeiten jedoch allein durch die Maschine erfolgt. HMI-Interaktionen in teilautomatisierten Produktionsprozessen sind zum Beispiel das Starten eines Kantprozesses nach dem Einlegen eines Bleches, die Erhöhung des Drahtvorschubs eines Schweißroboters, die Anpassung von Parametern eines Extrusion-Prozesses oder die Programmierung einer Werkzeugmaschine. Die vierte Kategorie bilden Produktionsprozesse der »*Vollautomatisierten Produktion*« in denen Maschinen über Minuten, Stunden, Tage oder gar Wochen vollständig autonom in den Hauptzeiten wertschöpfend arbeiten. Die Mitarbeitenden interagieren nur in Rüst- und Nebenzeiten mit der Maschine, um Halbzeuge in einen sogenannten »*Bunker*« oder »*Magazin*« einzulegen, um Werkzeuge einzusetzen oder zu entnehmen oder fertige Bauteile zu entnehmen.

Manuelle Produktion	Maschinelle Produktion	Teilautomatisierte Produktion	Vollautomatisierte Produktion
Einlegen von Produkten in einen Karton	Bohren eines Lochs mit einem Akkuschrauber	Start eines Kantprozesses nach dem Einlegen eines Bleches	Büroklammer-Herstellung in einem vollautomatisierten Umformprozess
Zusammensetzen einer Baugruppe aus Bauteilen	Anziehen einer Schraubverbindung mit einem Schlagschrauber	Erhöhung des Drahtvorschubs eines Schweißroboters	Vollautomatisiertes Roboter-Schweißen einer Fahrzeugkarosserie
Lösen einer Schraubverbindung mit einem Schraubenschlüssel	Schleifen einer Oberfläche mit einem Exzentschleifer	Anpassung von Parametern im Einfahrprozess eines Kunststoff-Extruders	Vollautomatisiertes Strangfließpressen von Kunststoffprofilen
Entgraten eines Bauteils mit einer Feile	Zuschneiden eines Blechs mit einem Trennschleifer	Programmierung eines Plandrehprozesses an einer Drehmaschine	Vollautomatisierte Fräsbearbeitung von Zylinderköpfen

Abbildung 2: Produktionsprozesse nach Mechanisierungsgrad

In dieser Arbeit wird eine Methode entwickelt, mit der sich das Effizienzpotenzial einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen, haptisch-visuellen

HMI bewerten lässt. Der Fokus liegt dabei auf den Nebenzeiten in teilautomatisierten Produktionsprozessen, in denen Mitarbeitende mit HMIs entlang des Produktionsablaufs interagieren. Auf Basis der Bewertung des Effizienzpotenzials einer Sprachsteuerung mithilfe der entwickelten Methode kann am Ende eine Prognose über die Reduzierung von Nebenzeiten zur Steigerung der Produktivität im gesamten Produktionsablauf gegeben werden. Die Einflussfaktoren auf das Effizienzpotenzial können dabei Bestandteile des Produktionsablaufs (**Art und Anzahl von Laufwegen** und **Anzahl simultan zu bedienender Maschinen**), aber auch Bestandteile des konventionellen, haptisch-visuellen HMIs (**Komplexität von Menüstrukturen**) sein (vgl. **Kapitel 2.5 & 4**).

2.2 Konventionelle haptisch-visuelle Human-Machine Interfaces

Konventionelle, haptisch-visuelle HMIs werden über speicherprogrammierbare Steuerungen (SPS) mit den Sensoren und Aktoren von verketteten Produktionsanlagen, Produktionsautomaten, Werkzeugmaschinen und Robotern verbunden (vgl. **Kapitel 4**). Jeder Steuerungshersteller liefert zu seinen SPS sogenannte Software Development Kits (SDKs) zur eigenständigen Entwicklung von grafischen Benutzeroberflächen (engl.: graphical user interface (GUI)) durch HMI-, Steuerungs- und Maschinenhersteller. In **Abbildung 3** wird exemplarisch die Softwaresuite »Studio 5000« des Steuerungsherstellers Rockwell Automation abgebildet. Mit dieser können grafische Benutzeroberflächen für Bildschirme und Touchscreens von Rockwell Automation Steuerungen entwickelt werden. Im rechten oberen Abbildungsbereich ist die Darstellung des HMIs als Applikationsansicht auf dem Touchscreen an der Maschine abgebildet. Auf dem Bildschirm im Zentrum der Abbildung ist die Softwaresuite »Studio 5000« zu sehen. Über den Datei-Explorer am linken Bildschirmrand in der Abbildung kann in den Projektdateien navigiert werden, während die Toolbox im unteren linken Bildschirmrand der Abbildung die einfache Integration von grafischen Elementen in dem GUI ermöglicht. Die Entwicklungsansicht auf dem PC-Bildschirm und die Applikationsansicht auf dem Touchscreen an der Maschine sind, wie bei vielen Steuerungsherstellern mittlerweile üblich, über eine proprietäre Datenschnittstelle miteinander verbunden, sodass auch bereits in der Entwicklungsansicht Echtzeit-Daten angezeigt werden können. [Inn16]



Abbildung 3: Softwaresuite »Studio 5000« von Rockwell Automation zur Entwicklung von HMIs für Produktionsmaschinen/-anlagen [Inn16]

Zusätzlich bieten Steuerungshersteller die Möglichkeit HMIs von Drittanbietern über standardisierte Schnittstellen mit ihren SPS zu verbinden, auch wenn sie diese Vorgehensweise nicht favorisieren und gegebenenfalls nicht oder mit erhöhtem Aufwand alle Funktionen der jeweiligen SPS über die HMIs von Drittanbietern adressierbar sind. Dies ist im Besonderen für die Integration einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen HMI relevant.

In **Abbildung 4** wird exemplarisch die aktuelle TNC7 Steuerung des deutschen Steuerungsherstellers Heidenhain für mehrachsige Dreh- und Fräsbearbeitungszentren als komplexester Vertreter konventioneller, haptisch-visueller HMIs dargestellt. In dieser Orientierungsansicht wird die relative Positionierung des HMIs zur Prozesskammer dargestellt. Dieses konventionelle HMI einer Werkzeugmaschine steht exemplarisch für alle haptisch-visuellen HMI an Produktionsmaschinen, die für den Vergleich mit einer Sprachsteuerung in dieser Arbeit relevant sind. Heidenhain Steuerungen werden vor allem für Werkzeugmaschinen wie Fräs-, Dreh- und Schleifbearbeitungszentren genutzt, wodurch sich entsprechend komplexe Menüstrukturen mit vielen Funktionen ergeben. Aus diesem Grund sind die HMIs dieses Steuerungsherstellers das ideale Beispiel für komplexe Menüstrukturen, die potenziell mit einer Sprachsteuerung durchbrochen werden können. Andere Steuerungshersteller, wie Siemens oder Fanuc, bieten für ihre Steuerungen von Werkzeugmaschinen sehr ähnliche HMIs an, sodass die Beschreibung des Aufbaus anhand des HMIs von nur einem Hersteller im Kontext aktueller HMI und komplexer Menüstrukturen in dieser Arbeit ausreichend ist. [Hei24]

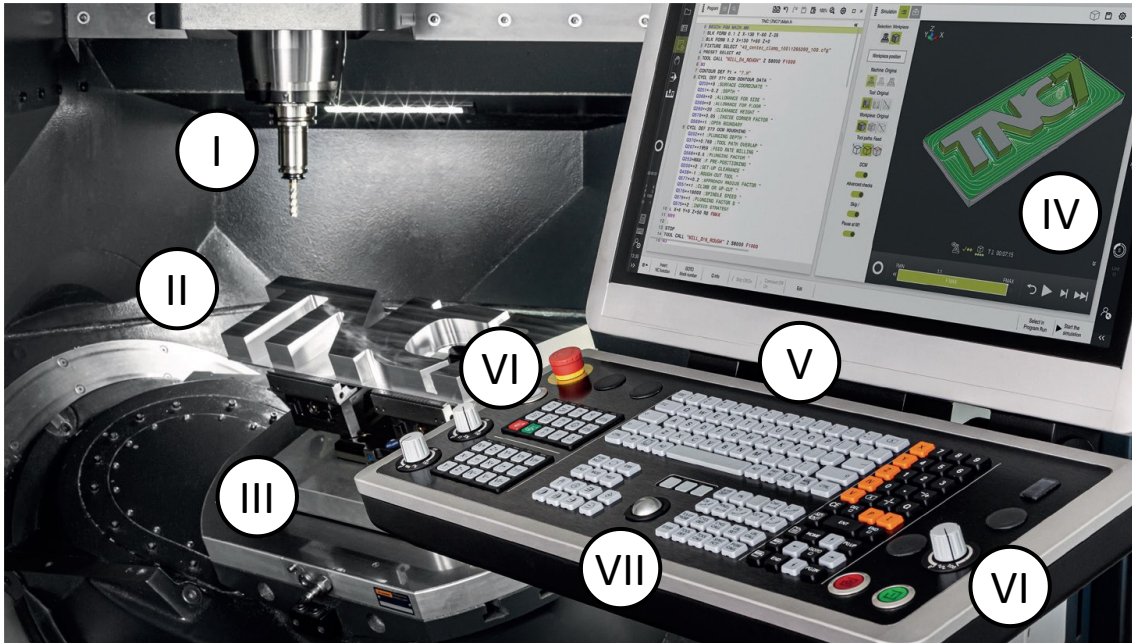


Abbildung 4: Orientierungsansicht des Human-Machine Interface einer Heidenhain TNC7 Steuerung eines 5-Achs-Fräsbearbeitungszentrums (I. Frässpindel mit eingespanntem Fräswerkzeug, II. Eingespanntes Bauteil, III. Mehrachsiger Schwenktisch mit Schraubstock, IV. Touchscreen, V. Tastaturen, VI. Schalter, Taster, Drehregler, VII. Maus in Form eines Trackballs mit drei zusätzlichen Funktionstasten) [Hei24]

Im linken Hintergrund der **Abbildung 4** ist ein spanend bearbeitetes, aufgespanntes Werkstück in Form des Schriftzugs des Namens der Steuerung »TNC7« (vgl. **Abbildung 4**, II) und ein eingespanntes, mehrschneidiges Fräsbearbeitungswerkzeug (vgl. **Abbildung 4**, I) innerhalb der Prozesskammer eines 5-Achs-Fräsbearbeitungszentrums zu sehen. Im Vordergrund ist das haptisch-visuelle HMI des Fräsbearbeitungszentrums zu sehen, das alle in dieser Arbeit betrachteten haptisch-visuellen HMI-Komponenten beinhaltet. Dazu zählen Taster, Schalter und Drehregler (vgl. **Abbildung 4**, VI), ein Touchscreen (vgl. **Abbildung 4**, IV), Tastaturen (vgl. **Abbildung 4**, VII) und eine Maus in Form eines sogenannten »Track-Balls«. Diese HMIs befinden sich immer neben den Prozesskammern, sodass der Bediener gleichzeitig den Fertigungsprozess innerhalb der Prozesskammer und die Maschinendaten auf dem Touchscreen oder Bildschirm im Blick hat. Über einen Schwenkarm lassen sich diese HMIs flexibel in ihrer Position einstellen, um ergonomisch der Anatomie und den Vorlieben des Maschinenbedieners zu entsprechen. [Hei24]

Da die Prozesskammern von Werkzeugmaschinen zum Teil auch mehrere Meter in der Länge, Breite und Höhe messen, besteht die Möglichkeit, an die Steuerung eine Fernbedienung anzuschließen, um vor allem im Einrichtbetrieb zentral vor der Prozesskammer oder gar in der Prozesskammer zu stehen. Die mobilen Handgeräte werden dabei entweder über ein Kabel oder,

bei größeren Entfernungen zum Bedienterminal, über eine Funkverbindung mit dem HMI verbunden. Sie bilden redundant zum Haupt-HMI einzelne Funktionen ab, die für den Einrichtbetrieb von Bedeutung sind. [Hei24]

In **Abbildung 5** wird das HMI der TNC7 von Heidenhain in einer Detailansicht dargestellt. Entlang des Rands des Bedienpanels unterhalb des Touchscreens sind Taster, Schalter und Drehregler (vgl. **Abbildung 5**, 12., 13. (rechts) und 14.) zu finden, die gleich mehrere Tastaturen (vgl. **Abbildung 5**, 7., 10., 11. Und 13. (links)) und einen Track-Ball (vgl. **Abbildung 5**, 9.), als gegenüber Spänen, Metallpartikeln und Schmierstoffen robusteres Äquivalent zu einer klassischen Maus, einrahmen. Darüber befindet sich ein Bildschirm, der zeitgleich auch ein Touchscreen ist (vgl. **Abbildung 5**, 1., 2., 3., 4. und 5.). [Hei24]



1. **TNC-Leiste** mit Zurück, Betriebsarten Statusübersicht, Taschenrechner, Bildschirmstastatur, Einstellungen, Datum und Uhrzeit
2. **Informationsleiste** mit aktiver Betriebsart, Benachrichtigungsmenü
3. **Anwendungsleiste** mit Reiter der geöffneten Anwendungen, Auswahlmenü für Arbeitsbereiche
4. **Arbeitsbereiche**
5. **Maschinenherstellerleiste**
6. **Funktionsleiste** mit Schaltflächen und Auswahlmenü für Schaltflächen
7. **Alpha-Tastatur** für Kommentare und PC-Tastensatz zum Bedienen von Betriebssystem-Funktionen
8. **USB-Anschluss** für zusätzliche Datenspeicher oder Zeigergeräte
9. **Trackball und Maustasten** für eine leichte Bedienung
10. **Achswahl-Tasten und Zehnerblock**
11. **Funktionstasten** für Programmierbetriebsarten, Maschinenbetriebsarten, TNC-Funktionen, Verwalten und Navigation
12. **Override-Potentiometer** für Vorschub und Spindeldrehzahl
13. **Maschinenbedienfeld** mit Clips-Tasten und Leuchtdioden
14. **Override-Potentiometer** für Eilgang

Abbildung 5: Detailansicht des Human-Machine Interface einer Heidenhain TNC7 Steuerung eines 5-Achs-Fräsbearbeitungszentrums [Hei24]

In **Abbildung 6** wird die Programmlaufansicht des HMI dargestellt. In dieser sieht der Maschinenbediener im oberen Teil des Arbeitsbereichs (vgl. **Abbildung 5**, 4.), im Fenster mit der Überschrift »Programm« den G-Code des aktuell auf der Maschine ausgewählten oder laufenden Programms. Darunter sind im Fenster mit der Überschrift »Positionen« die aktuellen Werkzeug-, Vorschub, und Spindeleinstellungen sowie die Ist-Positionen der Maschinenachsen zu sehen. Im rechten Teil des Arbeitsbereichs, im Fenster mit der Überschrift »GPS« werden unterschiedliche Positions- und Bewegungsinformationen sowie deren Korrekturen dargestellt. Über die

TNC-Leiste, Maschinenherstellerleiste und Funktionsleiste (vgl. **Abbildung 5**, 1., 5. und 6.) könnten unterschiedliche Funktionen des HMIs in mehrstufigen Menüstrukturen aufgerufen werden. [Hei24]

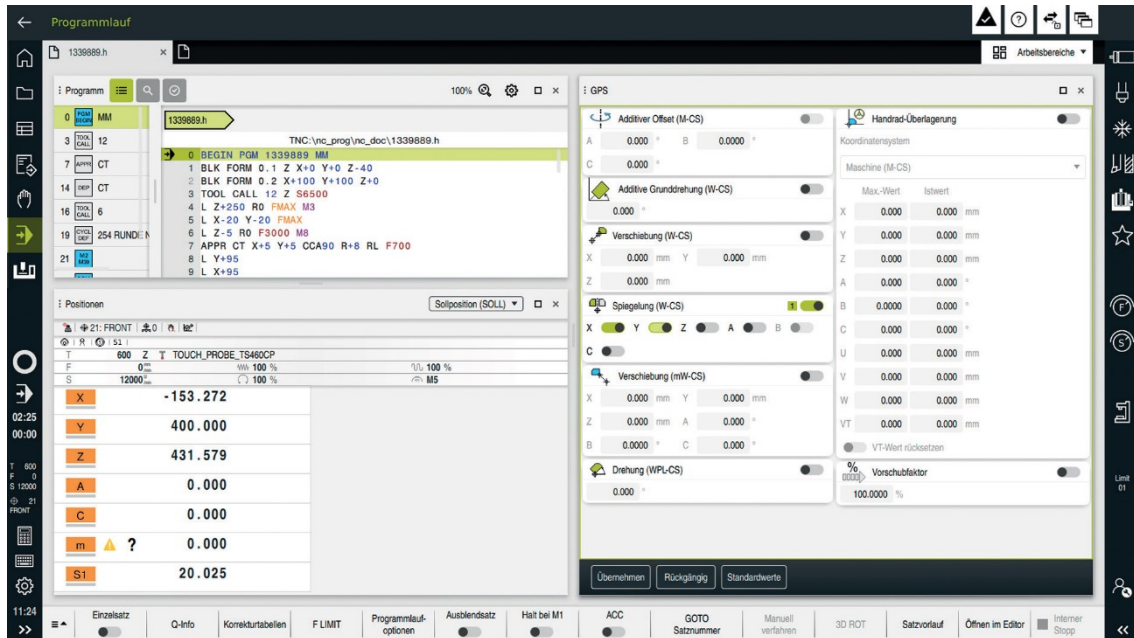


Abbildung 6: Programmablaufansicht des Human-Machine Interface einer Heidenhain TNC7 Steuerung eines 5-Achs-Fräsbearbeitungszentrums [Hei24]

Das Design und der Aufbau der Menüstrukturen dienen später in dieser Arbeit als Vorbild für die Entwicklung einer eigenen grafischen Benutzeroberfläche für die Evaluation der Effizienz von Sprachsteuerung im Vergleich zu einem konventionellen, haptisch-visuellen HMI (vgl. **Kapitel 5.2.2**).

2.3 Neuartige nutzerzentrierte Human-Machine Interfaces

Im Kontrast zu den in **Kapitel 2.2** vorgestellten konventionellen, haptisch-visuellen HMIs werden im Folgenden neuartige, nutzerzentrierte HMI-Modalitäten beschrieben. Im Fokus stehen dabei neuartige HMI-Ansätze wie Gestensteuerung, Eye Movement Control, BCIs und Sprachsteuerung, die bereits im Kontext der Steuerung von Produktionsmaschinen wissenschaftlich evaluiert und/oder sogar bereits prototypisch in der Industrie eingesetzt werden. Neben der grundsätzlichen Funktion dieser sogenannten natürlichen Benutzerschnittstellen (engl.: natural user interfaces (NUI)) wird auf die Historie der Entwicklung, Medizin- und Forschungsanwendungen

sowie abschließend auf industrielle Anwendungen in der Produktion eingegangen. Dadurch werden das Funktionsprinzip, die Vorteile und Herausforderungen sowie der technische Reifegrad im Verhältnis zur in dieser Arbeit untersuchten Sprachsteuerung eingeordnet.

In **Kapitel 2.3.1** wird mit der Gestensteuerung als die HMI-Modalität begonnen, die nach der Sprachsteuerung mit ihrer bereits mehrjährigen praktischen Anwendung in den Bereichen Automotive, Smartphone, Hausautomation und Gaming einer flächendeckenden Industrialisierung am zweitnächsten steht. Anschließend folgt in **Kapitel 2.3.2** das Eye Tracking mit vereinzelt praktischen Anwendungen in den Bereichen Automotive, Smartphone, Hausautomation und Gaming auf dem dritten Platz der neuartigen und nutzerzentrierten HMIs. Danach wird in **Kapitel 2.3.3** das BCI als die HMI-Modalität vorgestellt, die noch am weitesten von einer flächendeckenden Industrialisierung entfernt, aber dennoch intensiv in diesem Anwendungsbereich erforscht wird. Zuletzt wird in **Kapitel 2.3.4** die industrielle Sprachsteuerung als das in dieser Arbeit im Fokus stehende neuartige, nutzerzentrierte HMI vorgestellt.

2.3.1 Gestensteuerung / Gesture-based Interaction

Gesten sind eine nonverbale Form der Kommunikation, die von Menschen genutzt wird, um auf das Vorhandensein von Objekten hinzuweisen, auf Objekte oder Menschen zu zeigen oder die Einstellung zu etwas zum Ausdruck zu bringen. Eine Geste beginnt mit einer entspannten Haltung der Hand vor dem Körper. Der bedeutende Teil der Geste erfolgt anschließend durch eine von der Art und Bedeutung der Geste abhängigen Bewegung der Hand, bevor die Hand wieder in einer entspannten Haltung vor dem Körper positioniert wird. Gesten werden dabei nach autonomen Gesten, die Sprache vollständig ersetzen und co-verbalen Gesten, die mit anderen Modalitäten (in erster Linie mit Sprache) kombiniert werden. [BrLa11, Gra20] Nach Saffer wird die Gestensteuerung (engl.: gesture-based interaction (GBI)) in Touchscreen-basierte und free-form oder in-air Gesten kategorisiert. Beide Modalitäten nutzen ein Eingabegerät, wie entweder einen Touchscreen, spezielle GBI-Handschuhe oder Fernbedienungen sowie Kameras oder Infrarot Sensoren zur Erfassung von Körperbewegungen. Während die Gesten zur Interaktion mit Touchscreens auf Tippen (Betätigung von Interaktionsflächen mit einem Finger), Sliden (Hoch- und Runterscrollen entlang der vertikalen Achse in einer Menüebene), Swipen (Wechsel von Menüebenen in der horizontalen Achse) und Zoomen (Vergrößern oder Verkleinern von grafischen Inhalten auf einer Benutzeroberfläche) beschränkt sind, bieten die free-form oder in-air Gesten deutlich mehr Gestaltungsspielraum. [Gra20, Saf08] Zur Erfassung dieser trägt der Be-

diener spezielle GBI-Handschuhe mit piezoelektrischen Sensoren oder hält eine GBI Fernbedienung in der Hand [FMP23, Gra20]. Am flexibelsten und intuitivsten ist hingegen die Erfassung von Gesten über Kameras und/oder Infrarot Sensoren ohne zusätzliche Handschuhe oder Fernbedienungen, wobei diese Art der free-form Eingabe Einschränkungen hinsichtlich der Robustheit mit sich bringt. [Gra20, Saf08]

2.3.2 Eye Tracking / Gaze Control / Eye Movement Control

Die Aufnahme der Pupillenbewegung und die daraus resultierenden Blickrichtung von Menschen mit Sensoren, wie Kameras wird als Eye Tracking, Gaze Control oder Eye Movement Control bezeichnet. Im Folgenden wird für alle drei in der Praxis redundant verwendete Begriffe die Bezeichnung Eye Tracking verwendet. Ursprünglich stammt diese Methode aus der Psychologie zur Messung von kognitiver Belastung, wird aber aktuell auch in der Forschung zu Human-Computer Interfaces (HCI), bei der Softwareentwicklung sowie als HMI zur Steuerung von technischen (Assistenz-)Geräten für Menschen mit Behinderung eingesetzt. Die Überwachung kleiner Bewegungen der Iris, Pupille und Sklera im Auge erlaubt als direkte Schnittstelle zum Gehirn direkte Rückschlüsse auf die Anwesenheit, Aufmerksamkeit und den Fokus des Menschen. Die Ergebnisse des Eye Tracking werden meist in sogenannten Heat Maps oder Gaze Plots dargestellt. Diese zeigen sowohl die Häufigkeit von fokussierten Punkten als auch deren Verteilung in einem GUI an. Auf Basis dieser Daten können in einem Eye Tracking-basierten HMI-Aktionen zur Steuerung von technischen Geräten ausgelöst werden. Vorteile dieser HMI-Modalität sind der nutzerzentrierte Ansatz, die Frequenz von Steuerungsimpulsen sowie die Barrierefreiheit und die daraus resultierenden freien Hände von Maschinenbedienern. [CZL23, MLK20, RCC21]

Vereinzelt wird Eye Tracking bereits im industriellen Kontext untersucht. Chen et al. nutzen Eye Tracking exemplarisch als nutzerzentrierte Schnittstelle mit einem selbst entwickelten GUI für die Human-Robot Collaboration (HRC) (dt.: Mensch-Roboter-Interaktion (MRI)). Das GUI beinhaltet vier Menüebenen, in dem über eine an den PC angeschlossene Kamera vier unterschiedliche Werkzeuge (Schraubenzieher, Innensechskantschlüssel, Schraubenschlüssel und Seitenschneider) und Bauteile (Block, Dichtung, Schraube und Prisma) identifiziert und nach diesen beiden Klassen unterschieden werden. Ein Proband, der in dem Basis-Experiment 40 cm vor einer Webcam, mit einer Framerate von 30 fps und dem Bildschirm mit dem GUI platziert wird, steuert über vier verschiedene Blickrichtungen/Augenbewegungen (engl.: eye gaze) das GUI, um eines der vier Werkzeuge oder Bauteile auszuwählen und sich in der Zukunft, in einer nächsten Entwicklungsstufe durch einen Roboterarm anreichern zu lassen (vgl. **Abbildung 7**). [CZL23]

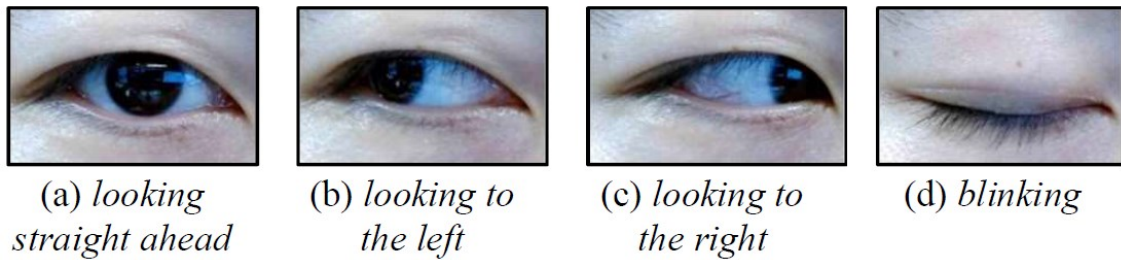


Abbildung 7: Eye Tracking – Vier verschiedene Blickrichtungen/Augenbewegungen zur Navigation in der grafischen Benutzeroberfläche, a) geradeaus schauen, b) nach links schauen, c) nach rechts schauen, d) blinzeln [CZL23]

In **Abbildung 8** wird die vierte und letzte Menüebene des GUI, in der die Auswahl eines Bauteils, über die Blickrichtung (geradeaus, links oder rechts) und Bestätigung durch ein Blinzeln erfasst wird. Zuvor wird in der dritten Menüebene durch einen Blick nach links oder rechts auf dem GUI zwischen den beiden Klassen »Werkzeuge« und »Bauteile« gewählt. [CZL23]

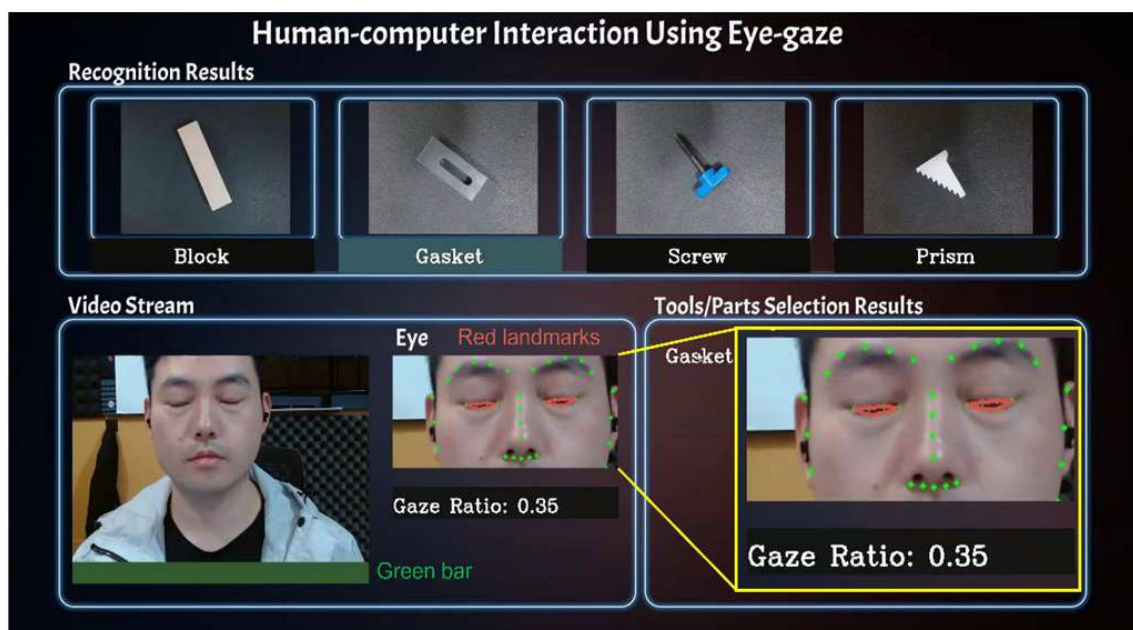


Abbildung 8: Eye Tracking - Grafische Benutzeroberfläche mit Auswahl eines Bauteils durch Auswertung der Blickrichtung und Bestätigung durch Blinzeln [CZL23]

Chen et al. berichten von Erkennungsraten zwischen ca. 94 % und ca. 100 % in Abhängigkeit vom Abstand des Probanden zur auf dem Bildschirm montierten Webcam (10-160 cm) und einer resultierenden durchschnittlichen Erkennungsrate von ca. 99 %. Dieses Praxisbeispiel demonstriert das grundsätzliche Potenzial des Eye Tracking als neuartige und nutzerzentrierte HMI-Modalität. Gleichzeitig konstatieren bereits Chen et al., dass in weiterführenden Studien Umgebungsbedingungen wie die Beleuchtungssituation, Reflexionen und Schatten sowohl im Training

als auch im eigentlichen Experiment berücksichtigt werden müssen. Zusätzlich wird die praktische Anwendbarkeit durch den Einsatz von Spezialkameras oder Brillen mit integrierten Kameras eingeschränkt. Des Weiteren wird die zeitliche Effizienz und Echtzeit-Fähigkeit durch die benötigte Rechenleistung von aktuellen Eye Tracking Modellen beeinflusst. [CZL23] Für ein aussagekräftiges Ergebnis sind mehr als ein Proband, eine Trennung zwischen Trainings- und Test-Datensets und für eine quantifizierte Untersuchung der Effizienz im Vergleich zum einem konventionellen HCI Zeitdaten erforderlich. Gleichwohl zeigen die guten Erkennungsraten in diesem exemplarischen Experiment, dass Eye Tracking eine mögliche Option für ein neuartiges und nutzerzentriertes sowie potenziell effizientes HMI in der Produktion ist.

2.3.3 Brain-Computer Interface

BCIs beschreiben Schnittstellen zwischen Mensch und Computer, die es Bedienern ermöglichen, auf Basis cerebraler Aktivitäten, die mittels Elektroenzephalografie (EEG) gemessen werden, Computer oder sogar technische Geräte mit physischen Aktoren zu steuern [NiGo12, SRA24]. Die Forschung in dieser Disziplin reicht bis in die 60er-Jahre des 20. Jahrhunderts zurück, wobei Jacques J. Vidal als Pionier auf diesem Gebiet gilt und mit seinem im Jahr 1973 erschienenen Artikel *»Toward Direct Brain-Computer Communication«* den Begriff *»Brain-Computer Interface«* etablierte [Vid73]. Praktische Anwendungen von BCIs haben ihren Ursprung in den Bereichen medizinische Rehabilitation, Neuroprothetik und assistiven Technologien. BCIs ermöglichen es beispielsweise Schlaganfallpatienten und Patienten mit Amyotropher Lateralsklerose (ALS) oder Rückenmarksverletzungen (engl.: spinal cord injury (SCI)) über cerebrale Aktivitäten im Gehirn, Exoskelette oder Neuroprothesen zu steuern, um Bewegungsabläufe zu trainieren und so Schritt für Schritt motorische Fähigkeiten zu rehabilitieren. Über visuelles, akustisches und taktilen Feedback erhalten die Patienten in Echtzeit eine Rückmeldung zu den aus den cerebralen Aktivitäten abgeleiteten Steuerungsbefehlen. Neben medizinischen Anwendungen werden BCIs in der Unterhaltungsindustrie bei der Entwicklung von Spielen eingesetzt. So können Spieler mit ihren eigenen Gedanken interagieren, was zu einem besonders immersiven Spielerlebnis führt. Des Weiteren werden BCIs in der neurowissenschaftlichen Forschung eingesetzt, um das Verständnis von Gehirnfunktionen und neuronalen Mechanismen zu vertiefen. So können unterschiedliche Stimuli des Gehirns in Echtzeit untersucht werden. [NiGo12, SRA24]

Bis jetzt werden BCIs nur im wissenschaftlichen Kontext in industriellen Produktionsszenarien eingesetzt und untersucht. Wesentliche Herausforderung ist aktuell die Signalqualität der mit

mobilen Endgeräten erhobenen EEG-Daten in Produktionsumgebungen mit vielen unterschiedlichen visuellen und akustischen Stimuli auf das Gehirn von Bedienern. Neben technischen Aspekten sind auch Einbußen im Komfort beim Tragen von mobilen EEG-Geräten und die Benutzerakzeptanz einer MMS, die Informationen direkt aus dem Gehirn des Bedieners auswertet, Hürden, die für eine Industrialisierung gemeistert werden müssen. Des Weiteren spielt auch die Gewährleistung des Datenschutzes in Produktionsumgebungen bei der Nutzung solch tiefgreifender Informationen aus dem menschlichen Gehirn eine wichtige Rolle. [SRA24] Zuletzt bestehen BCIs aus komplexen und damit kostenintensiven Elektronikkomponenten, deren Amortisierung nur durch starke Verbesserungen in der Gebrauchstauglichkeit, speziell der Effizienz in der Interaktion zwischen Mensch und Maschine in der Produktion erreicht wird. Trotz allem bietet die direkte Übersetzung von cerebralen Aktivitäten das Potenzial für ein hocheffizientes HMI in der Produktion, da keine zusätzliche Schnittstelle in Form der Bewegung von Extremitäten oder Augen oder die Sprache benötigt wird.

2.3.4 Sprachsteuerung / Voice Control

Für die Sprachsteuerung einer Produktionsmaschine mit akustischem Feedback werden über ein ortsfestes Mikrofon oder Mikrofon-Array mit zusätzlichen Lautsprechern oder über ein mobiles In-Ear oder Over-Ear Wearable analoge Sprachsignale erfasst und für die Verarbeitung in digitale Signale umgewandelt. Anschließend werden mithilfe einer Blockverarbeitung und Fourier Transformation die Frequenzspektren analysiert. Danach werden akustische Merkmale, wie Mel-Frequency Cepstral Coefficients (MFCCs), aus dem akustischen Signal extrahiert, um mit Sprachmodellen verglichen zu werden. Hidden Markov Modelle (HMM) oder neuronale Netze werden dazu eingesetzt, um die wahrscheinlichsten Wörter oder Phrasen zu identifizieren, bevor nach einem Abgleich mit den in der Steuerung hinterlegten Maschinenbefehlen eine Aktion der Maschine und optional ein simultanes Audio-Feedback ausgelöst wird. [JuMa08] Erst durch die Verbindung der ASR mit einer technischen Steuerung wie beispielsweise einer SPS, wird aus der ASR eine industrielle Sprachsteuerung.

Die Steuerung von technischen Systemen per Sprache ist eine HMI-Modalität, die in vielen Consumer-Geräten des Alltags bereits seit Jahren standardmäßig zur Verfügung steht. Sowohl die Interaktion mit dem Multimedia Interfaces (MMI) von Fahrzeugen als auch die Navigation in Smartphone-Menüs oder aber die Kommunikation mit smarten Hausautomatisierungen wird bereits über unterschiedliche Anbieter von Sprachsteuerungen ermöglicht. [Gei21, Gün22, KaMe20]

Die Kerntechnologie ist die sogenannte ASR, die die menschliche Sprache auf Basis einer Sprachsynthese in maschinenlesbare Strings transkribiert [PfKa17]. Die Motivationen zum Einsatz einer Sprachsteuerung in Consumer-Anwendungen ähneln oft denen in technischen Anwendungen. Deshalb werden im Folgenden zunächst die Vorteile einer Sprachsteuerung in Consumer-Anwendungen beschrieben, bevor in **Kapitel 2.5 & 2.6** gezeigt wird, wie diese bereits in wissenschaftlichen Untersuchungen genutzt wurden, um die Effizienz in Produktionsabläufen zu steigern.

Autofahrern ermöglicht die Sprachsteuerung eine anhaltende Fokussierung des Blicks auf den Straßenverkehr, mit beiden Händen am Lenkrad, bei kleinstmöglicher kognitiver Ablenkung durch die Anpassung von Klimafunktionen, Auswahl der Musik oder Eingaben in die Navigations-App des MMI. Während im Kontext des Autofahrens die Gewährleistung eines sicheren Hauptprozesses im Fokus steht, ermöglicht die Interaktion mit dem Sprachassistenten eines Smartphones die schnelle und damit effiziente Navigation durch komplexe Menüstrukturen trotz kleiner Bildschirmgröße. Auch in Wohnhäusern werden bereits Sprachsteuerungsdienste unterschiedlicher Anbieter genutzt, um über spezielle schaltbare (Relais-)Steckdosen oder direkte Verknüpfungen über eine Anwendungsprogrammierschnittstelle (engl.: application programming interface (API)) elektrische Geräte, wie Spül-/Waschmaschinen, Lampen, Garagentore, Heizungen, Rollläden oder ähnliches zu steuern. Diese Anwendung ermöglicht es den Hausbewohnern, ortsunabhängig, falls nötig weltweit, über ein mobiles Device, wie einen Homepod oder ein Smartphone, Haushaltsgeräte ein und auszuschalten, Türen und Tore zu öffnen und zu schließen sowie Statusmeldungen zu erhalten.

Obwohl die beschriebenen Vorteile einer Sprachsteuerung gegenüber einem konventionellem HMI auch in industriellen Anwendungen, abhängig vom Produktionsablauf, Vorteile sein können, sind sowohl die Anforderungen an ein HMI als auch die Auswahlkriterien im industriellen Kontext andere als in Consumer-Anwendungen [GhRe19].

Neben den in dieser Arbeit untersuchten Effizienzvorteilen zeigt das Anwendungsszenario der Steuerung einer Laserschmelzanlage (ugs.: Metall-3D-Drucker) (vgl. **Abbildung 9, I**) ergänzende Vorteile in Produktionsabläufen, wie die freihändige Steuerung der Maschine über Sprachbefehle (vgl. **Abbildung 9, V**) zur Vermeidung einer Kontamination mit gesundheitsschädlichem Metallpulver über konventionelle, haptisch-visuelle HMIs (vgl. **Abbildung 9, II**) in Rüst- und Nebenzeiten. Gleichzeitig ermöglicht die Sprachsteuerung die Kommunikation mit Kollegen (vgl. **Abbildung 9, VI**) in Umgebungen mit Störgeräuschen, wie laufender Maschinenperipherie (Kühlaggregat und explosionsgeschützter Pulverstaubsauger) bei gleichzeitig durch persönliche

Schutzausrüstung gedämpfter Stimme, wie den in **Abbildung 9** gezeigten Atemschutzhelmen (vgl. **Abbildung 9**, IV).

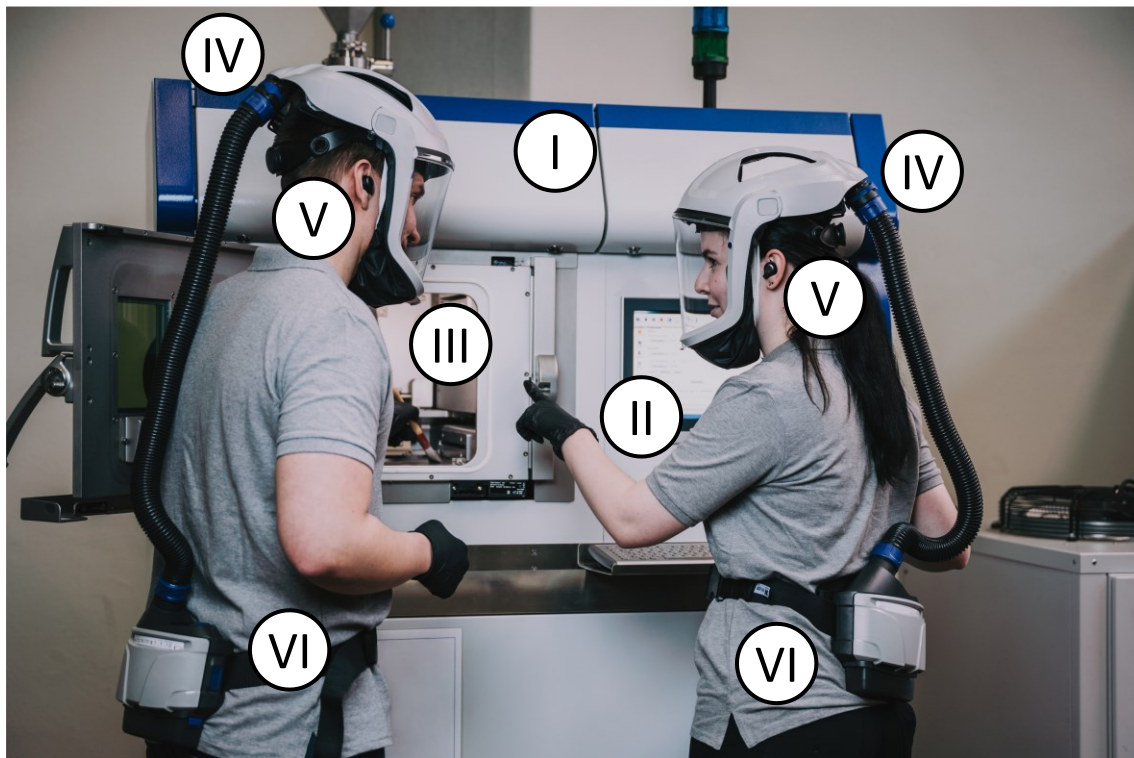


Abbildung 9: Sprachsteuerung über mobile In-Ear Wearables als Ergänzung zu einem konventionellen Human-Machine Interface einer industriellen Laserschmelzanlage, bestehend aus einem Touchscreen, Maus und Tastatur (I. Laserschmelzanlage (ugs.: Metall-3D-Drucker), II. haptisch-visuelles HMI (Touchscreen und leicht zu reinigende Edelstahl-Tastatur), III. Prozesskammer, IV. Atemschutz-Helme, V. In-Ear-Hedatasets (hearpieces), VI. Bediener/in) © Fraunhofer IDMT 2022 / Anika Bödecker

Obwohl Sprachsteuerungen bereits vereinzelt zur Steuerung von Maschinen in der Produktion eingesetzt werden, ist diese HMI-Modalität dort bei Weitem noch kein Standard. Vereinzelt wird sie jedoch weltweit in unterschiedlichen Kontexten erforscht. In den zugehörigen Veröffentlichungen wird zum Großteil von praktischen Herangehensweisen berichtet, in denen SDKs etablierter Consumer- oder Open-Source-Spracherkennungen genutzt werden, um mit einigen wenigen Befehlen eine Hand voll an Operationen an Produktionsmaschinen auszulösen. Der Fokus aller dieser Veröffentlichungen liegt auf der grundsätzlichen softwaretechnischen Verbindung einer Spracherkennung mit einer Maschinensteuerung.

So entwickelten Gayathri et al. 2022 eine dreiachsige Miniatur-CNC-Fräsmaschine, deren Arduino-Microcontroller-basierte Steuerung sie mit einem Spracherkennungsmodul (Elechouse Voice Recognition Module V3) koppelten. Das verwendete Spracherkennungsmodul basiert dabei auf der Google Spracherkennungs-API. Der Fokus des Projekts lag auf der Entwicklung der

Miniatur-3-Achs-CNC-Fräsmaschine sowie der grundsätzlichen Steuerungsmöglichkeit per Sprache. Motivation für die Integration einer Sprachsteuerung ist in dieser Untersuchung die Möglichkeit der freihändige Bedienung per Sprache als natürlichste Form der Kommunikation des Menschen [NMR22]. Besonders beliebt für prototypische Integrationen von Sprachsteuerungen sind Roboter, als weit verbreitete Vertreter der industriellen Produktionsmaschinen. Mehrere Veröffentlichungen beschreiben die Verbindung unterschiedlicher Spracherkennungen mit unterschiedlichen Robotern (-armen) [BiAy20, JHR21, JRV21, NER23, PYJ22, RPS22]. Etwas weniger Veröffentlichungen beschreiben die Verknüpfung einer Spracherkennung mit einem Hausautomatisierungssystem [YKK21], einem virtuellen Sprachassistenten an einem Drehbearbeitungszentrum [LoPa20], einem Datenmanagementsystem [AFA19] oder einer Aufzugssteuerung [LWL21]. Diese praxisnahen Ansätze demonstrieren die Integration von Sprachsteuerung in unterschiedlichen technischen Systemen und zeigen, dass die grundsätzliche Steuerung eines technischen Systems über Sprache möglich ist.

Neben der Schnittstelle ist die Robustheit der ASR gegenüber Störgeräuschen ein wichtiger Faktor für die Effektivität, als Teil der Gebrauchstauglichkeit dieser HMI-Modalität (vgl. **Kapitel 2.4.1**). In vereinzelt Studien wird die Leistungsfähigkeit der ASR und die Robustheit einer Sprachsteuerung in Störgeräuschumgebungen im Vergleich zu konventionellen HMI-Modalitäten untersucht.

Janiček et al. nutzten 2021 ein englisches Arduino-ASR-Kit mit 22 vordefinierten Sprachbefehlen aus dem Bereich der Hausautomation zur Steuerung eines Dobot Magician Lehrroboters. Ziel war es den Einsatz von Sprachsteuerung im Kontext der Interaktion mit kollaborativen Robotern hinsichtlich der Leistungsfähigkeit dieser neuartigen HMI-Modalität zu untersuchen. Insgesamt wurde jeder der 22 Sprachbefehle einhundertmal über ein Headset eingegeben, sodass insgesamt 2200 Befehlseingaben analysiert wurden. Während keiner der Sprachbefehle in 100 % der Fälle erkannt wurde, lag die Worterkennungsrates (engl.: word recognition rate (WRR)), in Abhängigkeit von der jeweiligen Art und Länge des Sprachbefehls zwischen 68 % und 98 %. Janiček et al. identifizierten sowohl die grundsätzliche Architektur der ASR als auch die Art und Länge der Sprachbefehle als wesentliche Einflussfaktoren auf die WRR und empfahlen einen weiteren Fokus auf diese Parameter in folgenden Untersuchungen der Leistungsfähigkeit einer ASR. [JRV21]

Ambadar et al. nutzten 2022 einen LEAP Motion Controller (LEAP) zur Gestensteuerung und eine Windows Speech Recognition API zur Entwicklung einer Vertrauensmetrik zur Bewertung der

Robustheit einer multimodalen Sprach-Gesten-Schnittstelle. Dazu wurden in 210 Testdurchläufen mit Störgeräuschen Befehle über Gesten und über Sprache eingegeben. Während die unimodale Sprachsteuerung bereits eine WRR von 96 % erreichte, konnte diese durch die Ergänzung einer Gestensteuerung auf bis zu 99 % gesteigert werden. [AmMa23]

Auf Basis der vorgestellten Veröffentlichungen ergeben sich die aufgelisteten fünf Hauptvorteile einer Sprachsteuerung im Verhältnis zu einem konventionellen haptisch-visuellen HMI [NMR22, SSK21].

- Fokus auf eine Haupttätigkeit zur Gewährleistung der Sicherheit des Prozesses
- Freihändige Steuerung zur effizienten Parallelisierung von Tätigkeiten
- Schnelle und damit effiziente Navigation durch komplexe Menüstrukturen
- Vermeidung von Laufwegen durch ortsunabhängige und dadurch effiziente Interaktion mit technischen Systemen
- Simultane und damit effiziente Interaktion mit mehreren technischen Systemen

Auf Basis dieser beschriebenen fünf Haupt-Vorteile, die in den Veröffentlichungen genannt und in Consumer-Anwendungen bereits genutzt werden, werden in **Kapitel 4** Einflussfaktoren auf die Effizienz einer Sprachsteuerung in teilautomatisierten Produktionsabläufen identifiziert (vgl. **Kapitel 2.5 & 2.6**).

2.4 Normen zur Ergonomie der Mensch-System- Interaktion, Ergonomischen Gestaltungsgrundsätzen und Evaluationsprinzipien

Der Großteil an wissenschaftlichen Veröffentlichungen zum Thema Sprachsteuerung von Produktionsmaschinen fokussiert sich auf die allgemeine Funktionalität einer Sprachsteuerung und/oder die technische Integration einer Sprachsteuerung in einer spezifischen Produktionsanwendung oder gibt Prognosen auf Basis existierender Literatur zur potenziellen Nutzung von Sprachsteuerung in der Produktionshalle der Zukunft (vgl. **Kapitel 2.3.4**) [AFA19].

Im Unterschied dazu demonstrieren Gross et al. zwar auch primär, dass die Integration einer Consumer-Sprachsteuerung (Amazon Echo Dot) in die Steuerung eines Arduino-Roboters technisch einfach möglich ist. Ergänzend konstatieren sie jedoch als eine der wenigen, dass die Entwicklung klarer Gestaltungsregeln für Sprachbefehle einen großen Einfluss auf die Gebrauchstauglichkeit (engl.: usability) des HMIs und damit auf die Effizienz der Interaktion zwi-

schen Mensch und Maschine hat. [GFS20] Auch Panter et al. machen auf einen Mangel an Anforderungen und Gestaltungsregeln für NUI in der Industrie 5.0 aufmerksam. Auf Grund der Komplexität von Produktionsabläufen und Produktionsmaschinen bedarf es Gestaltungsregeln zur praktischen Implementierung von NUIs, um gemäß den Paradigmen der Industrie 5.0 die Benutzererfahrung zu verbessern, die Produktivität zu steigern und nachhaltige und widerstandsfähige industrielle Ökosysteme mit dem Menschen im Mittelpunkt zu entwickeln. [BNP21, PLK24]

Die Ergonomie der Mensch-System-Interaktion oder umgangssprachlich »die Gebrauchstauglichkeit von Software« wird in der »DIN EN ISO 9241 – Ergonomie der Mensch-System Interaktion« beschrieben. Diese Norm besteht aus mehreren unterschiedlichen Teilen, die jeweils spezifische Bereiche wie Begriffe und Konzepte, Dialogführung mittels Menüs, Sprachdialogsysteme, die Menschzentrierte Gestaltung interaktiver Systeme und Prozesse zur Ermöglichung, Durchführung und Bewertung genauer betrachten. Die für diese Arbeit relevantesten Teile der DIN EN ISO 9241 sind im Folgenden aufgelistet.

- **DIN EN ISO 9241-11 Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018) [DIN18]**
- DIN EN ISO 9241-14 Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten – Teil 14: Dialogführung mittels Menüs (ISO 9241-14:1997) [DIN00]
- DIN EN ISO 9241-154 Ergonomie der Mensch-System-Interaktion – Teil 154: Sprachdialogsysteme (ISO 9241-154:2013) [DIN13]
- DIN EN ISO 9241-210 Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019) [DIN20b]
- DIN EN ISO 9241-220 Ergonomie der Mensch-System-Interaktion – Teil 220: Prozesse zur Ermöglichung, Durchführung und Bewertung (ISO 9241-220:2019) [DIN20a]

Neben den aufgelisteten Teilen der DIN EN ISO 9241 sind für diese Arbeit noch die »DIN EN 614-2 Sicherheit von Maschinen – Ergonomische Gestaltungsgrundsätze – Teil 2: Wechselwirkungen zwischen der Gestaltung von Maschinen und den Arbeitsaufgaben« im Kontext von Fehleingaben in HMIs (vgl. **Kapitel 2.4.6**) und die »ISO/IEC 25041:2012-10 System und Software-Engineering – Qualitätskriterien und Bewertung von System- und Softwareprodukten (SQuARE) – Evaluationshandbuch für Entwickler, Ankäufer und unabhängige Gutachter« im Kontext der systematischen Bewertung des Effizienzpotenzials einer Sprachsteuerung (vgl. **Kapitel 2.4.7**) relevant. [DIN08, ISIE12]

In den folgenden **Kapiteln 2.4.1 bis 2.4.7** werden die oben aufgelisteten Normen chronologisch vorgestellt und die Elemente, die in dieser Arbeit genutzt werden, um eine »*Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion*« zu entwickeln, beschrieben.

2.4.1 Definition von Gebrauchstauglichkeit, Effektivität, Effizienz und Zufriedenstellung im Kontext von Human-Machine Interfaces

Die Effizienz als Hebel zur Steigerung der Produktivität ist eine wesentliche Anforderung an NUIs im Kontext der Industrie 5.0 [BNP21, GFS20, PLK24]. In der »*DIN EN ISO 9241–11 Ergonomie der Mensch-System Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018)*« wird die Effizienz von MMS neben der Effektivität und Zufriedenstellung des Bedieners als Bestandteil der Gebrauchstauglichkeit über die Relation zwischen eingesetzten Ressourcen (Zeit, menschlicher Aufwand, Geld und Materialien) und den erreichten Ergebnissen definiert. Die Messgröße für die Messung der Effizienz von HMIs ist gemäß der DIN EN ISO 9241-11 die für die Ausführung einer Aufgabe benötigte Zeit inklusive der Korrektur etwaiger Eingabefehler. [DIN18]

In **Abbildung 10** wird im linken Kreis der Nutzungskontext, bestehend aus dem Benutzer, der Umgebung, den Ressourcen sowie den Zielen und Aufgaben in Relation zu einem System, einem Produkt oder einer Dienstleistung dargestellt [DIN18]. Für den in dieser Arbeit betrachteten Fall, das HMI einer Produktionsmaschine, das hinsichtlich seiner Effizienz untersucht wird, ist der Benutzer der Maschinenbediener, die Umgebung die Produktionshalle und die betrachtete Ressource die oben beschriebene Eingabezeit inklusive Korrektur etwaiger Eingabefehler. Die Ziele und Aufgaben leiten sich aus dem Produktionsablauf ab und werden mithilfe von Maschinenbefehlen erreicht. Im Zentrum des Nutzungskontexts steht, in dem in dieser Arbeit betrachteten Fall, ein HMI als System. Dessen Nutzung hat laut DIN EN ISO 9241-11 unterschiedliche Ergebnisse. Die Gebrauchstauglichkeit, bestehend aus der Effektivität, Effizienz und Zufriedenstellung wird als Hauptergebnis, neben der Zugänglichkeit, User Experience sowie der Vermeidung von nutzungsbedingten Schäden und gegebenenfalls weiteren kontextabhängigen Ergebnissen definiert [DIN18].

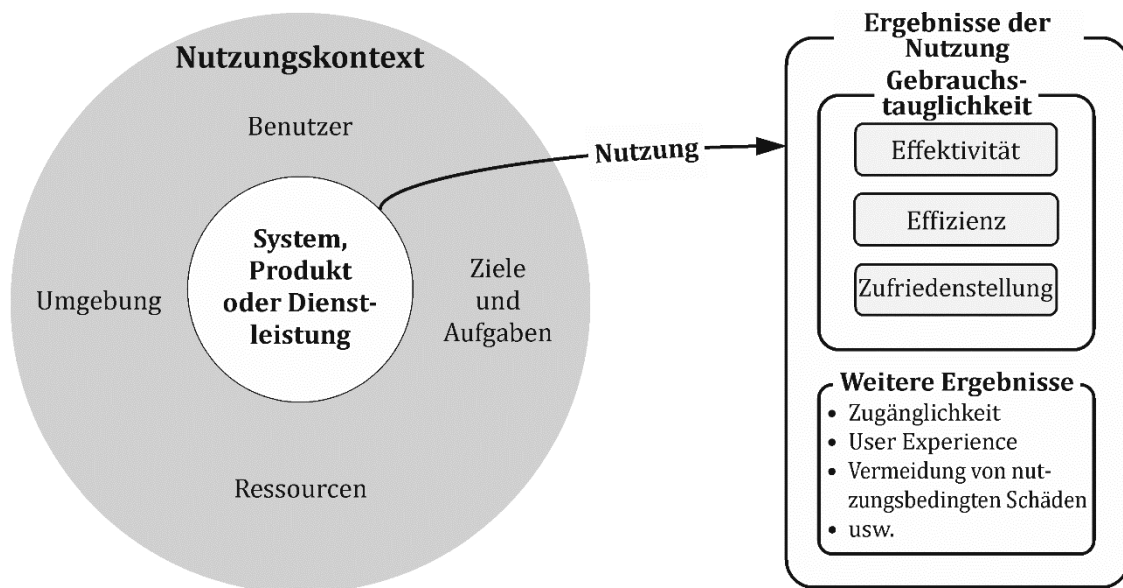


Abbildung 10: Gebrauchstauglichkeit als Ergebnis der Nutzung eines Systems, eines Produkts oder einer Dienstleistung in einem bestimmten Nutzungskontext [DIN18]

In dieser Arbeit liegt der Fokus auf der Betrachtung der Effizienz als Bestandteil der Gebrauchstauglichkeit. Diese ist damit Ergebnis der Interaktion eines Maschinenbedieners mit einem industriellen HMI als Komponente einer Produktionsmaschine.

2.4.2 Gestaltung von konventionelle haptisch-visuellen Human-Machine Interfaces als Ausgangsbasis

Die »DIN EN ISO 9241–14 Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten – Teil 14: Dialogführung mittels Menüs (ISO 9241-14:1997)« beschreibt die ergonomische Gestaltung von Menü-Systemen (Menü-Dialoge). Diese konventionelle, haptisch-visuelle Form der HMI dient in dieser Arbeit als Ausgangsbasis für die Entwicklung einer Methode zur Evaluation des Effizienzpotenzials einer Sprachsteuerung. Die DIN EN ISO 9241-14 besteht aus einer Auflistung von Empfehlungen zur Gestaltung von ergonomischen Menüsystemen. In dieser Arbeit werden diese Empfehlungen genutzt, um ein menübasierte grafische Benutzeroberfläche (engl.: graphical user interface (GUI)) als Pendant zu einer menschenzentrierten (im Folgenden »nutzerzentrierten« genannten) Sprachsteuerung zu entwickeln, um den Einflussfaktor **Komplexität von Menüstrukturen** in einem konventionellen haptisch-visuellen HMI auf das Effizienzpotenzial einer Sprachsteuerung zu quantifizieren (vgl. **Kapitel 5.2.2**). [DIN00]

2.4.3 Gestaltung von Spracheingaben eines nutzerzentrierten Human-Machine Interface

Für die allgemeine Gestaltung von Benutzerschnittstellen mit automatischer Spracherkennung (engl.: Automatic Speech Recognition (ASR)) als Teil nutzerzentrierter HMI existieren (noch) keine Normen und/oder Gestaltungsregeln. Dies ist die stärkste Motivation zur Entwicklung einer »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« in dieser Arbeit, um ein Werkzeug zu schaffen mit dem die Vorteile einer Sprachsteuerung systematisch in der Industrie aufgezeigt und genutzt werden können. Die »DIN EN ISO 9241–154 Ergonomie der Mensch-System-Interaktion – Teil 154: Sprachdialogsysteme« von 2013 fokussiert sich ausdrücklich auf Sprachdialogsysteme unter Anwendung eines Telefons als sogenannter »Informationsübertragungsmechanismus«. [DIN13] Nichtsdestotrotz können Empfehlungen und Gestaltungsgrundsätze, die die zu Grunde liegende ASR betreffen auch auf den Einsatz einer industriellen Sprachsteuerung übertragen werden. Dazu zählt unter anderem die Empfehlung zum gleichen Zeitpunkt in einem Produktionsablauf zur Verfügung stehende Optionen an Sprachbefehlen so zu gestalten, dass diese sich phonetisch größtmöglich unterscheiden, um die Effektivität der Spracherkennung und die daraus resultierende Effizienz in der Interaktion zu gewährleisten [DIN13]. Phoneme bilden die kleinsten bedeutungsunterscheidenden sprachlichen Einheiten einer Sprache. Mithilfe der sogenannten »Minimalpaaranalyse« können beispielsweise die Phoneme [h] und [m] wie im Beispiel Haus (phonetisch: [haʊ̯s]) und Maus (phonetisch: [maʊ̯s]) identifiziert werden. [PfKa17] Neben dem Phonem-Abstand sollten sich die Grammatik in den Sprachbefehlen je nach Maschinenzustand unterscheiden, um dem Bediener eine leichtere Orientierung im Arbeitsablauf mit der Sprachsteuerung zu ermöglichen [DIN13]. Gleichzeitig sollte der Aufbau der Sprachbefehle aber einem festen Schema folgen, damit Bediener sich diese auch logisch herleiten können, wenn sie den exakten Sprachbefehl vergessen haben. Synonyme Bestandteile in den Sprachbefehlen bei gleichzeitiger Gewährleistung des Phonem Abstands ermöglichen es dem Bediener den Sprachbefehl intuitiv zu formulieren [DIN13].

Diese Gestaltungsgrundsätze der größtmöglichen phonetischen Distanz zwischen Sprachbefehlen ist für die Forschungsstudie in **Kapitel 5** wichtig, in der eine selbst entwickelte Sprachsteuerung eingesetzt wird, um den Einflussfaktor Komplexität von Menüstrukturen auf die Effizienz einer Sprachsteuerung zu quantifizieren (vgl. **Kapitel 5.2.1 & 5.2.2**). Der Trade-Off zwischen grammatikalischen Unterschieden von Sprachbefehlen und dem Folgen eines festgelegten Sche-

mas zur intuitiven Identifikation von Sprachbefehlen wird ebenfalls in **Kapitel 5** benötigt. In **Tabelle 3** werden die Sprachbefehle für unterschiedliche Maschinenkomponenten und -zustände am Beispiel einer selbst entwickelten Technologieplattform gezeigt. Für jeden der in **Kapitel 5** eingesetzten Sprachbefehle stehen auch synonyme Sprachbefehle zur Verfügung, um eine größtmögliche Intuitivität in der Bedienung zu gewährleisten.

2.4.4 Gestaltung von nutzerzentrierten Human-Machine Interfaces zur Steigerung der Gebrauchstauglichkeit

Der Ansatz der Industrie 5.0, HMIs an den Bedienern sowie ihren Erfordernissen und Anforderungen auszurichten, um so gebrauchstaugliche HMIs und letztlich gebrauchstaugliche Produktionsmaschinen zu erhalten, wird in der »DIN EN ISO 9241–210 Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019)« aufgegriffen. Die nutzerzentrierte Gestaltung von HMIs und Produktionsmaschinen verbessert die Gesamtqualität von Produktionsabläufen durch die im Folgenden aufgelisteten Faktoren. [DIN20b]

- **Steigerung der Gebrauchstauglichkeit (Effektivität, Effizienz und Zufriedenstellung)**
- **Steigerung der Produktivität der Bediener und der Wirtschaftlichkeit von Unternehmen**
- Steigerung der Barrierefreiheit (für Menschen aus einer Bevölkerungsgruppe mit dem weitesten Umfang an Benutzererfordernissen, Merkmalen und Fertigkeiten)
- Verbesserung der User Experience
- Reduzierung der Kosten für Schulung und Betreuung, durch leichteres Verständnis
- Reduzierung von Unbehagen und Stress
- Schaffung eines Wettbewerbsvorteiles durch Schärfen des Markenbilds des Systems
- Beiträge zum Erreichen von Nachhaltigkeitszielen

Im Besonderen die Steigerung der Gebrauchstauglichkeit, bestehend aus Effektivität, Effizienz und Zufriedenstellung, von HMIs sowie die daraus resultierende Steigerung der Produktivität von Bedienern und Wirtschaftlichkeit von Unternehmen ist eine Motivation den systematischen Einsatz einer nutzerzentrierten Sprachsteuerung zu untersuchen. Zusätzlich ermöglicht die Steigerung der Barrierefreiheit Menschen mit Behinderung die Teilhabe am Arbeitsleben und gegebenenfalls sogar den Zugang zum ersten Arbeitsmarkt. Dadurch kann der Druck auf den Arbeitsmarkt reduziert werden, da Unternehmen ein neues Segment an Arbeitnehmern eröffnet wird. Die sogenannte »User Experience« wird als »Wahrnehmungen und Reaktionen einer Person, die

aus der tatsächlichen und/oder der erwarteten Benutzung eines Systems, eines Produkts oder einer Dienstleistung resultieren« definiert. Eine positive User Experience gewährleistet die Zufriedenstellung des Bedieners und sichert damit mittel- und langfristig die Effizienz der Interaktion mit einem HMI (vgl. **Kapitel 2.4.1**). Intuitive und selbsterklärende HMIs reduzieren nicht nur die Kosten für die Schulung und die Betreuung von Bedienern, sondern auch Unbehagen und Stress, der zu Fehlern und damit einer geminderten Effizienz führen kann. Die Schärfung eines Markenbilds durch nutzerzentrierte HMI zur Schaffung eines Wettbewerbsvorteils sowie Beiträge nutzerzentrierter HMIs zum Erreichen von Nachhaltigkeitszielen haben keinen wesentlichen Einfluss auf das in dieser Arbeit untersuchte Effizienzpotenzial einer Sprachsteuerung. Trotzdem werden die beiden Punkte der Vollständigkeit halber mit aufgeführt. [DIN20b]

2.4.5 Prozesse zur Gestaltung von nutzerzentrierten Human-Machine Interfaces

In der »DIN EN ISO 9241–220 Ergonomie der Mensch-System-Interaktion – Teil 220: Prozesse zur Ermöglichung, Durchführung und Bewertung menschenzentrierter Gestaltung für interaktive Systeme in Hersteller- und Betreiberorganisationen« werden Prozesse, die bewährte projektinterne und projektübergreifende Verfahrensweisen zur nutzerzentrierten Gestaltung beinhalten, beschrieben. Damit baut die DIN EN ISO 9241-220 mit strukturierten Prozessen zur Umsetzung auf die in DIN EN ISO 9241-210 beschriebenen Grundsätze eines nutzerzentrierten Ansatzes auf. Zur Implementierung von nutzerzentrierten Grundsätzen in einem Unternehmen wird ein Modell aus vier unterschiedlichen Prozessstufen empfohlen, die im Folgenden aufgelistet sind. [DIN20a]

1. Sicherstellen der Unternehmensausrichtung auf nutzerzentrierte Qualität
2. Ermöglichen einer projekt- und systemübergreifenden menschenzentrierten Gestaltung
- 3. Durchführung menschenzentrierter Gestaltung innerhalb eines Projektes**
4. Einführung, Betrieb und Ende der Lebensdauer eines Systems

Im ersten Schritt ist es notwendig übergeordnet im Unternehmen ein Bewusstsein für nutzerzentrierte Ansätze zur Gestaltung von Systemen zu wecken und zu verankern. Dies ist der Grundstein, um ein ganzheitliches und damit projekt- und systemübergreifendes Bewusstsein bei Mitarbeitenden in der Gestaltung von Produktionsabläufen zu etablieren. Die in dieser Arbeit entwickelte Methode ist ein Werkzeug um den dritten Schritt, die exemplarische Gestaltung eines nutzerzentrierten Produktionsablaufs, mit einer Evaluation des Effizienzpotenzials einer Sprachsteuerung und einer darauf aufbauenden Prognose der Reduzierung von Nebenzeiten zu unterstützen (vgl. **Kapitel 3**). Zusätzlich kann die entwickelte Methode auch im vierten Schritt

im Rahmen des Betriebs genutzt werden, um das prognostizierte Effizienzpotenzial nach der Integration einer Sprachsteuerung im realen Betrieb zu validieren. [DIN20a]

2.4.6 Ergonomische Gestaltungsgrundsätze auf Basis der Sicherheit von Produktionsmaschinen

Nachdem die DIN EN ISO 9241 mit ihren verschiedenen für diese Arbeit relevanten Teilen die Ergonomie der Mensch-System-Interaktion aus der Sicht der Interaktion des Bedieners mit einem HMI beschreibt, widmet sich die »DIN EN 614-2 Sicherheit von Maschinen – Ergonomische Gestaltungsgrundsätze – Teil 2: Wechselwirkungen zwischen der Gestaltung von Maschinen und den Arbeitsaufgaben« ergonomischen Gestaltungsgrundsätzen aus der Sicht der Maschinensicherheit. Unter anderem verweist die DIN EN 614-2 auf ausreichende, für den Bediener sinnvolle Rückmeldungen von Produktionsmaschinen, um so Fehleingaben schnellstmöglich zu erkennen und zu korrigieren (idealerweise mit Vorschlägen an den Bediener). [DIN08] Dieses Feedback der Maschine ist wichtig, um die Effizienz der Eingabe zu gewährleisten (vgl. **Kapitel 2.4.1**) [DIN18]. Zusätzlich wird in der DIN EN 614-2 empfohlen eine Über- oder Unterforderung sowie repetitive Arbeitsaufgaben im Produktionsablauf zu vermeiden, um Fehler zu vermeiden und die Effizienz zu gewährleisten [DIN08]. Dazu zählen auch komplexe oder gar willkürliche Menüstrukturen sowie das immer wiederkehrende Scrollen durch Menüebenen. [DIN00] Dieser Gestaltungsgrundsatz ist eine weitere Motivation zum Einsatz einer nutzerzentrierten Sprachsteuerung zum Durchbrechen komplexer Menüstrukturen (vgl. **Kapitel 5**).

2.4.7 Evaluation von Human-Machine Interfaces in industriellen Produktionsabläufen

Die sogenannten »System- and Software Quality Requirements« (SQuaRE) (dt.: System- und Software Qualitätsanforderungen) beziehen sich auf eine Reihe von Normen aus der ISO/IEC-25000-Serie. Diese widmen sich der Qualitätsbewertung von Software und softwarebasierten Systemen. Die »ISO/IEC 25041:2012-10 System und Software-Engineering – Qualitätskriterien und Bewertung von System- und Softwareprodukten (SQuaRE) – Evaluierungshandbuch für Entwickler, Ankäufer und unabhängige Gutachter« beinhaltet generelle Anforderungen und Empfehlungen zur Bewertung von Software und softwarebasierten Systemen sowie Beispiele für übergeordnete Bewertungskonzepte. Zu den softwarebasierten Systemen zählen sowohl konventionelle, haptisch-visuelle als auch neuartige, nutzerzentrierte HMIs. Die Zielgruppe der ISO/IEC 25041:2012-10 sind Software-Entwickler, Einkäufer von Software in Unternehmen und

unabhängige Gutachter. Gemäß ISO/IEC 25041:2012-10 wird die Evaluation als **»systematische Ermittlung, inwieweit eine Einheit ihre festgelegten Merkmale erfüllt«** definiert. [ISIE12] In der in dieser Arbeit entwickelten Methode werden Einheiten, die in dieser Arbeit als Einflussfaktoren bezeichnet werden daraufhin überprüft, inwieweit sie ihre festgelegten Merkmale, die in dieser Arbeit als Parameter bezeichnet werden, erfüllen (vgl. **Kapitel 3**).

Der Evaluationsprozess kann laut ISO/IEC 25041:2012-10 die im Folgenden aufgelisteten Ressourcen einschließen.

- Anwendbare Messwerkzeuge und Methoden
- Anwendbare Normen aus der SQuaRE-Serie
- Personalressourcen für die Evaluation
- Finanzressourcen für die Evaluation
- Informationssysteme für die Evaluation
- Wissensdatenbanken für die Qualitätsbewertung

Messwerkzeuge und Methoden, die für die Evaluation des Effizienzpotenzials einer Sprachsteuerung genutzt werden können, werden in **Kapitel 2.7** beschrieben. Die in dieser Arbeit entwickelte Methode (vgl. **Kapitel 3.2**), die unter anderem auf den in **Kapitel 2.7** beschriebenen Messwerkzeugen und Methoden basiert, zählt zusammen mit den zugehörigen Messwerkzeugen (vgl. **Kapitel 3.2.8**) ebenfalls zum ersten Punkt der aufgelisteten Ressourcen. Die entwickelte Methode basiert dabei, wie in **Kapitel 3.1** beschrieben nicht nur auf anwendbaren Normen der SQuaRE-Serie, sondern auch auf den in diesem **Kapitel 2.4** beschriebenen Normen.

2.5 Einflussfaktoren auf die Gebrauchstauglichkeit industrieller Human-Machine Interfaces

Außerhalb der in **Kapitel 2.3.4** aufgeführten praktischen Untersuchungen zu Schnittstellen zwischen unterschiedlichen Spracherkennungen und Maschinensteuerungen sowie der Leistungs- und Robustheitsbewertung industrieller Sprachsteuerungen werden nur in wenigen Studien übergeordnete Einflussfaktoren auf HMIs oder gar konkrete Einflussfaktoren speziell auf die Effizienz einer Sprachsteuerung in der Produktion untersucht [AFA19]. Im Folgenden werden die wenigen Veröffentlichungen zu Einflussfaktoren auf die Gebrauchstauglichkeit einer Sprachsteuerung, die existieren, in chronologischer Reihenfolge, beginnend mit der Neuesten

vorgestellt. Auffallend ist, dass je neuer die Veröffentlichung ist, desto konkreter werden Einflussfaktoren und der Bedarf nach quantitativen Methoden zur Evaluation dieser artikuliert. Dies spricht für eine steigende Bedeutung des Themas.

Im Kontrast zur Untersuchung der technischen Verbindung einer Sprachsteuerung mit einer Produktionsmaschine untersuchte Nzuva die Effizienz aktueller HMIs an industriellen Produktionsmaschinen in Kenia. Dazu führte er eine strukturierte Befragung unter 15 Maschinenbedienern, die in Kenia arbeiten, mindestens zwei Jahre Berufserfahrung und umfangreiches Wissen über HMIs haben sowie aktuell als Maschinenbediener, Vorarbeiter oder technischer Leiter arbeiten durch. 26,67 % der Befragten gaben an, dass die Navigation durch die Menüstrukturen kompliziert ist, während 53,33 % angaben, dass für den Produktionsablauf wichtige Funktionen nur schwer identifizierbar sind. Neben der leichten Identifikation von benötigten Funktionen gaben 33,33 % der Befragten an, dass eine Optimierung des HMIs zur Steigerung der Effizienz notwendig ist, während 20 % eine nutzerzentrierte HMI-Modalität zum Durchbrechen von komplexen Menüstrukturen als notwendig erachteten. Die durchgeführte strukturierte Befragung der 15 Maschinenbediener zeigt, dass die **Komplexität von Menüstrukturen** ein Einflussfaktor auf die Effizienz von HMIs in Produktionsabläufen ist. Des Weiteren sehen bis zu einem Drittel der Befragten ein Potenzial in nutzerzentrierten HMIs wie Sprachsteuerung zur Steigerung der Produktivität. Die geringe Anzahl an Befragten sowie die ausschließliche Fokussierung auf das subjektive Empfinden der Bediener über eine strukturierte Erfahrung zeigen zwar einen Trend. Für aussagekräftigere Ergebnisse sind jedoch quantitative Untersuchungen notwendig. [Nzu21]

Schulte et al. 2021 nutzten eine Spracherkennungs-API (Vosk/Alpha Cephei) an einem manuellen Montagearbeitsplatz, um die Betriebsdatenerfassung, als nicht wertschöpfende Dokumentationstätigkeit in der manuellen Montage, zum eigentlichen Montageprozess zu parallelisieren. Zum einen soll diese direkte Dokumentation Eingabe- und Übertragungsfehler reduzieren und zum anderen die Effizienz des Montageablaufs steigern. In dem Anwendungsszenario von Schulte et al. wurde die Sprachsteuerung eingesetzt, um Bedienern die parallele Interaktion mit einem Montagearbeitsplatz und einem System zur Betriebsdatenerfassung (BDE), mit dem Ziel der Steigerung der Datenqualität und Reduzierung von nicht direkt wertschöpfenden Nebenzeiten. Die Testphase zeigt, dass sowohl die **Anzahl an simultan zu bedienenden Maschinen** als auch die direkte Spracheingabe ohne die **Komplexität von Menüstrukturen** ein Einflussfaktor auf die Effizienz von Produktionsabläufen ist. Zur quantitativen Bewertung des Effizienzpotenzials einer Sprachsteuerung in der beschriebenen Art von Produktionsabläufen sind jedoch auch hier weitere Untersuchungen notwendig. [SSK21]

Villani et al. entwickeln eine Methode zur Auslegung adaptiver, industrieller HMIs für komplexe Automatisierungssysteme in Abhängigkeit von Einflussfaktoren des Bedieners als eine der zwei Seiten eines HMIs (vgl. **Kapitel 4.2**). Dies sind dauerhafte oder zumindest längerfristige Einflussfaktoren, wie Erfahrungsmangel oder Behinderungen, die die Nutzung der HMI-Modalität einschränken, aber auch temporäre Einflussfaktoren, wie Müdigkeit oder hektische Arbeitsabläufe. Ergänzend werden auf Basis einer Literaturrecherche allgemeine Richtlinien für das Design eines adaptiven HMIs innerhalb einer Methode zusammengefasst. Der Fokus liegt dabei auf der Auswahl einer visuellen, haptischen oder auditiven HMI-Modalität in Abhängigkeit des Alters, des Bildungsniveaus, der Computerkenntnisse und möglichen Behinderungen von Bedienern. Villani et al. haben erkannt, dass der Einsatz von nutzerzentrierten HMIs die Produktivität in Arbeitsabläufen verbessern kann, es jedoch Methoden zu der systematischen Nutzung der Vorteile der jeweiligen HMI-Modalität bedarf. Mit ihrer Arbeit zeigen sie, dass es aktuell noch an Einflussfaktoren mangelt, die sich über konkrete Parameter quantifizieren lassen, mangelt. [VSL21]

Lodgaard et al. untersuchten organisatorische Einflussfaktoren auf die Effizienz von nutzerzentrierten HMIs in Produktionsabläufen. Dazu erhoben sie im Rahmen einer einjährigen Fallstudie mithilfe von teilstrukturierten/freien Befragungen, Shopfloor-Besuchen, Workshops und Shadowing sowie ergänzenden Dokumenten Daten zu konventionellen und neuartigen HMIs. Ziel war es, mithilfe der Daten aus dem norwegischen first-tier Automobilzulieferer, mit ca. 600 Mitarbeitern organisatorische Einflussfaktoren zu identifizieren, die bei der Einführung neuartiger HMIs im Austausch zu konventionellen HMI berücksichtigt werden müssen, um die Gesamtanlageneffektivität (engl.: overall equipment effectiveness (OEE)) zu steigern. In der durchgeführten Fallstudie wurde die schnelle, ortsunabhängige Bereitstellung von Informationen zu Maschinenausfällen und Werkzeugbeschädigungen als wesentlicher Einflussfaktor auf die Effektivität identifiziert. Diese digitalen Informationen unterstützen Maschinenbediener dabei, Entscheidungen zu treffen und gewährleisten effiziente Produktionsabläufe. In dem von Lodgaard et al. beschriebenen Anwendungsszenario wird gezeigt, dass sowohl die **Art und Anzahl von Laufwegen** als auch die **Anzahl an simultan zu bedienenden Maschinen** (oder Systemen) Einflussfaktoren auf das Effizienzpotenzial einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen HMI sind. Auch in dieser Untersuchung wird klar, dass der Bedarf zum Einsatz von nutzerzentrierten HMIs zur Steigerung der Produktivität vorhanden ist, es jedoch an Methoden zur quantitativen Bewertung der bereits qualitativ identifizierten Einflussfaktoren mangelt. [LoDr20]

Longo et al. untersuchten das Konzept einer sozial-smarten Fabrik (engl.: social smart factory) auf Basis eines multimodalen virtuellen Assistenten (engl.: virtual assistant) mit Touch- und Sprachsteuerung. Dieser wurde in eine CNC-Drehmaschine integriert, um dem Bediener zum einen Informationen über den Betriebszustand der Maschine sowie Fehlermeldungen und zum anderen Schritt-für-Schritt Hilfe für die Behebung letzterer bereitzustellen. Sie identifizierten ein besseres Lernverhalten, den schnelleren Abruf von Informationen, eine höhere Maschinensicherheit sowie eine reduzierte kognitive Belastung als Vorteile neuartiger HMI, wie Sprachsteuerung. Gleichzeitig betonten sie, dass die Adaptivität, mit der sich ein HMI an die Fähigkeiten und Bedürfnisse des Maschinenbedieners anpasst, ein wesentlicher Einflussfaktor auf die Effizienz eines HMIs ist. Auch Longo et al. sehen die **Komplexität von Menüstrukturen** in konventionellen, haptisch-visuellen HMIs als Einflussfaktor auf das Effizienzpotenzial einer Sprachsteuerung. Zur quantitativen Bewertung fehlt es jedoch auch in diesem Anwendungsszenario an Methoden und ausreichend Probanden. [LoPa20]

Lorenz beschreibt die Entwicklung eines mehrschichtigen Bewertungssystems für neuartige und nutzerzentrierte HMIs, um diese hinsichtlich ihrer Gebrauchstauglichkeit, Produktivität und Sicherheit in industriellen Anwendungen vergleichbar zu machen. Dabei betrachtet er HMIs in einem Dreieck zwischen Mensch, Maschine und Anwendung/Szenario/Umwelt. Lorenz Ansatz zur Quantifizierung von Einflussfaktoren ist einer der wenigen existierenden konkreten Vorschläge, um unter anderem die Effizienz von NUIs zu bewerten. Leider gibt es zum aktuellen Zeitpunkt noch keine Veröffentlichung zum fertig entwickelten Bewertungssystem. [Lor20]

Ghofrani et al. untersuchten den Einsatz intelligenter Assistenzsysteme aus dem Home- und Unterhaltungsbereich (Amazon Alexa, IBM Watson, etc.) zur Sprachinteraktion mit kollaborativen Robotern. Im Speziellen untersuchten sie, welche die aktuellen Hürden sind, die eine flächendeckende Integration von Sprachsteuerung in der Produktion verhindern, welche Möglichkeiten existieren, um diese zu überwinden und welche Aspekte bei dem Austausch oder der Ergänzung eines konventionellen HMIs durch eine Sprachsteuerung berücksichtigt werden sollten. Neben der Untersuchung der Robustheit der Spracherkennung war es das Ziel, die Präzision von kollaborativen Robotern mit der Intelligenz des Bedieners in seiner Funktion als Supervisor, über ein digitales Assistenzsystem zu verbinden. Ghofrani et al. identifizierten ebenfalls die Vermeidung der **Komplexität von Menüstrukturen** in konventionellen haptisch-visuellen HMIs von Robotern über direkte Sprachbefehle als Einflussfaktor. Eine quantitative Bewertung des Effizienzpotenzials einer Sprachsteuerung wurde nicht durchgeführt. [GhRe19]

Ahamed et al. fassten in ihrem Artikel den aktuellen Stand der Forschung aus den Bereichen HCI, Interaktionsmethoden sowie Taxonomie der HCI zusammen. Ihr Fokus richtete sich darauf, aus den unterschiedlichen, wissenschaftlich untersuchten Einflussfaktoren allgemeine Prinzipien und Gestaltungsrichtlinien für industrielle HMI abzuleiten. Dazu nutzten Sie sowohl die Activity Theory als auch User-Centered Design Methoden. Konkrete Einflussfaktoren auf die Effizienz von HMI werden nicht genannt. [AhZA17]

Suarez et al. untersuchen multimodale HMIs zur Steuerung von Drohnen hinsichtlich ihrer Benutzerfreundlichkeit. Dazu nutzen Sie unterschiedliche NUIs, wie Hand- und Körpergesten, die über visuelle Marker erfasst werden, konventionelle Touchsteuerung und Sprachsteuerung sowie verschiedene visuelle und haptische Feedbacksysteme. Sowohl für die Gesten- als auch für die Sprachsteuerung sehen Suarez et al. klare Vorteile in der Effizienz im Vergleich zu konventionellen HMI-Modalitäten. Gleichzeitig weisen sie aber auf Herausforderungen bei der Robustheit durch Störgeräusche bei der Sprachsteuerung und Behinderungen der Sichtachse der Kamera bei der Gestensteuerung. Auch Suarez et al. wiesen bisher nur die grundsätzliche Möglichkeit zum Einsatz nutzerzentrierter HMIs zur Steuerung von technischen Systemen nach. Eine quantitative Bewertung erfolgte noch nicht. [SSS16]

Kumar et al. untersuchten in einer Probandenstudie die unterschiedliche kognitive Belastung bei der Bearbeitung von mathematischen Aufgaben über eine auditive oder visuelle Schnittstelle. Sie stellten eine Korrelation zwischen einer geringen kognitiven Belastung und einer effizienten Interaktion fest, konstatieren jedoch gleichzeitig auf Grund ihrer Versuchsplanung (engl.: design of experiment (DoE)), dass die tatsächlichen Effizienzvorteile kontextabhängig von der konkreten Aufgabenstellung sein könnten. Auch wenn die Ergebnisse von Kumar et al. darauf hinweisen, dass eine Sprachsteuerung auf Grund ihrer Intuitivität zu einer geringeren kognitiven Belastung und damit einer effizienteren Eingabe führt, so wird kein ausreichender quantitativer Nachweis dafür erbracht. [KuKu16]

Gorecky et al. beschreiben klar den Wandel der Rolle des Maschinenbedieners hin zum Supervisor in der Mehrmaschinenbedienung und die damit verbundene Restrukturierung von HMIs. Sie konstatieren, dass in der Zukunft die Flexibilität und Nutzerzentrierung moderner HMIs mit einem Fokus auf die Bedürfnisse und Fähigkeiten der Bediener wesentlicher Einflussfaktor der Effizienz von Produktionsabläufen sind. Auch wenn Gorecky et al. die Flexibilität und Nutzerzentrierung als wichtige Auswahlkriterien für HMIs der Zukunft benennen, so werden keine konkreten Einflussfaktoren genannt, die sich anhand von abgeleiteten Parametern quantifizieren lassen. [GSL14]

2.6 Effizienzpotenzial einer industriellen Sprachsteuerung in der Produktion

Die Sprache bildet als natürlichste Form der Kommunikation des Menschen ein großes Potenzial HMIs hinsichtlich ihrer Effizienz zu optimieren und konventionelle User Interfaces zu ersetzen [AhZA17, BiAy20]. Der Maschinenbediener kann mittels Sprachsteuerung nicht nur mit mehreren Maschinen unabhängig von der Position eines Bedienterminals interagieren, sondern hat auch beide Hände frei, um parallele Tätigkeiten durchzuführen.

Das von der Europäischen Kommission in 2021 beschriebene Konzept der Industrie 5.0, das die Verschiebung des Fokus in der HMI-Entwicklung von der Maschine zum Bediener in Form eines nutzerzentrierten Designs (engl.: user-centered design (UCD)) beinhaltet, ist nicht neu [BNP21]. Suárez et al. untersuchten bereits 2016 unterschiedliche, multimodale NUIs wie Sprachsteuerung und Gestensteuerung im Vergleich zu konventionellen HMIs hinsichtlich ihrer Gebrauchstauglichkeit zur Steuerung von Drohnen [SSS16]. Des Weiteren bietet eine Sprachsteuerung das Potenzial, die kognitive Belastung, vor allem in einer **Mehrmaschinenbedienung**, bei einem parallelen Mangel an hochqualifizierten Arbeitskräften zu reduzieren. [GRC19, Hig13, LRP22] Der wichtigste Aspekt ist jedoch das Effizienzpotenzial einer Sprachsteuerung gegenüber konventionellen HMIs zur Durchbrechung der **Komplexität von Menüstrukturen** durch die direkte Adressierung mit Sprachbefehlen ohne das Scrollen durch Menüebenen und die **Vermeidung von Laufwegen** (auch in der **Mehrmaschinenbedienung**) [IoSc21, MLA22, SSS16].

Zwar wird das Thema Sprachsteuerung in einigen der in den vorherigen Kapiteln vorgestellten Studien, in unterschiedlichen Kontexten diskutiert, jedoch wird die quantitative Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion nicht berücksichtigt.

Linares-Garcia et al. hingegen untersuchen in den U.S.A. den Einsatz eines selbst entwickelten Sprachassistenten mit dem Namen »Voice-Based Intelligent Virtual Agents« (VIVA) zur Unterstützung von Mitarbeitenden in der Baubranche. Durch den Einsatz des Sprachassistenten wollten sie die Effizienz in der Montage von Baugruppen aus Stahlträgern auf Basis von technischen Zeichnungen steigern. Ihre Studie zeigte, dass die Mitarbeitenden, die konventionell auf Basis von gedruckten Bauplänen die Baugruppen zusammensetzten ca. 18 % langsamer als diejenigen waren, die durch den Sprachassistenten mit den jeweils nächsten Bauschritten und weiteren Informationen unterstützt wurden. Linares-Garcia et al. zeigten damit, dass die Mitarbeitenden

sich durch die effiziente Interaktion über eine Sprachsteuerung mit Audio-Feedback auf den eigentlichen Montageprozess fokussieren und die Produktivität steigern konnten. Die Einsparungen in der Montagezeit erfolgten hier durch den sprachbasierten Dialog mit dem digitalen Assistenzsystem im Vergleich zum händischen Blättern in Bauplänen und Suchen der zur Montage benötigten Informationen. [LRP22]

Ionescu et al. untersuchten die Effizienz in der Programmierung eines Roboterarms für Pick-and-Place Aufgaben über eine händische Positionierung mit sprachbasierter Programmierung von Positionen im Vergleich zu einem konventionellen HMI, bestehend aus Maus und Tastatur. Dazu untersuchten sie zehn Sprachbefehle über vier Probanden hinweg und identifizierten ein Effizienzpotenzial von ca. 46 % in ihrem spezifischen Anwendungsszenario. Die Steigerung der Effizienz erreichten sie durch die händische Manipulation des Roboterarms und parallele, freihändige Programmierung von Positionen per Sprache. [IoSc21]

Wang et al. untersuchen quantitativ die Effektivität und Effizienz, als Bestandteile der Gebrauchstauglichkeit anhand von drei HMIs (Eye Tracking, Gesten- und Sprachsteuerung) und auf Basis von zwölf Probanden und zwei Aufgaben in fünf teilweise multimodalen Szenarien. Dazu entwickelten sie ein multimodales HMI, das sie selber als »*Gaze-Gesture-Speech Augmented Reality*« (GGS-AR) bezeichnen. In ihrer Studie gelangen sie zu der Erkenntnis, dass ein multimodales HMI mit Eye Tracking, Gesten- und Sprachsteuerung effizienter als jede einzelne dieser HMI-Modalitäten ist. [WWY21]

Die wenigen existierenden, in diesem Kapitel vorgestellten Studien verfolgten das Ziel eine Sprachsteuerung zur Steigerung der Effizienz in der Kollaboration zwischen Mensch und Maschine zu verbessern, um die Produktivität von Produktionsabläufen zu steigern. Die Komplexität des jeweiligen konventionellen HMIs und des entsprechenden Sprachbefehls wird jedoch nicht ausreichend und vor allem nicht einheitlich, wie zum Beispiel über die **Anzahl an Menüebenen** und **Anzahl an Interaktionen** im konventionellen HMI quantifiziert (vgl. **Kapitel 5**). Ergänzend war die Anzahl an Probanden und untersuchten/verglichenen Befehlen in allen vorgestellten Studien gering. Die Studien liefern zwar erste Anhaltspunkte für ein ungenutztes Effizienzpotenzial in unterschiedlichen Produktionsabläufen, jedoch sind für reproduzierbare quantitative Aussagen umfangreichere Studien mit einem eindeutigen DoE, über Parameter quantifizierbare Einflussfaktoren sowie mehr Probanden und Durchläufe notwendig.

Die Verbreitung neuartiger, nutzerzentrierter HMI-Modalitäten, wie die Sprachsteuerung in der Produktion bleibt trotz des technischen Reifegrads dieser bisher aus. Lodgaard et al. konstatieren, dass der Fokus der Industrie auf der Vollautomatisierung von Produktionsprozessen mithilfe

von Werkzeugen der Digitalisierung aus der Industrie 4.0 liegt und deshalb das Effizienzpotenzial neuartiger nutzerzentrierter HMI-Modalitäten aus der Industrie 5.0 erst systematisch nachgewiesen werden muss [LoDr20]. Die technischen Hürden, wie die Entwicklung robuster Spracherkennung auch in Störgeräuschumgebungen sowie die Entwicklung von Schnittstellen für die Adaption an moderne industrielle SPS, sind zum aktuellen Zeitpunkt ausreichend gut gelöst. Der technische Reifegrad eilt hier der praktischen Anwendung und Verbreitung voraus. Vorteile einer Sprachsteuerung aus dem Consumer-Bereich, wie freie Hände während der Interaktion, der Fokus auf eine Haupttätigkeit oder das Vermeiden komplexer Menüstrukturen auf kleinen Bildschirmen sind auch in der Produktion von Bedeutung. Die entscheidende Motivation für den Einsatz einer Sprachsteuerung als nutzerzentrierte HMI-Modalität ist jedoch die Steigerung der Effizienz in der Kollaboration zwischen Mensch und Maschine. Dadurch werden Nebenzeiten reduziert und die Produktivität durch die Einsparung von Zeit und damit Geld gesteigert. [GFS20, LoDr20, NAL24, SSK21] Jetzt bedarf es vom Produktionsprozess unabhängige, allgemeingültige Methoden zur Bewertung der Effizienz, als Teil der Gebrauchstauglichkeit von Sprachsteuerung, um das Potenzial sichtbar zu machen und einen Market-Pull durch die potenziellen Endanwender zu generieren.

2.7 Etablierte Konzepte, Methoden und Werkzeuge zur Evaluation der Effizienz von Produktionsabläufen

Im Folgenden werden etablierte Konzepte, Methoden und Werkzeuge aus der HCI-Forschung und industriellen Praxis zur Evaluation der Effizienz von Produktionsabläufen mit ihrem jeweiligen Zielen sowie ihren jeweiligen Vor- und Nachteilen vorgestellt. Anschließend werden die Bestandteile herausgestellt, auf Basis derer im Hauptteil dieser Arbeit eine »*Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion*« entwickelt wird. Begonnen wird mit den drei übergeordneten industriellen Konzepten Lean Production, REFA und Root-Cause-Analysis (RCA). Danach wird das Methods-Time Measurement (MTM) als industrielle Werkzeuge zur Erfassung von Soll-Zeiten in Produktionsabläufen vorgestellt. Abschließend werden die Think Aloud Method und NASA Task Load Index (TLX) vorgestellt. Diese beiden HCI-Methoden kommen der in dieser Arbeit entwickelten quantitativen Evaluationsmethode von ihren Ansätzen her am nächsten, trotzdem sie ausschließlich auf der subjektiven Bewertung von Probanden basieren.

2.7.1 Lean Production

Lean Production (dt.: schlanke Produktion) ist eine japanische Unternehmensphilosophie, deren Ursprünge bis in die 1950er Jahre zurückreichen. Sie soll das Management von Unternehmen hinsichtlich der Aspekte Qualität, Produktivität, Flexibilität und Mitarbeitermotivation unterstützen. Die Kernthesen sind die Vermeidung jeglicher Art von Verschwendung und der Fokus auf die Bedürfnisse des Kunden. [KaBr03, Sys06] Daraus ergeben sich im Kontext von Lean Production folgende Paradigmen [Sys06]:

- **Keine Verschwendung von Ressourcen**, die im Rahmen der Wertschöpfung zum Einsatz kommen
- **Vermeidung von hohen Beständen** durch einen effektiven Materialfluss im Unternehmen
- **Vermeidung von Fehlern** anstelle zeit- und kostenintensiver Nacharbeit
- **Förderung von Eigenverantwortlichkeit und Selbstständigkeit** der Mitarbeitenden im Kontrast zu Fremdbestimmung und Monotonie

Mit diesen Paradigmen sollen folgende drei Hauptziele in den Produktionsabläufen erreicht werden [Mäh95, Sys06]:

- **Qualitätserhöhung:** Null-Fehler-Strategie, Entwicklungsqualität, Produkt- und Prozessqualität
- **Zeitreduzierung:** kurze Durchlaufzeiten, zuverlässige Prozesse, Vermeiden von Verschwendung
- **Kostenreduzierung:** effiziente Wertschöpfung, Vermeiden von Gemeinkosten, niedrige Herstellkosten

Alle drei Lean Production Hauptziele werden durch den Einsatz einer Sprachsteuerung als HMI adressiert. Eine robuste und intuitive Sprachsteuerung unterstützt die **Null-Fehler-Strategie** und sichert **zuverlässige Prozesse**, während die Gebrauchstauglichkeit und damit auch die Effizienz der Sprachsteuerung **kurze Durchlaufzeiten** und damit eine **effiziente Wertschöpfung** gewährleistet. In dieser Arbeit werden die sogenannten »*Spaghetti*«-Diagramme aus dem Werkzeugkasten der Lean Production genutzt, um Laufwege aus untersuchten Produktionsszenarien zwischen verschiedenen HMIs und zwischen HMIs und Arbeitsplätzen zu visualisieren (vgl. **Kapitel 6**).

2.7.2 Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung e. V. (REFA)

Der »Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung e. V.« (REFA) ist die älteste Organisation zur Steigerung der Effizienz in Produktionsunternehmen in Deutschland. Der 1924 unter dem Namen »Reichsausschuß für Arbeitszeitermittlung« gegründete Verein stellt Methoden zur Optimierung von betrieblichen Produktionsabläufen zur Verfügung, mit derer sich durch die Reduzierung von Verschwendung die Effizienz steigern lässt. [BBA21] Dabei unterscheidet REFA nach verdeckter Verschwendung (Verschwendungstyp 1) durch **nicht wertschöpfende, aber noch notwendige Tätigkeiten**, wie Datenablage, Rüsten, Transportieren, und offensichtlicher Verschwendung (Verschwendungstyp 2) durch **nicht wertschöpfende und überflüssigen Tätigkeiten**, wie Suchen, Warten, Beobachten [REF21]. Durch die gezielte Reduzierung von Verschwendung soll unter Berücksichtigung der Leistungsfähigkeit und der Bedürfnisse der einzelnen Mitarbeitenden die Produktivität von Unternehmen erhöht werden [BBA21].

Nach REFA werden die folgenden drei Arten von Systemen bei der Herstellung von Produkten unterschieden [BBA21]:

- Technische Systeme (Maschinen-Systeme)
- Soziale Systeme (Systeme von Menschen)
- Soziotechnische Systeme (Mensch-Maschinen-Material-Mitwelt-Systeme)

In **technischen Systemen** stehen Betriebsmittel und Werkzeuge in einer Beziehung zueinander, so wie es beispielsweise in einem Drehbearbeitungszentrum der Fall ist. Die Systembezeichnung kann in diesem Beispiel »Werkzeugmaschine« sein, während der Systemzweck die »Herstellung von rotationssymmetrischen Bauteilen« ist. In **sozialen Systemen** stehen Menschen in Beziehung zueinander, so, wie es beispielsweise in einer Abteilung der Fall ist. Die Systembezeichnung kann in diesem Beispiel »Arbeitsvorbereitung« sein, während der Systemzweck die »Organisation von erforderlichem Material und Arbeitsmitteln« ist. In **soziotechnischen Systemen** stehen Menschen, Betriebsmittel, Material und Mitwelt in einer Beziehung zueinander, so, wie es beispielsweise an Arbeitsplätzen und HMIs in der Produktion der Fall ist. Die Systembezeichnung kann in diesem Beispiel »HMI des Produktionsautomaten« sein, während der Systemzweck die »Anpassung von Prozessparametern« ist. [BBA21]

In dieser Arbeit liegt der Fokus auf den soziotechnischen Systemen, speziell auf den HMIs als Schnittstelle zwischen Mensch und Maschine. Mithilfe der unterschiedlichen REFA-Methoden

werden Produktionsabläufe auf ihre kleinsten Bestandteile heruntergebrochen und mit unterschiedlichen Berechnungsformeln für das jeweilige Anwendungsszenario Soll-Zeiten berechnet [BBA21]. Die unterschiedlichen Methoden aus dem REFA-Werkzeugkasten, im Besonderen die »REFA-Standard Arbeitsdatenermittlung«, werden in dieser Arbeit als Vorbild für die Entwicklung einer »Methode zur Evaluation der Effizienz von Human-Machine Interfaces in der Produktion« genutzt (vgl. **Kapitel 3**).

2.7.3 Root Cause Analysis (RCA)

Die Root Cause Analysis (RCA) ist eine 2008 erstmals veröffentlichte britische Methode des Unternehmens »Quality Management & Training Limited« mit der auf Probleme oder sogenannte Nichtkonformitäten in der Produktion reagiert wird. Sie hilft dabei, die grundlegende Ursache (engl.: »root cause«) eines Problems zu identifizieren. Die RCA beschreibt die systematische Anwendung bekannter und etablierter Ansätze zur Identifikation von Einflussfaktoren auf ein Problem oder eine Nichtkonformität. Das Grundprinzip der RCA wird in **Abbildung 11** dargestellt. Nachdem im ersten Schritt das Problem identifiziert wird, wird dieses im zweiten Schritt klar definiert. Die RCA empfiehlt hier die Verwendung von S.M.A.R.T. Prinzipien (Specific; Measurable; Actions oriented, Realistic, Time constrained). [VoBu08]

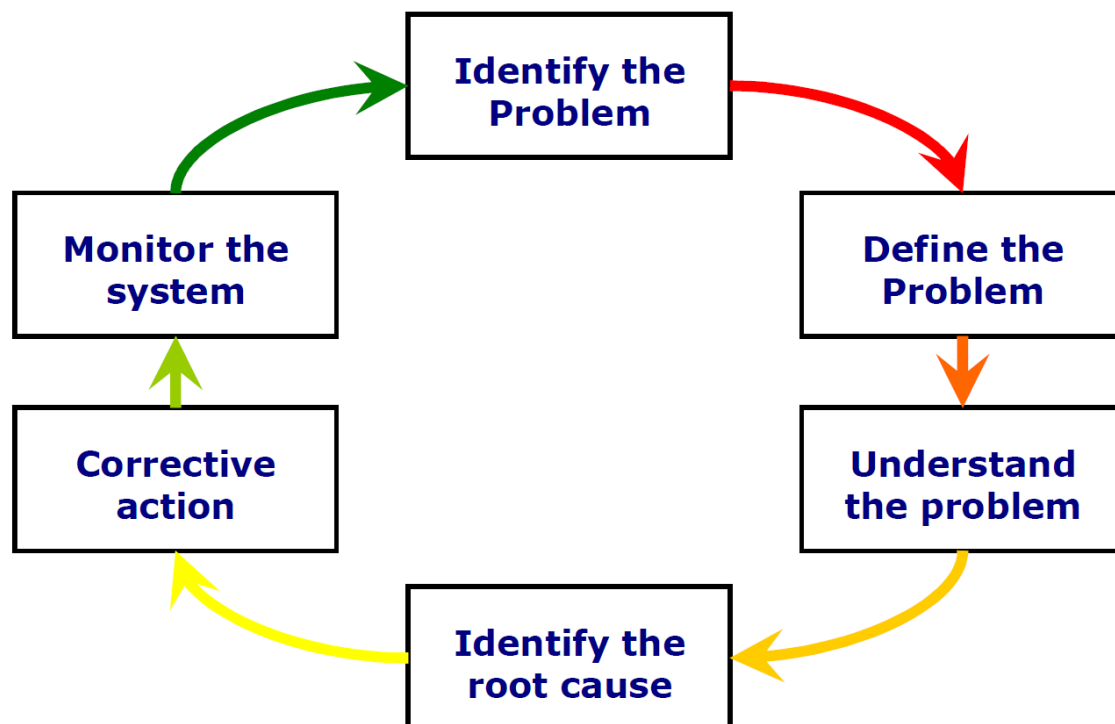


Abbildung 11: Root Cause Analysis Methode [VoBu08]

Auf Basis dieser erhobenen Daten wird im dritten Schritt dann mit unterschiedlichen Techniken versucht das Problem zu verstehen, um die grundlegende Ursache auszumachen. Danach werden Handlungen zur Abmilderung oder, im besten Fall, Eliminierung dieser durchgeführt. Im letzten Schritt wird das untersuchte System überwacht, um die Auswirkungen der Handlungen zu evaluieren. [VoBu08]

Im Kontext dieser Arbeit werden die Methoden der RCA genutzt, um Einflussfaktoren auf die Effizienz eines HMIs, im Speziellen auf Sprachsteuerung als HMI-Modalität, in Industrie- und Probandenstudien systematisch zu identifizieren. Dies ist Voraussetzung für die Evaluation der Effizienz von Sprachsteuerung und Handlungsempfehlungen für den effizienten Einsatz von Sprachsteuerung in der Praxis (vgl. **Kapitel 4.1 & 4.2**).

2.7.4 Methods-Time Measurement (MTM)

Das Methods-Time Measurement (MTM) ist eine Methode zur Vorgabezeitermittlung in der Produktion, die empirisch erhobene Zeiten in Tabellen für die Durchführung von Tätigkeiten innerhalb von Produktionsabläufen nutzt. Diese Methode wird vornehmlich in der industriellen Massenfertigung und im Kontext von Lean Production eingesetzt. Der empirische Ansatz steht dabei in Konkurrenz zu softwarebasierten Optimierungsalgorithmen zur Produktivitätssteigerung. Durch die Zerlegung von Produktionsabläufen in kleinste Schritte und die Messung der zugehörigen Zeiten in sogenannten Time Measurement Units (TMU), die dem Tausendstel einer Stunde (0,036 s) entsprechen, wird versucht, deterministische Gesetzmäßigkeiten in menschlichen Bewegungsabläufen zu identifizieren. [Bul95, MTM19, Sys06] Die Gesamtzeit wird durch die Addition aller Teilzeiten innerhalb eines Produktionsablaufs berechnet [Sys06]. Nach MTM werden manuelle Tätigkeiten in drei Arten von Arbeitsbewegungen kategorisiert [Sys06]:

- **Elementar-Bewegungen:** Hinlangen, Bewegen, Drehen, Greifen und Loslassen
- **Zusammengesetzte Bewegungen:** Einführen, Lösen, Bein- und Fußbewegungen und Körperbewegungen
- **Hilfsbewegungen:** Drücken, Augenbewegungen

In dieser Arbeit wird die MTM zur Quantifizierung von Laufwegzeiten konventioneller und sprachbasierter HMI im Kontext des effizienten Einsatzes von Sprachsteuerung genutzt. Obwohl die MTM laut Literatur sehr präzise Ergebnisse liefert [Sys06], werden die empirisch ermittelten MTM-Zeiten für Bewegungsabläufe zusätzlich mit eigenen Probandenstudien validiert. Der Ansatz der MTM zur empirischen Ermittlung von Prognosefaktoren wird zusätzlich in der entwickelten Methode für die sogenannten »Zeitäquivalente« genutzt (vgl. **Kapitel 3.2.5 & 5.4.6**).

2.7.5 Think Aloud Method & NASA Task Load Index (NASA-TLX)

Die Think Aloud Method (dt.: Methode des lauten Denkens) wurde ursprünglich für Psychologen und Sozialwissenschaftler entwickelt, um Gedankengänge und Strategien zur Problemlösung von Menschen nachvollziehen zu können. Im industriellen Kontext wird die Methode unter anderem eingesetzt, um Probleme in der Gebrauchstauglichkeit von HMIs zu identifizieren. Dabei interagiert der Bediener über ein HMI mit einem technischen System und spricht laut aus, was er gerade macht und denkt. Parallel dokumentiert ein Studienleiter in einem sogenannten »Think-Aloud-Protocol« die Aussagen und Verhaltensweisen des Bedieners und wertet diese im Anschluss aus. Alternativ kann die Interaktion auch über Video-/Audioaufnahmen erfasst werden, um im Anschluss ausgewertet zu werden. [LFH17, SBS94] In dieser Arbeit werden Think-Aloud-Protocols in einer Industriestudie zur Quantifizierung der Einflussfaktoren **Art und Anzahl von Laufwegen** sowie **Anzahl der simultan zu bedienenden Maschinen** genutzt, um den Produktionsablauf zu verstehen und die Art der Laufwege im Rahmen der entwickelten Methode zu kategorisieren (vgl. **Kapitel 6 & 7.3**).

Der NASA-Task Load Index (NASA-TLX) ist eine quantitative Methode zur Bewertung der subjektiven Arbeitsbelastung von Mitarbeitern bei der Durchführung von Aufgaben. Sie wurde im Jahr 1986 von der National Aeronautics and Space Administration (NASA) vorgestellt. [Har06, Har86] **Unter der Prämisse, dass eine erhöhte Arbeitsbelastung von Mitarbeitenden zu Fehlern und zu einer Reduzierung der Effizienz von Arbeitsabläufen führt, kann der NASA-TLX eingesetzt werden, um Rückschlüsse auf die Effizienz von Arbeitsabläufen zu ziehen.**

Die mehrdimensionale Bewertungsskala des NASA-TLX basiert auf den sechs Parametern Mental Demands, Physical Demands und Temporal Demands sowie Frustration, Effort und Performance (vgl. **Abbildung 12**) [Nat22]. Dabei beziehen sich die ersten drei Parameter (Mental Demands, Physical Demands und Temporal Demands) auf die Bedürfnisse des Probanden und die letzten drei Parameter (Frustration, Effort und Performance) auf die Interaktionen des Probanden bei der Durchführung der jeweiligen Aufgabe. Jeder Parameter wird auf einer 21-Punkte Sub-Skala, bestehend aus drei Teilbereichen mit jeweils 7 Punkten (high, medium & low) bewertet (vgl. **Abbildung 12**). [Har06, Har86]

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date
------	------	------

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low Very High

Subject ID: _____ Task ID: _____

WEIGHTED RATING WORKSHEET			
Scale Title	Weight	Raw Rating	Adjusted Rating (Weight X Raw)
MENTAL DEMAND			
PHYSICAL DEMAND			
TEMPORAL DEMAND			
PERFORMANCE			
EFFORT			
FRUSTRATION			

Sum of "Adjusted Rating" Column = _____

WEIGHTED RATING =
[i.e. (Sum of Adjusted Ratings)/15]

Abbildung 12: NASA-TLX Paper and Pencil (dt.: Papier und Stift) Bewertungsskala [Nat22]

Abbildung 13: NASA-TLX Arbeitsblatt zur gewichteten Bewertung der Arbeitsbelastung [Nat22]

Zur Verbesserung der Sensitivität der Bewertungsmethode hinsichtlich der Art der Aufgabe werden im ursprünglichen Vorgehen der NASA-TLX Methode die sechs Sub-Skalen zusätzlich gewichtet. Dazu werden den Probanden vor dem eigentlichen Experiment fünfzehn Paarvergleiche der sechs Parameter vorgelegt, in dem sie jeweils den Parameter einkreisen, der auf die zu untersuchende Aufgabe den größten Einfluss hat. Danach wird für jeden Parameter ausgewertet, wie häufig er markiert wurde. Die Anzahl der Übereinstimmung der Markierungen mit einem der sechs Parameter, die der Gewichtung der Parameter entspricht, kann dabei Werte von null (nicht relevant) bis fünf (relevanter als alle anderen Parameter) annehmen. Zur Berechnung der Arbeitsbelastung in der untersuchten Aufgabe werden die subjektiven Bewertungen der Probanden auf allen sechs Sub-Skalen mit der jeweiligen Gewichtung multipliziert, anschließend aufsummiert und durch 15 geteilt, um eine Bewertung der Arbeitsbelastung von 0 bis 100 zu erhalten (vgl. **Abbildung 13**). Die 15 ergibt sich dabei durch die Addition aller natürlichen Zahlen der Folge 0, 1, 2, 3, 4, 5. [Har86]

Für die Durchführung des Experiments gibt es zwei Ansätze. Entweder geben die Probanden parallel zur Bearbeitung der Aufgabe die Bewertung ihrer subjektiven Arbeitsbelastung auf Basis der NASA-TLX Bewertungsskala an oder die Bearbeitung der Aufgabe wird auf Video aufgenommen und im Nachhinein retrospektiv durch die Probanden bewertet. Während die erste Variante dafür sorgt, dass keine Informationen verloren gehen, bietet die zweite Option den Vorteil, dass die Probanden nicht durch die Bewertungsmethode von ihrer Aufgabe abgelenkt werden. [Har86]

Folgende Vorteile des NASA-TLX im Kontext der Evaluation des Effizienzpotenzials einer Sprachsteuerung existieren:

- Quantitative Auswertung des subjektiven Empfindens der Arbeitsbelastung von Mitarbeitenden auf einer Skala von 0 bis 100 [Har06, Har86]
- Der NASA-TLX ist unabhängig vom Arbeitsablauf und den daraus resultierenden Aufgaben (Zur Beschreibung des Einsatzspektrums werden einfache Labortätigkeiten in Relation zum Fliegen eines Flugzeugs genannt) [Har06, Har86]

Trotz der genannten Vorteile und der Tatsache, dass der NASA-TLX schon die Bewertungsmethode ist, die der in dieser Arbeit entwickelten, quantitativen Evaluationsmethode am nächsten kommt, existieren folgende Nachteile des NASA-TLX im Kontext der Evaluation des Effizienzpotenzials einer Sprachsteuerung:

- Der NASA-TLX wird zur subjektiven Bewertung der Arbeitsbelastung bei der Bearbeitung einer Aufgabe eingesetzt, die nur indirekt Rückschlüsse auf die Gebrauchstauglichkeit oder gar die Effizienz der Interaktion mit einem HMI zulässt
- Komplexe Bewertungsskala, die nach jedem Schritt im Arbeitsablauf beantwortet werden muss und so gegebenenfalls zu einer Ermüdung oder Verdrossenheit der Probanden führt [Lau18]
- Das Ausfüllen der sechs Sub-Skalen nach jedem Schritt greift in den Arbeitsablauf ein, wodurch dieser nicht mehr seiner ursprünglichen Geschwindigkeit und Ablauf entspricht [Lau18]
- Der Studienleiter muss die komplexe Bewertungsskala den Probanden mehrfach erklären (im Besonderen den Unterschied zwischen ähnlichen Einflussfaktoren wie »*effort*« und »*mental demand*«) [Lau18]
- Die Probanden bewerten ihre subjektive Arbeitsbelastung selbst entlang der sechs Sub-Skalen. Dies dazu führen, dass unterschiedliche Probanden die subjektive Arbeitsbelastung für die gleiche Aufgabe unterschiedlich bewerten [Lau18]

3 Entwicklung einer Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion

Ziel ist es HMI-, Steuerungs- und Maschinenherstellern sowie Softwareentwicklern und Industrieunternehmen als Betreiber von Produktionsmaschinen und letztlich Endanwender mithilfe einer Methode die quantitative Bewertung ihrer Produktionsabläufe im Hinblick auf das Effizienzpotenzial einer Sprachsteuerung im Vergleich zu einem konventionellen, haptisch-visuellen HMI zu ermöglichen, ohne dass diese zunächst in eine Produktionsmaschine integriert werden muss (vgl. **Kapitel 1.3 & 1.4**).

In diesem Kapitel wird die entsprechend der Anforderungen entwickelte »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« vorgestellt (vgl. **Kapitel 1.4**). Begonnen wird in **Kapitel 3.1** mit der Beschreibung wie etablierte Konzepte, Methoden und Werkzeuge aus der HCI-Forschung und industriellen Praxis, der Stand der Forschung zu Evaluationsmethoden sowie konventionellen und neuartigen HMIs zusammen mit empirischen Forschungsergebnissen aus Industrie- und Forschungsstudien zur Entwicklung beitragen. Anschließend wird in **Kapitel 3.2** die entwickelte Methode mit ihren einzelnen Schritten in den **Kapiteln 3.2.1 bis 3.2.7** vorgestellt. Abschließend werden in **Kapitel 3.2.8** die zugehörigen Werkzeuge für den Einsatz in der industriellen Praxis erläutert.

3.1 Vorgehensweise bei der Entwicklung

Die Vorgehensweise zur Entwicklung der »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« wird in **Abbildung 14** dargestellt. Zur Evaluation des Effizienzpotenzials einer Sprachsteuerung als Austausch oder Ergänzung zu einem konventionellen, haptisch-visuellen HMI ist es im ersten Schritt notwendig den gesamten Produktionsablauf zu erfassen und zu verstehen. Dazu werden die Bestandteile etablierter und praxiserprobter Forschungsmethoden wie Think Aloud Method und NASA-TLX sowie Shadowing und Experteninterviews identifiziert, die für diese Aufgabe eingesetzt werden können (vgl. **Kapitel 2.7.5**). Parallel werden Erkenntnisse aus der aktuellen Forschung zu Evaluationsmethoden und HMIs genutzt, um zum einen den Aufbau der Methode zu planen und zum anderen die Einflussfaktoren auf das Effizienzpotenzial einer Sprachsteuerung zu identifizieren (vgl. **Kapitel 2.5 & 2.6**). Anschließend werden mithilfe von Werkzeugen industrieller Konzepte und Methoden wie LEAN, MOST, MTM und REFA sowie den Normen DIN EN ISO 9241-11,

-14, -154, -210, -220, DIN EN 614-2 und ISO/IEC 25041:2012-10 von den Einflussfaktoren Parameter abgeleitet und zum Teil bereits quantifiziert (vgl. **Kapitel 2.4 & 2.7**).

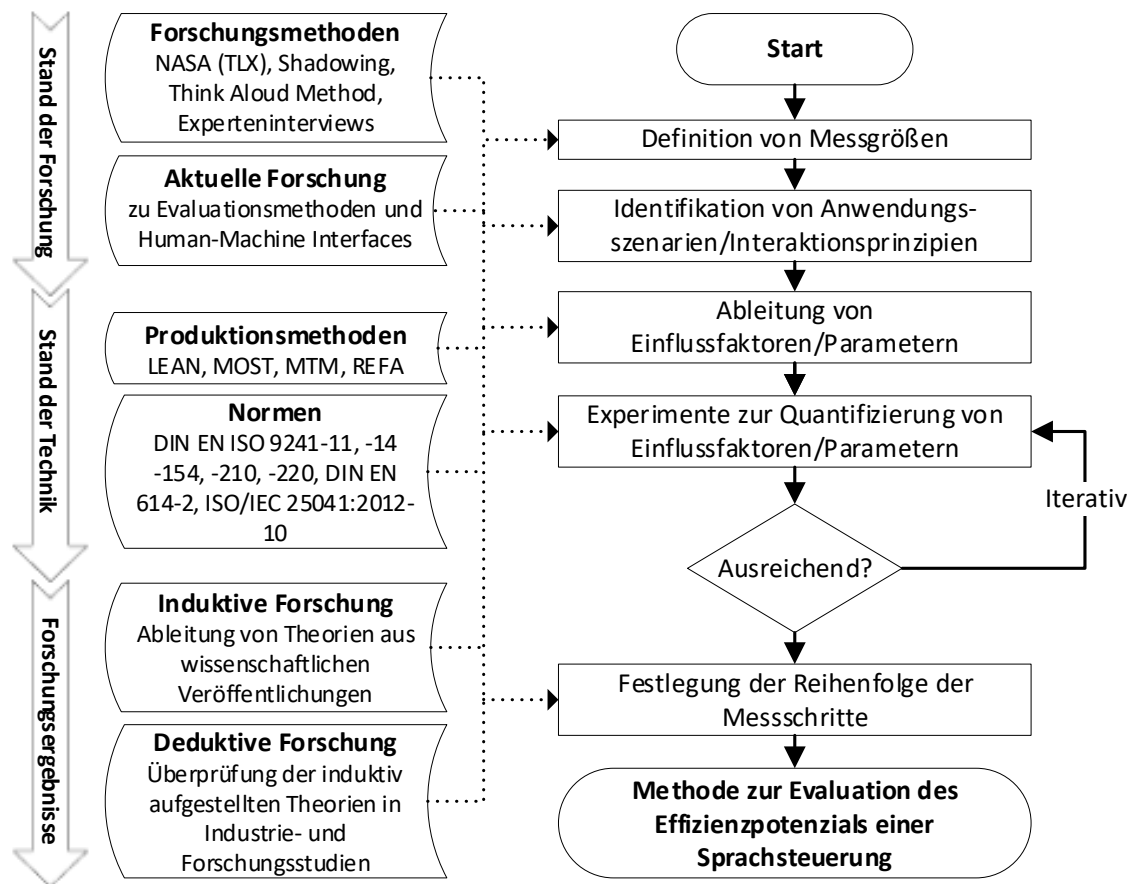


Abbildung 14: Vorgehensweise zur Entwicklung einer »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion«

In einer induktiven Vorgehensweise werden anschließend konkrete Forschungsthemen zur quantitativen Untersuchung der Einflussfaktoren, die sich nicht bereits über etablierte Konzepte, Methoden und Werkzeuge oder Normen quantifizieren lassen, aufgestellt. Zuletzt werden diese in einer deduktiven Vorgehensweise in empirischen Experimenten in Form von Probanden- und Industriestudien quantifiziert (vgl. **Kapitel 5 & 6**).

3.2 Die entwickelte Methode

In diesem Kapitel wird nun auf Basis der in **Kapitel 3.1** beschriebenen Inhalte eine allgemeingültige und prozessunabhängige »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« entwickelt (vgl. **Abbildung 15**).

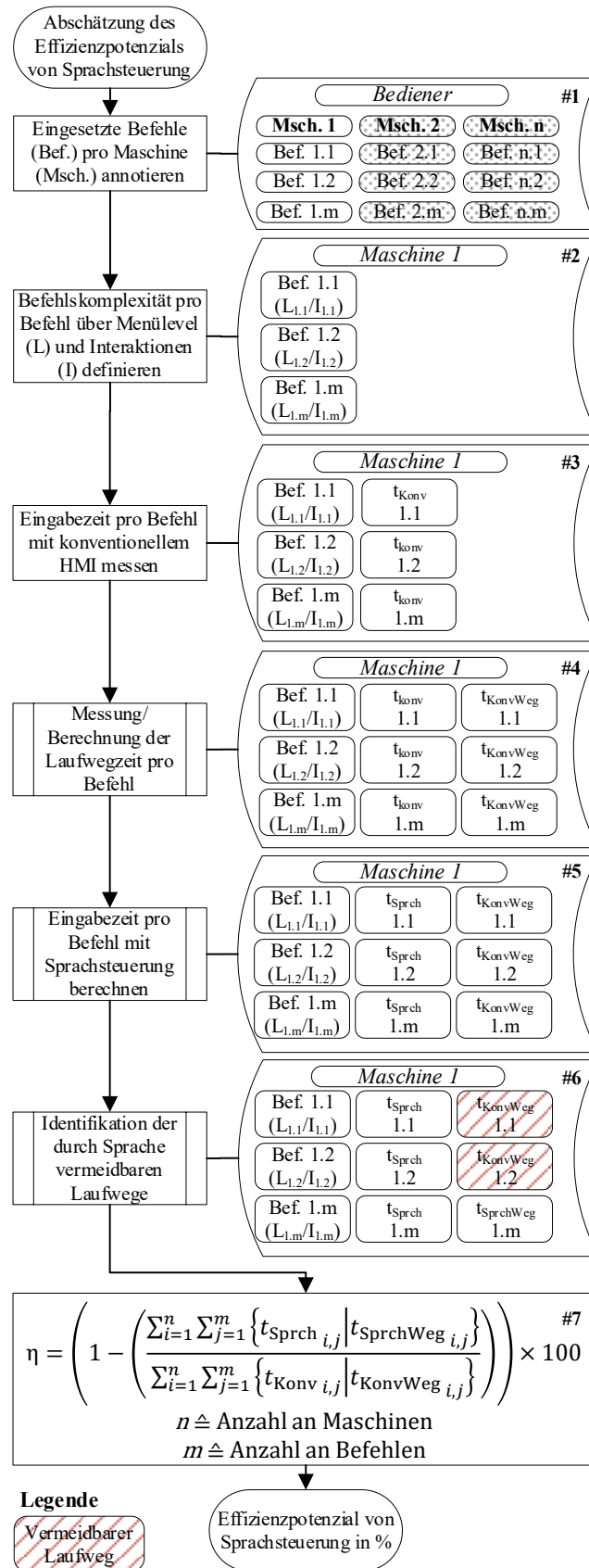


Abbildung 15: Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion [NAL24]

Diese entwickelte Methode basiert auf den drei Haupt-Einflussfaktoren **Anzahl der simultan zu bedienenden Maschinen, Komplexität von Menüstrukturen** und **Art und Anzahl von Laufwegen** [NAL24]. Diese Einflussfaktoren sind unabhängig vom wertschöpfenden Produktions-/Montageprozess in den Hauptzeiten des Produktionsablaufs, wodurch die Allgemeingültigkeit der Methode sichergestellt und die **Anforderung 4** aus dem Lastenheft erfüllt wird (vgl. **Kapitel 1.4**). Der signifikante Einfluss der drei ausgewählten Einflussfaktoren auf die Effizienz einer Sprachsteuerung im Vergleich zu einer Touchsteuerung wird in den **Kapiteln 5 & 6** anhand von abgeleiteten Parametern quantitativ evaluiert.

Das Effizienzpotenzial einer Sprachsteuerung wird gemäß der **Anforderung 3** des Lastenhefts und der DIN EN ISO 9241-11 auf Basis der aufgewendeten Interaktionszeit, bestehend aus Eingabe- und Laufwegzeiten inklusive etwaiger Fehleingaben gemessen (vgl. **Kapitel 1.4**).

Mithilfe der entwickelten Methode können Unternehmen auch ohne die Implementierung einer Sprachsteuerung in ihre Produktionsmaschinen das Potenzial zur Effizienzsteigerung im Vergleich zu einem konventionellen, haptisch-visuellen HMI evaluieren und eine potenzielle Reduzierung von Nebenzeiten prognostizieren. Dies ist notwendig zur Erfüllung der **Anforderungen 1 & 2** des Lastenhefts (vgl. **Kapitel 1.4**). Die Bewertung des Effizienzpotenzials der Sprachsteuerung erfolgt dabei anhand eines Bedieners in einem Produktionsablauf, in dem dieser mit einem haptisch-visuellen HMI interagiert und/oder Laufwege zurücklegt und/oder mehrere Maschinen simultan bedient. [NAL24] Der Bediener wird durch einen Versuchsleiter, der zum Beispiel ein Kollege sein kann, begleitet. Dieser Kollege dokumentiert den Produktionsablauf des Bedieners entlang der in **Abbildung 15** abgebildeten Methode mithilfe des in **Tabelle 1 & Tabelle 2** zur Verfügung gestellten Templates. In den folgenden **Kapiteln 3.2.1 bis 3.2.7** werden die einzelnen Schritte der Methode beschrieben, bevor in **Kapitel 3.2.8** die Berechnung des Effizienzpotenzials einer Sprachsteuerung gemäß **Schritt #7** in **Abbildung 15** beschrieben wird.

3.2.1 Anzahl der simultan zu bedienenden Maschinen (Mehrmaschinenbedienung)

Die **Anzahl der simultan zu bedienenden Maschinen** skaliert die Effekte der anderen beiden Haupt-Einflussfaktoren **Komplexität der Menüstruktur** und **Art und Anzahl der Laufwege**. Das bedeutet, dass wenn der Bediener mehrere gleiche Maschinen simultan in einem Produktionsablauf bedient, das aus der **Komplexität der Menüstruktur** und **Art und Anzahl der Laufwege** resultierende Effizienzpotenzial mit der **Anzahl der simultan zu bedienenden Maschinen** steigt. Durchbricht der Bediener über direkte Sprachbefehle komplexe Menüstrukturen und vermeidet

Laufwege an der ersten Maschine, so wird diese Einsparung an weiteren gleichen Maschinen bei gleichem Arbeitsablauf gleich sein.

In **Schritt #1** der in **Abbildung 15** dargestellten Methode wird die Anzahl der simultan zu bedienenden Maschinen bestimmt. Diese Anzahl setzt sich aus den Maschinen zusammen, mit denen ein Bediener in einem Anwendungsszenario im Rahmen eines Arbeitstages in unterschiedlichen Bediensituationen interagiert. Danach werden alle Befehle, die der Bediener in diesem Zeitraum innerhalb des Produktionsablaufs eingibt, geordnet nach der jeweiligen Maschine, erfasst (vgl. **Tabelle 1 & Tabelle 2**). Für diesen Vorgang stehen mehrere Vorgehensweisen zur Verfügung. [NAL24]

Manuelle Erfassung

Zum einen können die durch den Bediener eingesetzten Befehle händisch durch eine weitere Person, die den Bediener mindestens einen Arbeitstag lang im Rahmen eines »Shadowings« begleitet, erfasst werden [LFH17]. Alternativ können die Produktionsabläufe an den entsprechenden Maschinen, nach Zustimmung der betroffenen Mitarbeitenden und gegebenenfalls weiterer zustimmungspflichtiger, unternehmensinterner Gremien, mindestens einmal entlang eines Arbeitstages bei der Interaktion mit den zu untersuchenden Maschinen gefilmt werden. Anschließend können zeitversetzt in einer Retrospektive, die im HMI durch den Maschinenbediener eingegebenen Befehle auf Basis dieser Videoaufnahmen durch einen weiteren Mitarbeiter erfasst werden.

Automatisierte Erfassung

Moderne Maschinen dokumentieren oft bereits automatisch in einer Protokolldatei (engl.: log file) alle eingegebenen Befehle inklusive der exakten Uhrzeit der Eingabe oder Ausführung sowie weitere Metadaten. Diese Protokolldatei kann direkt aus der SPS der Maschine oder einem übergeordneten Supervisory Control and Data Acquisition (SCADA)-System auf Prozessebene oder Manufacturing Execution System (MES) auf Betriebsebene ausgelesen werden. In selteneren Fällen sind diese Protokolldateien auf Unternehmensebene in ERP-Systemen oder auf firmeninternen Dateiservern zu finden. Eine Auswertung der automatisch generierten Protokolldateien kann entweder ebenfalls händisch oder aber durch automatisierte Skripte erfolgen.

Das Erfassen der simultan zu bedienenden Maschinen und jeweiligen Maschinenbefehls im Produktionsablauf bildet die Basis zur Erfüllung der **Anforderung 1** aus dem Lastenheft, die quantitative Evaluation des Effizienzpotenzials einer nutzerzentrierten Sprachsteuerung im Vergleich zu einem konventionellen, haptisch-visuellen Human-Machine Interface (vgl. **Kapitel 1.4**).

3.2.2 Komplexität der Menüstruktur im konventionellen Graphical User Interface

In **Schritt #2** wird in dem konventionellen GUI des HMI für jeden in **Schritt #1** erfassten Befehl die Anzahl an Menüebenen und Interaktionen annotiert, die zur Eingabe notwendig sind (vgl. **Abbildung 15**). Über diese beiden Parameter **Anzahl an Menüebenen** und **Anzahl an Interaktionen** wird der Einflussfaktor **Komplexität der Menüstruktur** kategorisiert. Viele zu überwindende Menüebenen und/oder eine große Anzahl an Interaktionen in einzelnen Menüebenen führen zu einer hohen Komplexität der Menüstruktur. Über die Anzahl an Interaktionen können auch Handlungen wie das Scrollen in einer Menüebene oder das Ausfüllen eines Eingabefelds erfasst werden. [NAL24] Der Grad der Komplexität des GUI wird über die Anzahl der Interaktionen als untergeordnete und feinere Messgröße der Anzahl an Menüebenen beschrieben [NAL24, NER23].

Durch die Einflussfaktoren und abgeleiteten Parameter aus dem konventionellen Produktionsablauf wird der **Anforderung 2** aus dem Lastenheft, der Prognose der potenziell zu reduzierenden Nebenzeiten auf Basis von Einflussfaktoren des konventionellen Produktionsablaufs ohne die Notwendigkeit der Integration einer Sprachsteuerung, entsprochen (vgl. **Kapitel 1.4**)

3.2.3 Eingabezeiten im konventionellen Graphical User Interface

In **Schritt #3** werden für jeden der erfassten und kategorisierten Befehle die benötigten Eingabezeiten des Bedieners in der konventionellen haptischen HMI-Modalität gemessen (vgl. **Abbildung 15**). Im Vergleich zum Erfassen und Kategorisieren der Befehle ist dieser Schritt schon herausfordernder, da der Versuchsleiter, der wie in Kapitel 3.2 auch ein Kollege sein kann, hier mit einer Stoppuhr die Eingabezeiten für jeden erfassten Befehl messen muss. [NAL24] Selbst, wenn die untersuchten Maschinen automatisiert Protokolldateien generieren, so dokumentieren diese nur den Zeitpunkt der Ausführung des Befehls und nicht den Beginn der Interaktion des Bedieners mit dem HMI. Der Versuchsleiter muss also für jeden eingegebenen Befehl die Eingabedauer mit der Stoppuhr messen und dem zuvor erfassten und kategorisierten Befehl zuordnen. Auch hier sind Videoaufnahmen der jeweiligen Interaktionen des Bedieners mit dem HMI nach den zuvor beschriebenen Zustimmungen möglich, um eine Auswertung im Nachgang zeitversetzt durchzuführen.

Alternativ lassen sich die Eingabezeiten, über die in **Schritt #2** annotierten Eingabe-Uhrzeiten und die daraus resultierenden Abstände zwischen den Eingaben einzelner Maschinenbefehle berechnen. Dies ist aber nur möglich, wenn alle Befehle direkt hintereinander ohne Pause eingegeben werden, da wie oben beschrieben der Beginn der Interaktion mit dem HMI nicht automatisiert von aktuellen Produktionsmaschinen erfasst wird. Werden die Befehle hingegen direkt hintereinander ohne Pause eingegeben, so beschreibt das Ende des ersten Befehls den Beginn der Eingabe des zweiten Befehls. Das genaue Messverfahren zur automatisierten Erfassung von Eingabezeiten wird in **Abbildung 22** beschrieben.

Die Messung der Eingabezeiten inklusive etwaiger Fehleingaben, als Bestandteil der Interaktionszeit mit dem konventionellen, haptischen-visuellen HMI dient der Erfüllung der **Anforderung 3** des Lastenhefts, der Evaluation des Effizienzpotenzials auf Basis der aufgewendeten Interaktionszeit inklusive etwaiger Fehleingaben (gemäß DIN EN ISO 9241-11:2018) (vgl. **Kapitel 1.4**).

3.2.4 Messung/Berechnung von Laufwegzeiten im Produktionsablauf

Sprachsteuerungen bieten die Möglichkeit Laufwege in industriellen Anwendungen zu vermeiden (vgl. **Kapitel 4**). Damit sind Laufwege ein Einflussfaktor auf das Effizienzpotenzial einer Sprachsteuerung im Produktionsablauf. In dieser Arbeit wird der Einflussfaktor über die Parameter »Art« und »Anzahl« von Laufwegen in einer Industriestudie in **Kapitel 6** quantifiziert. In **Schritt #4** der entwickelten Methode werden für jeden erfassten, kategorisierten, mit Eingabezeit gemessenen Befehl die zugehörigen Laufwege (wenn vorhanden) des Bedieners in Metern gemessen (vgl. **Abbildung 15**). Sollte der Maschinenbediener entlang seines Arbeitstages keine Laufwege zwischen Arbeitsplätzen und HMIs oder unterschiedlichen HMIs zurücklegen, so müssen auch keine Laufwege erfasst werden. Für den Fall, dass nach einem Laufweg mehrere Befehle nacheinander am HMI eingegeben werden, wird ausschließlich der Laufweg des ersten Befehls, für den der Bediener initial das HMI aufgesucht hat, zugeordnet. Auf Basis der ermittelten Laufweglängen werden die Laufwegzeiten berechnet. Hierbei kann zwischen drei unterschiedlichen Herangehensweisen ausgewählt werden. [NAL24]

Die **erste Option** ist die Messung der tatsächlichen Laufwegzeiten der Maschinenbediener mit einer Stoppuhr durch eine weitere Person für jeden einzelnen realen Laufweg. Die **zweite Option** ist die Messung der durchschnittlichen Laufgeschwindigkeiten der Maschinenbediener im eigenen Unternehmen per Stoppuhr und das Anschließende Dividieren der einzelnen Laufweg-

längen durch die durchschnittliche Laufgeschwindigkeit, um die Laufwegdauer für jeden einzelnen Laufweg zu erhalten. [NAL24] Die **dritte Option** ist die Berechnung der Laufwegdauer jedes einzelnen Laufwegs mithilfe des Methods-Time Measurement Universal Analysis System (MTM-UAS) der MTM Association e.V.. Laut diesem benötigen Mitarbeiter 0,9 s oder 25 sogenannte Time-Measurement-Units (TMU), um einen Meter Laufweg zurückzulegen, was einer Laufgeschwindigkeit von 1,11 m/s entspricht. [MTM19, NAL24, NVE23, Sys06, Vee23]

1. Messung der Laufwegzeiten für jeden einzelnen untersuchten Laufweg
2. Messung der durchschnittlichen Laufgeschwindigkeit der Maschinenbediener und anschließendes Dividieren der einzelnen Laufweglängen durch die durchschnittliche Laufgeschwindigkeit
3. Berechnung der Laufwegdauer jedes einzelnen Laufwegs mittels MTM

Für die Abschätzung des Effizienzpotenzials mithilfe der in dieser Arbeit entwickelten und vorgestellten Methode die Genauigkeit der dritten Option, der Berechnung der Laufwegdauer auf Basis einer empirisch ermittelten durchschnittlichen Laufwegzeit hinreichend genau (vgl. **Kapitel 7**).

Die Messung/Berechnung der Laufwegzeiten, als Bestandteil der Interaktionszeit mit dem konventionellen, haptischen-visuellen HMI dient der Erfüllung der **Anforderung 3** des Lastenhefts, der Evaluation des Effizienzpotenzials auf Basis der aufgewendeten Interaktionszeit inklusive etwaiger Fehleingaben (gemäß DIN EN ISO 9241-11:2018) (vgl. **Kapitel 1.4**).

3.2.5 Berechnung der Eingabedauer einer Sprachsteuerung über Zeitäquivalente

Das Durchbrechen von komplexen Menüstrukturen mit ihren Menüebenen und der benötigten Anzahl an Interaktionen über das direkte Ausführen mittels Sprachbefehlen kann in Abhängigkeit der Parameter »Anzahl an Menüebene« und »Anzahl an Interaktionen« zu einer Reduzierung der Eingabezeit und daraus resultierenden Effizienzsteigerung des Produktionsablaufs führen. Das Verhältnis zwischen der Eingabezeit eines konventionellen haptisch-visuellen HMIs und einer Sprachsteuerung wird über sogenannte Zeitäquivalente beschrieben. Im **Kapitel 5** werden in **Tabelle 4** empirisch ermittelte Zeitäquivalente für den Einsatz einer Sprachsteuerung als Austausch zu einer Touchsteuerung mit einer state-of-the-art (SOTA) GUI, bestehend aus bis zu fünf Menüebenen und zehn Interaktionen, gezeigt. Die Eingabezeiten beider HMI-Modalitäten (Sprache/Touch), die als Berechnungsgrundlage für die Zeitäquivalente dienen, werden in der in **Kapitel 5** beschriebenen Probandenstudie ermittelt. In **Schritt #5** wird die Eingabedauer der

Sprachsteuerung für jeden einzugebenden Befehl über die Multiplikation dieser Zeitäquivalente mit den gemessenen Eingabedauer des konventionellen haptisch-visuellen HMIs abgeschätzt (vgl. **Abbildung 15**). [NAL24]

Die Berechnung der Eingabezeiten dient der Erfüllung der **Anforderung 2** des Lastenhefts, der Prognose der potenziell zu reduzierenden Nebenzeiten auf Basis von Einflussfaktoren des konventionellen Produktionsablaufs ohne die Notwendigkeit der Integration einer Sprachsteuerung (vgl. **Kapitel 1.4**).

3.2.6 Kategorisierung von Laufwegen

In **Schritt #6** werden die ermittelten Laufwege entsprechend der in **Kapitel 6** vorgestellten vier Arten von Laufwegen im Kontext des Effizienzpotenzials von Sprachsteuerung kategorisiert (vgl. **Abbildung 15**). Die Kategorisierung wurde im Rahmen einer Industriestudie in neun deutschen Produktionsunternehmen entwickelt (vgl. **Kapitel 6**). [NAL24, NVE23, Vee23]

1. Unvermeidbare Laufwege
2. Vermeidbar durch optimierte Arbeitsorganisation
3. Vermeidbar durch optimiertes HMI-Konzept
4. **Vermeidbar durch Sprachsteuerung**

Für die Evaluation des Effizienzpotenzials einer Sprachsteuerung sind nur Laufwege der Kategorie 4 relevant, die ausschließlich durch den Einsatz einer Sprachsteuerung vermieden werden können. Durch die Kategorisierung werden die Laufwege identifiziert, die durch eine Sprachsteuerung vermieden werden und deren Laufwegzeiten damit eingespart werden.

Die Kategorisierung der Laufwege dient der Erfüllung der **Anforderung 2** des Lastenhefts, der Prognose der potenziell zu reduzierenden Nebenzeiten auf Basis von Einflussfaktoren des konventionellen Produktionsablaufs ohne die Notwendigkeit der Integration einer Sprachsteuerung (vgl. **Kapitel 1.4**).

3.2.7 Berechnung des Effizienzpotenzials

In **Schritt #7** wird das Effizienzpotenzial einer Sprachsteuerung im Vergleich zu einer konventionellen Touchsteuerung, in Abhängigkeit der drei Einflussfaktoren **Komplexität der Menüstrukturen, Art und Anzahl von Laufwegen** sowie **Anzahl der simultan zu bedienenden Maschinen** berechnet. Dazu werden die Interaktionszeiten, bestehend aus Eingabe- und Laufwegzeiten des konventionellen HMIs über alle Befehle und Maschinen aufsummiert. Anschließend werden

diese durch die Summe aller mithilfe der vorherigen **Schritte #1 bis #6** der Methode ermittelten Eingabe- und Laufwegzeiten einer potenziellen Sprachsteuerung dividiert. Im letzten Schritt wird auf Basis des Quotienten ein prozentuales Effizienzpotenzial η berechnet (vgl. **Formel 3.1**). [NAL24]

Das Effizienzpotenzial η wird im Excel-Template (vgl. **Kapitel 3.2.8**) automatisch über die **Formel 3.1** aus der entwickelten »Methode zur Evaluation der Effizienz einer Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« berechnet (vgl. **Abbildung 15**).

$$\eta = \left(1 - \left(\frac{\sum_{i=1}^n \sum_{j=1}^m \{t_{\text{Sprch } n,m} | t_{\text{SprchWeg } n,m}\}}{\sum_{i=1}^n \sum_{j=1}^m \{t_{\text{Konv } n,m} | t_{\text{KonvWeg } n,m}\}} \right) \right) \times 100 \quad \mathbf{3.1}$$

$\eta \triangleq$ Effizienzpotenzial einer Sprachsteuerung in %

$n \triangleq$ Nummer der Maschine

$m \triangleq$ Nummer des Befehls

Alternativ können händisch erfasste Werte im ausgedruckten Excel-Template auch dazu genutzt werden, um das Effizienzpotenzial manuell zu berechnen. Positive Prozentwerte des Effizienzpotenzials η zeigen eine Verringerung der Eingabe- und/oder Laufwegzeiten durch den Einsatz einer Sprachsteuerung und damit eine gesteigerte Effizienz im Vergleich zu einem konventionellen HMI an. Negative Prozentwerte des Effizienzpotenzials η zeigen hingegen eine Erhöhung von Eingabe- und/oder Laufwegzeiten durch den Einsatz einer Sprachsteuerung und damit eine verringerte Effizienz im Vergleich zu einem konventionellen HMI an. [NAL24]

3.2.8 Template zur Anwendung in der industriellen Praxis

Nachdem im vorherigen **Kapitel 3.2** die entwickelte »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« vorgestellt wurde, wird im Folgenden ein tabellarisches Microsoft Excel-Template vorgestellt, dass die einfache Anwendung der Methode in der industriellen Praxis ermöglicht. Dieses Template besteht aus einem Bereich zur Erfassung der in **Kapitel 3.2** beschriebenen Parameter zur Berechnung des Effizienzpotenzials einer Sprachsteuerung und einem Infokasten zur schnellen Orientierung des Zeitnehmers im Template (vgl. **Tabelle 1 & Tabelle 2**).

Header und Maschinen-/Befehlsnummer

Im oberen Bereich der **Tabelle 1** befindet sich der »Header«, in dem der Name des Unternehmens, der Name des Zeitnehmers, die Bezeichnung des Produktionsablaufs, das Datum und der Ort sowie eine Kurzbeschreibung des Prozesses vom Zeitnehmer eingetragen werden. Dies kann entweder digital direkt im Excel-Template oder händisch in der ausgedruckten Tabelle erfolgen. Mithilfe dieser Informationen können die erfassten Laufwege und Eingabezeiten eindeutig zugeordnet werden. Das Template beginnt mit der Spalte »B« am linken Rand der Tabelle, in der die Befehlsnummer der jeweiligen Produktionsmaschine nach dem Schema »*Maschinennummer.Befehlsnummer*« (n.m), analog zur entwickelten Methode, erfasst wird (vgl. **Abbildung 15**). Der erste Maschinenbefehl an der ersten Maschine startet mit der Nummer 1.1, gefolgt von der Nummer 1.2 für den zweiten Befehl der ersten Maschine oder der Nummer 2.1 für den ersten Befehl der zweiten Maschine. Aufgrund dieser Nummerierung sind die Befehle auch bei einer großen Anzahl an Maschinenbefehlen und Maschinen noch eindeutig zuordenbar. Gleichzeitig ist der Nummerierung direkt zu entnehmen, um den wievielten Maschinenbefehl an dieser Maschine es sich handelt.

Tabelle 1: Template zur Erfassung von Eingabezeiten (Konventionell/Sprachsteuerung) und Laufwegen zwischen Maschinen und Arbeitsplätzen (auch in der Mehrmaschinenbedienung)

[illegible]

Befehlskomplexität und Eingabe- und Laufwegzeiten für Touch- und Sprachsteuerung

Im nächsten Schritt wird in der Sektion »Befehlskomplexität« der **Tabelle 1** sowohl die Komplexität des Touchbefehls über die Menüebenen »E« und Anzahl an Interaktionen »I« als auch die Komplexität des Sprachbefehls über Hotword, Attribut, Parameter und ggf. zusätzliche Wörter sowie die Gesamtlänge definiert. In der zweiten Sektion »Konventionell« werden die Eingabezeit t_k und die Laufweglänge s_k sowie die Laufwegzeit t_{kw} des konventionellen HMIs für jeden einzelnen Maschinenbefehl erfasst. Analog erfolgt in der dritten Sektion »Sprachsteuerung« die Erfassung der Eingabezeit t_s und der Laufweglänge s_s sowie der Laufwegzeit t_{sw} der Sprachsteuerung. Die Laufweglänge s_s und die Laufwegzeit t_{sw} sind dabei immer entweder so groß wie die Laufweglänge s_k und die Laufwegzeit t_{kw} des konventionellen HMIs oder, wenn die Laufwege durch eine Sprachsteuerung vermieden werden können, null. Grund dafür sind die konstante Laufweglänge s_k und die davon abgeleitete Laufwegzeit t_{kw} , die unabhängig von der HMI-Modalität ist.

In **Tabelle 2** des Excel-Templates werden dem Zeitnehmer zusätzliche Informationen für die einzelnen oben beschriebenen Parameter zur Verfügung gestellt.

Tabelle 2: Zusätzliche Informationen für den Zeitnehmer

Kategorie	Bezeichnung	Beschreibung	Eingabe	Berechnung	Einheit
Befehlskomplexität	Befehl B	Maschinennummer.Befehlsnummer	zB. »2.1«	NaN	NaN
	Ebenen E	Anzahl der Ebenen Beschränkt	0/1/3/5	NaN	NaN
	Interaktionen I	Anzahl der Interaktionen Beschränkt (Entweder Ebenen oder Interaktionen!)	1/3/5/6/7/10	NaN	NaN
	Sprachbefehl	Genaue Bezeichnung des Befehls	Name	NaN	NaN
Konventionell K	Eingabezeit t_k	Dauer der Touchbedienung Messen	Wert	NaN	s
	Laufweglänge s_k	Nötiger Laufweg Messen	Wert	NaN	m
	Laufwegzeit t_{kw}	Dauer des Laufwegs Berechnen	Wert	$s_k/1,11$	s
Sprachsteuerung S	Eingabezeit t_s	Dauer der Sprachsteuerung Berechnen	Wert	$t_k * \text{Zeitäquivalent}$	s
	Laufweglänge s_s	Nötiger Laufweg »0«, wenn vermeidbar!	Wert	NaN	m
	Laufwegzeit t_{sw}	Dauer des Laufwegs Berechnen	Wert	$s_s/1,11$	s
	Effizienzpotenzial	Endgültiges Effizienzpotenzial Aus den vorherigen Werten berechnen	Wert	$\left(1 - \frac{\sum t_s + \sum t_{sw}}{\sum t_k + \sum t_{kw}}\right) * 100$	%

In **Kapitel 5** wird der Einflussfaktor **Komplexität von Menüstrukturen** und in **Kapitel 6** die Einflussfaktoren **Art und Anzahl von Laufwegen** sowie **Anzahl simultan zu bedienender Maschinen quantifiziert**. Anschließend wird die Methode in einer Probandenstudie in **Kapitel 7** anhand der Anforderungen des Lastenhefts validiert. Dies im Speziellen für die Anforderung 5, eine höhere Präzision der Prognose als interindividuelle Unterschiede in Interaktionszeiten von Bedienern zur Gewährleistung der praktischen Nutzbarkeit, wichtig (vgl. **Kapitel 1.4**).

4 Anwendungsszenarien, Interaktionsprinzipien und Einflussfaktoren für eine Sprachsteuerung

In den folgenden Kapiteln werden zunächst Anwendungsszenarien mit unterschiedlichen Interaktionsprinzipien beschrieben, die auf Basis der Literatur potenziell von einer Sprachsteuerung profitieren (vgl. **Kapitel 2.5 & 2.6**). Die Anwendungsszenarien werden exemplarisch an den vier Produktionsmaschinentypen **Produktionsautomaten** (vgl. **Kapitel 4.1.1**), **Roboter** (vgl. **Kapitel 4.1.2**), **Werkzeugmaschinen** (vgl. **Kapitel 4.1.3**) und **Produktionsanlagen** (vgl. **Kapitel 4.1.4**) beschrieben. Auf Basis der Anwendungsszenarien werden übergeordnete und vom produktionsprozessunabhängige Einflussfaktoren auf die Effizienz einer Sprachsteuerung identifiziert (vgl. **Kapitel 4.2**). Diese werden anschließend in Industrie- und Probandenstudien anhand von abgeleiteten Parametern quantifiziert (vgl. **Kapitel 5 & 6**).

Der Einsatz eines effizienten HMIs ist nur sinnvoll in Produktionsabläufen, die eine Interaktion zwischen Mensch und Maschine über eine Schnittstelle erfordern. Vollautomatisierte oder vollständig manuelle Produktions- und Montageprozesse profitieren nicht von einem effizienten HMI. Dennoch können für vollautomatisierte Prozesse vor- oder nachgelagerte Rüst- oder Wartungsprozesse notwendig sein, die teilautomatisiert, kollaborierend durch Mensch und Maschine oder vollständig manuell durchgeführt werden. Dies zeigt, dass ein Produktionsablauf aus einem vollautomatisierten Produktionsprozess und ergänzenden teilautomatisierten oder manuellen Rüst- oder Wartungsprozessen bestehen kann. Für die Evaluation der Effizienz eines HMI sind nur die Teile des Produktionsablaufs relevant, in denen auch eine Interaktion zwischen Mensch und Maschine stattfindet, weshalb ausschließlich teilautomatisierte Produktionsabläufe betrachtet werden.

Für den quantitativen Nachweis des Effizienzpotenzials einer Sprachsteuerung als industrielles HMI ist es notwendig die Produktionsabläufe in der Praxis zu identifizieren, die hinsichtlich der Effizienz der Interaktion zwischen Mensch und Maschine und damit auch hinsichtlich ihrer Produktivität profitieren. Anschließend ist es notwendig übergeordnete Einflussfaktoren abzuleiten, die einen Einfluss auf die Effizienz einer Sprachsteuerung haben. In dieser Arbeit liegt der Fokus auf der Sprachsteuerung von Produktionsmaschinen, also der Bewegung von Aktoren der Produktionsmaschine nach erfolgreich erkanntem Sprachbefehl. Dies liegt daran, dass dieser Anwendungsbereich auf Grund seiner potenziellen Gefahren für Mensch und Maschine bisher deutlich weniger als Werkerassistenzsysteme in Form von digitalen sprachbasierten Assistenten

untersucht wurde. Die Evaluation des Effizienzpotenzials einer Sprachsteuerung von Produktionsmaschinen soll die Hemmschwelle zum Einsatz dieser HMI-Modalität in der Praxis senken. Natürlich muss für eine ganzheitliche Betrachtung neben Gestaltungsregeln für den effizienten Einsatz einer Sprachsteuerung stets auch die Robustheit und damit verbundene Sicherheit der Produktionsmaschine berücksichtigt werden [GFS20].

4.1 Steigerung der Produktivität von Produktionsabläufen durch effiziente Sprachsteuerung

In **Kapiteln 2.3.1 bis 2.3.4** werden bereits unterschiedliche Anwendungsszenarien von Sprachsteuerung, aber auch von anderen NUIs in der Produktion vorgestellt. Neben einer Literaturrecherche werden in einer Industriestudie neun Unternehmen mit unterschiedlichen Produktionsabläufen aus den Branchen Kunststoffrecycling, Automotive, Papierproduktion, Schiffbau, Flechtmaschinenherstellung und Folienkondensatorherstellung untersucht (vgl. **Kapitel 6**). Dabei fällt auf, dass branchenübergreifend Produktionsabläufe zu einem großen Teil vollautomatisiert oder vollständig manuell durchgeführt werden. Nur ein kleiner Teil der untersuchten Produktionsabläufe sind teilautomatisiert. Laut den Mitarbeitenden liegt dies auch daran, dass konventionelle HMI für die entsprechenden Produktionsabläufe nicht geeignet sind. [NVE23, Vee23]

Die Produktionsabläufe, die hinsichtlich der Effizienz in der Kollaboration zwischen Mensch und Maschine von einer Sprachsteuerung profitieren, lassen sich nach den eingesetzten Maschinenarten, nach aufsteigender Komplexität in folgende Kategorien einteilen:

- Produktionsautomaten
- Roboter
- Werkzeugmaschinen
- Produktionsanlagen

Diese vier Arten von Produktionsmaschinen werden in der aktuellen Forschung besonders oft zur Untersuchung des industriellen Einsatzes von Sprachsteuerung eingesetzt (vgl. **Kapitel 2.3.4, 2.5 & 2.6**). Produktionsautomaten werden besonders oft in der **Mehrmaschinenbedienung** eingesetzt, während Roboter in der aktuellen Forschung die am häufigsten im Kontext neuartiger, nutzerzentrierter HMIs untersuchten Produktionsmaschinen sind (vgl. **Kapitel 2.3.4**). Werkzeugmaschinen sind aufgrund ihrer vielen Funktionen die Produktionsmaschinen mit dem größten

Einfluss der **Komplexität von Menüstrukturen** auf die Effizienz eines HMIs (vgl. **Kapitel 2.2**). Produktionsanlagen bestehen hingegen oft aus vielen einzelnen Maschinenkomponenten und HMIs, sodass sich die **Art und Anzahl von Laufwegen** als Einflussfaktor an diesen besonders gut darstellen lässt. Die Anwendungsszenarien lassen sich mit ihren jeweiligen Interaktionsprinzipien jedoch auch auf weitere Arten von Produktionsmaschinen transferieren. Im Folgenden werden die jeweiligen Produktionsabläufe, in denen die Maschinen eingesetzt werden und das konkrete Effizienzpotenzial einer Sprachsteuerung beschrieben.

4.1.1 Produktionsautomaten

Produktionsautomaten sind Produktionsmaschinen, die Halbzeuge und Bauteile nach den folgenden sechs Hauptgruppen der Fertigungsverfahren gemäß DIN 8580:2022-12 bearbeiten können [DIN22]:

- Urformen
- Umformen
- Trennen
- Fügen
- Beschichten
- Stoffeigenschaften ändern

Produktionsautomaten sind also zum Beispiel Spritzguss-, Walz-, Bohr-, Schweiß- oder Beschichtungs- oder Schmiedeautomaten. Für das Effizienzpotenzial einer Sprachsteuerung bei der Bedienung eines Produktionsautomaten ist weniger die konkrete Art der Maschine und auch nicht der konkrete Produktionsprozess, den die Maschine in den Hauptzeiten verrichtet, entscheidend. Viel wichtiger ist die Art der Kollaboration zwischen Mensch und Maschine im Rahmen des gesamten Produktionsablaufs. Bediener arbeiten mit Produktionsautomaten sequenziell in einem teilautomatisierten Produktionsablauf zusammen. Der Bediener bereitet Bauteile vor, die er anschließend zur Bearbeitung in den Produktionsautomaten einlegt. Anschließend gibt er keine oder nur wenige Parameter über ein konventionelles HMI ein, um anschließend den Bearbeitungsprozess zu starten. Klassischerweise ist dieser Bearbeitungsprozess eines Produktionsautomaten sehr kurz und dauert wenige Sekunden oder maximal ein paar Minuten. Innerhalb dieser Zeit bereitet der Bediener entweder ein weiteres Bauteil vor oder bestückt weitere Produktionsautomaten. Da Produktionsautomaten repetitive, vollautomatisierte Fertigungsprozesse immer nach dem gleichen Schema durchführen, ist in den meisten Fällen kein Eingriff des Bedieners innerhalb der Hauptzeiten notwendig. Ausnahmen sind Störungen oder Anpassungen

von Parametern bei Veränderungen in der Bauteilqualität. Aus diesem Grund sind bei der Interaktion mit Produktionsautomaten die **Art und Anzahl der Laufwege** zwischen dem Arbeitsplatz des Bedieners und dem HMI oder Laufwege zwischen den HMIs unterschiedlicher Produktionsautomaten entscheidender für die Effizienz einer Sprachsteuerung als die **Komplexität von Menüstrukturen** im GUI. Auch die im Folgenden beschriebenen Roboter und Werkzeugmaschinen können als Produktionsautomaten zum Einsatz kommen.

4.1.2 Roboter

Im Kontext der Interaktion mit Robotern spielt vor allem die Programmierung eine zentrale Rolle. Eine Sprachsteuerung ermöglicht es dem Bediener manuell, mit beiden Händen den Endeffektor eines Roboters im Raum zu positionieren und parallel, ohne den Einsatz seiner Hände und ohne den Blick auf den Endeffektor zu verlieren, Positionen per Sprache in einem sogenannten »*teach-in*«-Prozess in der Steuerung zu speichern [GFS20, IoSc21, MVG15]. In **Abbildung 16** wird ein Bediener gezeigt (vgl. **Abbildung 16**, VI), der einen kollaborativen Roboter (Cobot) (vgl. **Abbildung 16**, I) über eine kabelgebundene haptisch-visuellen Fernbedienung (vgl. **Abbildung 16**, V) eines konventionellen HMI (vgl. **Abbildung 16**, IV) programmiert.

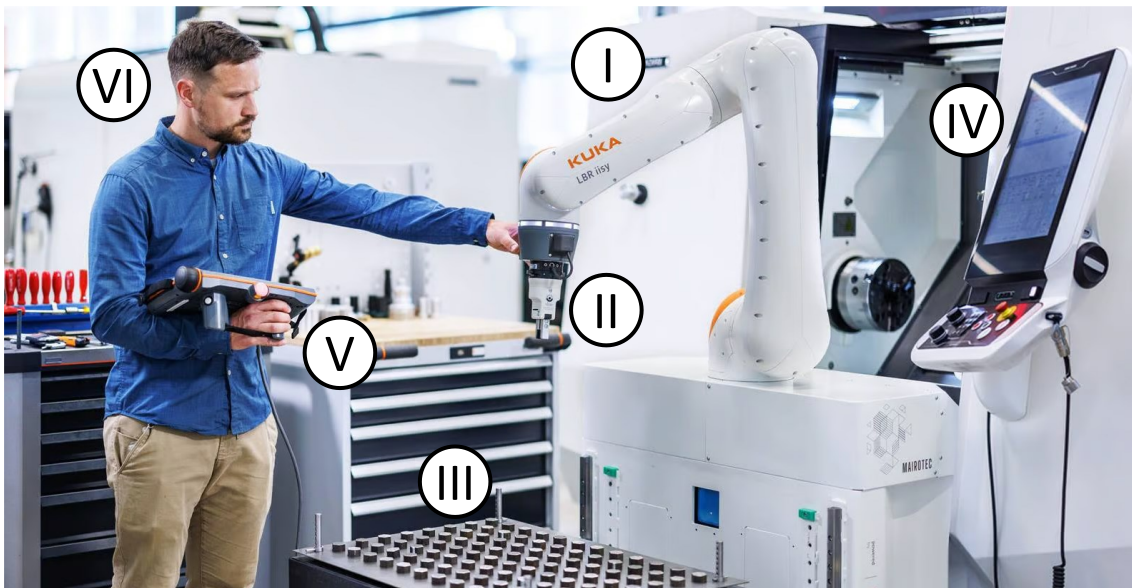


Abbildung 16: Programmierung eines KUKA 6-Achs-Cobots über eine kabelgebundene haptisch-visuelle Fernbedienung eines konventionellen HMIs (I. KUKA 6-Achs-Cobot, II. Greifer am Tool-Center-Point (TCP), III. Tisch mit Bauteilen, IV. haptisch-visuelles HMI mit Touchscreen, Tastatur, Schaltern, Tastern und Drehreglern, V. kabelgebundenes Handheld (Fernbedienung) in Form eines Industrie-Tablets, VI. Bediener) [KUK24]

Dabei hält der Bediener mit der rechten Hand die Fernbedienung, während er mit der linken Hand den Endeffektor des Cobots (vgl. **Abbildung 16, II**) im Arbeitsraum positioniert. Durch eine Sprachsteuerung werden sowohl die Navigation durch **komplexe Menüstrukturen** als auch **potenzielle Laufwege** zwischen Endeffektor und Bedienterminal vermieden. Bei der Verwendung eines mobilen Gerätes, wie einem Tablet oder der in **Abbildung 16** gezeigten kabelgebundenen Fernbedienung werden der Wechsel zwischen haptisch-visueller Eingabe und händischer Manipulation des Endeffektors vermieden. Dadurch verliert der Bediener nicht den Fokus auf den zu programmierenden Bewegungsablauf des Roboters. [GFS20, MVG15] Besonders Produktionsabläufe mit Cobots profitieren von der Interaktion per Sprache als natürlichste Form der Kommunikation. Cobots werden im Gegensatz zu klassischen Robotern nicht nur einmal programmiert und arbeiten dann in vollautomatisierten Produktionsprozessen, sondern sie arbeiten Hand in Hand mit ihren menschlichen Kollegen. Für eine hohe Produktivität ist es notwendig, dass Mensch und Maschine so effizient wie möglich miteinander kommunizieren, da die Anzahl an Interaktionen innerhalb der Kollaboration hoch ist. [IoSc21]

4.1.3 Werkzeugmaschinen

Werkzeugmaschinen wie Dreh-, Fräs- und Schleifmaschinen sind komplexe Produktionsmaschinen, mit denen subtraktiv in einem spannenden Fertigungsprozess aus oft metallischen Halbzeugen Bauteile hergestellt werden. Sie unterscheiden sich hinsichtlich des HMIs in ihrer Komplexität gegenüber Robotern und vor allem Produktionsautomaten. In **Kapitel 2.2** wird das haptisch-visuelle HMI einer 5-Achs-Fräsmaschine als komplexester Vertreter konventioneller HMI vorgestellt. Moderne Werkzeugmaschinen werden mithilfe von automatischen Werkzeugwechslern sowie automatischer Materialzufuhr und Bauteilentnahme vornehmlich im Automatikbetrieb genutzt. Bediener von Werkzeugmaschinen sind hochqualifizierte Fachkräfte, die oft mehrere Werkzeugmaschinen simultan als Supervisor bedienen. In **Abbildung 17** wird ein Bediener (vgl. **Abbildung 17, VI**) vor der Prozesskammer (vgl. **Abbildung 17, II**) einer industriellen 5-Achs-Fräsmaschine (vgl. **Abbildung 17, I**) gezeigt. Im Hintergrund ist die Prozesskammer mit der Spindel (vgl. **Abbildung 17, III**) zu sehen, während in der rechten Bildhälfte das schwenkbare konventionelle Bedienpanel (vgl. **Abbildung 17, IV**) der Fräsmaschine zu sehen. Obwohl die in der Abbildung gezeigte Steuerung eine Siemens 840D ist, ähnelt sie vom Aufbau sehr der moderneren in **Kapitel 2.2** vorgestellten Heidenhain TNC7. Ergänzend zum haptisch-visuellen HMI (vgl. **Abbildung 17, IV**) trägt der Bediener ein Over-Ear-Headset (vgl. **Abbildung 17, V**), das über Bluetooth mit einem Industrie-PC (IPC) als Teil der SPS verbunden ist.



Abbildung 17: Sprachsteuerung einer Werkzeugmaschine als Ergänzung zu einem konventionellen, haptisch-visuellen HMI (I. 5-Achs Fräsmaschine, II. Prozesskammer, III. Frässpindel mit eingespanntem Werkzeug, IV. haptisch-visuelles HMI mit Bildschirm, Tastaturen, Schaltern, Drehreglern und Tastern, V. Over-Ear-Headset (Hearpiece), VI. Proband) ©Fraunhofer IDMT 2022 / Anika Bödecker

Die **Komplexität der Menüstrukturen** im GUI ist bei Werkzeugmaschinen der größte Einflussfaktor auf die Effizienz einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen, haptisch-visuellen HMI. Nach dem bereits etablierten Funktionsprinzip des MMI im Auto kann der Bediener per Sprache Befehle direkt adressieren, die konventionell nur über viele Menüebenen und Interaktionen im GUI erreichbar sind. [NMR22, NAL24, NER23] Neben dem Durchbrechen von komplexen Menüstrukturen kann der Bediener mithilfe der Sprachsteuerung Werkzeugmaschinen ortsunabhängig mit freien Händen bedienen. Dies ermöglicht es ihm auch bei Arbeiten in der Prozesskammer mit dem HMI zu interagieren, um beispielsweise pneumatische oder hydraulische Spannvorrichtungen auf einem Schwenktisch oder in einer Spindel zu öffnen oder zu schließen.

Vor allem bei großen Werkzeugmaschinen werden zusätzlich **Laufwege** zwischen HMI und Arbeitsplatz in der Prozesskammer durch eine ortsunabhängige Sprachsteuerung **vermieden**. In

der **Mehrmaschinenbedienung** als Supervisor werden **Laufwege** zwischen den HMIs unterschiedlicher Maschinen **vermieden**. Werkzeugmaschinen können genau wie Produktionsautomaten und Roboter auch in Fertigungsstraßen von Produktionsanlagen integriert werden.

4.1.4 Produktionsanlagen

Produktionsanlagen bestehen aus vielen einzelnen, verketteten Produktionsmaschinen. Zusammen bilden diese Produktionsmaschinen entlang einer Produktionsstraße eine Produktionsanlage. Diese Art von komplexen Maschinen wird in unterschiedlichen Branchen genutzt. In der in **Kapitel 6** beschriebenen Industriestudie wurden Produktionsanlagen sowohl in der Papierherstellung, der Verpackungsindustrie, der Automobilindustrie und im Kunststoffrecycling hinsichtlich ihres Effizienzpotenzials durch die **Vermeidung von Laufwegen** untersucht.

Produktionsanlagen wurden in der Vergangenheit noch oft von mehreren Mitarbeitenden bedient. Nicht zuletzt durch den Kostendruck der Globalisierung im Hochlohnland Deutschland werden neue Produktionsanlagen vorwiegend auf die Bedienung durch nur einen einzelnen Mitarbeitenden ausgelegt. Im Folgenden werden Produktionsabläufe an Produktionsanlagen beschrieben, die hinsichtlich des Effizienzpotenzials einer Sprachsteuerung profitieren.

Anpassung von Parametern im Automatikbetrieb

Die **Vermeidung von Laufwegen** bietet bei diesen oft langen und zum Teil auch mehrstöckigen und verwinkelten Maschinen das größte Effizienzpotenzial. Entlang der Produktionsanlage sind an strategischen Punkten konventionelle, haptisch-visuelle HMIs verteilt. Da Produktionsanlagen im Regelfall in einem vollautomatisierten Betrieb laufen, stellt dies im Regelbetrieb kein großes Problem dar. Je nach Reifegrad des Produktionsprozesses muss der Bediener während des Betriebs lediglich Parameter überwachen oder aber einzelne Parameter nachjustieren. Innerhalb der in **Kapitel 6** beschriebenen Industriestudie fielen oft Laufwege zwischen dem Beobachtungspunkt des Bedieners und dem nächstliegenden HMI an. Eine ortsunabhängige Sprachsteuerung per Headset bietet in dieser Situation das Potenzial, Parameter direkt an der entsprechenden Stelle der Produktionsanlage anzupassen und gleichzeitig die Auswirkungen der Anpassung direkt überwachen zu können.

Rüstprozesse zur flexiblen Umstellung von Produktionsprozessen

Die in **Kapitel 1** beschriebene Hyper-Individualisierung führt zusätzlich dazu, dass auch Produktionsanlagen in kürzeren Abständen auf andere/neue Produkte angepasst werden müssen. Speziell der Anteil an Rüstprozessen, der nicht manuell, sondern teilautomatisiert erfolgt, profitiert

hinsichtlich seiner Effizienz von einer ortsunabhängigen und freihändigen Sprachsteuerung per Headset. Bediener können so an beliebigen Stellen der Produktionsanlage Umrüstarbeiten durchführen und gleichzeitig **ohne Laufwege** mit dem HMI per Sprache kommunizieren. In diesem Anwendungsszenario können **viele wiederholte Laufwege** zum HMI **vermieden werden**.

Akute Störungsbeseitigung und geplante Wartungsarbeiten

Zuletzt bietet eine Sprachsteuerung mit Audio-Feedback das Potenzial, Bediener bei der Beseitigung von Störungen und geplanten Wartungsarbeiten zu unterstützen. Zum einen bekommt der Bediener Störungen über ein ortsunabhängiges Audio-Feedback über ein Headset schneller mit als bei einem statischen konventionellen HMI. Zum anderen kann der Bediener unabhängig davon, wo er sich an der Produktionsanlage befindet, zusätzliche Informationen wie Betriebszustände einzelner Maschinenkomponenten abrufen. Dadurch können **Laufwege** zum HMI **vermieden werden**, weil der Bediener durch detaillierte Störungsinformationen entweder direkt zum betreffenden Teil der Produktionsanlage laufen oder sogar direkt benötigtes Werkzeug und gegebenenfalls passende Ersatzteile aus der Werkstatt/dem Lager holen kann. Ähnlich wie bei Rüstprozessen profitiert der Bediener bei geplanten Wartungsarbeiten von der ortsunabhängigen und freihändigen Sprachsteuerung. Auch hier vermeidet er Laufwege vom Arbeitsort zum HMI und zurück zur Abfrage von Parametern der Sensoren oder zur Steuerung von Aktoren.

Ergänzend zum Effizienzpotenzial einer Sprachsteuerung zur **Vermeidung von Laufwegen** beim Betrieb von Produktionsanlagen bietet die Sprachsteuerung auch in diesem Anwendungsszenario das Potenzial die **Komplexität von Menüstrukturen** zu durchbrechen (vgl. **Kapitel 5 und 6**) [NER23, NVE23]. Gleichzeitig beinhaltet die Bedienung von Produktionsanlagen immer eine **Mehrmaschinenbedienung**, die potenziell hinsichtlich ihrer Effizienz durch eine Sprachsteuerung profitiert, da sich diese per Definition aus mehreren Produktionsmaschinen zusammensetzen (vgl. **Kapitel 3.2**) [NAL24].

Aus den untersuchten Produktionsabläufen werden im folgenden **Kapitel 4.2** die übergeordneten Einflussfaktoren auf die Effizienz einer industriellen Sprachsteuerung von Produktionsmaschinen zusammengefasst und kategorisiert, bevor diese in **Kapitel 5 und 6** in Industrie- und Probandenstudien anhand von Parametern quantifiziert. Abschließend wird in **Kapitel 7** die in **Kapitel 3** entwickelte Methode in einer Probandenstudie validiert.

4.2 Einflussfaktoren auf die Effizienz einer Sprachsteuerung in industriellen Produktionsabläufen

Grundsätzlich kann ein effizientes HMI in einem Produktionsablauf, in dem seine Vorteile genutzt werden, zu einer Reduzierung von Nebenzeiten und zu einer Steigerung der Produktivität beitragen (vgl. **Kapitel 2.4.1**). Die Effizienz eines HMIs besteht dabei aus zwei wesentlichen Bestandteilen. Zum einen aus der Leistungsfähigkeit des Bedieners und zum anderen aus der Leistungsfähigkeit der jeweiligen HMI-Modalität im konkreten Produktionsablauf. [DIN18]

Die Leistungsfähigkeit des Bedieners setzt sich sowohl aus eher konstanten Faktoren wie dem Ausbildungsgrad, speziellem Prozess- und Maschinenwissen als auch aus tagesindividuellen Faktoren, wie der Motivation, physischen Verfassung und kognitiven Kapazität des Maschinenbedieners, zusammen. Diese Kombination aus vielen individuellen und zum Teil tagesformabhängigen Faktoren hat einen wesentlichen Einfluss auf die Interaktionsgeschwindigkeit zwischen Mensch und Maschine und damit auch auf die Effizienz des HMI. Da menschliche Einflussfaktoren, wie der Ausbildungsgrad und spezielles Prozess- und Maschinenwissen vor allem in Zeiten des Fachkräftemangels immer schwerer beeinflussbar sind, liegt in dieser Arbeit der Fokus auf den technischen Einflussfaktoren unter der Prämisse konstanter menschlicher Einflussfaktoren im Betrachtungszeitraum.

Konventionelle HMIs an aktuell eingesetzten Produktionsmaschinen bestehen in den meisten Fällen noch aus haptischen-visuellen Komponenten (vgl. **Kapitel 2.2**). Die Einflussfaktoren auf die Effizienz einer Sprachsteuerung im Vergleich zu solch einem HMI sind vielfältig und komplex. Sowohl die Art und der Aufbau eines konventionellen haptisch-visuellen HMIs mit dem eine Sprachsteuerung verglichen wird als auch der Produktionsablauf, in dem die jeweilige HMI-Modalität eingesetzt wird, hat einen Einfluss auf das Effizienzpotenzial einer Sprachsteuerung. Primär hat die Art und die Komplexität der Bedienungsaufgabe (Einschalten einer Prozesskammerbeleuchtung vs. Setzen eines neuen Werkstücknullpunkts), genau wie das Design des Sprachbefehls (Auswahl und Anordnung von Hotwords, Parametern und Attributen im Sprachbefehlssatz) einen Einfluss auf die Effizienz. Zusätzlich hat die **Komplexität von Menüstrukturen** (Menüebenen und Anzahl an Interaktionen im konventionellen visuellen HMI) und die Bedienungsart (Automatikbetrieb, Teilautomatikbetrieb, Handbetrieb) der Maschine einen Einfluss auf die Effizienz. Des Weiteren spielen der Produktionsablauf mit der **Art und Anzahl an Laufwegen**, die **Anzahl der simultan zu bedienenden Maschinen** sowie die Auswahl der Audioschnittstelle

(Headset, Mikrofon oder Mikrofon-Array) und deren Positionierung in der Produktionsumgebung eine wichtige Rolle. Nicht alle Einflussfaktoren haben in jedem Produktionsablauf den gleichen Einfluss auf die Effizienz der Kollaboration zwischen Mensch und Maschine, weshalb eine Quantifizierung von Einflussfaktoren über Parameter notwendig ist (vgl. **Kapitel 5 & 6**). Die für die Evaluation des Effizienzpotenzials einer Sprachsteuerung relevanten Einflussfaktoren werden mit ihren jeweiligen Parametern (in Klammern), in absteigender Reihenfolge, im Folgenden aufgelistet. Begonnen wird mit den Einflussfaktoren, die laut dem aktuellen Stand der Forschung am relevantesten sind (vgl. **Kapitel 2.3.4, 2.5 & 2.6**).

- **Komplexität von Menüstrukturen (Menüebenen/Interaktionen) [IoSc21, Nzu21]**
- **Laufwege (Art und Anzahl) [LoDr20]**
- **Mehrmaschinenbedienung (Anzahl der simultan zu bedienenden Maschinen) [LoDr20, SSK21]**
- Befehlsart (Rüst-, Werkzeug-, Positions-, Programmier-, Referenzierungsbefehle sowie Material-Handling und Peripherie-Steuerung) [LoPa20]
- Aufgabenkomplexität (leichte, mittlere, schwere Maschinensteuerung) [LoPa20]
- Befehlsdesign (Auswahl/Anordnung von Hotwords, Parametern und Attributen sowie Befehlslänge) [LoPa20, SSK21]
- Dialogdesign (direkte Befehle vs. Chat-Bot) [LRP22, LoPa20]
- Bedienungsart (Automatikbetrieb, Teilautomatikbetrieb, Handbetrieb) [IoSc21, LRP22]
- Auswahl und Positionierung der Audioschnittstelle im Produktionsablauf (Headset, Mikrofon oder Mikrofon-Array) [MLA22]
- Robustheit (Effektivität als Voraussetzung für ein effizientes HMI) [GhRe19, JHR21, JRV21, MLA22]

Auf Basis des aktuellen Stands der Forschung und den in den **Kapiteln 5 und 6** beschriebenen empirischen Forschungs- und Industriestudien werden die drei Einflussfaktoren **Komplexität von Menüstrukturen, Art und Anzahl von Laufwegen** sowie die **Anzahl simultan zu bedienender Maschinen** als die drei Haupt-Einflussfaktoren identifiziert (vgl. **Kapitel 2.5 & 2.6**). Auf Basis dieser Einflussfaktoren wird deshalb in **Kapitel 3** eine »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« entwickelt.

5 Quantifizierung des Einflussfaktors »Komplexität von Menüstrukturen«

In der wissenschaftlichen Literatur herrscht die übereinstimmende Meinung, dass die Sprachsteuerung eine besonders gebrauchstaugliche HMI-Modalität ist, die es Maschinenbedienern ermöglicht, Maschinen intuitiv und damit effizient zu bedienen. Besonders häufig wird die **Komplexität der Menüstrukturen** im konventionellen, haptisch-visuellen HMI als Einflussfaktor auf das Effizienzpotenzial einer Sprachsteuerung als Ergänzung oder Austausch zu diesem genannt (vgl. **Kapitel 2.5, 2.6 & 4.2**).

Aus diesem Grund wird in der in diesem Kapitel beschriebenen Probandenstudie das quantitative Verhältnis zwischen einer Touch- und einer Sprachsteuerung bestimmt. Dieses über Zeitäquivalente beschriebene Verhältnis dient in **Schritt #5** der in **Abbildung 15** dargestellten Methode dazu, auf Basis der in **Schritt #3** gemessenen Eingabezeiten des konventionellen, haptisch-visuellen HMIs die Eingabezeiten einer Sprachsteuerung abzuschätzen. Nur über diese Zeitäquivalente ist es möglich, das Effizienzpotenzial einer Sprachsteuerung im Hinblick auf den Einflussfaktor **Komplexität von Menüstrukturen** zu prognostizieren, ohne eine Sprachsteuerung tatsächlich in eine Produktionsmaschine zu integrieren. Neben dieser Quantifizierung des Verhältnisses zwischen Touch- und Sprachsteuerung ist auch die Quantifizierung der Einflussfaktoren **Art und Anzahl von Laufwegen** sowie **Mehrmaschinenbedienung** in **Kapitel 6** notwendig, um die **Anforderung Nr. 2** im Lastenheft zu erfüllen. Diese fordert die unabhängige Prognose der Reduzierung von Nebenzeiten ohne die Notwendigkeit der Integration einer Sprachsteuerung (vgl. **Kapitel 1.4**). Im Folgenden wird zunächst in einem induktiven Vorgehen eine Theorie auf Basis des oben beschriebenen Stands der Forschung aufgestellt, die dann in einem deduktiven Vorgehen empirisch in einer Probandenstudie quantifiziert wird. Dadurch wird eine Datengrundlage in Form von Zeitäquivalenten geschaffen, die dann in der entwickelten Methode genutzt wird (vgl. **Kapitel 3.2**).

Theorie

Eine Sprachsteuerung reduziert im Vergleich zu einer konventionellen haptisch-visuellen Touchsteuerung die Eingabezeiten von Maschinenbefehlen an Produktionsmaschinen.

Aus dieser Theorie werden im Folgenden die folgenden Null- und Alternativhypothesen (H0 & H1) abgeleitet.

Nullhypothese H0

Die **Anzahl der Menüebenen** und die zur Navigation benötigte **Anzahl an Interaktionen** in einem GUI **hat keinen Effekt** auf die Eingabedauer einer **Touchsteuerung** und das daraus resultierende Effizienzpotenzial einer **Sprachsteuerung**, die die Möglichkeit bietet, komplexe Menüstrukturen zu durchbrechen und Maschinenbefehle direkt auszuführen

Alternativhypothese H1

Die **Anzahl der Menüebenen** und die zur Navigation benötigte **Anzahl an Interaktionen** in einem GUI **hat einen Effekt** auf die Eingabedauer einer **Touchsteuerung** und das daraus resultierende Effizienzpotenzial einer **Sprachsteuerung**, die die Möglichkeit bietet, komplexe Menüstrukturen zu durchbrechen und Maschinenbefehle direkt auszuführen

Zur Überprüfung der induktiv aufgestellten Theorie und Quantifizierung der abgeleiteten Hypothesen wird in einem deduktiven Vorgehen ein Experiment in Form einer Probandenstudie zum empirischen Vergleich der Effizienz einer nutzerzentrierten Sprach- und einer konventionellen, haptisch-visuellen Touchsteuerung anhand der jeweiligen Eingabezeiten durchgeführt. Die Quantifizierung erfolgt dabei anhand des Einflussfaktors **Komplexität der Menüstrukturen** in dem GUI der Touchsteuerung über die beiden Parameter **Anzahl an Menüebenen** und **Anzahl an Interaktionen** (vgl. Kapitel 4.2).

5.1 Die Wizard-of-Oz-Methode im Vergleich zur Implementierung einer Sprachsteuerung

Da keine Produktionsmaschinen existieren, die standardmäßig über ein konventionelles, haptisch-visuelles HMI in Form eines menübasierten GUI auf einem Touchscreen und gleichzeitig über eine industrielle Sprachsteuerung verfügen, gibt es für die Durchführung der Probandenstudie hinsichtlich der Sprachsteuerung nur zwei Optionen (vgl. **Kapitel 2.2 & 2.3**) [NER23]. Entweder wird die sogenannte »*Wizard-of-Oz-Methode*« (Ursprünglich: »*OZ[sic]-Paradigma*«) nach John F. Kelly zur Simulation einer Sprachsteuerung bei vorhandenem konventionellen HMI angewendet oder die beiden HMI-Modalitäten (Sprach- & Touchsteuerung) werden in eine Produktionsmaschine integriert [Kel83].

Die erste Option ist die Anwendung der »*Wizard-of-Oz-Methode*«, welche dem Probanden den Eindruck vermittelt, mit einem HMI zu interagieren, ohne dass die zu untersuchende HMI-Mo-

dalität tatsächlich für das spezifische Programm oder die Maschine implementiert wird. Der Studienleiter übernimmt die Rolle des »Assistenten« und führt die verstandenen Sprachbefehle des Teilnehmers auf dem konventionellen, bereits implementierten HMI-System aus. [Kel83]

Dadurch entfällt die Notwendigkeit der technischen Integration einer Sprachsteuerung in eine Produktionsmaschine mit bereits bestehendem konventionellem HMI. Stattdessen wird aber eine unabhängige Variable in Form des Studienleiters (»Wizard«), in das Studiendesign integriert. Während der technische Aufwand beim Einsatz der »Wizard-of-Oz-Methode« stark reduziert wird, entsteht durch die unabhängige Variable der Latenz der Schnittstelle zwischen »Wizard« und konventionellem, haptisch-visuellem HMI eine zusätzliche Messungenauigkeit. Der Anteil dieser am gemessenen Gesamtergebnis ist besonders für Befehle mit kurzen Eingabezeiten kritisch hoch. [FeSe10, LWL06]

Bereits eine geringe Variabilität der »Wizard«-Schnittstelle könnte zu einer Verschiebung des Break-Even-Points (BEP) zwischen der zeitlichen Effizienz der Sprachsteuerung und dem konventionellen HMI führen. Zusätzlich zur nicht konstanten Latenz des »Wizards« unterscheidet sich diese zur Latenz einer ASR. Zudem unterscheiden sich ein menschlicher »Wizard« und eine echte ASR in der Art und Weise ihrer Erkennungsfehler. Während beide Optionen unweigerlich Fehler machen werden (insbesondere unter schwierigen akustischen Bedingungen), liegt der Unterschied darin, dass die Erkennungsrate und die daraus resultierende Befehlsverlässlichkeit bei einer echten ASR konstanter ist und zumindest pro Teilnehmer gemessen werden kann. [NER23]

Aktuell genutzte ASRs erreichen bis zu 90 % Erkennungsraten, in speziellen Szenarien sogar bis zu 100 %, was vor allem vom vorherrschenden Signal-Rausch-Verhältnis (engl.: signal-to-noise-ratio (SNR)) abhängig ist [BiAy20, SKM18]. Im Allgemeinen ist die Erkennungsrate eines Menschen, insbesondere in komplexen akustischen Szenarien, besser als die einer ASR [SKM18]. Gleichzeitig ist diese Erkennungsrate aber nicht konstant und hängt von der physischen und kognitiven Verfassung des »Wizards« während der Durchführung des Experiments ab [NER23].

Zuletzt kann auch die ungewohnte Interaktionssituation mit einem menschlichen »Wizard« als zusätzliche, vorgeschaltete Schnittstelle zwischen Probanden und eigentlichem HMI zu einer Voreingenommenheit der Probanden führen. Die ist auch dann der Fall, wenn diese nicht explizit darüber aufgeklärt werden, dass sie nicht mit einer echten Sprachsteuerung interagieren. [LFH17] Da in dieser Arbeit vergleichend die Eingabezeiten und die daraus resultierende zeitliche Effizienz einer Sprachsteuerung im Vergleich zu einem konventionellen HMI ohne laute Störgeräusche untersucht wird, ist die Kontinuität der Sprachsteuerung in dieser Probandenstudie wichtiger als die Erkennungsraten in lauten akustischen Szenarien [NER23].

5.2 Implementierung einer Sprach- und Touchsteuerung in eine Produktionsmaschine

Für die Durchführung von Probandenstudien wurde aus den oben genannten Gründen ein Fräsbearbeitungszentrum mit automatisiertem Materialhandling als Technologieplattform mit dem Namen »Voice-Controlled Production (VCP)-Technologieplattform I« entwickelt (vgl. **Abbildung 18**) [NER23]. Die Unterscheidung nach »VCP-Technologieplattform I« und »VCP-Technologieplattform II« in dieser Arbeit ist notwendig, da in **Kapitel 7** zur Validierung der entwickelten Methode ein anderes, weiterentwickeltes Fräsbearbeitungszentrum mit automatisiertem Materialhandling, mit dem Namen »VCP-Technologieplattform II« eingesetzt wird.

Sowohl eine 3-Achsen-Fräsmaschine (vgl. **Abbildung 18**, I) als auch ein 4-Achsen-Roboter (vgl. **Abbildung 18**, II) wurden in der Prozesskammer installiert. Vor der Prozesskammer sind ein Touchscreen (vgl. **Abbildung 18**, III) und ein omnidirektionales Mikrofon (vgl. **Abbildung 18**, IV) zur Nutzung der HMI-Modalitäten Touch- und Sprachsteuerung montiert. Über zwei Lautsprecher (vgl. **Abbildung 18**, V) unterhalb der Prozesskammer wird ein Audio-Feedback an den Bediener abgegeben. Ergänzend bieten diese Lautsprecher die Möglichkeit zur Simulation von Produktionsumgebungen als Störgeräuschen. Zusätzlich zu der 3-Achs Fräsmaschine und dem 4-Achs Roboter stehen ein Rohteil- und ein Fertigteilmagazin zur Verfügung. Mithilfe des Roboters kann der Maschinenbediener Aluminium- Rohteile, die als »Halbzeuge« bezeichnet werden, aus dem Halbzeugmagazin in das pneumatische Spannfutter der Fräsmaschine laden. Anschließend kann der Bediener das gewünschte Bearbeitungsprogramm der Fräsmaschine auswählen, das Halbzeug spanend bearbeiten lassen und zuletzt das fertige Aluminium-Bauteil entnehmen und im Bauteilmagazin ablegen. Neben diesen Hauptfunktionen kann der Bediener das Licht der Prozesskammer sowie das Kühlschmiermittelsystem der Fräsmaschine ein- und ausschalten, jede Achse der Werkzeugmaschine und des Roboters einzeln verfahren oder aber eine der beiden Maschinen zu gespeicherten Positionen bewegen. [NER23]

Vier verschiedene Sicherheitsstufen der Sprach- und Touchsteuerung innerhalb der Technologieplattform bieten dem Maschinenbediener die Möglichkeit, das Audio-Feedback und/oder Verzögerungen in der Ausführung von Befehlen einzustellen. In der Sicherheitsstufe 1 (engl.: Security level 1 (SL1)) erhält der Maschinenbediener nur kurze Signaltöne als Bestätigung der erfolgreichen Erkennung und Ausführung eines Maschinenbefehls. Die Sicherheitsstufe 2 (SL2) gibt ein wirkliches Text-to-Speech (TTS) Audiofeedback in Form der Wiedergabe des per Sprach- oder Touchsteuerung ausgelösten Befehls. In der Sicherheitsstufe 3 (SL3) werden zusätzlich fünf

Sekunden Wartezeit für einen möglichen Abbruch des Maschinenbefehls ergänzt. Die höchste Sicherheitsstufe 4 (SL4) bedarf der zusätzlichen Bestätigung jedes ausgesprochenen Befehls durch den Maschinenbediener, um Fehleingaben korrigieren zu können. [NER23]

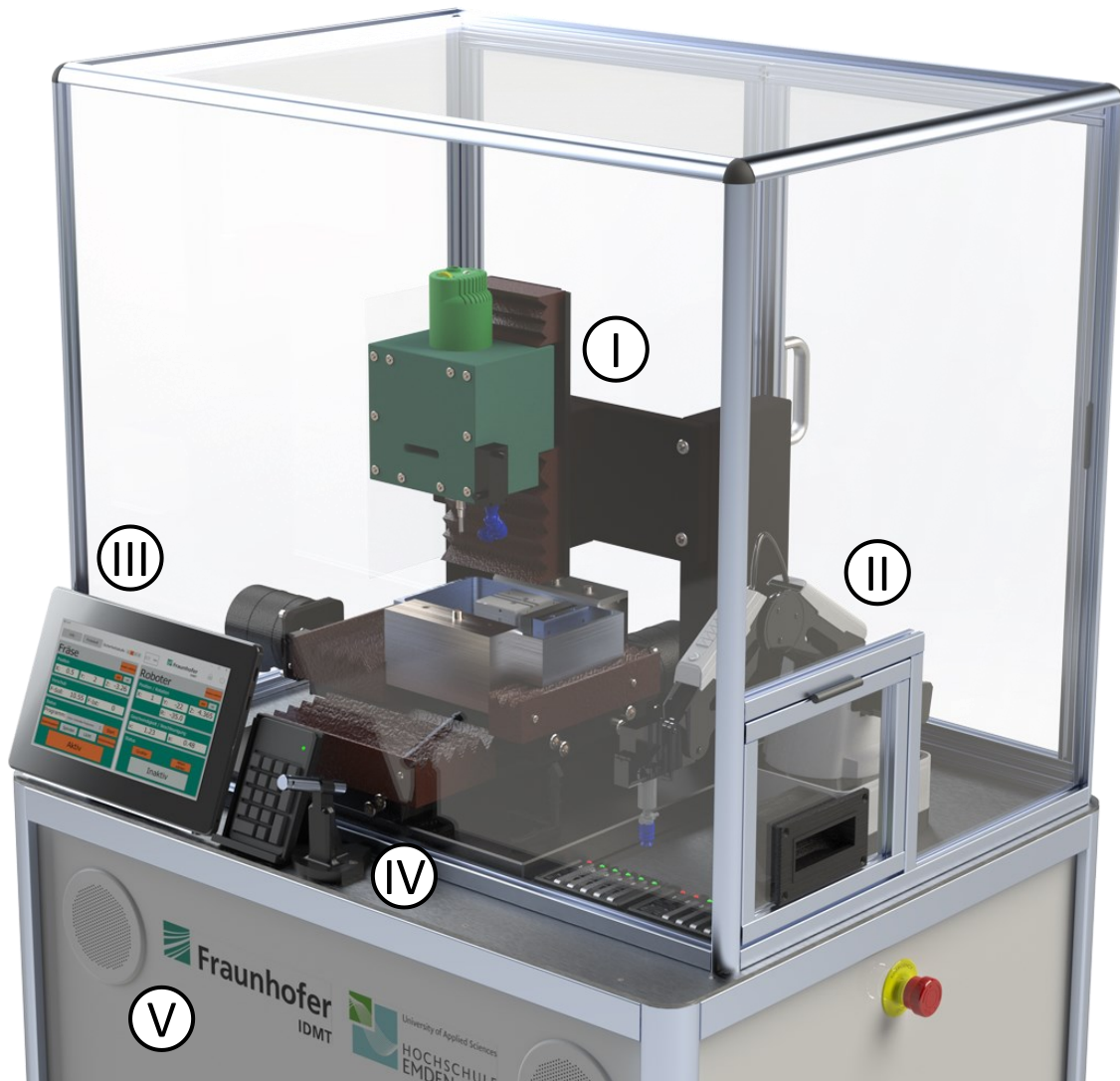


Abbildung 18: Rendering des CAD-Modells der »VCP-Technologieplattform I« mit sprach- und touchbasiertem HMI (I. 3-Achs Fräsmaschine, II. 4-Achs Roboter, III. Touchscreen, IV. Omnidirektionales Mikrofon, V. Lautsprecher) [NER23] © 2023 IEEE

Mit den beschriebenen Funktionen der Technologieplattform kann der Maschinenbediener in zwei verschiedenen Arbeitsmodi die Maschine nutzen. Zum einen kann er über Sprach- oder Touchbefehle konventionell einzelne Aktionen der unterschiedlichen Maschinenkomponenten auslösen. Gleichzeitig hat er aber auch die Möglichkeit, über zusammengesetzte Befehle simultane Aktionen unterschiedlicher Maschinenkomponenten zu triggern und so als Supervisor übergeordnet den Bearbeitungsprozess zu regeln.

Die Technologieplattform verfügt neben einer voll funktionsfähigen, am Fraunhofer-Institut für Medientechnologie IDMT in Oldenburg entwickelte ASR als Teil einer Sprachsteuerung über ein eigens entwickeltes GUI, das über einen Touchscreen bedient wird. Die einzelnen Befehle sind dabei in bis zu fünf Menüebenen geordnet und mit bis zu zehn Interaktionen ausführbar. Die Sprachbefehle für das Gesamtsystem der Technologieplattform I, die 3-Achs-Fräsmaschine und den 4-Achs-Roboter, sind über das jeweilige Hotword und die entsprechende Befehlsphrase ausführbar (vgl. **Tabelle 3**).

5.2.1 Sprachsteuerung

Die beiden in dieser Probandenstudie eingesetzten HMI-Modalitäten der VCP-Technologieplattform I bieten auf der technischen Seite der Schnittstelle eine Befehlssicherheit von 100 %. Das bedeutet, dass wenn der Maschinenbediener den richtigen Button in dem GUI auf dem Touchscreen drückt oder den richtigen Sprachbefehl ausspricht, wird auch der korrekte Maschinenbefehl erkannt und auf der VCP-Technologieplattform I ausgeführt. Speziell für die ASR als Teil der Sprachsteuerung ist dies nur durch ein dediziertes Training auf ein spezifisches Befehlsset (vgl. **Tabelle 3**) und eine nachgelagerte Synonymerkennung möglich. Grundsätzlich kann die verwendete ASR auf jedes Standard-Sprachmodell und zugehörige, spezifische Befehle trainiert werden. Das Training maximiert die Wahrscheinlichkeit, dass die spezifischen Sprachbefehle aus **Tabelle 3** der VCP-Technologieplattform I anstelle anderer Wörter aus dem verwendeten Sprachmodell erkannt werden. [NER23]

Ein weiterer, wesentlicher Einflussfaktor auf die Leistungsfähigkeit der eingesetzten ASR ist die Lautstärke von Umgebungsgeräuschen in der Laborhalle, die für diese Probandenstudie konstant unter 40 dB (A) gehalten wurde. Dadurch wird sichergestellt, dass der SNR selbst bei Probanden mit einer sehr leisen Stimme mindestens 0 dB beträgt. In der Realität sind 40 dB (A) Umgebungslautstärke unter dem Pegel der meisten Produktionsumgebungen, wodurch sichergestellt wird, dass diese keinen Einfluss auf die Eingabezeiten der Sprachsteuerung hat. Gleichzeitig kann so das omnidirektionale Mikrofon an der VCP-Technologieplattform I genutzt werden. [NER23] In einer realen Produktionsumgebung kann ein Headset mit einem Mikrofon nah am Mund des Maschinenbedieners den SNR deutlich verbessern, um den Einfluss eines erhöhten Störgeräuschpegels auszugleichen.

Gleichwohl sind Headsets nicht für alle Produktionsunternehmen geeignet, da ein Headset für jeden Maschinenbediener teurer als ein zentrales stationäres Mikrofon oder ein Mikrofonarray ist. Zusätzlich können Headsets während der Arbeit vom Kopf des Maschinenbedieners in die

Maschine fallen, verloren gehen oder durch einen nicht geladenen Akku ausfallen. Des Weiteren kann ein Headset dem Maschinenbediener beim permanenten Tragen ein Gefühl der Überwachung geben, was bei einem ortsfesten Mikrofon an der Maschine gegebenenfalls weniger der Fall ist. So gut Headsets aus audiotechnologischer Sicht zur Steigerung des SNRs und damit der Robustheit einer Sprachsteuerung auch sind, gibt es trotzdem organisatorische Gründe, die für ein stationäres Mikrofon oder ein Mikrofonarray sprechen. Für die Untersuchungen der Effizienz von Sprachsteuerung in der Produktion wird eine robuste und damit effektive Sprachsteuerung vorausgesetzt.

Akustisches Modell und trainiertes Sprachmodell

In dieser Arbeit wurde eine hybride ASR aus einem generellen akustischen Modell und einem auf die Aufgabe zur Steuerung von Maschinen angepassten Sprachmodell verwendet. Das akustische Modell basiert auf einem Time Delay Neural Network (TDNN) und einem Long Short-Term Memory (LSTM), wie das von Peddinti et al. genutzte System mit sieben versteckten Schichten von 800 Neuronen und Rectified Linear Unit (ReLU) Aktivierungsfunktion, woraus sich ein Netzwerk aus insgesamt ca. 25 Millionen Parametern ergibt. [GaPe18, PPK15] Als Optimierungskriterium wurde das sogenannte »*lattice-free maximum mutual information criterion*« verwendet [PPG16]. Um das akustische Modell robust gegen Störgeräusche zu machen, wird ein interner Korpus von 1.000 Stunden deutscher Sprache mit unterschiedlichen Raumimpulsantworten, Störgeräuschen und SNRs auf insgesamt 8.000 h augmentiert. Aus diesen 16 kHz Audiodaten werden anschließend 40-dimensionale log-Mel Filter Bank Merkmale für das Training des Deep Neural Network (DNN) erstellt. [NER23]

Zur Generierung des 4-Gram Sprachmodells wurde ein Trainingsset erstellt, das alle in dieser Probandenstudie eingesetzten 33 Sprachbefehle sowie deren leicht angepasste Varianten (insgesamt 2270 Varianten) enthält. Ergänzend wurde zur Steigerung der Robustheit gegenüber Hintergrundsprache und Wörtern, die nicht in der Liste an Sprachbefehlen vertreten sind, das Sprachbefehl-spezifische Sprachmodell mit einem generellen hauseigenen 4-Gram Sprachmodell für eine effiziente Verarbeitung beschnitten. [NER23]

5.2.2 Touchsteuerung

Die Touchsteuerung der Technologieplattform basiert auf einem GUI, das den HMIs gängiger SPS- und Produktionsmaschinenhersteller entspricht, um valide Forschungsergebnisse zu gewährleisten (vgl. **Kapitel 2.2 & Abbildung 19**). [NER23]

Allgemeine Menüfunktionen

Oben links befindet sich ein Button mit der Aufschrift »Info«, der ein Untermenü öffnet, in dem die zur Verfügung stehenden Sprachbefehle der VCP-Technologieplattform I für die 3-Achs-Fräsmaschine, den 4-Achs-Roboter und die Peripherie aufgelistet sind. Die daneben liegende Schaltfläche »Protokoll« zeigt dem Bediener in einem Untermenü die bereits ausgeführten Sprach- und Touchbefehle seit dem Einschalten der VCP-Technologieplattform I sowie aufgetretene Fehlermeldungen. Über den angrenzenden Button »Sicherheitsstufe« kann der Maschinenbediener zwischen den zuvor beschriebenen vier Sicherheitsstufen wechseln. Über das Lautsprechersymbol kann der Maschinenbediener die Lautstärke des Audiofeedback der VCP-Technologieplattform I in den Stufen 1 bis 10 einstellen. Der Button mit dem Zahnradsymbol direkt rechts neben dem Fraunhofer IDMT-Logo öffnet die Einstellungen zur Anpassung des JOG-Vorschubs, des JOG-Inkrementes, des Werkzeugschneiden Radius und der Werkzeuglänge sowie die Werkzeug- und Magazinauswahl. [NER23]

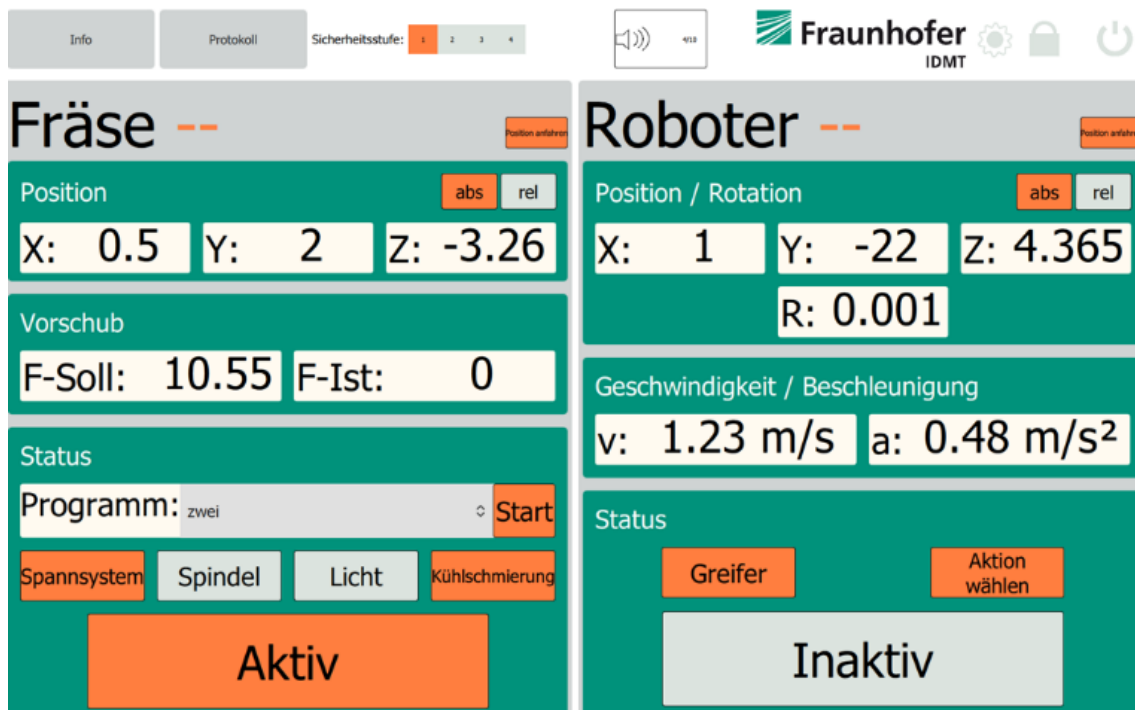


Abbildung 19: Selbst entwickeltes GUI auf dem Touchscreen der »VCP-Technologieplattform I« [NER23]

© 2023 IEEE

Der Button in Form eines Schlosses in der rechten oberen Ecke sperrt die GUI gegen ungewollte Sprach- und Touchbefehle. Über den »Power«-Button in der rechten oberen Ecke kann die VCP-Technologieplattform sowohl gestartet als auch heruntergefahren werden. [NER23]

3-Achs-Fräsmaschine

Die linke Bildschirmhälfte mit dem Titel »Fräse« stellt alle Touch-Befehle für die 3-Achs-Fräsmaschine zur Verfügung. In der grauen Kopfzeile des Menübereichs der Fräse können über den orangenen Button »Position anfahren« zuvor definierte Positionen, wie die Home-, die Referenzierungs- oder die Ladeposition, angefahren werden. In der ersten grünen Sektion darunter kann die 3-Achs-Fräsmaschine sowohl absolut als auch relativ entlang ihrer X-, Y- und Z-Achse verfahren werden. Die zweite grüne Sektion zeigt den Soll-Wert des Vorschubs aus dem Bearbeitungsprogramm und den aktuellen Ist-Vorschub in der Bearbeitung an. Die dritte und größte grüne Sektion stellt die Auswahl und den Start des Bearbeitungsprogramms sowie Peripheriefunktionen, wie die Ansteuerung des pneumatischen Spannfutters, der Frässpindel, der Prozesskammerbeleuchtung und des Kühlschmiermittelsystems, zur Verfügung. In der Mitte des großen grünen unteren linken Feldes zeigt das Feld »Aktiv/Inaktiv« mit den Farben Orange und Grau den aktuellen Betriebszustand der 3-Achs-Fräsmaschine an. [NER23]

4-Achs-Roboter

Die rechte Hälfte des GUI stellt alle Touch-Befehle für den 4-Achs-Roboter zur Verfügung. Analog zur 3-Achs-Fräsmaschine ermöglicht die orange Schaltfläche »Position anfahren« im grauen Kopfbereich das Anfahren zuvor definierter Positionen wie der Referenzierungs- oder Einrichtungsposition. In den darunter liegenden ersten grünen Sektionen können die vier Achsen des Roboters X, Y, Z und R sowohl relativ als auch absolut verfahren werden. Die darunter liegende grüne Sektion zeigt im Feld »Geschwindigkeit/Beschleunigung« die aktuellen Geschwindigkeits- und Beschleunigungswerte des Roboters an. Die dritte grüne Sektion ganz unten rechts mit dem Titel »Status« ermöglicht die Aktivierung und Deaktivierung des pneumatischen Vakuumsaugers am Ende des Roboterarms. Zuletzt kann der Maschinenbediener über die Schaltfläche »Aktion wählen« Rohteile, sogenannte Halbzeuge in das pneumatische Spannfutter der Fräsmaschine beladen und fertig bearbeitete Aluminiumbauteile entladen. Genau wie bei der 3-Achs-Fräsmaschine zeigt das Feld »Aktiv/Inaktiv« in orange und grau den aktuellen Betriebszustand des Roboters an. [NER23]

5.3 Design of Experiment

Zur Aufnahme der Eingabezeiten der beiden HMI-Modalitäten gaben in dieser Probandenstudie fünf weibliche und zehn männliche Probanden im Alter zwischen 22 und 39 Jahren 33 Maschinenbefehle in einer festen Reihenfolge (vgl. **Tabelle 3**) jeweils drei Mal per Sprache und per Touch ein. Daraus resultierten insgesamt 2970 Befehle (jeweils 1485 Sprach-/Touchbefehle) und

45 Eingabezeit-Messwerte für jeden einzelnen Sprach- und Touchbefehl. Der Aufbau des Experiments entsprach hierbei einem Within-Group Design, da jeder der Probanden die Maschinenbefehle über beide HMI-Modalitäten eingegeben hat. Durch dieses bewusst gewählte Studiendesign wurde die Anzahl an Durchläufen für die Probanden zwar verdoppelt, jedoch konnten interindividuelle Differenzen zwischen den Probanden, wie die grundsätzliche Fähigkeit einzelner Probanden, HMIs jeglicher Form besonders schnell bedienen zu können, berücksichtigt werden. Dadurch wird der Einfluss einer ungleichen Verteilung von langsamen und schnellen Probanden auf die erhobenen Daten im Verhältnis zu einem Between-Group Design, bei dem jeder Proband die Maschinenbefehle nur mithilfe einer der beiden HMI-Modalitäten eingeben würde, minimiert. Auf der Gegenseite kann ein Within-Group Design zu Lerneffekten führen, die die Leistung der Probanden verbessern, wenn diese eine Aufgabe mehrfach wiederholen oder die grundsätzliche Aufgabenstellung bereits aus den Durchläufen der ersten HMI-Modalität kennen. [LFH17]

Zum Ausgleich dieser Lerneffekte beginnen sieben Probanden mit der Sprach- und acht Probanden mit der Touchsteuerung. Die Auswahl der jeweiligen Probanden erfolgt dabei randomisiert, um valide kausale Zusammenhänge identifizieren zu können [LFH17, Oeh00]. Alle Probanden sind deutsche Muttersprachler sowie Studierende der Ingenieurwissenschaften oder wissenschaftliche Mitarbeiter mit Erfahrung in der Bedienung von Produktionsmaschinen. Gleichwohl hat keiner der Probanden im Vorhinein eine Produktionsmaschine oder die eingesetzte VCP-Technologieplattform I per Sprache bedient, wodurch die eingesetzte Sprach- und Touchbefehlsliste (vgl. **Tabelle 3**) für alle Probanden neu ist. Jeder Proband erhielt die gleiche Einführung mit einer anschließenden Testphase, in der er die Möglichkeit hatte, sowohl Sprach- als auch Touchbefehle an der VCP-Technologieplattform I einzugeben. [NER23]

Alle in dieser Probandenstudie eingesetzten Befehle sind repräsentativ für die Interaktionen mit CNC-gesteuerten Produktionsmaschinen wie automatisierten Bearbeitungszentren sowie Fräsmaschinen und Robotern. Drei verschiedene Hotwords (»RIVA«, »Roboter« und »CNC«) wurden in Kombination mit unterschiedlichen Befehlen verwendet, um Aktionen an der VCP-Technologieplattform I auszulösen. [NER23] Das Hotword »RIVA« wird mit Befehlen kombiniert, die entweder die gesamte VCP-Technologieplattform I, bestehend aus 3-Achs-Fräsmaschine und 4-Achs-Roboter, oder aber Peripheriefunktionen wie das Prozesskammerlicht, das Spannsystem oder das Kühlschmiermittelsystem adressieren. »RIVA« ist dabei ein frei gewähltes Akronym für »Robust Industrial Voice Application«. Die Hotwords »Roboter« und »CNC« werden mit Befehlen

zur Steuerung des 4-Achs-Roboters und der 3-Achs-CNC-Fräsmaschine kombiniert. Die drei unterschiedlichen eingesetzten Hotwords ermöglichen nicht nur eine Zuordnung der Befehle zur jeweiligen Maschinenkomponente, sondern reduzieren auch eine Beeinflussung der Daten durch ein bestimmtes Hotword.

Tabelle 3: Maschinenbefehle für Sprach- und Touchsteuerung nach [NER23] © 2023 IEEE

#	Sprachbefehl	Touchbefehl	
		Menüebenen	Interaktionen
1	RIVA Licht an	0	1
2	Sicherheitsstufe 2	0	1
3	RIVA lauter	1	3
4	Roboter Position X 12	1	3
5	Roboter Position Y 14	1	3
6	Roboter Position Z 23	1	3
7	CNC Position Y 34	1	3
8	CNC Position Z 4	1	3
9	CNC Position X 55	1	3
10	Sicherheitsstufe 1	0	1
11	Roboter Position X 71	1	3
12	Roboter Position Y 13	1	3
13	Roboter Einrichteposition anfahren	3	5
14	CNC Position X 49	1	3
15	CNC Position Z 13	1	3
16	CNC Einrichteposition anfahren	3	5
17	RIVA Programm 17 laden	3	6
18	RIVA Kühlschmierung an	0	1
19	RIVA Kühlschmierung aus	0	1
20	RIVA Werkzeug 20 auswählen	3	7
21	RIVA Programm 18 laden	3	6
22	RIVA Programm 18 auswählen	3	6
23	RIVA Magazin 20 auswählen	3	6
24	RIVA Nullpunkt X 12 Y 17 Z 31	3	7
25	Roboter Einrichteposition anfahren	3	5
26	CNC Einrichteposition anfahren	3	5
27	Roboter JOG Vorschub 10	5	7
28	CNC JOG Vorschub 100	5	7
29	CNC JOG Inkrement 100	5	7
30	Roboter JOG Inkrement 1000	5	7
31	RIVA Werkzeug 20 Schneidenradius 12,125	5	10
32	RIVA Werkzeug 20 Ausspannlänge 122,45	5	10
33	RIVA ausschalten	0	1

Über die jeweils drei Durchläufe per Sprache und per Touch wird die Komplexität der jeweils 33 Maschinenbefehle sukzessiv, gemessen an den Menüebenen und der benötigten Anzahl der Interaktionen auf dem GUI sowie dem resultierenden Sprachbefehl, gesteigert (vgl. **Tabelle 3**). Für die Probandenstudie werden ausschließlich die Sicherheitsstufen 1 und 2 der VCP-Technologieplattform I genutzt, um ein ungewünschtes und befehlsabhängig langes Audio-Feedback, das zu einer Verfälschung der gemessenen Eingabezeiten führt, zu vermeiden (vgl. **Kapitel 5.2.2**). Die Reihenfolge der Maschinenbefehle innerhalb der Probandenstudie bildet einen Kompromiss zwischen einem realistischen Produktionsablauf und der beschriebenen, sukzessiven Steigerung der Befehlskomplexität. Zur Durchführung des Experiments erhalten alle Probanden eine Liste mit allen Maschinenbefehlen, wie sie in **Tabelle 3** zu sehen sind. [NER23]

Im ersten Schritt starten die Probanden die VCP-Technologieplattform I (vgl. **Tabelle 3, Befehl #1**), wählen die Sicherheitsstufe 2 mit einem Text-To-Speech Audio-Feedback aus (**Befehl #2**) und erhöhen die Lautstärke des Audio-Feedbacks (**Befehl #3**). [NER23]

Anschließend positionieren die Probanden sowohl den 4-Achs-Roboter als auch die 3-Achs-Fräsmaschine mit absoluten Positionierungsbefehlen und eingespeicherten Positionierungsbefehlen (**Befehl #4-9, #11-16**). [NER23] Relative Positionierungen werden aus organisatorischen Gründen nicht genutzt, obwohl die VCP-Technologieplattform diese Möglichkeit bietet. So wurde gewährleistet, dass alle Maschinenkomponenten am Anfang des folgenden Durchlaufs wieder einen definierten Status haben, auch nach Fehleingaben oder dem Auslassen von Maschinenbefehlen. Der Einfluss einer relativen Positionierung im Verhältnis zu einer absoluten Positionierung im Kontext der zu untersuchende These ist unerheblich.

Wie in einem realen Produktionsablauf ändern die Probanden zwischen den Roboter- und Fräsmaschinenbefehlen erneut die Sicherheitsstufe des HMIs (**Befehl #10**), aktivieren oder deaktivieren das Kühlschmiermittelsystem der Fräsmaschine (**Befehl #18**) und wählen sowohl Fräswerkzeuge als auch Fräsprogramme aus (**Befehl #17, #20-23**). Anschließend setzen sie einen neuen Nullpunkt (**Befehl #24**) und fahren sowohl Roboter als auch Fräsmaschine in ihre Ausgangsposition (**Befehl #25, #26**). Mit den folgenden Maschinenbefehlen werden sowohl der Vorschub als auch das Inkrement für den Handbetrieb, der sogenannte JOG-Vorschub und das JOG-Inkrement für den Roboter und die Fräsmaschine in der VCP-Technologieplattform verändert (**Befehl #27-30**). Danach werden für ein Werkzeug der Schneidenradius und die Ausspannlänge eingegeben (**Befehl #31-32**), bevor die VCP-Technologieplattform I mit einem finalen Maschinenbefehl heruntergefahren wird (**Befehl #33**). [NER23]

Die Komplexität eines Maschinenbefehls wird über die Anzahl an Menüebenen (vgl. **Tabelle 3**, »Menüebenen«) und die Anzahl der Interaktionen (Tastendrucke) (vgl. **Tabelle 3**, »Interaktionen«), die auf dem GUI notwendig sind, um diesen auszuführen, quantifiziert. Eine Sprachsteuerung kann nicht äquivalent nach diesen beiden Parametern quantifiziert werden, auch wenn die Komplexität der Sprachbefehle, gemessen an der inhaltlichen Komplexität und der Länge dieser proportional zum Touchbefehl, zunimmt. Während des Experiments steigen die Menüebenen von null bis fünf über die 33 einzugebenden Maschinenbefehle an, wobei die Menüebene null, die sogenannte »Root«-Ebene oder den Startbildschirm beschreibt. Befehle auf dieser Ebene sind ohne Untermenüebenen oder Pop-Up-Fenster erreichbar. Daraus folgt, dass die einfachsten einzugebenden Befehle auf der Menüebene null nur eine Interaktion zur Ausführung benötigt, wohingegen die komplexesten Befehle in dieser Probandenstudie zehn Interaktionen auf bis zu fünf Menüebenen benötigen. [NER23]

Die zentrale Messgröße in dieser Probandenstudie ist entsprechend der DIN EN ISO 9241-11 (vgl. **Kapitel 2.4.1**) und der Anforderung 3 des Lastenhefts (vgl. **Kapitel 1.4**) die Eingabezeit inklusive etwaiger Eingabefehler. Die gemessenen Eingabezeiten für die **Anzahl an Menüebenen** sind dabei dieselben wie die für die **Anzahl an Interaktionen**. Die **Anzahl an Interaktionen** ist lediglich eine feinere Skalierung der **Anzahl an Menüebenen**, die detailliertere Korrelationen zwischen der Befehlskomplexität und der benötigten Eingabedauer ermöglichen, da in einer Menüebene auch mehrere Interaktionen erfolgen können. Daraus ergeben sich präzisere Ergebnisse für Aussagen zur Effizienz von Sprach- und Touchsteuerung. [NER23]

Reale Produktionsabläufe können noch komplexere Maschinenbefehle beinhalten, die dafür aber nur sehr selten genutzt werden. Um ein repräsentatives Studiendesign zu erhalten, wurden sowohl Gespräche mit Produktionsexperten von HMI-, Steuerungs- und Maschinenhersteller als auch Betreibern von Produktionsmaschinen sowie erfahrenen Maschinenbedienern geführt. [NER23]

In **Abbildung 20** wird die Korrelation zwischen Sprach- und Touchbefehlen in der jeweiligen HMI-Modalität exemplarisch anhand von vier zufällig ausgewählten Maschinenbefehlen dargestellt. Die Spalten zeigen die jeweilige Menüebene im GUI an, während die Zeilen die benötigte Anzahl an Interaktionen zur Eingabe des Befehls repräsentieren. Eine Interaktion wird durch einen dünnen schwarzen Pfeil dargestellt. Die großen Pfeile darunter zeigen den entsprechenden Sprachbefehl an (vgl. **Tabelle 3**). [NER23]

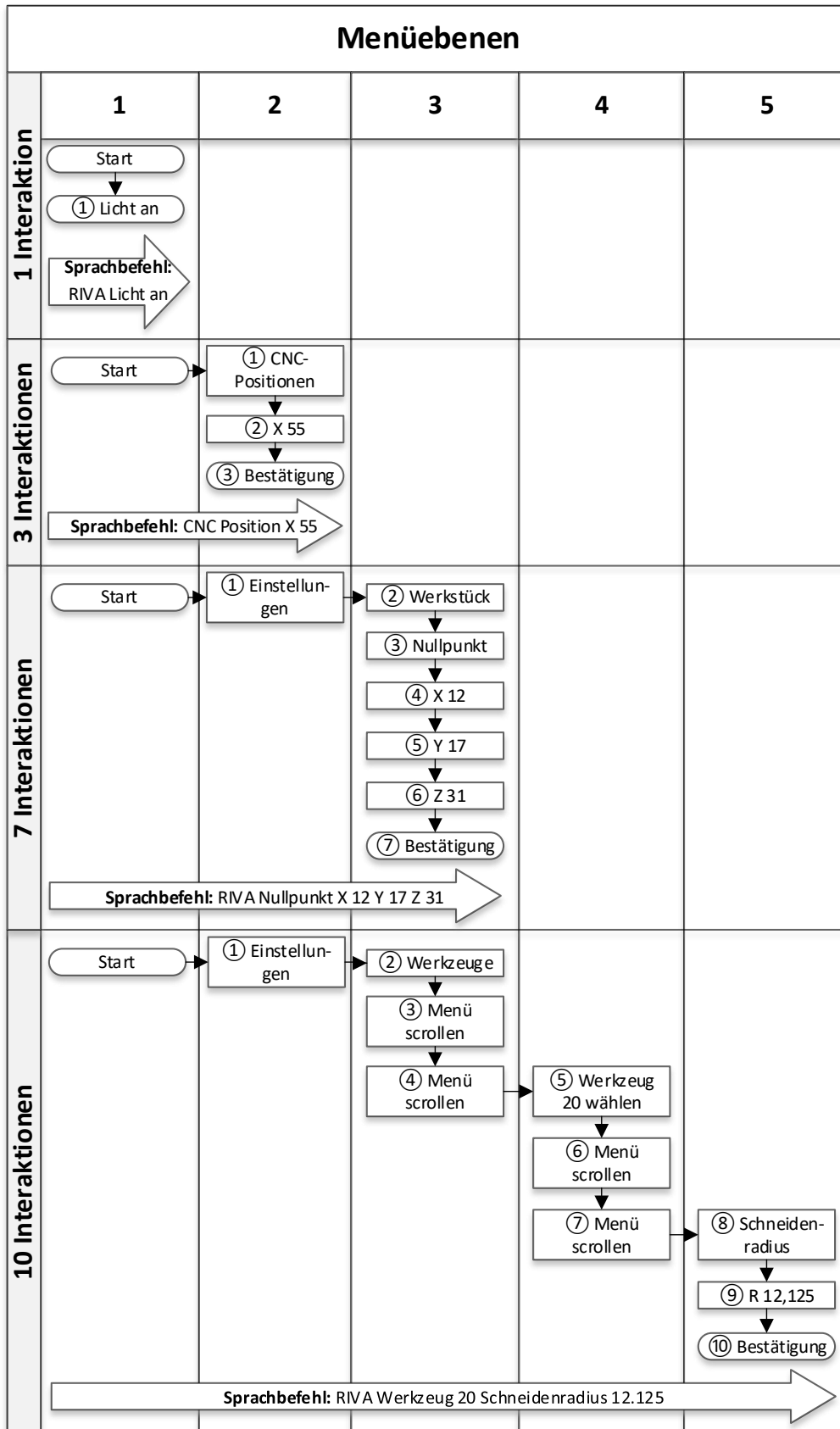


Abbildung 20: Korrelation der Komplexität von Sprach- und Touchbefehlen, kategorisiert nach Menüebenen und der Anzahl an Interaktionen innerhalb des HMI nach [NER23] © 2023 IEEE

Für Befehle, die mehr als eine Interaktion innerhalb einer Menüebene benötigen, stehen die korrespondierenden Interaktionen in der Spalte derselben Menüebene. Das beschriebene Studiendesign mit seinem möglichst realitätsnahe Ablauf und Auftreten von Befehlen soll eine bestmögliche Konzentration und Fokussierung der Probanden ermöglichen [DIN18]. Die gesamte Durchführung des Experiments dauert für die Probanden jeweils 90 Minuten, was der empfohlene maximale Dauer einer Probandenstudie in dieser Form entspricht [Nie05]. In dieser Zeit ist auch die Einführung in das Experiment durch den Studienleiter, eine vorangehende Testphase der Probanden, um das Verhalten beider HMI-Modalitäten kennen zu lernen sowie das Ausfüllen eines kurzen abschließenden Fragebogens zum subjektiven Empfinden der Probanden enthalten. In **Abbildung 18** wird eine Probandin (vgl. **Abbildung 18**, VII) vor der zuvor beschriebenen VCP-Technologieplattform I gezeigt. [NER23]



Abbildung 21: Interaktion der Probanden mit der »VCP-Technologieplattform I« innerhalb der Forschungsstudie (I. 3-Achs Fräsmaschine, II. 4-Achs Roboter, III. Touchscreen, IV. Omnidirektionales Mikrofon, V. Lautsprecher, VI. In-Ear-Headset (Hearpiece), VII. Probandin) [NER23] © 2023 IEEE

Die präzise und vor allem konsistente Messung der Eingabezeiten des jeweiligen HMI ist nicht so trivial, wie es klingen mag. Die größte Hürde besteht darin, den Beginn der Interaktion des

Probanden mit dem HMI messtechnisch zu erfassen, weshalb für diese Probandenstudie ein spezielles Vorgehen genutzt wird, was in **Abbildung 22** dargestellt wird.

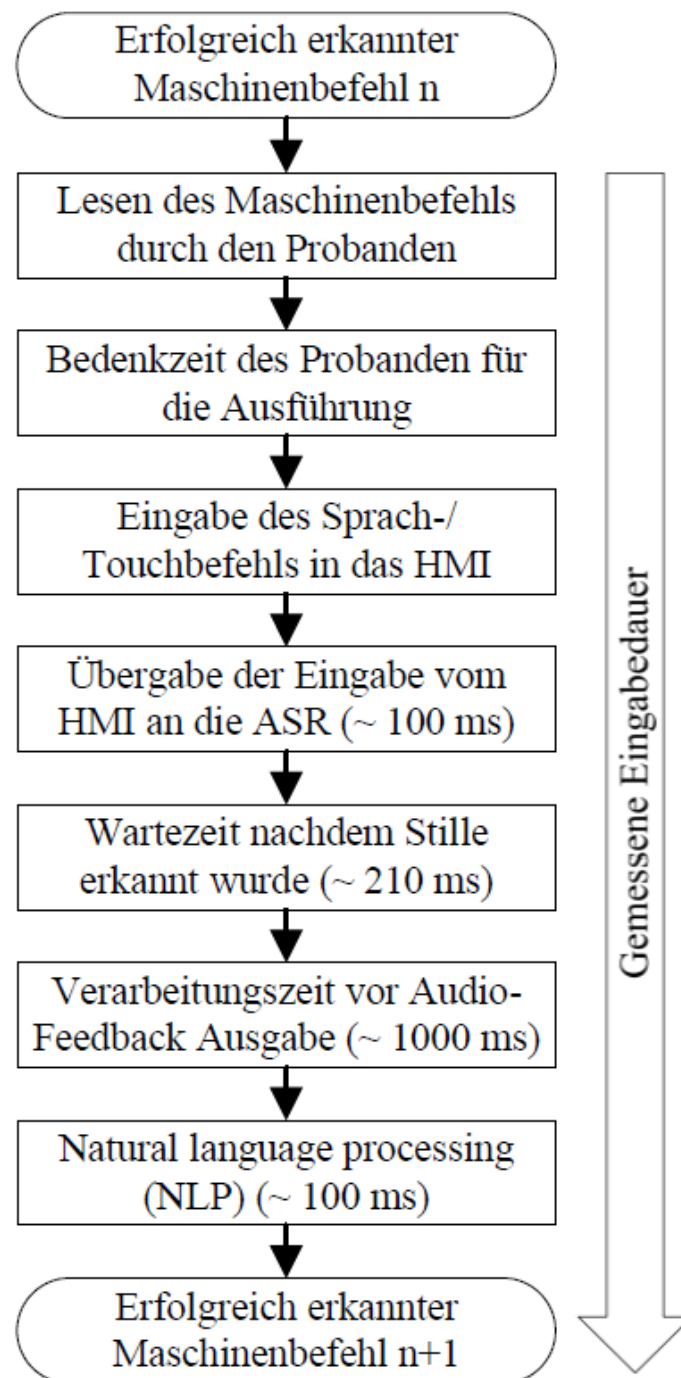


Abbildung 22: Messmethode für die Eingabezeiten inklusive der ASR der VCP Technologieplattform [NER23] © 2023 IEEE

Die Messung der Eingabezeit beginnt direkt, nachdem der vorherige Maschinenbefehl erfolgreich vom System erkannt wurde. Dadurch beinhalteten die Eingabezeiten sowohl das Ablesen des Folgebefehls (vgl. **Abbildung 22** – »Lesen des Maschinenbefehls durch den Probanden«) von

der Befehlsliste (vgl. **Tabelle 3**) als auch die Zeit zum Überlegen, wie der Maschinenbefehl mit der jeweiligen HMI-Modalität (Touch/Sprache) ausgeführt wird (vgl. **Abbildung 22** – »Bedenkzeit des Probanden für die Ausführung«). Danach beginnen die Probanden mit Aktionen auf dem Touchscreen oder beginnen den Sprachbefehl auszusprechen (vgl. **Abbildung 22** – »Eingabe des Sprach-/Touchbefehls in das HMI«). Diese Methode zur Messung der Eingabezeiten wird gewählt, da die Zeit vom Beginn des Lesens des aktuellen Maschinenbefehls über das Nachdenken und über die Ausführung bis hin zur ersten Berührungsinteraktion oder zur ersten gesprochenen Silbe nicht einfach gemessen werden kann. Gegebenenfalls wäre dies über EEG möglich. Zur Reduzierung dieses Aufwands wird einfach die Zeit von der erfolgreichen Ausführung des vorherigen Befehls bis zur erfolgreichen Ausführung des darauffolgenden Befehls gemessen (vgl. **Abbildung 22** – »Gemessene Eingabedauer«). Zur Wahrung der Konsistenz der Messwerte werden die Probanden gebeten die Maschinenbefehle zwar in einer natürlichen Art und Weise, jedoch so schnell, wie es einem realen Produktionsablauf entsprechen würde, über die jeweilige HMI-Modalität einzugeben. [NER23] Trotz dieser Anweisung beinhalten alle gemessenen Eingabezeiten sowohl die interindividuelle Bedenkzeit des jeweiligen Probanden für den jeweiligen Befehl als auch die für Fehleingaben verstrichene Zeit gemäß DIN EN ISO 9241-11 (vgl. **Kapitel 2.4.1**) [DIN18]. Da die Bedenkzeit jedoch sowohl für die Sprach- als auch für die Touchsteuerung gemessen wird, sollte diese keinen signifikanten Einfluss auf die gemessenen Eingabezeiten haben [NER23]. Die Fehleingaben führen gemäß der DIN EN ISO 9241-11 gleichermaßen für beide HMI-Modalitäten zu erhöhten Eingabezeiten (vgl. **Kapitel 2.4.1**) [DIN18].

5.4 Ergebnisse

Im Folgenden werden die gemessenen Eingabezeiten analysiert. Zusammenfassend zeigen die Anzahl an Menüebenen, Interaktionen und Wiederholungen in der jeweiligen HMI-Modalität eine signifikante Korrelation mit den gemessenen Eingabezeiten. Aus diesem Grund ist das **Kapitel 5.4** nach diesen drei Parametern gegliedert. Vorangestellt ist das **Kapitel 5.4.1** mit einer statistischen Analyse der Messwerte. [NER23]

5.4.1 Statistik

Die gemessenen Eingabezeiten wurden mithilfe des lme4-Pakets für statistische Berechnungen in der Software »R« analysiert [BMB15]. Dazu wurden drei Linear Mixed-Effect Models (LMM), jeweils eins für den zu untersuchenden Parameter (Menüebenen – vgl. **Kapitel 5.4.2**, Interakti-

onen – vgl. **Kapitel 5.4.3**, Wiederholungen – vgl. **Kapitel 5.4.4**), genutzt, um die abhängige Variable Eingabedauer mithilfe einer Reihe von experimentellen Manipulationen vorauszusagen. Alle drei Modelle beinhalteten die Probanden als Zufallsvariable. [NER23] Mithilfe der R package performance wurden die Effekte »linearity«, »homogeneity of variance, collinearity«, »influential observations«, »normality of residuals« und »normality of random effects« visuell überprüft [LBP21]. Da die visuelle Analyse ergibt, dass die Residuen aller LMMs nicht normalverteilt sind, werden die Ergebnisse mit dem natürlichen Logarithmus transformiert [NER23]. Zuletzt wird mit dem modellbasierten Paket eine Kontrastanalyse mit Holm Korrekturen durchgeführt [MBP20]. Für alle statistischen Tests wurde der Alpha-Fehler auf 0,05 gesetzt [NER23].

5.4.2 Einfluss der Menüebenen auf die zeitliche Effizienz

Maschinenbediener müssen an moderne GUIs von Produktionsmaschinen für Befehle, die täglich genutzt werden, durch bis zu fünf Menüebenen navigieren. Die begrenzte Größe von Bildschirmen zwingt Softwareentwickler dazu, Menüebenen in ihren GUI einzuführen. Auch wenn darauf geachtet wird, dass die grundsätzliche Architektur aller Menüebenen in einem GUI gleich ist, stellt jede einzelne Menüebene trotzdem Informationen und Funktionen in einer einzigartigen Struktur dar. Dies führt dazu, dass der Bediener eine gewisse Zeit benötigt, um sich zu orientieren und den gewünschten Befehl zu finden. Sowohl ein einfaches und gut strukturiertes GUI als auch die Erfahrung des Bedieners mit dem jeweiligen GUI ermöglichen eine effiziente Orientierung und Ausführung des benötigten Touchbefehls. Analog führen ein kurzes Hotword gepaart mit kurzen Sprachbefehlen innerhalb einer Sprachsteuerung zu einer effizienten Eingabe von Maschinenbefehlen. Gleichzeitig kann die Genauigkeit der Spracherkennung und die daraus resultierende Befehlssicherheit unter besonders kurzen Sprachbefehlen leiden. Deshalb stellt die Länge von Sprachbefehlen immer einen Kompromiss zwischen Effizienz und Befehlssicherheit dar. Je mehr Sprachbefehle ein trainiertes Sprachmodell für die Steuerung einer Produktionsmaschine beinhaltet, desto wichtiger ist es, dass sich diese eindeutig voneinander unterscheiden, um eine hohe Befehlssicherheit zu erreichen (vgl. **Kapitel 2.4.3**). [NER23]

Die mithilfe des natürlichen Logarithmus transformierte abhängige Ergebnisvariable Eingabezeit wurde mittels des ersten LMM vorausgesagt. Das Modell schätzt die sogenannten »Fixed-Effects« für die Menüebenen, die HMI-Modalität (Sprache/Touch) und deren Interaktion. Eine parametrisierte Varianzanalyse (engl.: analysis of variance (ANOVA)) dieser Modelle zeigte die folgenden statistisch signifikanten Effekte: Menüebenen ($F(3,2948) = 974.8$, $p < 0.001$), HMI Mo-

alität ($F(1,2948) = 1019.8$, $p < 0.001$) und deren Interaktion ($F(3,2948) = 646$, $p < 0.001$). Das bedeutet, dass die gemessene Eingabezeit der Probanden abhängig von der Menüebene, der HMI Modalität und deren Interaktion ist. Hohe F-Werte und niedrige p-Werte zeigen, dass diese Effekte statistisch signifikant sind. Die signifikante Interaktion wurde mit einer Kontrastanalyse bewertet, wobei sich alle Kontraste als hoch signifikant herausstellten ($p < 0.001$). In **Abbildung 23** werden die zurücktransformierten Ergebnisse der Eingabedauer mit den beschrifteten Pfeilen dargestellt. [NER23] Dabei werden auf der Abszissenachse sowohl die Menüebenen von null bis fünf als auch die jeweilige HMI-Modalität für jede Menüebene dargestellt. Auf der Ordinate ist die Eingabedauer für die jeweilige HMI-Modalität über alle Probanden und alle Wiederholungen zu sehen. Die rote Linie stellt jeweils den Median pro Menüebene und HMI-Modalität über alle Probanden und Durchläufe dar, während die grüne/blau Box die Interquartil Range (IQR) zwischen unterem und oberem Quartil, in der sich 50 % der Daten befinden, einrahmt. Die grüne Box rahmt die Daten der Sprachsteuerung ein, während die blaue Box die Daten der Touchsteuerung beinhaltet.

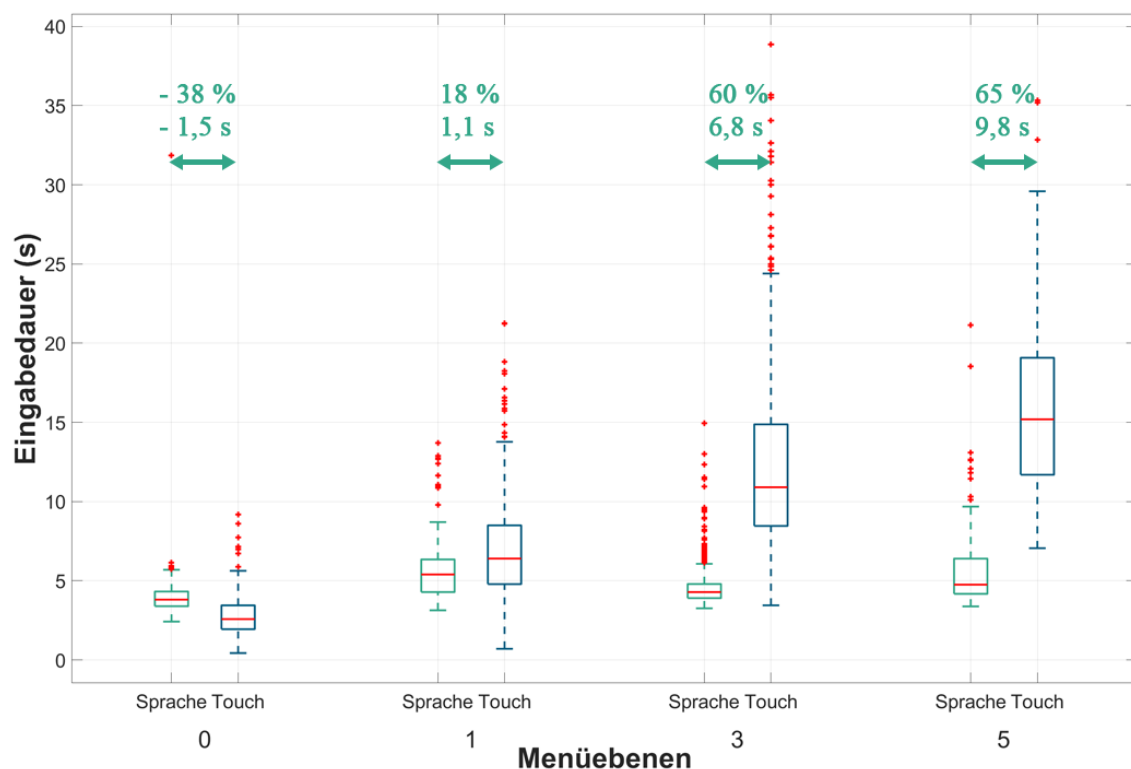


Abbildung 23: Einfluss der Menüebenen auf die Eingabedauer. Jeder Boxplot repräsentiert die Verteilung der Eingabezeit über alle 33 Befehle, 15 Probanden und drei Wiederholungen für eine spezifische Menüebene (0, 1, 3 oder 5) und Modalität (Sprach-/Touchsteuerung) [NER23] © 2023 IEEE

Die gestrichelten grünen/blauen Linien bilden die Längen der Whisker mit $1,5 \cdot \text{IQR}$. Alle Messwerte außerhalb der Whisker werden als Ausreißer mit einem roten Kreuz gekennzeichnet. Ausreißer oberhalb von $1,5 \cdot \text{IQR}$ werden dabei als »milde« Ausreißer, solche oberhalb von $3,0 \cdot \text{IQR}$ als »extreme« Ausreißer bezeichnet. Die grünen Pfeile zeigen die Unterschiede der Eingabezeiten für die jeweilige Modalität für jede Menüebene in Prozent und Sekunden.

Für die Touchsteuerung nimmt die Eingabedauer monoton proportional zum Anstieg der Menüebenen zu. Einfache Touchbefehle auf der Menüebene 0 werden von den Probanden im Median mit 2,5 s Eingabedauer ausgeführt. Für komplexe Touchbefehle der Menüebene 5 benötigen die Probanden hingegen im Median 15 s Eingabedauer. Die Eingabedauer der Sprachsteuerung pendelt hingegen im Median um 5 s Eingabedauer, unabhängig von der jeweiligen Menüebene. Im Vergleich ist die Touchsteuerung nur auf der Menüebene 0 schneller als die Sprachsteuerung. Für die in dieser Probandenstudie genutzten Befehle beträgt die Differenz der Medianwerte der beiden HMI-Modalitäten -1,5 s (-38 %). Die Eingabedauer der Sprachsteuerung ist hingegen für die Menüebene 1 bereits 1,1 s (18 %) schneller als die Touchsteuerung. Für die Menüebene 3 und 5 steigen die Median-Differenzen sogar auf 6,8 s (60 %), bzw. 9,8 s (65 %) an. Des Weiteren weisen sowohl der Interquartilsabstand als auch die oberen und unteren Whisker der Touchsteuerung eine viel breitere Streuung als die Pendants der Sprachsteuerung auf. Diese Streuung nimmt zusätzlich auch noch proportional zum Anstieg der Menüebenen zu. Auch die Ausreißer zeigen ein ähnliches Bild. Zwar gibt es sowohl für die Sprach- als auch für die Touchsteuerung Ausreißer, jedoch sind die der Sprachsteuerung über alle Menüebenen hinweg deutlich konstanter als die zum Teil weit verstreuten Ausreißer der Touchsteuerung. In Summe zeigt die **Abbildung 23**, dass die Eingabezeiten der Sprachsteuerung unabhängig von der an den Menüebenen gemessenen Komplexität der Touchbefehle sind. Gleichzeitig zeigen die Daten einen erheblichen Zeitvorteil der Sprachsteuerung im Vergleich zu der Touchsteuerung bei der Bedienung einer Produktionsmaschine. [NER23]

5.4.3 Einfluss der Anzahl an Interaktionen auf die zeitliche Effizienz

Mithilfe des natürlichen Logarithmus transformierte abhängige Ergebnisvariable Eingabezeit wurde mittels des ersten LMM vorausgesagt. Das zweite Modell schätzt die Fixed-Effects für die Anzahl an Interaktionen, die HMI-Modalität (Sprache/Touch) und deren Interaktion. Eine ANOVA dieser Modelle zeigt die folgenden statistisch signifikanten Effekte: Anzahl der Interakti-

onen ($F(5,2944) = 738.3$, $p < 0.001$), HMI-Modalität ($F(1,2944) = 1925.1$, $p < 0.001$) und deren Interaktion ($F(5,2944) = 454.5$, $p < 0.001$). Die signifikante Interaktion wurde mit einer Kontrastanalyse analysiert, wobei sich alle Kontraste als hoch signifikant herausstellten ($p < 0.001$). In **Abbildung 24** werden die zurücktransformierten Ergebnisse der Eingabedauer mit den beschrifteten Pfeilen dargestellt. [NER23]

Analog zum Anstieg der Menüebenen führt auch ein Anstieg der Anzahl an Interaktionen zu einer deutlichen Steigerung der Medianwerte der Eingabedauer für die Touchsteuerung von ca. 2,5 s (eine Interaktion) bis hin zu ca. 20 s (zehn Interaktionen). Mit Medianwerten zwischen 4 s (eine Interaktion) und 7 s (zehn Interaktionen) ist die Varianz der Eingabedauer für die Sprachsteuerung erheblich geringer im Vergleich zur Touchsteuerung. Ähnlich wie in **Abbildung 23** für die Menüebenen dargestellt, ergibt sich auch für die Anzahl an Interaktionen bereits sehr früh ein BEP ab dem die Eingabe per Sprachsteuerung schneller als per Touchsteuerung ist. [NER23]

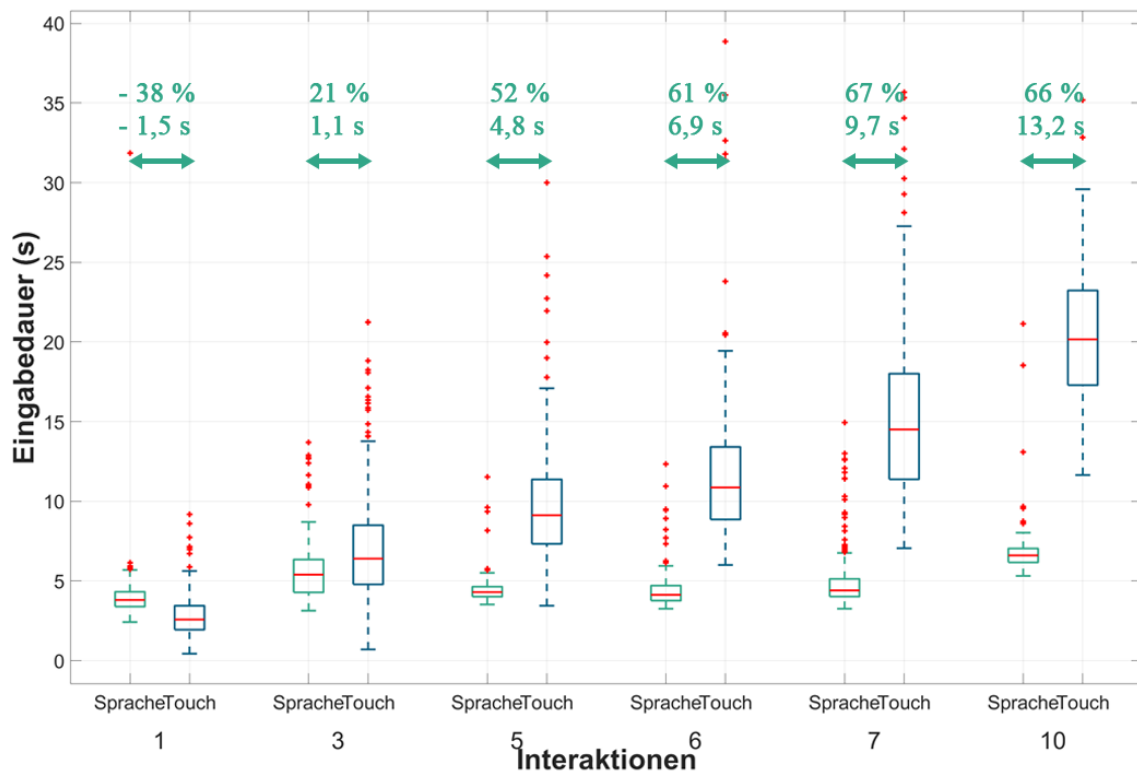


Abbildung 24: Einfluss der Anzahl an Interaktionen auf die Eingabedauer. Jeder Boxplot repräsentiert die Verteilung der Eingabezeit über alle 33 Befehle, 15 Probanden und drei Wiederholungen für eine spezifische Anzahl an Interaktionen (1, 3, 5, 6, 7 oder 10) und Modalität (Sprach-/Touchsteuerung) [NER23]

© 2023 IEEE

Für eine einzelne Interaktion entsprechen die Daten den vorherigen Daten der Menüebene 0, da eine einzelne Interaktion zwangsläufig auf der Root-Ebene ausgeführt wird (vgl. **Kapitel 5.4.2**). Hier waren der Medianwert der Eingabezeiten der Sprachsteuerung genau wie in **Abbildung 23** -1,5 s (- 38 %) geringer als der Medianwert der Eingabezeiten der Touchsteuerung. Bereits ab drei Interaktionen war die Sprachsteuerung jedoch der Touchsteuerung um 1,1 s (21 %) voraus. Dieser Trend setzt sich weiter fort, sodass der Medianwert der Eingabezeit für die Sprachsteuerung für fünf Interaktionen 4,8 s (52 %), für sechs Interaktionen 6,9 s (61 %), für sieben Interaktionen 9,7 s (67 %) und für das Maximum von zehn Interaktionen 13,2 s (66 %) geringer ist als der jeweilige Medianwert der Touchsteuerung. Ein leichter Bruch in diesem Trend ist beim Übergang der Befehle mit sieben hin zu den komplexesten Befehlen mit zehn Interaktionen zu sehen. Hier sinkt der zeitliche Effizienzvorteil um ein Prozent von 67 % auf 66 %. Grund dafür sind die Sprachbefehle 31 und 32 (vgl. **Tabelle 3**), deren Komplexität auf Grund der Länge und der mehrstelligen Kommazahlen im Verhältnis zu den vorherigen Sprachbefehlen überdurchschnittlich hoch ist, sodass sich der Abstand des Medianwerts zur Touchsteuerung nicht in dem bekannten Maße fortsetzt. Zusätzlich sieht man, dass der Abstand der Ausreißer zum Medianwert der Eingabedauer für die Sprachsteuerung hier im ganzen Diagramm am höchsten ist. [NER23]

5.4.4 Lerneffekte

Für jede der beiden HMI-Modalitäten (Sprache/Touch) gaben die Probanden jeden Befehl jeweils drei Mal ein. Dies ermöglicht die Untersuchung möglicher Lerneffekte, wie sie in **Kapitel 5.3** beschrieben sind. [NER23]

Das dritte LLM schätzt die Fixed-Effects für die Anzahl an Wiederholungen, die HMI-Modalität (Sprache/Touch) und deren Interaktion. Eine ANOVA dieser Modelle zeigt die folgenden statistisch signifikanten Effekte: Anzahl der Wiederholungen ($F(2,2950) = 59.4, p < 0.001$), HMI-Modalität ($F(1,2950) = 496.8, p < 0.001$) und deren Interaktion ($F(2,2950) = 31.4, p < 0.001$). Die signifikante Interaktion wurde mit einer Kontrast Analyse analysiert, wobei sich alle Kontraste als hoch signifikant herausstellten ($p < 0.001$). In **Abbildung 25** werden die zurücktransformierten Ergebnisse der Eingabedauer mit den beschrifteten Pfeilen dargestellt. [NER23]

Auffällig ist, dass der Median der Eingabedauer für die Sprachsteuerung über alle drei Wiederholungen, alle Probanden und alle Befehle konstant knapp unter 5 s liegt, während der Median der Eingabedauer für die Touchsteuerung von ca. 11 s in der ersten Wiederholung auf 7,5 s in

der dritten Wiederholung abnimmt. Jeder Medianwert beider HMI-Modalitäten basiert dabei auf jeweils knapp 500 analysierten Befehlen. [NER23]

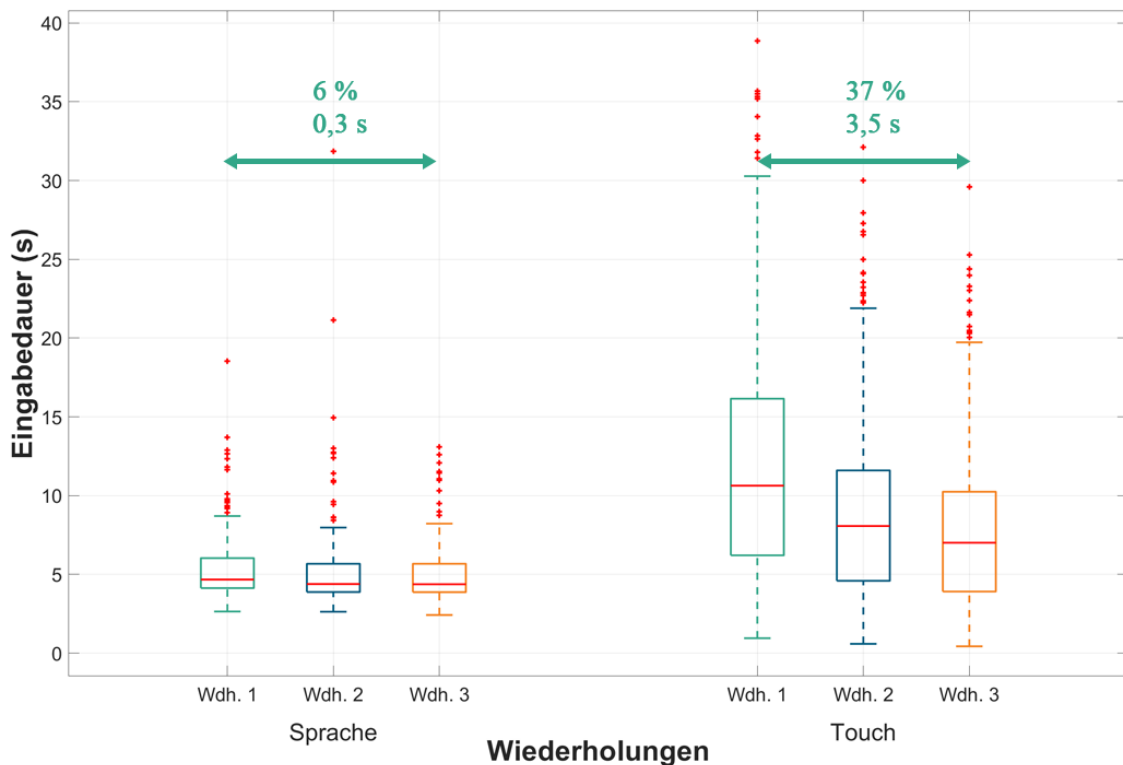


Abbildung 25: Einfluss der Wiederholungen auf die Eingabedauer. Jeder Boxplot repräsentiert die Verteilung der Eingabezeit über alle 33 Befehle, 15 Probanden, 1-5 Menüebenen und 1-10 für eine spezifische Wiederholung (1, 2 oder 3) und Modalität (Sprach-/Touchsteuerung) [NER23] © 2023 IEEE

Über die drei Wiederholungen verringert sich der Median der Eingabedauer nur um 0,3 s (6 %), während sich der Median der Eingabedauer für die Touchsteuerung über die drei Wiederholungen um 3,5 s (37 %) verringert [NER23].

5.4.5 Subjektive Einschätzung der Probanden (Fragebogen)

Neben der quantitativen Messung der Eingabedauer zur Bewertung der Effizienz füllten alle Probanden im Anschluss an ihren Studiendurchlauf einen Fragebogen aus, um ihr subjektives Empfinden in der Interaktion mit den beiden HMI-Modalitäten darzustellen. Die Intuitivität der Sprachsteuerung als HMI-Modalität bewerteten die Probanden mit durchschnittlich 9,2 Punkten auf einer Skala von 1 bis 10. Währenddessen bewerteten sie die Intuitivität der Touchsteuerung um ca. ein Drittel (-36,52 %) geringer mit nur durchschnittlich 6,9 Punkten. Das subjektive Empfinden der Probanden hinsichtlich ihrer kognitiven Belastung während der Teilnahme an der

Probandenstudie bewerteten diese für die Touchsteuerung (8,4/10 Punkte) im Durchschnitt doppelt so hoch wie für die Sprachsteuerung (4,1/10 Punkte), unabhängig mit welcher HMI-Modalität sie starten. Obwohl keiner der Probanden im Vorhinein eine Produktionsmaschine per Sprache bedient hat, bewerteten sie im Durchschnitt das Wohlempfinden in der Interaktion mit dem HMI in Anwesenheit Dritter für die Sprachsteuerung mit 8,8/10 Punkten und für die Touchsteuerung mit 8,1/10 Punkten und damit nahezu gleich. Am Ende des Experiments gaben acht der 15 Probanden an, eine Sprachsteuerung als HMI in einem realen Produktionsszenario zu bevorzugen (53,3 %), während vier Probanden sich eine Sprachsteuerung als Ergänzung zu einem konventionellen HMI gut vorstellen können (26,7 %). Lediglich drei Teilnehmer würden lieber bei einem konventionellen HMI bleiben (20,0 %). [NER23]

5.4.6 Zeitäquivalente als quantifizierte Effizienzfaktoren

Zur Entwicklung ästhetischer, konsistenter und intuitiver (und damit effizienter) HMI sind Gestaltungsregeln notwendig. Sowohl in Normen als auch in Maschinenrichtlinien und HMI-Style-Guides sind diese zu finden. [Fri20] Zur praktischen Umsetzung von Gestaltungsregeln mit dem Ziel einer effizienten Mensch-Technik-Interaktion (MTI) bedarf es allgemeingültiger und quantifizierbarer Einflussfaktoren. Für die Abschätzung des Effizienzpotenzials einer Sprachsteuerung als Ergänzung Austausch zu einem konventionellen, haptisch-visuellen HMI werden aus den gemessenen Eingabedauern aus dieser Probandenstudie maschinen- und prozessunabhängige relative Zeitäquivalente abgeleitet. Diese beschreiben das Verhältnis zwischen der Eingabedauer einer Sprachsteuerung im Vergleich zu einer Touchsteuerung, in Abhängigkeit der Parameter **Anzahl an Menüebenen** und **Anzahl an Interaktionen**. Die Zeitäquivalente werden in **Schritt #5** der entwickelten Methode eingesetzt, um auf Basis der gemessenen Eingabedauer des konventionellen haptisch-visuellen HMIs die Eingabedauer einer Sprachsteuerung abzuschätzen (vgl. **Kapitel 3.2.5**). Erst durch diese empirisch in der Probandenstudie in **Kapitel 5** ermittelten Zeitäquivalente ist eine Prognose der potenziell zu reduzierenden Nebenzeiten auf Basis der Evaluation des Effizienzpotenzials einer Sprachsteuerung möglich. Dadurch werden die **Anforderungen 1 & 2** des Lastenhefts erfüllt (vgl. **Kapitel 1.4**).

In **Tabelle 4** sind die errechneten Zeitäquivalente für bis zu fünf Menüebenen und zehn Interaktionen zu sehen. Die zu Grunde liegende Annahme ist, dass das Verhältnis zwischen der Eingabedauer einer Touchsteuerung im Verhältnis zu einer Sprachsteuerung als konstant angenommen werden kann, während die absolute Eingabedauer von Maschinenbefehlen abhängig vom jeweiligen Produktionsablauf und dem zugehörigen Befehlsdesign ist. Auf Basis dieser Annahme

kann mithilfe der Zeitäquivalente die Eingabedauer eines entsprechenden Sprachbefehls abgeschätzt werden. [NER23]

Tabelle 4: Zeitäquivalente [NER23] © 2023 IEEE

Menüebenen	Zeitäquivalent	Interaktionen	Zeitäquivalent
0	1,38	1	1,38
1	0,82	3	0,79
3	0,40	5	0,48
5	0,35	6	0,39
		7	0,33
		10	0,34

Mithilfe der errechneten Eingabedauer kann dann ein zeitlicher Effizienzvorteil oder -nachteil bewertet werden. Die Zeitäquivalente basieren dabei auf den Daten für die beiden in dieser Probandenstudie genutzten HMI-Modalitäten Sprach- und Touchsteuerung, die in der **Abbildung 23** und **Abbildung 24** dargestellt werden. Zur Berechnung der Zeitäquivalente wird zunächst die durchschnittliche Eingabedauer aller Befehle für jede untersuchte Menüebene und Anzahl an Interaktionen für beide HMI-Modalitäten berechnet. Anschließend werden die berechneten Eingabedauern der Touchsteuerung durch die entsprechenden Eingabedauern der Sprachsteuerung dividiert, um die Zeitäquivalente für die jeweiligen Menüebenen und Anzahl an Interaktionen zu erhalten. Multipliziert man in der Anwendung Zeitäquivalente über eins mit der gemessenen Eingabedauer einer konventionellen Touchsteuerung, so ergibt sich eine längere Eingabedauer und damit eine schlechtere zeitliche Effizienz für eine potenzielle Sprachsteuerung. Für Zeitäquivalente, die kleiner als eins sind, ergibt sich aus der Multiplikation mit einer gemessenen Eingabedauer einer konventionellen Touchsteuerung eine kürzere Eingabedauer und damit eine verbesserte zeitliche Effizienz einer potenziellen Sprachsteuerung (vgl. **Kapitel 3.2.5**). [NER23]

Sowohl Hersteller als auch Nutzer von Maschinen können mithilfe dieser Zeitäquivalente das Effizienzpotenzial einer Sprachsteuerung als Ergänzung oder Austausch zu einem vorhandenen konventionellen HMI abschätzen. Dazu werden die gemessenen Eingabedauern des konventionellen HMIs für alle Befehle, die an einem repräsentativen Arbeitstag eingegeben werden, mit den Zeitäquivalenten der jeweiligen Menüebenen oder Anzahl an Interaktionen multipliziert. Am Ende ergibt ein Aufsummieren der Eingabedauern, die mit den Zeitäquivalenten multipliziert werden, die Zeit, die im Vergleich zu dem konventionellen HMI mit einer Sprachsteuerung be-

nötigt werden würde (vgl. **Kapitel 3.2.5**). Dieser Vergleich ermöglicht eine quantitative Bewertung beider HMI-Modalitäten im Kontext des betrachteten Produktionsablaufs. **Die Zeitäquivalente zeigen den Break-Even-Point für Menüebenen über Null und mehr als eine Interaktion, ab dem eine Sprachsteuerung hinsichtlich der zeitlichen Effizienz einen Vorteil bietet** (vgl. **Tabelle 4**). Die genaue Berechnung mithilfe der Zeitäquivalente wird in **Kapitel 3.2.8** beschrieben. [NER23]

Die Quantifizierung des Einflussfaktors **Komplexität von Menüstrukturen** über die beiden Parameter **Anzahl an Menüebenen** und **Anzahl an Interaktionen** eröffnet die Möglichkeit die qualitativ geprägte Diskussion vorheriger Arbeiten (vgl. **Kapitel 2.5 & 2.6**) um eine quantitative Dimension zu erweitern. Dadurch kann neben subjektiven Eindrücken von Maschinenbedienern auch ein datengetriebener Vergleich der HMI-Modalitäten erfolgen. [NER23]

6 Quantifizierung der Einflussfaktoren »Laufwege« und »Mehrmaschinenbedienung«

Neben der **Komplexität von Menüstrukturen** auf GUIs haben die Laufwege, die Mitarbeitende zwischen Arbeitsplätzen und HMI oder unterschiedlichen HMIs einen wesentlichen Einfluss auf die Effizienz der Interaktion und in kritischen Pfaden damit auch auf die Effizienz/Produktivität des gesamten Produktionsablaufs. In den folgenden Kapiteln werden die Eigenschaften von Laufwegen und das Effizienzpotenzial einer Sprachsteuerung betrachtet, durch die gewisse Laufwege vermieden werden können. Im ersten Schritt werden dazu die unterschiedlichen Eigenschaften von Laufwegen in der Produktion beschrieben.

Lean Production beschreibt sieben Arten der Verschwendung, also nichtwertschöpfende Tätigkeiten, wovon eine Bewegung, zu der auch Laufwege zählen, darstellt (vgl. **Kapitel 2.7.1**) [Bal15, Vee23]. Ein nutzerzentriertes HMI reduziert diese, um so die Effizienz von Produktionsabläufen zu steigern [PeHi13a, PeHi13b, PKO21, Vee23]. Ein Ansatz zur nutzerzentrierten Optimierung von HMIs bietet die Bedienung via Sprache, als natürlichste Form der Kommunikation [RBR20, Vee23]. Neben ihrer Intuitivität und der Möglichkeit über ein Headset oder Mikrofon Arrays ortsunabhängig mit Maschinen und Systemen zu interagieren, bietet sie das Potenzial gewisse Laufwege in Produktionsabläufen zu vermeiden.

6.1 Eigenschaften von Laufwegen

Laufwege in der Produktion werden primär über die beiden Parameter **Länge** und **Häufigkeit** quantifiziert. Zur Bewertung einer möglichen Vermeidung von Laufwegen durch den Einsatz einer Sprachsteuerung als HMI ist es jedoch zusätzlich notwendig, dass die **Arten von Laufwegen** als Einflussfaktor kategorisiert wird (vgl. **Kapitel 4.2**). In **Abbildung 26** wird eine Methode zur quantifizierten Abschätzung des Effizienzpotenzials einer Sprachsteuerung im Vergleich zu einem konventionellen HMI, hinsichtlich der Vermeidung von Laufwegen dargestellt. Dazu werden die Parameter Länge, Häufigkeit und Art von Laufwegen genutzt. Diese fließt in **Kapitel 3.2** in die übergeordnete »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« ein (vgl. **Kapitel 3.2.4 & 3.2.6**).

Das Effizienzpotenzial einer Sprachsteuerung zur Vermeidung von Laufwegen in der Produktion steigt proportional zur absoluten **Anzahl an Laufwegen** als Einflussfaktor (vgl. **Kapitel 4.2**). Zur in **Abbildung 26** dargestellten quantifizierten Abschätzung des realen Effizienzpotenzials ist es jedoch notwendig, dass die im ersten Schritt in ihrer Länge und Häufigkeit erfassten Laufwege

qualitativ entlang der Kategorisierung bewertet werden. Produktionsabläufe, in denen jedoch sowieso nur wenig Laufwege anfallen, bieten im Umkehrschluss nur ein geringes absolutes Effizienzpotenzial auf Basis der Vermeidung von Laufwegen durch Sprachsteuerung. Selbst bei prozentual hoher Zeiteinsparung durch die Vermeidung einzelner Laufwege ergeben sich für den gesamten Produktionsablauf nur geringe Effizienzsteigerungen. Für die Betrachtung im Kontext der Effizienzsteigerung von Produktionsabläufen durch den Einsatz einer Sprachsteuerung werden die in der Produktion auftretenden Laufwege in vier **Arten von Laufwegen** kategorisiert (vgl. **Abbildung 26**). [NVE23, Vee23]

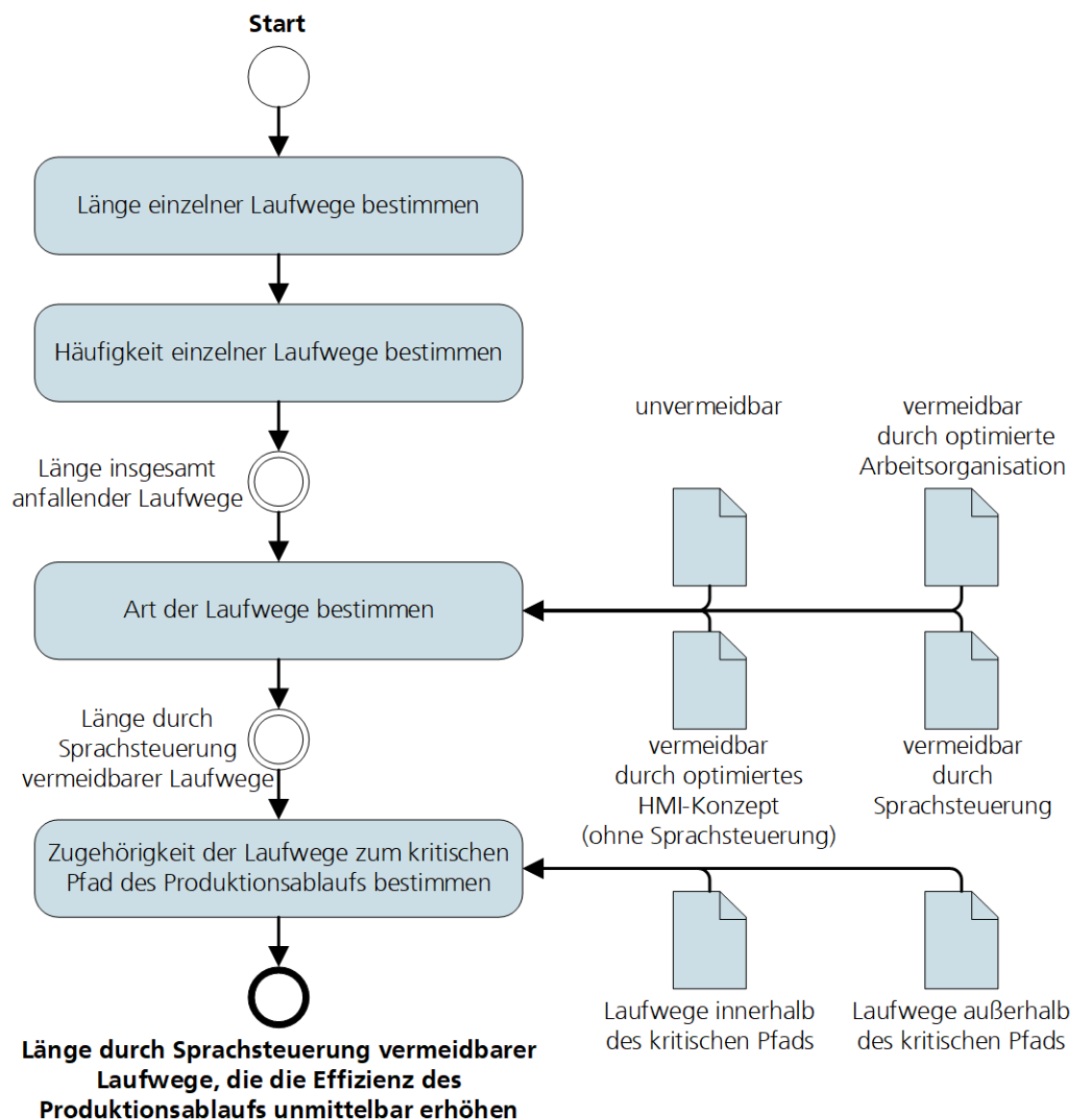


Abbildung 26: Methode zur quantifizierten Abschätzung des Effizienzpotenzials von Sprachsteuerung zur Reduzierung von Laufwegen [NVE23, Vee23]

- **Unvermeidbare Laufwege:** Laufwege, die trotz Optimierung der Arbeitsorganisation, des HMI-Konzepts oder der Einführung einer Sprachsteuerung anfallen. Dazu zählen Laufwege zu Wartungs- oder Handlungstätigkeiten, wie der Reinigung von Maschinenkomponenten oder das Rüsten einer Maschine [NVE23, Vee23]
- **Vermeidbar durch optimierter Arbeitsorganisation:** Laufwege, die durch eine optimierte Arbeitsorganisation, wie die Aufbewahrung von Werkzeug und Materialien zur Störungsbeseitigung an einem definierten Ort, vermeidbar sind [NVE23, Vee23]
- **Vermeidbar durch optimiertes HMI-Konzept:** Laufwege, die durch die Optimierung des aktuellen HMI-Konzepts, wie die Einführung eines zentralen Leitstands anstatt einzelner verteilter HMIs, vermeidbar sind [NVE23, Vee23]
- **Vermeidbar durch Sprachsteuerung:** Laufwege, die nur durch die Integration einer Sprachsteuerung zur Durchführung von Steuerungs-, Kontroll- oder Dokumentationstätigkeiten, vermeidbar sind [NVE23, Vee23]

Eine Sprachsteuerung via Headset bietet Mitarbeitenden die Möglichkeit, mit freien Händen Maschinen zu steuern, ortsunabhängig Statusmeldungen zu kontrollieren oder parallel zu wertschöpfenden Tätigkeiten Informationen in einem ERP-System zu dokumentieren. Bei der Kategorisierung der Laufwege ist die Einschätzung erfahrener Maschinenbediener ein wichtiges Element. Besonders beim Aktivieren von Maschinenfunktionen ohne direkten Sichtkontakt muss gewährleistet sein, dass eine sichere Bedienung möglich und sinnvoll ist. Dies führt dazu, dass Laufwege nur im Kontext des gesamten Produktionsablaufs kategorisiert werden können. Während in einem Produktionsablauf zum Beispiel die Vermeidung von Laufwegen zur Eingabe von Prozessparametern möglich ist, da die Bedienung aus der Ferne gefahrlos möglich ist, kann ein anderer Produktionsablauf die Anwesenheit des Bediener an der Maschine erfordern, um den aus den neu eingegebenen Parametern resultierenden Prozess zu bewerten. Des Weiteren können Laufwege, die zur Überwachung des Prozesses notwendig sind, nur durch den Einsatz von zusätzlichen Sensoren, die Prozesszustände erkennen und diese über die Speicherprogrammierbare Steuerung (SPS) der Maschine an ein HMI weitergeben, vermieden werden.

6.2 Eigenschaften des Produktionsablaufs

Nach REFA werden die Sollzeiten für Produktionsabläufe in Hauptzeiten (unmittelbarer Fortschritt durch den laufenden Produktionsprozess), Nebenzeiten (Werkzeuge/Werkstücke einspannen/spannen, Laufwege zu Bedienterminals bestreiten) und Brachzeiten (Maschine verfü-

bar, aber nicht genutzt) unterschieden, wobei Nebenzeiten nur indirekt zum Fortschritt des Produktionsprozesses beitragen (vgl. **Kapitel 2.1 & 2.7.2**). Die Art des HMIs hat dabei nur einen Einfluss auf die Effizienz der Kollaboration zwischen Mensch und Maschine mit dem Ziel, die Nebenzeiten zu reduzieren. Der eigentliche Produktionsprozess (Hauptzeiten) wird durch die Effizienz des HMI nicht beeinflusst. [BuSc21, NVE23, Tsc83, Vee23] Wie in **Kapitel 2.4.1** beschrieben wird in dieser Arbeit die benötigte Interaktionszeit, bestehend aus Eingabe- und Laufwegzeiten, als Ressource ins Verhältnis zum erreichten Ziel gesetzt, um die Effizienz zu untersuchen. Die Reduzierung der Laufwegzeiten führt, bei gleichbleibenden anderen Parametern, zu einer Reduzierung der Nebenzeiten und damit zu einer Steigerung der Effizienz des Produktionsablaufs. [Bec05, DIN18]

Nachdem alle Laufwege identifiziert wurden, deren einzelne **Längen** gemessen und deren **Art** hinsichtlich ihrer Eigenschaften kategorisiert wurden, erfolgt im letzten Schritt die Analyse, ob diese im kritischen Pfad des Produktionsablaufs liegen (vgl. **Abbildung 26 – »Zugehörigkeit der Laufwege zum kritischen Pfad des Produktionsablaufs bestimmen«**) [NVE23, Vee23]. Ein Laufweg, unabhängig davon, mit welcher der in **Kapitel 6.1** vorgestellten Maßnahmen er vermieden wurde, senkt nur dann die Nebenzeiten des Produktionsablaufs und steigert damit die Effizienz, wenn er sich im sogenannten *»kritischen Pfad«* befindet. Dieser beschreibt alle Handlungen und deren Abfolge, die die Dauer des Produktionsablaufs im Kontext der Herstellung oder Verarbeitung eines Produkts beschreiben (vgl. **Kapitel 2.1**). Die Vermeidung oder Reduzierung von Tätigkeiten im kritischen Pfad, führt zu einer Verkürzung der Nebenzeiten. [AiHe22, Bec05] Ein Beispiel dafür ist das Starten eines Prozesses einer Maschine per Sprachsteuerung aus der Ferne, ohne dass der Bediener den Laufweg zum konventionellen HMI bestreiten muss. Wenn Mitarbeitende die eingesparte Zeit anderweitig wertschöpfend nutzen können, kann es dennoch sinnvoll sein, Laufwege zu vermeiden, die außerhalb des kritischen Pfads liegen.

6.3 Teilautomatisierte Produktionsabläufe mit Laufwegen in der Produktion

Im Rahmen einer Industriestudie wurden neun Unternehmen aus unterschiedlichen Produktionsbranchen (Papierherstellung, Verpackungsindustrie, Automobilindustrie & Kunststoffrecycling) im Kontext des Potenzials einer Sprachsteuerung zur Vermeidung von Laufwegen in ihren Produktionsabläufen analysiert. Die in **Kapitel 6.1** beschriebene Vorgehensweise wird im Folgenden beispielhaft anhand von vier realen Produktionsabläufen mit insgesamt 18 resultierenden Bediensituationen demonstriert. Dabei beschreibt der Produktionsablauf die Gesamtheit

der Interaktionen, die als Bediensituationen definiert werden, zwischen Mitarbeitendem und Maschine oder Maschinenkomponente in einem Anwendungsszenario. Mittels des sogenannten »Shadowings« (vgl. **Kapitel 3.2.1 & 3.2.6**) werden die Mitarbeitenden im alltäglichen Produktionsablauf begleitet und die anfallenden Laufwege dokumentiert. Anschließend werden die Mitarbeitenden interviewt, um die Gründe für jeden einzelnen Laufweg zu erfahren, um diese kategorisieren zu können. Die gewonnenen Informationen über die Produktionsabläufe werden in Produktionsablaufdiagrammen dokumentiert. Im nächsten Schritt wird das Layout der Produktionshalle mit einem Laserentfernungsmessgerät erfasst und digitalisiert. Speziell die Laufwege zwischen Arbeitsplätzen und HMIs, die mittels Shadowing erfasst wurden, sind hier entscheidend. Diese werden auf Basis des digital erfassten Produktionshallenlayouts berechnet. Weiterführend werden dann auf Basis der Länge der einzelnen Laufwege die benötigten Laufwegzeiten für jeden Laufweg mithilfe der MTM bestimmt (vgl. **Kapitel 2.7.4**). [NVE23, Vee23] Laut MTM wird die Bewegung von Mitarbeitenden im Produktionskontext in einzelne Basisbewegungen, wie z.B. »Hilfsmittel handhaben«, »etwas Betätigen« oder »Gehen« zerlegt. Die Dauer dieser Grundbewegungen wird dabei in sogenannten »Time Measurement Units« (TMU), die dem Hunderttausendstel einer Stunde (0,036 s) entsprechen, gemessen. Die »Methods-Time Measurement Universal Analysis System« (MTM-UAS)-Datenkarte der MTM Association e. V. definiert die Laufgeschwindigkeit eines Mitarbeitenden in der Produktion mit 25 TMU, was 0,9 s oder einer Laufgeschwindigkeit von 1,11 m/s entspricht. [MTM19, Sys06] Die Nutzung der MTM mit seinen TMU stellt sicher, dass Ergebnisse unterschiedlicher Unternehmen, mit unterschiedlichen Mitarbeitern und Laufgeschwindigkeiten, über eine konstante fest definierte Laufgeschwindigkeit vergleichbar sind. Die Start- und Endpunkte der Laufwege sowie die zur Kategorisierung notwendigen Kontextinformationen zu den Laufwegen innerhalb des Produktionsablaufs werden dabei mithilfe des zuvor beschriebenen »Shadowings« erfasst. [NVE23, Vee23]

Im Folgenden werden zwei der insgesamt neun für die Entwicklung der Methode untersuchten Unternehmen mit ihren unterschiedlichen Produktionsabläufen und Bediensituationen dargestellt. Das erste Unternehmen ist tätig im Kunststoffrecycling und produziert sogenannte Regranulate (recycelte (Kunststoff-)Granulate) aus Industrierückständen in Großserien-Fließfertigung. Die anfallenden Laufwege sind hier besonders lang, treten dafür aber im Verhältnis seltener auf. Im Kontrast dazu ist das zweite betrachtete Unternehmen ein Automobilzulieferer, der Exterieur-Kunststoffverkleidungen in Großserien-Reihenfertigung produziert. Die anfallenden Laufwege sind hier eher kurz, treten dafür aber im Verhältnis sehr häufig auf. [TAG17, Vee23] Die beiden ausgewählten Unternehmen zeigen dabei die beiden in der Produktion auftretenden Extreme von wenigen langen und vielen kurzen Laufwegen. Die Industriestudie zeigt, dass die

Mehrheit der Laufwege in den untersuchten deutschen Produktionsunternehmen entweder unvermeidbar oder nur durch eine Sprachsteuerung vermeidbar sind. In **Abbildung 27** wird das Effizienzpotenzial von Sprachsteuerung hinsichtlich der Vermeidung von Laufwegen pro analysiertem Anwendungsszenario für einen Mitarbeitenden in einer achtstündigen Schicht in Stunden pro Jahr (h/a) dargestellt.

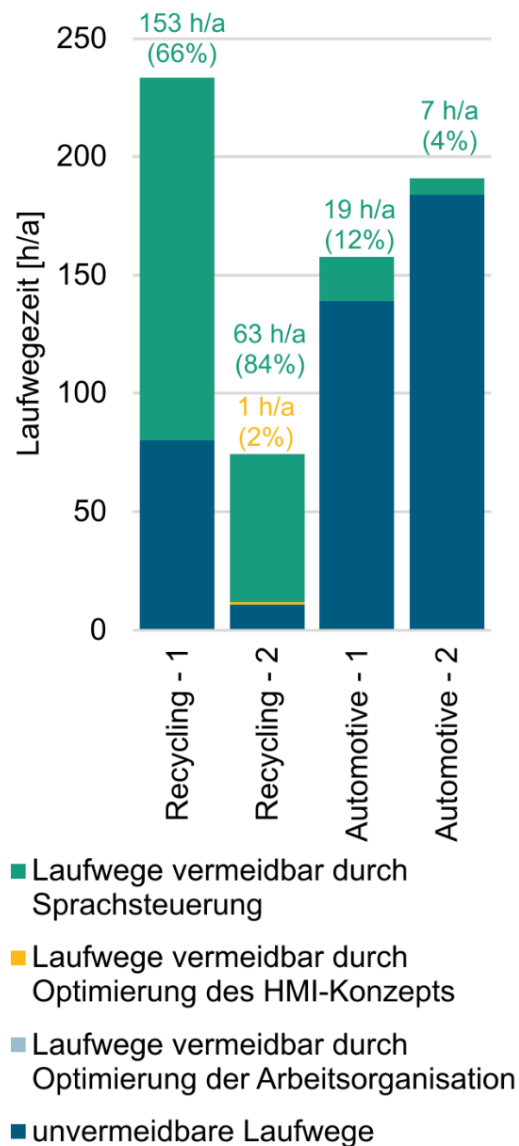


Abbildung 27: Einsparpotenzial durch Sprachsteuerung in unterschiedlichen Bediensituationen [NVE23]

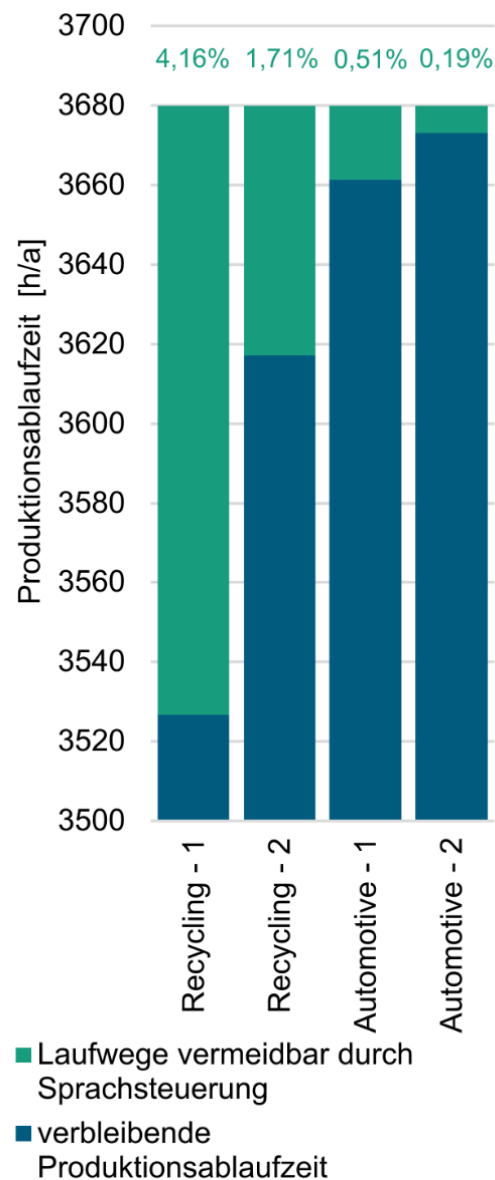


Abbildung 28: Einsparpotenzial durch Sprachsteuerung in Bezug auf die verbleibende Produktionsablaufzeit [NVE23]

Dabei wird links mit dem in grün dargestellten höchsten absoluten Effizienzpotenzial begonnen. Unvermeidbare Laufwege werden in blau, während Laufwege, die durch Optimierung der Arbeitsorganisation oder des HMI-Konzepts vermieden werden können in hellblau bzw. gelb dargestellt werden. Die beiden letztgenannten Arten traten in dieser Industriestudie nur zu einem sehr geringen Teil auf, was laut Aussage der Unternehmen daran liegt, dass der hohe Kostendruck der Produktion in Deutschland schon mehrere Optimierungswellen ausgelöst hat und Produktionsabläufe bereits iterativ optimiert wurden. [NVE23, Vee23]

In dieser Industriestudie werden die Ergebnisse auf zehn Schichten/Woche und 46 Arbeitswochen/Jahr normiert, um eine Vergleichbarkeit zwischen allen betrachteten Unternehmen zu schaffen. Das hinsichtlich der Laufwege identifizierte Einsparpotenzial wird anschließend auf ein Jahr hochgerechnet. In **Abbildung 28** wird das absolute Einsparpotenzial durch den Einsatz einer Sprachsteuerung als HMI für alle vier Bediensituationen in Bezug auf den gesamten Produktionsablauf über alle Arbeitstage und Schichten im Jahr dargestellt. Dabei können in den vier exemplarischen Anwendungsszenarien zwischen ca. 0,19 % und ca. 4,16 % pro Mitarbeitenden in der jeweiligen Bediensituation eingespart werden. Daraus resultiert ein absolutes Einsparpotenzial pro Mitarbeitenden im Anwendungsszenario von ca. 7 h/a bis ca. 153 h/a. Das zeitliche Einsparpotenzial durch die Vermeidung von Laufwegen durch Sprachsteuerung skaliert dabei proportional zu der Anzahl an gleichen Anwendungsszenario im Unternehmen. [NVE23, Vee23]

Anwendungsszenarien Kunststoffrecycling I & II

In dem untersuchten Kunststoffrecyclingunternehmen werden in zwei voneinander unabhängigen Hallen mit unterschiedlichen Produktionshallenlayouts und Produktionsabläufen in Regranulierungsprozessen verkettete und automatisierte Anlagen von Mitarbeitenden bedient (vgl. **Abbildung 29 & Abbildung 30**). Im Zentrum der verketteten Produktionsanlage steht jeweils eine große Extruderanlage, die zusammen mit ihrer Peripherie von Mitarbeitenden in ihrer Rolle als Supervisor während des Automatikbetriebs mit Material befüllt und überwacht wird. Gleichzeitig sind die Mitarbeitenden für die Störungsbeseitigung, die Nachregelung des Produktionsprozesses und den Transport von mit Kunststoffgranulat gefüllten Big Bags (flexibler Schüttgutbehälter mit Schlaufen) mit Flurförderfahrzeugen (Gabelstaplern) in das Warenlager zuständig. Die Laufwege in diesem Anwendungsszenario fallen zwischen einzelnen Komponenten der verketteten Produktionsanlage und verschiedenen HMIs an, um Prozessparameter, wie z.B. den Extruder-Durchsatz, Rotorumdrehungen eines vorgelagerten Schredders zur Zerkleinerung des Ausgangsmaterials sowie die Drehgeschwindigkeit des nachgelagerten Messerkopfs zum Zerteilen des Extrusionsstrangs in Granulatkörner anzupassen. [NVE23, Vee23]

Mithilfe der zuvor beschriebenen Vorgehensweise werden die Laufwege in den vier Anwendungsszenarien erfasst und kategorisiert. Für das Anwendungsszenario »Recycling I« ergibt sich durch den Einsatz einer Sprachsteuerung als Substitution zum konventionellen HMI ein Einsparpotenzial von ca. 612.775 m/a, 153 h/a (15,3 Mio. TMU), 66 % (vgl. **Abbildung 27**). Auf Grund von hohen Aufwärm- und Abkühlzeiten von Extrudern im Kunststoffrecycling laufen diese meist in einem dauerhaften Automatikbetrieb, während nur Prozessparameter nachgeregelt oder Wartungstätigkeiten und Störungsbeseitigungen sowie Befüll- und Einlagerungstätigkeiten verrichtet werden. Auf Grund dessen liegen alle Laufwege in diesem Anwendungsszenario »Recycling I« außerhalb des kritischen Pfads des Produktionsablaufs. [NVE23, Vee23]

In **Abbildung 29** werden alle im Anwendungsszenario »Recycling I« vorhandenen Laufwege der unterschiedlichen Bediensituationen in Form eines sogenannten Spaghetti-Diagramms dargestellt (vgl. **Kapitel 2.7.1**). Die vier Kategorien an Laufwegen in der Produktion im Kontext des effizienten Einsatzes von Sprachsteuerung zur Vermeidung von Laufwegen werden in Form von unterschiedlichen Linientypen dargestellt (vgl. **Abbildung 26/Abbildung 29**). Die Häufigkeit der Laufwege im Betrachtungszeitraum wird ergänzend über farbige (hellblau, grün, dunkelblau und orange) Pfeile dargestellt. In dem Spaghetti-Diagramm werden sowohl Laufwege zwischen unterschiedlichen HMIs als auch Laufwege zwischen HMIs und Arbeitsplätzen abgebildet. In diesem Anwendungsszenario fallen alle Laufwege aller Bediensituationen bis auf einen jeweils nur einmal im Betrachtungszeitraum an. Gleichzeitig ist die Anzahl der unterschiedlichen Laufwege aber hoch und die Länge der einzelnen Laufwege groß. Sowohl in **Abbildung 27** als auch in **Abbildung 29** ist erkennbar, dass im Anwendungsszenario »Recycling I« ausschließlich Laufwege der Kategorien »Unvermeidbar« und »Vermeidbar durch Sprachsteuerung« vorkommen.

Das Effizienzpotenzial durch die Vermeidung von Laufwegen liegt im Anwendungsszenario »Recycling II« bei 251.146 m/a, 63 h/a (6,3 Mio. TMU), 84 % (vgl. **Abbildung 27**). Im Kontrast zum Anwendungsszenario »Recycling I« liegen hier jedoch insgesamt vier von sieben Bediensituationen mit ca. 31.384 m/a, 8 h/a, 11 % im kritischen Pfad des Produktionsablaufs. Diese vier Bediensituationen finden vor dem Anlaufen des Produktionsprozesses im Rahmen von vorbereitenden Tätigkeiten, wie unter anderem einer dokumentationspflichtigen Filterprüfung, statt. [NVE23, Vee23]

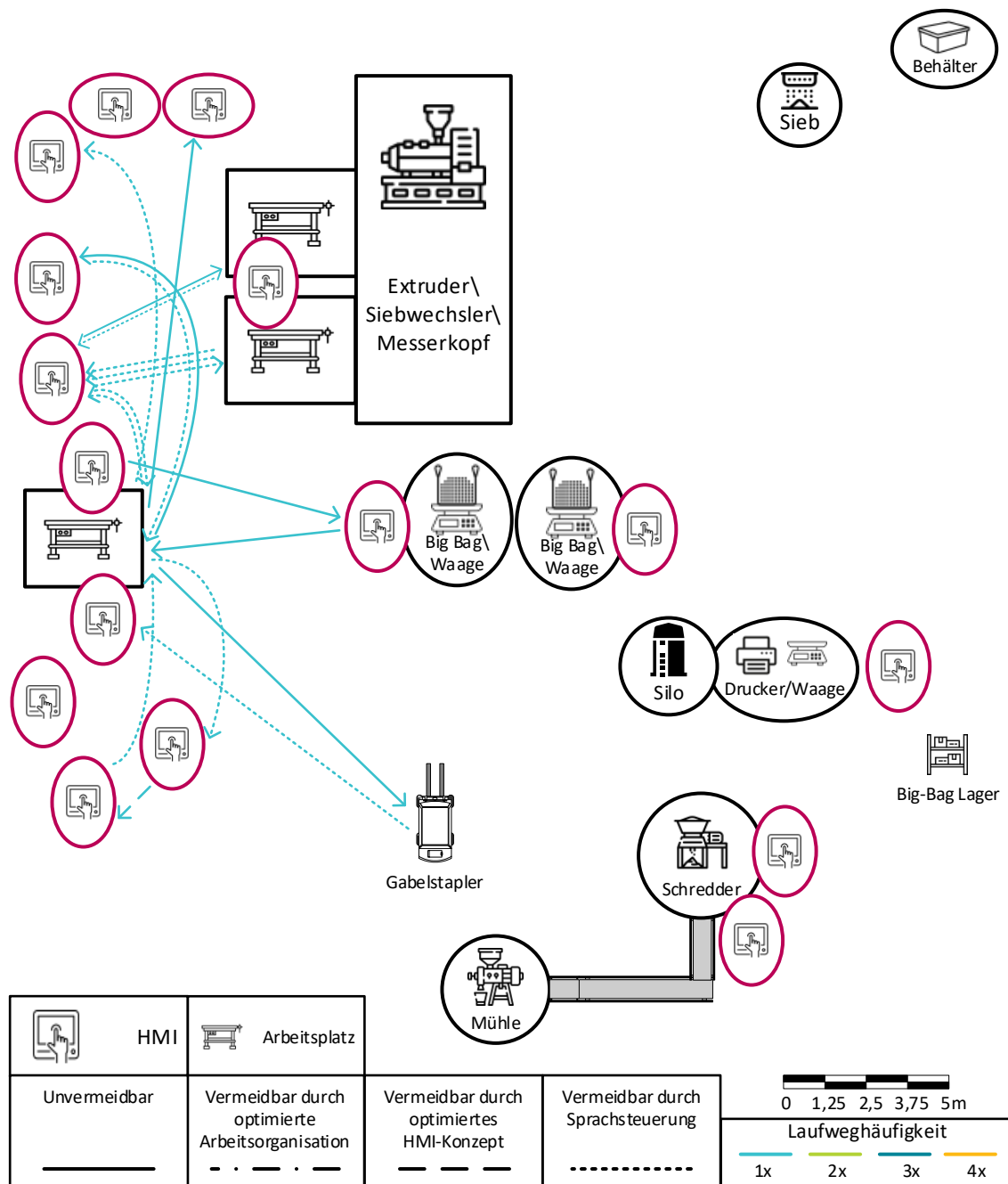


Abbildung 30: Spaghetti-Diagramm des Anwendungsszenarios »Recycling II« (Hersteller für Kunststoff-Regranulate aus Industrierückständen) nach [Vee23]

In diesem Anwendungsszenario fallen alle Laufwege aller Bediensituationen nur einmal im Betrachtungszeitraum an. Gleichzeitig ist die Anzahl der unterschiedlichen Laufwege aber hoch und die Länge der einzelnen Laufwege groß. Sowohl in **Abbildung 27** als auch in **Abbildung 29** ist erkennbar, dass im Anwendungsszenario »Recycling II«, wie im Anwendungsszenario »Recycling I« fast ausschließlich Laufwege der Kategorien »Unvermeidbar« und »Vermeidbar durch

Sprachsteuerung« vorkommen. Lediglich ein Laufweg entspricht der Kategorie »*Vermeidbar durch optimiertes HMI-Konzept*« (vgl. **Abbildung 30**).

Anwendungsszenarien Automotive I & II

Das Anwendungsszenario »*Automotive I*« beschreibt einen automatisierten kombinierten Bearbeitungs- und Montageprozess zum Stanzen von Löchern zur späteren Montage von Sensoren und dem zeitgleichen Einclippen von Zierelementen in die Front-Stoßstange eines Autos. Dabei transportiert ein Mitarbeitender das unbearbeitete Bauteil von einem Transportwagen in die Prozesskammer der Maschine, legt den Laufweg zum HMI zurück, startet den Bearbeitungsprozess mittels Knopfdrucks, legt den Laufweg zur Prozesskammer zurück und entnimmt das fertig bearbeitete Bauteil, bevor er es auf einem zweiten Transportwagen ablegt. Durch den Einsatz einer Sprachsteuerung anstatt des konventionellen, haptischen-visuellen HMI können in diesem Anwendungsszenario ca. 74.483 m/a, 19 h/a (1,9 Mio. TMU), 12 % (vgl. **Abbildung 27**) eingespart werden. Dabei liegen ca. 63.885 m/a, ca. 16 h/a, 10 % im kritischen Pfad des Produktionsablaufs (vgl. **Abbildung 31**). [NVE23, Vee23]

In **Abbildung 31** werden alle im Anwendungsszenario »*Automotive I*« vorhandenen Laufwege der unterschiedlichen Bediensituationen in Form eines Spaghetti-Diagramms dargestellt (vgl. **Kapitel 2.7.1**). Im Gegenteil zu den Spaghetti-Diagrammen der Anwendungsszenarien »*Recycling I*« und »*Recycling II*« in **Abbildung 29** und **Abbildung 30** fallen in diesem Anwendungsszenario alle Laufwege bis auf einen mehrfach an. Zehn Laufwege fallen viermal an, während drei Laufwege dreimal und nur ein Laufweg einmal im Betrachtungszeitraum anfallen. Gleichzeitig sind die Laufwege in diesem Anwendungsszenario jedoch deutlich kürzer und auch in ihrer aufsummierten Gesamtlänge etwas kürzer als in den Anwendungsszenarien »*Recycling I*« und »*Recycling II*«. Sowohl in **Abbildung 27** als auch in **Abbildung 31** ist erkennbar, dass im Anwendungsszenario »*Automotive I*« wie im Anwendungsszenario »*Recycling I*« ausschließlich Laufwege der Kategorien »*Unvermeidbar*« und »*Vermeidbar durch Sprachsteuerung*« vorkommen.

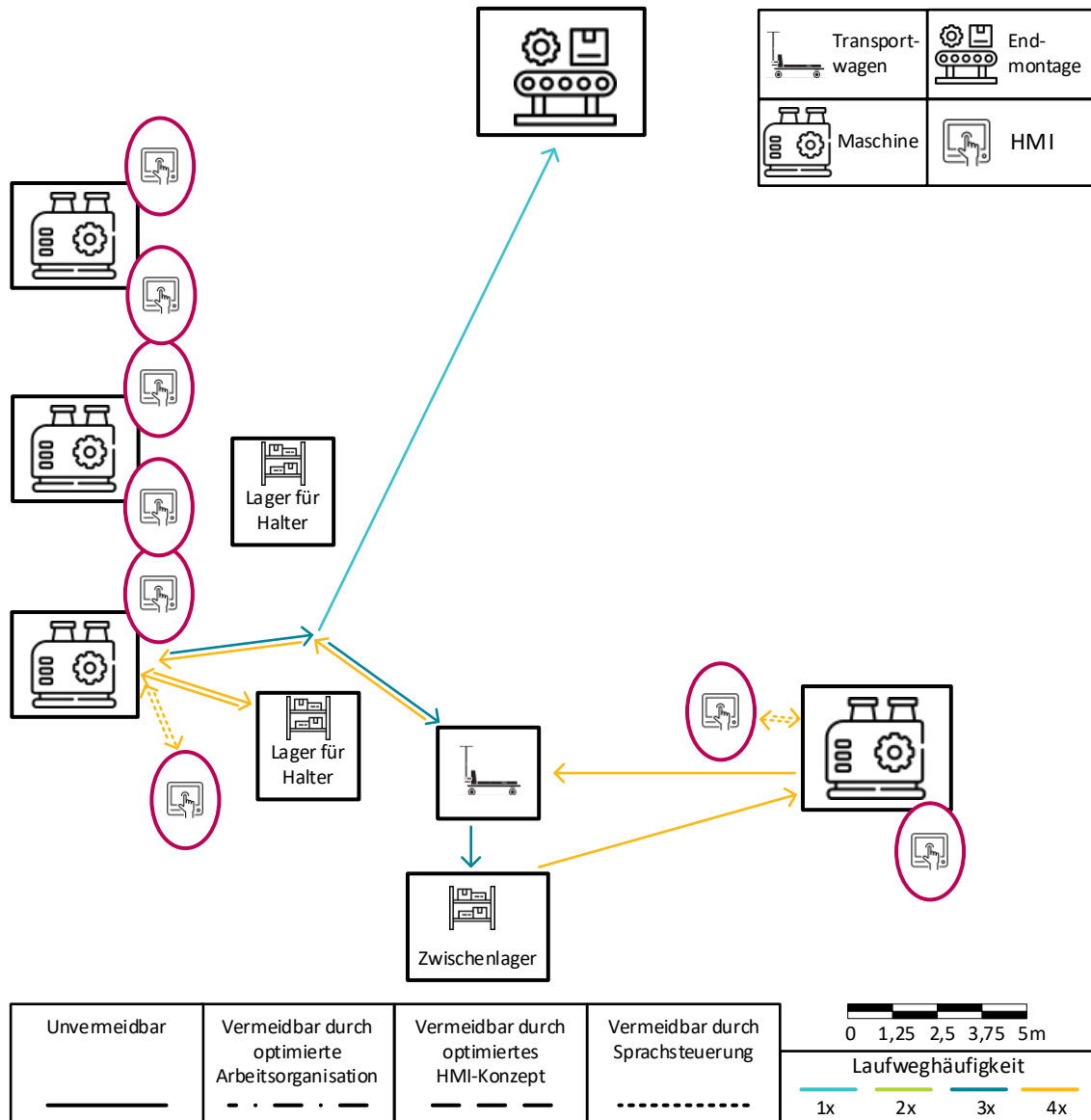


Abbildung 31: Spaghetti-Diagramm des Anwendungsszenarios »Automotive I« (Hersteller für Automobil-Kunststoffinterieurteile) nach [Vee23]

Das Anwendungsszenario »Automotive II« beschreibt einen teilautomatisierten Prozess, in dem zunächst Zierteile an der Fronstoßstange eines Autos vorausgerichtet werden, bevor diese vom Maschinenbediener in eine Maschine zur Montage eingelegt wird. Anschließend wird mittels Knopfdrucks der Prozess gestartet. In diesem Anwendungsszenario bedient der Mitarbeiter im Rahmen einer Mehrmaschinenbedienung alternierend zwei identische Maschinen. Durch den Einsatz einer Sprachsteuerung als Substitution zum konventionellen haptischen HMI können hier 27.738 m/a, 7 h/a (0,7 Mio. TMU), 0,19 % (vgl. **Abbildung 27**) eingespart werden. Dabei liegen in diesem Anwendungsszenario alle Laufwege des Produktionsablaufs im kritischen Pfad (vgl. **Abbildung 32**). [NVE23, Vee23]

In **Abbildung 32** werden alle im Anwendungsszenario »Automotive II« vorhandenen Laufwege der unterschiedlichen Bediensituationen in Form eines Spaghetti-Diagramms dargestellt (vgl. **Kapitel 2.7.1**). Genau wie im Anwendungsszenario »Automotive I« in **Abbildung 31** fallen in diesem Anwendungsszenario ein Großteil der Laufwege mehrfach an. Sechs Laufwege fallen viermal an, während neun Laufwege zweimal und nur drei Laufwege einmal im Betrachtungszeitraum anfallen. Gleichzeitig sind die Laufwege in diesem Anwendungsszenario »Automotive II« wie im Anwendungsszenario »Automotive I« deutlich kürzer und auch in Summe etwas weniger als in den Anwendungsszenarien »Recycling I« und »Recycling II«. Sowohl in **Abbildung 27** als auch in **Abbildung 32** ist erkennbar, dass im Anwendungsszenario »Automotive II« wie in den Anwendungsszenarien »Recycling I« und »Recycling II« ausschließlich Laufwege der Kategorien »Unvermeidbar« und »Vermeidbar durch Sprachsteuerung« vorkommen.

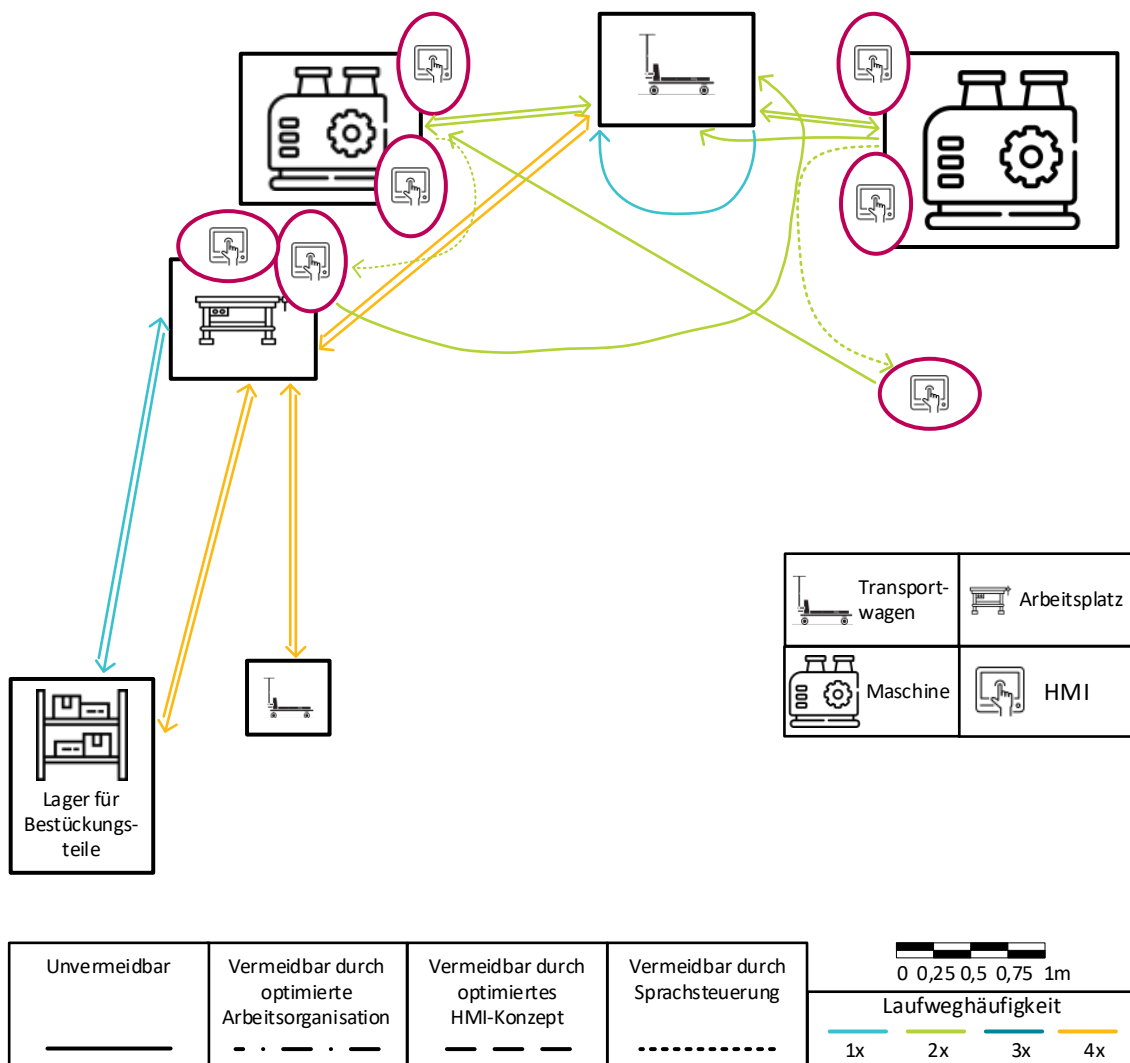


Abbildung 32: Spaghetti-Diagramm des Anwendungsszenarios »Automotive II« (Hersteller für Automobil-Kunststoffinterieurteile) nach [Vee23]

7 Validierung der entwickelten Methode

In diesem Kapitel wird die entwickelte Methode auf Basis der im Lastenheft beschriebenen Anforderungen validiert (vgl. **Kapitel 1.4**). In den ersten vier Anforderungen wird die Gestaltung und das Prinzip der entwickelten Methode beschrieben. So soll diese eine quantitative Evaluation des Effizienzpotenzials einer Sprachsteuerung und eine darauf basierende Prognose der Reduzierung von Nebenzeiten zur Steigerung der Produktivität ermöglichen (vgl. **Kapitel 2.4.7**). Gleichzeitig soll die Bewertung des Effizienzpotenzials auf Basis der aufgewendeten Interaktionszeit inklusive etwaiger Fehleingaben gemäß der DIN EN ISO 9241-11:2018 erfolgen (vgl. **Kapitel 2.4.1**). Die Evaluation soll dabei über identifizierte Einflussfaktoren und abgeleitete Parameter unabhängig vom wertschöpfenden Produktions-/Montageprozess in der Hauptzeit des Produktionsablaufs und damit allgemeingültig sein (vgl. **Kapitel 4**). Die Erfüllung der **Anforderungen 1 bis 4** wurden in den vorangegangenen Kapiteln bereits an verschiedenen Stellen beschrieben. Die **Anforderung 5** an die entwickelte Methode ist eine höhere Präzision der Prognose als interindividuelle Unterschiede in Interaktionszeiten von Bedienern zur Gewährleistung der praktischen Nutzbarkeit. Der Fokus dieses Kapitels liegt auf der Überprüfung dieser **Anforderung 5**, auch wenn die **Anforderungen 1 bis 4** ebenfalls nochmals validiert werden (vgl. **Kapitel 1.4**).

Dazu erhält ein Kollege in seiner Funktion als Versuchsleiter die Aufgabe, die Methode in einer komplett neuen Probandenstudie anzuwenden. Dabei werden sowohl die Auswahl der Anwendungsszenarien und Bediensituationen auf Basis realer Industrie-Anwendungsszenarien als auch das DoE der Probandenstudie überwacht. Der Versuchsleiter war an den vorangegangenen Industrie- und Probandenstudien in **Kapitel 5 & 6**, jedoch nicht an der Entwicklung der finalen Methode in **Kapitel 3** beteiligt. Das bedeutet, dass er grundsätzlich Erfahrungen mit Industrie- und Probandenstudien im Kontext der Effizienz von Sprachsteuerung als industrielles HMI in der Produktion hat, jedoch die vollständige Methode erstmalig nutzt. Die Validierung wird in einem Metaebenen-Experiment durchgeführt damit nicht ein und dieselbe Person die Methode sowohl entwickelt als auch validiert. Dadurch werden sowohl Erkenntnisse zur Allgemeingültigkeit und Abweichung der Methode von der Realität als auch Erkenntnisse zur Gebrauchstauglichkeit der Methode in der Praxis auf einer übergeordneten Ebene gewonnen. Zuerst wird in **Kapitel 7.1** die Versuchsplanung der Probandenstudie zur Validierung der entwickelten Methode beschrieben. Anschließend werden in den **Kapiteln 7.2 und 7.3** die identifizierten Einflussfaktoren und abge-

leiteten Parameter der entwickelten Methode validiert. Danach wird in **Kapitel 7.4** auf individuelle Lerneffekte der Probanden eingegangen und wie diese übergeordnet die Eingabezeiten eines konventionellen HMIs und einer nutzerzentrierten Sprachsteuerung beeinflussen.

7.1 Design of Experiment Probandenstudie Validierung

Auch in dieser Probandenstudie zur Validierung der entwickelten Methode wird eine Technologieplattform genutzt, da es aktuell, wie in Kapitel 5.2 beschrieben, keine Produktionsmaschinen gibt, die standardmäßig mit einer Sprach- und einer Touchsteuerung ausgestattet sind. In dieser Probandenstudie wird zur Validierung eine zweite, optimierte VCP-Technologieplattform II genutzt. Diese besteht ebenfalls aus einer beleuchteten Prozesskammer, in der ein 4-Achs-Roboter montiert ist, jedoch ist die VCP-Technologieplattform II im Gegensatz zu der VCP-Technologieplattform I (vgl. **Abbildung 18**) mit einer 5-Achs-Fräsmaschine und einer überarbeiteten GUI auf dem Touchscreen sowie äquivalenten Sprachbefehlen ausgestattet (vgl. **Kapitel 5.2**). Die VCP-Technologieplattform II ermöglicht die Simulation von noch mehr Anwendungsszenarien und die Untersuchung unterschiedlicher Produktionsabläufe und Bediensituationen. Sowohl die Prozesskammer mit integriertem stufenweisem kippbarem Touchscreen (0° , 15° , 30°) als auch der 4-Achs-Roboter und die 5-Achs Fräsmaschine werden in **Abbildung 33** dargestellt.

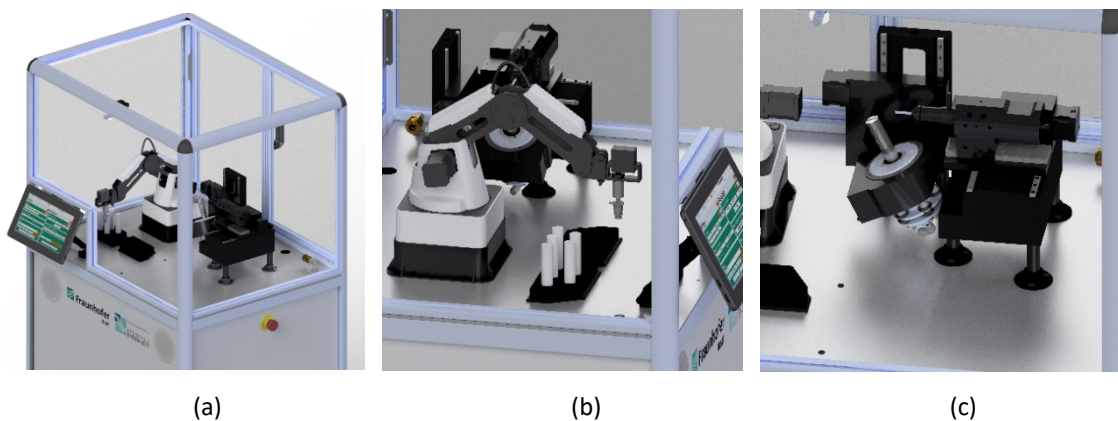


Abbildung 33: Rendering des CAD-Modells der »VCP-Technologieplattform II« mit sprach- und touchbasiertem HMI, (a) Isometrische Ansicht der Prozesskammer von oben rechts, (b) 4-Achs-Roboter mit Halbzeugen sowie Halbzeug- und Bauteilmagazin, (c) 5-Achs Fräsmaschine mit eingespanntem Aluminiumhalbzeug im Spannfutter und eingespanntem Fräswerkzeug in der Spindel

Zur Gewährleistung eines durchgängig gleichbleibenden SNRs, trotz der Untersuchung von Laufwegen, wird in dieser Probandenstudie nicht das oben in der Prozesskammer eingebaute, vom Fraunhofer IDMT in Oldenburg selbst entwickelte omnidirektionale Mikrofon, sondern ein handelsübliches Consumer Ein-Ohr-Headset vom Typ »Jabra Evolve2 65« verwendet. Dadurch bleibt

der Abstand zwischen dem Mund des Probanden und dem Mikrofon sowie der daraus resultierende SNR konstant.

Im ersten Schritt werden mithilfe der entwickelten »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« für die 16 in **Tabelle 5** dargestellten Maschinenbefehle **ohne Laufweg** und die vier in **Tabelle 6** abgebildeten Maschinenbefehle **mit Laufweg** die Eingabedauer/Laufwegdauer abgeschätzt/berechnet. Dazu werden alle Maschinenbefehle wieder hinsichtlich des Einflussfaktors der **Komplexität der Menüstruktur** anhand der beiden Parameter **Anzahl an Menüebenen** und der **Anzahl an Interaktionen** kategorisiert (vgl. **Kapitel 3.2.2**). Gleichzeitig erhalten die Maschinenbefehle mit einem Laufweg eine nach der MTM berechnete Laufwegdauer (vgl. **Kapitel 6.3**). Jeder Maschinenbefehl erhält so einen Soll-Wert für die Eingabedauer und, falls vorhanden, für die Laufwegdauer. Die Soll-Werte werden später mit den real gemessenen Eingabe- und Laufwegdauern aus der Probandenstudie verglichen. [Eng25]

Tabelle 5: Sprach- und Touchbefehle **ohne Laufweg** zur Validierung der Methode [Eng25]

Nr. #	Maschinenbefehl	Menüebenen	Interaktionen
1	Maschine Licht einschalten	0	1
2	Roboter X plus 21 verfahren	1	3
3	Roboter Y minus 14 verfahren	1	3
4	Roboter Z plus 12 verfahren	1	3
5	CNC Kühlung ausschalten	1	3
6	Maschine Licht einschalten	0	1
7	CNC Magazin 4 auswählen	3	5
8	CNC Jog Vorschub 80 eintragen	5	7
9	Maschine Licht einschalten	0	1
10	Roboter Z minus 11 verfahren	1	3
11	Roboter Rotation plus 14,5 verfahren	1	3
12	CNC Jog Inkrement 1.000 auswählen	5	7
13	CNC Werkzeug 5 Schneidenradius 12,15 eintragen	5	10
14	CNC Werkzeug 5 Ausspannlänge 122,3 eintragen	5	10

Tabelle 6: Sprach- und Touchbefehle **mit 10 m Laufweg** zur Validierung der Methode [Eng25]

Nr. #	Maschinenbefehl	Menüebenen	Interaktionen
1	CNC Kühlung einschalten	0	1
2	CNC Einrichtposition anfahren	3	5
3	Roboter Halbzeug 3 rüsten	2	3
4	CNC Programm 2 auswählen	1	3
5	CNC Werkzeug 5 laden	5	7
6	Maschine Licht ausschalten	0	1

Für die genutzten Befehle ohne Laufweg ergibt sich ein rechnerischer Effizienzvorteil von 62,17 %, während sich für die genutzten Befehle mit Laufweg ein rechnerischer Effizienzvorteil von 74,66 % ergibt. Dabei setzt sich der rechnerische Effizienzvorteil für die Befehle mit Laufwegen in der Mehrmaschinenbedienung zu 21,53 % aus dem Durchbrechen komplexer Menüstrukturen und zu 53,13 % durch die Vermeidung von Laufwegen zusammen. [Eng25]

Die für die Validierung ausgewählten Maschinenbefehle bestehen dabei aus Befehlen unterschiedlicher Komplexität für den 4-Achs Roboter als erste Maschine und für die 5-Achs Fräsmaschine als zweite Maschine innerhalb der Prozesskammer der »VCP-Technologieplattform II«. Die Probanden verfahren in der Studie einzelne Achsen beider Maschinen und verändern Parameter von Werkzeugen und Achsenantriebe sowie Halbzeug- und Bauteilmagazinen. [Eng25]

Zur Validierung geben fünf männliche Probanden und eine weibliche Probandin die insgesamt 20 Befehle in jeweils sechs Durchläufen per Sprachsteuerung und in sechs Durchläufen per Touchsteuerung ein (vgl. **Tabelle 7**). So wird jeder Maschinenbefehl in Summe jeweils 36-mal per Sprach- und Touchsteuerung von den Probanden eingegeben. Insgesamt ergeben sich somit jeweils 720 analysierte Sprach- und Touchbefehle. Die erhöhte Anzahl an Durchläufen erlaubt später eine qualifiziertere Aussage hinsichtlich möglicher Lerneffekte über die jeweils sechs Durchläufe, während durch die reduzierte Anzahl an Befehlen die Dauer des Experiments nicht verlängert wird. Dadurch bleiben die Probanden weiterhin konzentriert, obwohl sie insgesamt 240 Befehle eingeben müssen (120 Sprach- und 120 Touchbefehle). [Eng25]

Tabelle 7: Probanden-/Befehlsdaten der Probandenstudie zur Validierung der Methode nach [Eng25]

#	männlich/weiblich	Alter	Start HMI	HMI-Modalitäten	Befehle	Wdh.	Gesamt-Befehle
1	männlich	24	Touch	2	20	6	240 (120/120)
2	männlich	26	Sprache	2	20	6	240 (120/120)
3	männlich	24	Touch	2	20	6	240 (120/120)
4	männlich	27	Sprache	2	20	6	240 (120/120)
5	männlich	24	Touch	2	20	6	240 (120/120)
6	weiblich	26	Sprache	2	20	6	240 (120/120)

Zur Randomisierung beginnen die Probanden alternierend mit Touch- und Sprachsteuerung, wobei der erste Proband mit sechs Durchläufen der Touchsteuerung beginnt, bevor er die sechs Durchläufe per Sprachsteuerung antritt. In **Abbildung 34** wird das zur Validierung genutzte GUI der VCP-Technologieplattform II als konventionelle, haptisch-visuelle HMI-Modalität dargestellt. Dieses ähnelt sehr dem in **Kapitel 5** in der Probandenstudie eingesetzten und in Kapitel **5.2.2** in **Abbildung 19** gezeigten GUI. Die Unterschiede des zur Validierung der entwickelten Methode

eingesetzten GUI, was in **Abbildung 34** gezeigt wird, werden im Folgenden beschrieben. Zum einen befindet sich der Bereich zur Steuerung des 4-Achs-Roboters jetzt auf der linken und der Bereich zur Steuerung der 5-Achs-Fräsmaschine jetzt auf der rechten Seite des GUI. Das liegt daran, dass die Produktionsmaschinen in der Prozesskammer der VCP-Technologieplattform II gleichermaßen positioniert sind. Im Bereich »Fräse« befinden sich in der Sektion »Drehzahl/Vorschub« in der GUI zusätzlich zwei Schieberegler, um sowohl die Drehzahl der Spindel als auch den Vorschub der fünf Achsen der 5-Achs-Fräsmaschine zu regeln.

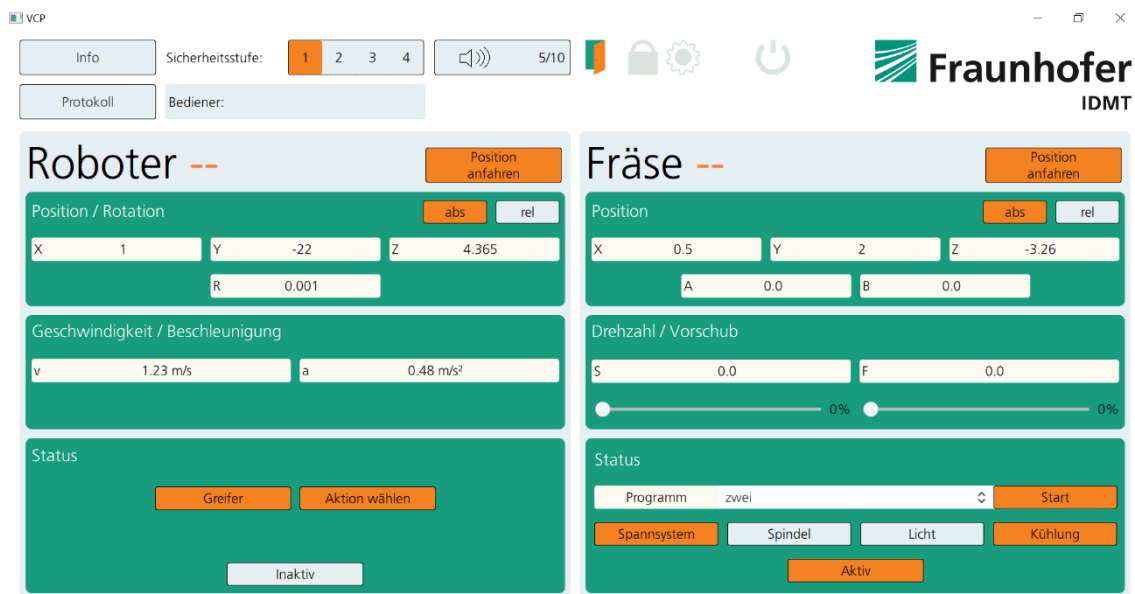


Abbildung 34: »VCP-Technologieplattform II« Graphical User Interface [Eng25]

Ansonsten unterscheidet sich lediglich die Position und Größe der Buttons und die zusätzliche Anzeige des durch die Echtzeit-Sprechererkennung der Sprachsteuerung erkannten Bedieners im »Header«.

7.2 Komplexität von Menüstrukturen

Im Folgenden wird der **Schritt #5** der »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« validiert (vgl. **Kapitel 3.2**). In diesem werden auf Basis, der zuvor in **Schritt #3** am konventionelle HMI für jeden Befehl gemessenen Eingabezeiten über die empirisch ermittelten Zeitäquivalente die Eingabezeiten einer potenziellen Sprachsteuerung abgeschätzt (vgl. **Kapitel 5**). Zunächst werden für alle 20 Befehle die Eingabezeiten für alle Probanden und Durchläufe am konventionellen HMI gemessen. Zur Validierung werden die 14 Befehle ohne Laufwege auf Basis der Menüebenen und Anzahl an Interaktionen des untersuchten GUI kategorisiert und die benötigten Eingabezeiten einer

Sprachsteuerung berechnet. In **Abbildung 35** werden diese zusammen mit den Eingabezeiten des konventionellen HMIs in Form von Boxplots dargestellt. Dabei werden die Eingabezeiten des GUI über T wie Touch und die Eingabezeiten der Sprachsteuerung unter V wie Voice für die Menüebenen 0, 1, 3 und 5 aufgetragen. Die blauen/grünen Kreuze sind Ausreißer für Datenpunkte, die größer als das 1,5-fache der IQR sind. Die schwarzen Pfeile zeigen die Unterschiede der Eingabezeiten für die jeweilige HMI-Modalität (Touch/Voice) für jede Menüebene in Prozent und Sekunden. Die magentafarbenen Sterne stellen für jede Menüebene die mithilfe der »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« berechnete Eingabezeit dar (vgl. **Kapitel 3**).

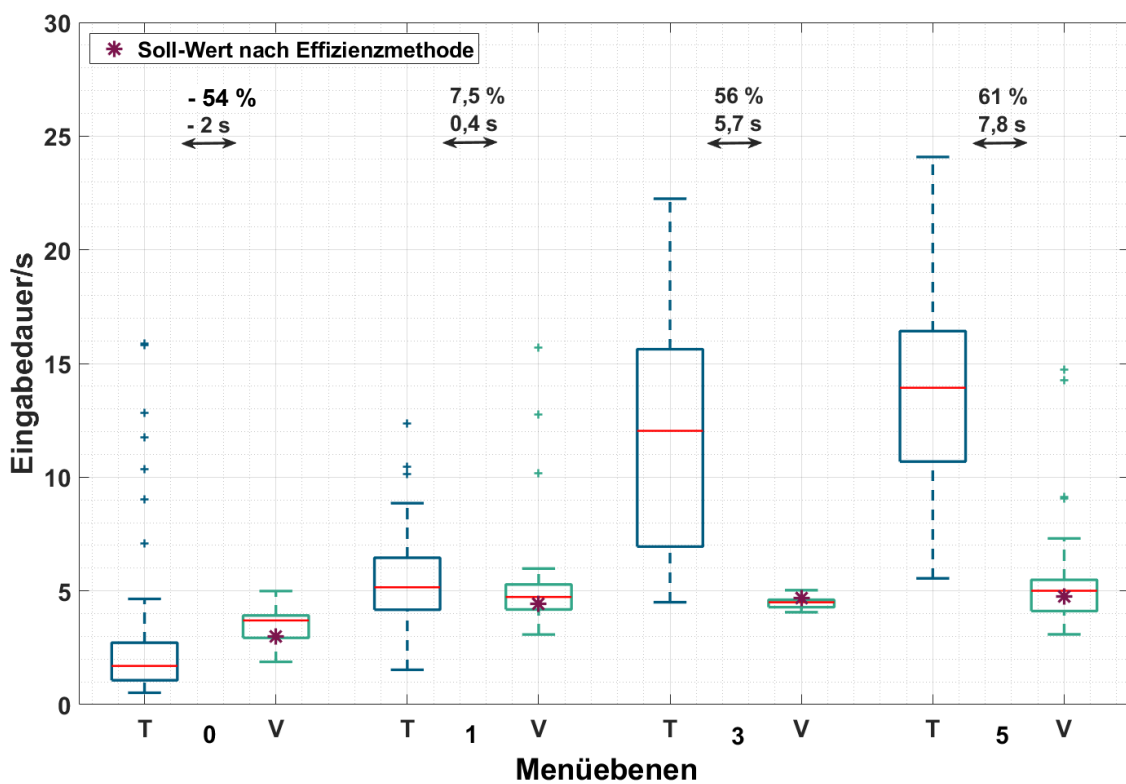


Abbildung 35: Eingabedauer aller Probanden, über alle Durchläufe für die 16 Befehle **ohne Laufwege**, für beide HMI-Modalitäten (Touch/Voice), über die Menüebenen 0, 1, 3 und 5 [Eng25]

Ähnlich zur vorherigen Probandenstudie in **Kapitel 5** steigt der Median der Eingabezeiten der Sprachsteuerung von ca. 3,7 s in der Root-Ebene, über ca. 4,7 s für eine Menüebene und ca. 4,5 s für drei Menüebenen auf ca. 5 s für fünf Menüebenen. Das bedeutet, dass die Eingabezeiten der Sprachsteuerung genau wie bei der vorherigen Probandenstudie von der Root-Ebene auf eine Menüebene leicht ansteigen, bevor sie um ca. fünf Sekunden Eingabezeit pendeln.

Die Mediane der Touchsteuerung steigen von ca. 1,7 s Eingabezeit in der Root-Ebene, über ca. 5,2 s für eine Menüebene und ca. 12 s für drei Menüebenen auf ca. 13,9 s für fünf Menüebenen

an. Ähnlich stiegen die Mediane der Eingabezeiten auch in der vorherigen Probandenstudie an. Hier wuchsen die Mediane jedoch für fünf Menüebenen auf knapp über 15 s Eingabezeit. Die absoluten Werte der Mediane sind hierbei stark abhängig von der Auswahl der Befehle und der Anzahl an Befehlen mit gleicher Anzahl an Menüebenen.

Für die Methode ist es entscheidend, dass der relative Abstand zwischen der Eingabezeit der Touchsteuerung und der der Sprachsteuerung konstant ist. Dieser Zusammenhang wird über die magentafarbenen Sterne dargestellt, die über die in der vorherigen Probandenstudie empirisch ermittelten Zeitäquivalente berechnet werden. Über alle Menüebenen hinweg liegen die berechneten Eingabezeiten der Sprachsteuerung in den (sehr flachen) Boxen der gemessenen Eingabezeiten (vgl. **Abbildung 35**).

In **Abbildung 36** werden die Eingabezeiten beider HMI-Modalitäten in doppelten Boxplots über die Anzahl an Interaktionen aufgetragen.

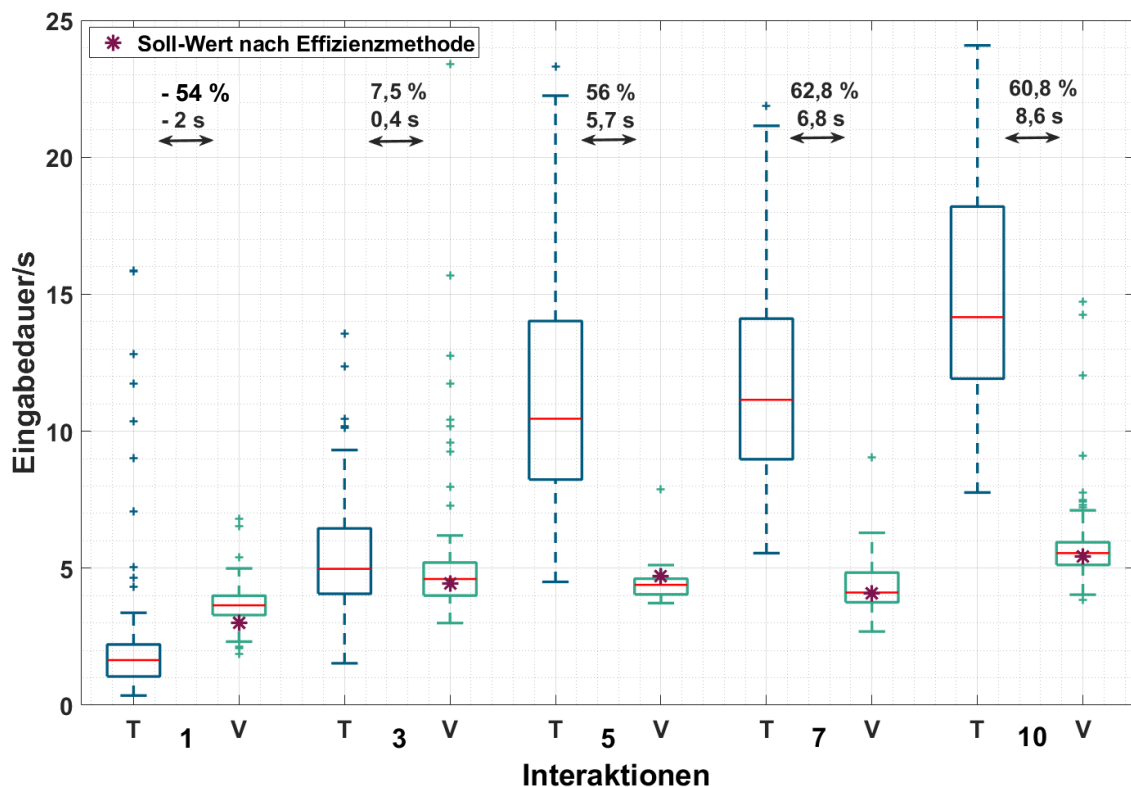


Abbildung 36: Eingabedauer aller Probanden, über alle Durchläufe für die 16 Befehle **ohne Laufwege**, für beide HMI-Modalitäten (Touch/Voice), über die Anzahl an Interaktionen 1, 3, 5 und 7 [Eng25]

Die Betrachtung der Anzahl an Interaktionen ermöglicht einen detaillierten Vergleich der Eingabezeiten der Touchsteuerung mit denen der Sprachsteuerung, da in einer Menüebene in Abhängigkeit vom Design des GUI unterschiedlich viele Interaktionen stattfinden können (vgl. **Kapitel**

5). Auch für die Anzahl der Interaktionen beginnt der Median der Eingabezeit der Sprachsteuerung für eine Aktion bei ca. 3,7 s. Dies entspricht einer Menüebene in **Abbildung 35**. Proportional zu den Menüebenen steigen die Eingabezeiten anschließend über ca. 4,6 s für drei Interaktionen, ca. 4,4 s für fünf Interaktionen, ca. 4,1 s für sieben Interaktionen auf ca. 5,6 s für zehn Interaktionen an.

Die Eingabezeiten für sieben und zehn Interaktionen, die jeweils aus Maschinenbefehlen mit fünf Menüebenen stammen, zeigen deutlich den Zusammenhang zwischen Menüebenen und der Anzahl an Interaktionen. Die 4,1 s Eingabezeit für sieben Interaktionen und die 5,6 s Eingabezeit für zehn Interaktionen ergeben im Mittel die ca. 5 s Eingabezeit für fünf Menüebenen. Der genaue Wert hängt von der Anzahl an Maschinenbefehlen mit sieben Interaktionen im Verhältnis zu denen mit zehn Aktionen ab. Zusätzlich hat die interindividuelle Eingabezeit der Probanden einen Einfluss auf die Eingabezeiten. In dieser Probandenstudie zur Validierung der Methode werden für die Befehle ohne Laufwege zwei Befehle mit sieben und zwei mit zehn Interaktionen genutzt (vgl. **Tabelle 5**). Daraus ergibt sich ein Verhältnis von 50:50.

Die Mediane der Eingabezeiten über die Anzahl der Interaktionen in der Touchsteuerung starten genau wie die der Menüebenen bei ca. 1,6 s, weil für einen Befehl in der Root-Ebene mindestens eine Interaktion notwendig sind (vgl. **Abbildung 36**). Für drei Interaktionen beträgt der Median der Eingabezeit der Touchsteuerung ca. 5 s, für fünf Interaktionen ca. 10,5 s, für sieben Interaktionen ca. 11,1 s und für zehn Interaktionen ca. 14,2 s. Damit steigen auch diese ähnlich proportional zur Anzahl an Interaktionen an, wie die Eingabezeiten in der vorherigen Probandenstudie.

Die magentafarbenen Sterne zeigen auch für die Betrachtung der Eingabezeiten über die Anzahl an Interaktionen, dass die über die empirisch ermittelten Zeitäquivalente berechneten Eingabezeiten der Sprachsteuerung sich ziemlich genau mit den gemessenen Eingabezeiten decken. Dies ist daran zu erkennen, dass die magentafarbenen Sterne entweder in oder zumindest sehr nah an den grünen flachen Boxen der Sprachsteuerung liegen (vgl. **Abbildung 36**).

Trotz anderer Maschinenbefehle, anderer Probanden und einer erhöhten Anzahl an Durchläufen ähneln die Mediane der Eingabezeiten über die beiden Parameter **Anzahl an Menüebenen** und **Anzahl an Interaktionen** (vgl. **Abbildung 35** & **Abbildung 36**) denen der vorherigen Probandenstudie (vgl. **Abbildung 23** & **Abbildung 24**). Die genaue Abweichung der gesamten Methode von der Realität wird in **Kapitel 7.5** beschrieben. Zuvor werden die Ergebnisse der anderen Einflussfaktoren **Art und Anzahl von Laufwegen** und **Anzahl simultan zu bedienender Maschinen** vorgestellt.

7.3 Laufwege und Mehrmaschinenbedienung

In **Kapitel 6** werden im Rahmen einer Industriestudie in neun unterschiedlichen Unternehmen in unterschiedlichen Anwendungsszenarien mit unterschiedlichen Produktionsabläufen und Bediensituationen Laufwege untersucht. Diese werden dabei in die vier Kategorien »Unvermeidbare Laufwege«, »Vermeidbar durch optimierte Arbeitsorganisation«, »Vermeidbar durch optimiertes HMI-Konzept« und »Vermeidbar durch Sprachsteuerung« unterteilt (vgl. **Kapitel 6.2**). Mithilfe von Laserentfernungsmessgeräten werden die unterschiedlichen Layouts in den Produktionshallen der Unternehmen erfasst, um daraus im Anschluss die Längen der anfallenden Laufwege der Maschinenbediener abzuleiten. Nachdem die Längen aller Laufwege innerhalb eines Produktionsablaufs mit seinen Bediensituationen erfasst sind, können die benötigte Laufwegzeiten mit den folgenden drei, in **Schritt #4** der »Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion« beschriebenen, Verfahren berechnet/ermittelt werden (vgl. **Kapitel 3.2**). [NAL24, NVE23, Vee23]

- Messung der Laufwegdauer des Maschinenbedieners für jeden einzelnen Laufweg per Stoppuhr
- Messung der durchschnittlichen Laufgeschwindigkeit der Maschinenbediener im Unternehmen per Stoppuhr und anschließende Berechnung der Laufwegdauer für jeden einzelnen Laufweg durch Division der durchschnittlichen Laufweggeschwindigkeit durch die Laufweglänge
- Berechnung der Laufwegdauer auf Basis des MTM

Zur Validierung des **Schritt #4** zur Berechnung/Ermittlung von Laufwegzeiten, der in **Kapitel 3.2** entwickelten Methode, werden in dieser Probandenstudie die Laufwegzeiten für jeden einzelnen Laufweg jedes einzelnen Probanden automatisiert über ein Skript der VCP-Technologieplattform II und zusätzlich mit einer Stoppuhr gemessen. Für die Berechnung der Laufwegzeiten werden die mit der Stoppuhr gemessenen Werte genutzt, während die automatisierte skriptbasierte Erfassung nur der Überprüfung und Identifikation von Messfehlern dient. Auch wenn die automatisierte Erfassung geringfügig präziser als die Messung mit der Stoppuhr ist, spiegelt sie nicht die Realität wider, da Maschinen in der Realität nicht mit einer Funktion zur automatischen Ermittlung von Laufwegzeiten zwischen einzelnen Maschinen ausgestattet sind, wie es bei der VCP-Technologieplattform II der Fall ist.

Auf Basis der gemessenen Laufwegzeiten für die jeweils sechs Befehle pro Durchlauf in der HMI-Modalität Touchsteuerung ergeben sich für jeden der sechs Probanden insgesamt 36 untersuchte Laufwege mit einer Länge von 10 Metern. Für die Sprachsteuerung fallen keine Laufwege an, da in diesem DoE alle Laufwege durch Sprachbefehle vermieden werden können. Nur so kann das Effizienzpotenzial hinsichtlich der Laufwege analysiert werden. Die einzelnen gemessenen Laufwegzeiten werden in **Abbildung 37** in Form von sechs Boxplots, jeweils einem Boxplot pro Probanden, aufgetragen. [Eng25]

Nach MTM beträgt die Laufwegzeit für zehn Meter geraden Laufweg, bei der Annahme von 1,11 m/s durchschnittlicher Laufgeschwindigkeit, entsprechend 25 TMU oder 0,9 s/m, ca. neun Sekunden. Auf Grund des DoEs dieser Probandenstudie zur Validierung der Methode legen die Probanden zunächst fünf Meter Laufweg zurück, führen dann eine 180° Drehung durch und legen erneut fünf Meter Laufweg zurück. Dadurch ergeben sich nach MTM 5,4 s Laufwegzeit für die fünf Meter Hinweg, plus 0,54 s für die 180° Drehung, plus 5,4 s Laufwegzeit für die fünf Meter Rückweg. Die daraus resultierende Laufwegzeit von 11,34 s ist mit einer orangen gestrichelten Linie in **Abbildung 36** aufgetragen.

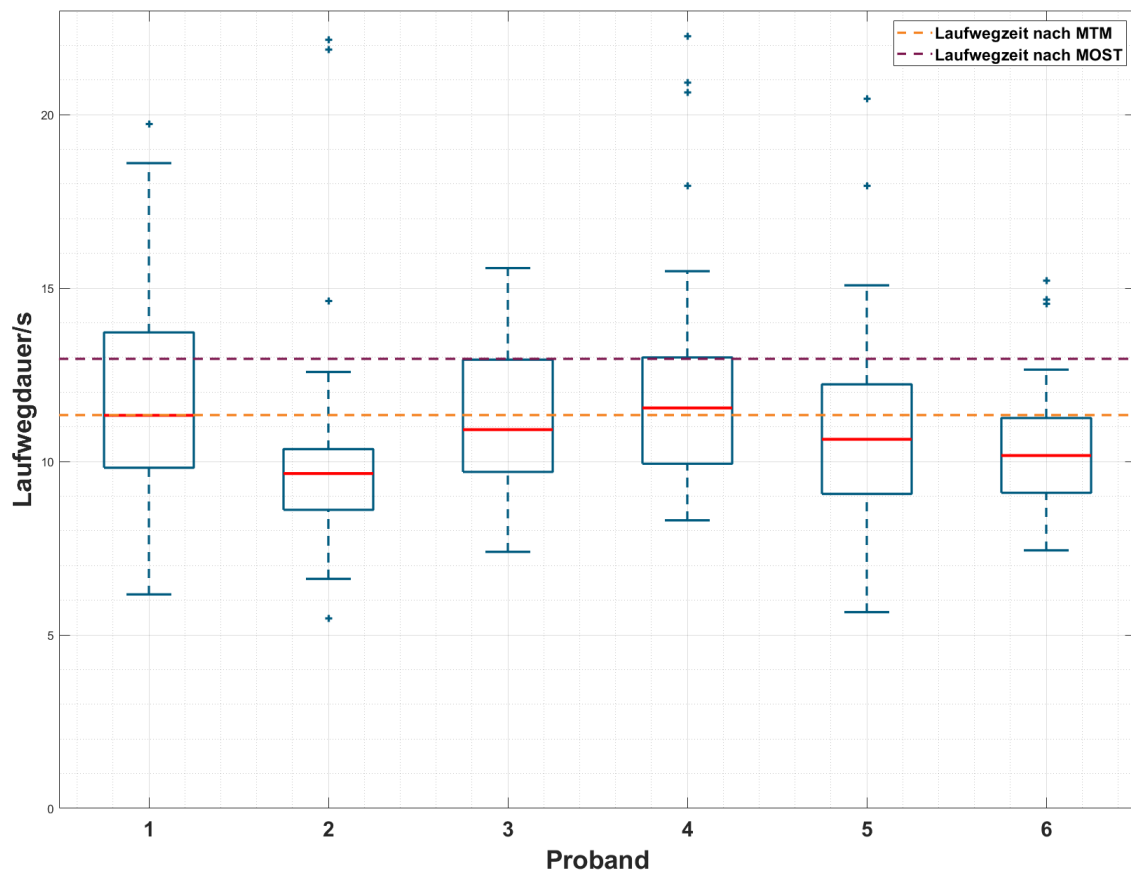


Abbildung 37: Laufwegzeitberechnung nach MTM und MOST im Vergleich mit den realen Laufwegzeiten der sechs Probanden [Eng25]

Zusätzlich wird die benötigte Laufwegzeit nach einem alternativen Verfahren, der »Maynard Operation Sequence Technique« (MOST) berechnet. [Eng25]

- A (Action Distance): 5m Laufweg (Indexwert = 10)
- B (Body Motion): 180°-Drehung (Indexwert = 6)
- G (Gain Control): Keine (Indexwert = 0)
- P (Placement): Keine (Indexwert = 0)
- A (Rückweg): 5m Rückweg (Indexwert = 10)
- Der Indexwert ist mit 10 zu multiplizieren, um die TMU nach MOST zu erhalten
- Gesamtdauer = $100 + 60 + 100 \text{ TMU} \cdot 0,036 = 12,96 \text{ s}$

Die mittels MOST-Verfahren berechnete Laufwegzeit von 12,96 s wird in **Abbildung 37** mit einer lila gestrichelten Linie aufgetragen. Die real gemessenen Laufwegzeiten werden für die sechs Probanden über die sechs Befehle mit Laufweg und die jeweils sechs Durchläufe in Boxplots abgebildet.

Für die Probanden 2-5 ist die Interquartil Range und die Länge der Whisker sehr ähnlich, während Proband 1 mit der höchsten Interquartil Range und den längsten Whiskern heraussticht. Für die Probanden 1-2 und 4-6 sind fast ausschließlich Ausreißer nach oben zu sehen, während sich für Proband 3 alle Daten innerhalb der Länge der Whisker oder der Interquartil Range befinden. Bei den Probanden 2, 4, 5 gibt es jeweils zwischen zwei und drei sogenannte »extreme« Ausreißer von bis knapp über 22 s Eingabezeit, während es nur einen einzigen »milden« Ausreißer nach unten beim Proband 2 bei 5,5 s Laufwegzeit gibt. Insgesamt hat der Proband 2 die Laufwege im Median am schnellsten zurückgelegt.

In **Abbildung 38** wird das Effizienzpotenzial durch die Vermeidung von Laufwegen im Verhältnis zum Effizienzpotenzial durch das Durchbrechen komplexer Menüstrukturen gezeigt. Auf der Abszisse werden die sechs Befehle mit Laufweg dargestellt, während auf der Ordinate die Eingabe-/Laufwegzeit angegeben wird (vgl. **Tabelle 6**). Die grünen und blauen gefüllten Balken geben jeweils den Median über alle Durchläufe und Probanden an. Die gestrichelten blauen Balken zeigen die mittels Sprachsteuerung eingesparten Laufwegzeiten durch vermiedene Laufwege.

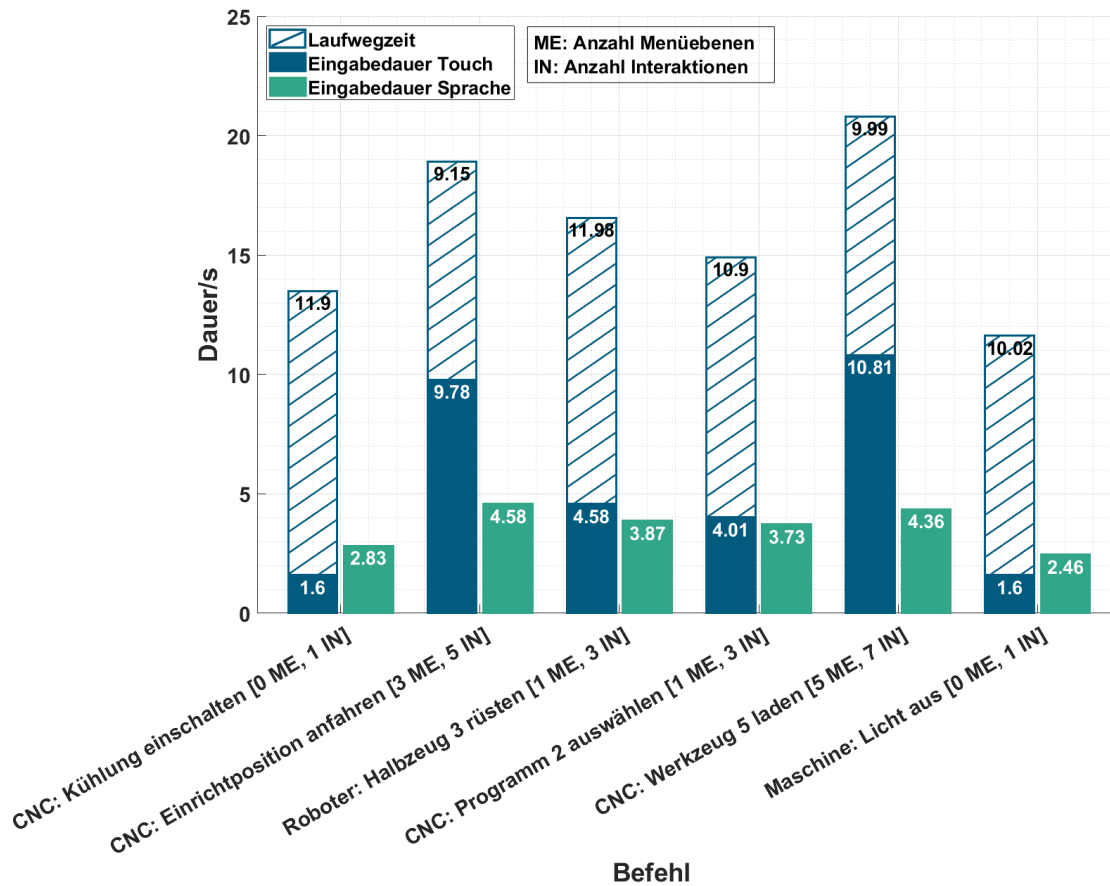


Abbildung 38: Effizienzpotenzial durch die Vermeidung von Laufwegen im Verhältnis zum Effizienzpotenzial durch das Durchbrechen komplexer Menüstrukturen

Wenig komplexe Befehle, die über das Durchbrechen von Menüstrukturen per Sprache keinen Effizienzvorteil haben, können wenn sie einen Laufweg aufweisen, trotzdem von einer Sprachsteuerung profitieren. Komplexe Befehle, die bereits vom Durchbrechen komplexer Menüstrukturen per Sprache profitieren, profitieren zusätzlich von der Vermeidung von Laufwegen durch Sprachsteuerung.

7.4 Lerneffekte

Bereits in den Probandenstudien zur Quantifizierung der Einflussfaktoren auf die Effizienz einer Sprachsteuerung sind in der HMI-Modalität Touchsteuerung signifikante Lerneffekte zu beobachten, während die Eingabezeiten der Sprachsteuerung nahezu konstant über alle Probanden und Durchläufe verliefen (vgl. **Kapitel 5.4.4**). Das bedeutet, dass ein Proband im ersten Durchlauf bei der Eingabe von Maschinenbefehlen in der Menüstruktur des GUI noch verhältnismäßig lange braucht, während in den folgenden Durchläufen die Interaktionsdauer sinkt und vor allem bei Befehlen mit nur wenig Menüebenen und notwendigen Interaktionen sich den

kürzeren Interaktionsdauern einer Sprachsteuerung annähert. Zur Untersuchung der **Komplexität von Menüstrukturen** werden in den in **Kapitel 5** beschriebenen Probandenstudien lediglich drei Durchläufe pro Proband und HMI-Modalität durchgeführt. Grund dafür ist der Fokus auf eine höhere Anzahl an Maschinenbefehlen zur detaillierten Untersuchung des Einflussfaktors **Komplexität von Menüstrukturen** unter Berücksichtigung einer maximalen Durchlaufzeit pro Proband von nicht mehr als 90 Minuten, um Ermüdungserscheinungen der Probanden vorzubeugen. Für diese Probandenstudie zur Validierung der entwickelten Methode wurde die Anzahl an Maschinenbefehlen von 33 in den Probandenstudien in **Kapitel 5** auf 20 (14 ohne Laufweg und sechs mit Laufweg) reduziert, während die Anzahl der Durchläufe von drei auf sechs erhöht wurde (vgl. **Kapitel 7.1**).

Lerneffekte in der Touchsteuerung reduzieren das Effizienzpotenzial einer Sprachsteuerung im Vergleich zur konventionellen, haptisch-visuellen HMI-Modalität. Die Erwartung ist, dass sich diese Lerneffekte nach einer gewissen Anzahl an Durchläufen oder in der Praxis nach einer gewissen Arbeitszeit mit einer Produktionsmaschine und dem zugehörigen HMI asymptotisch einem Grenzwert nähern. In **Abbildung 39** werden die Eingabedauern aller Probanden für die 14 Maschinenbefehle ohne Laufwege für die beiden HMI-Modalitäten (Touch/Voice), kategorisiert nach ihren Durchläufen, dargestellt.

Während die IQR und die Länge der Whisker für die Sprachsteuerung über alle sechs Durchläufe analog zu den in **Kapitel 5** beschriebenen Probandenstudien sehr konstant sind, sinken die Mediane der Eingabedauern für die Touchsteuerung monoton von anfänglich ca. 5,75 s im ersten Durchlauf bis zu ca. 3,75 s im sechsten und damit in dieser Probandenstudie letzten Durchlauf. Die Mediane der Sprachsteuerung pendeln hingegen für das in dieser Probandenstudie genutzte Befehlsset über alle Durchläufe hinweg um eine Eingabedauer von 3,75 s.

Einfluss des untersuchten Befehlskorpus auf die Ausprägung von Lerneffekten

Die untersuchten Maschinenbefehle mit der Kategorisierung der **Komplexität der Menüstruktur** nach **Anzahl an Menüebenen** und **Anzahl an Interaktionen** haben einen wesentlichen Einfluss auf den Grenzwert, dem sich die Eingabedauern asymptotisch nähern. Sowohl die Probandenstudien aus **Kapitel 5** als auch die Probandenstudie zur Validierung in **Kapitel 7.2** zeigen, dass Maschinenbefehle mit weniger als einer Menüebene und/oder drei benötigten Interaktionen über eine Touchsteuerung immer eine geringere Eingabedauer als eine Sprachsteuerung haben. Gleichzeitig zeigen die Daten aller Probandenstudien signifikant, dass Maschinenbefehle mit mehr als einer Menüebene und/oder drei Menüebenen, unter Verwendung einer Sprachsteuerung immer schneller sind als eine vergleichbare Touchsteuerung.

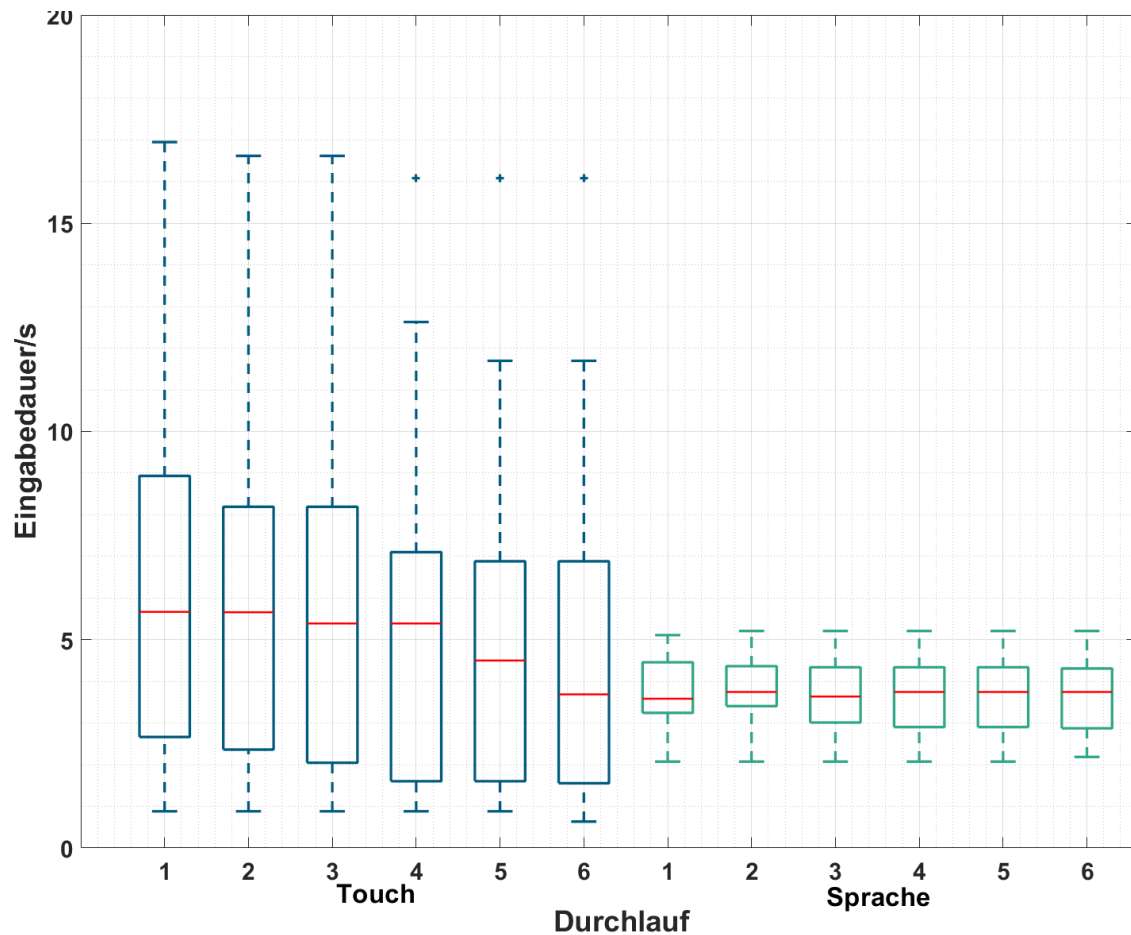


Abbildung 39: Einfluss von Lerneffekten auf die Eingabezeiten der sechs Durchläufe Touchsteuerung (**ohne** Laufwege) und sechs Durchläufe Sprachsteuerung (**ohne** Laufwege). Die blauen/grünen Kreuze sind Ausreißer für Datenpunkten, die größer als das 1,5-fache der Interquartil Range (IQR) sind [Eng25]

Das führt dazu, dass wenn vornehmlich Maschinenbefehle mit wenig Menüebenen und wenig Anzahl an Interaktionen betrachtet werden, der Lerneffekt stärker zum Tragen kommt. Zusätzlich sinken die Mediane der Eingabedauern für eine Touchsteuerung über die Durchläufe schneller. Gleichzeitig liegt der Startmedian der Touchsteuerung nur geringfügig höher oder gar bereits unter dem nahezu konstanten Median der Eingabedauer einer potenziellen Sprachsteuerung. Besteht das untersuchte Befehlsset hingegen vornehmlich aus Befehlen höherer Komplexität mit einer großen Anzahl an Menüebenen und Interaktionen, liegt der Median der Eingabedauer beim ersten Durchlauf signifikant höher als der Median der Eingabedauer eines entsprechenden Sprachsteuerungs-Pendants. In der Praxis kennen die Bediener das konventionelle, haptisch-visuelle HMI ihrer Produktionsmaschine. Aus diesem Grund werden die Lerneffekte in der Praxis niedriger als in Probandenstudien sein. Im Gegensatz dazu sind die Lerneffekte für eine neuartige, nutzerzentrierte Sprachsteuerung sowieso vernachlässigbar gering (vgl. **Abbildung 39**).

Einfluss von Befehlen mit Laufwegen auf die Ausprägung von Lerneffekten

In **Kapitel 7.3** wird die **Art und Anzahl von Laufwegen** als Einflussfaktor auf das mittels der entwickelten Methode berechnete Effizienzpotenzial einer industriellen Sprachsteuerung validiert. Dazu werden die Laufwege der Probanden für Maschinenbefehle der Touchsteuerung mit einer Stoppuhr gemessen (vgl. **Kapitel 7.1**). Zur Identifikation von potenziellen Messfehlern werden die händisch erfassten Eingabe- und Laufwegzeiten in dieser Probandenstudie zusätzlich noch automatisiert und skriptbasiert erfasst (vgl. **Kapitel 7.1**). Im Anschluss werden die gemessenen Laufwegzeiten der sechs Durchläufe zum einen mit den berechneten Laufwegzeiten nach der MTM und zum anderen mit den berechneten Laufwegzeiten nach der MOST verglichen.

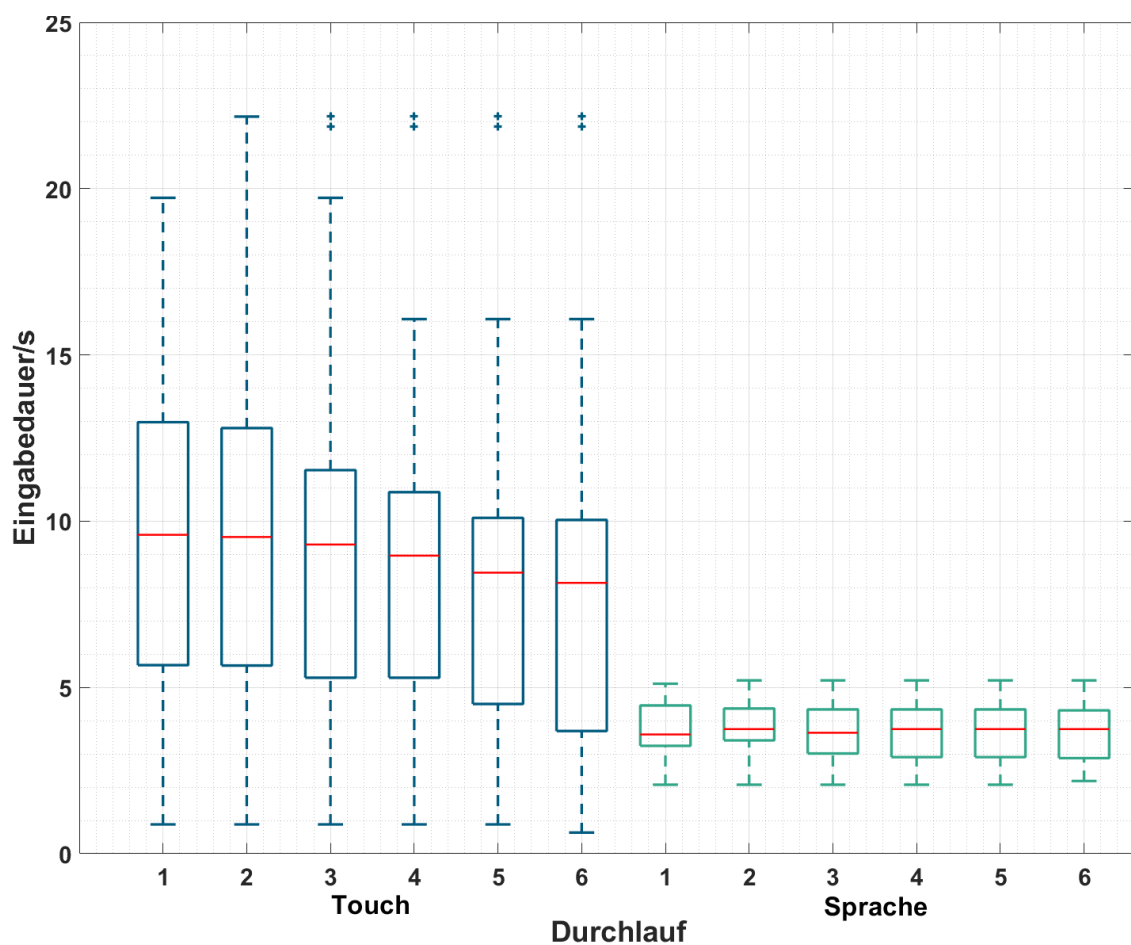


Abbildung 40: Einfluss von Lerneffekten auf die Eingabezeiten der sechs Durchläufe Touchsteuerung (mit Laufwegen) und sechs Durchläufe Sprachsteuerung (ohne Laufwege) [Eng25]

In **Abbildung 37** ist deutlich zu sehen, dass die gemessenen Laufwegzeiten knapp unter den berechneten Laufwegzeiten nach MTM und schon deutlich weiter unter den berechneten Laufwegzeiten nach MOST liegen. Gleichzeitig ist jedoch auch zu beobachten, dass die Mediane der Laufwegzeiten über alle Probanden und Maschinenbefehle der HMI-Modalität Touchsteuerung

nahezu konstant sind. Dies bedeutet im Umkehrschluss, dass Laufwege die Eingabezeiten von Maschinenbefehlen zwar mit einem Offset belegen, diese jedoch keinen Einfluss auf die Ausprägung von Lerneffekten haben. In **Abbildung 40** werden die Eingabedauern aller Probanden, für die sechs Maschinenbefehle mit Laufwegen für beide HMI-Modalitäten (Touch/Voice), nummeriert nach den Durchläufen in der Probandenstudie, dargestellt.

Während die interindividuellen Laufwegzeiten der Probanden abhängig von ihrer Physionomie, aber dennoch konstant sind, haben Ermüdungserscheinungen in der Praxis einen Einfluss auf diese. Viele und lange Laufwege führen dazu, dass vor allem weniger trainierte Mitarbeitende langsamer laufen. In diesen Fällen bietet die Sprachsteuerung ein zusätzliches Effizienzpotenzial, das nicht durch Lerneffekte ausgeglichen werden kann. Vor allem in einer alternden Gesellschaft führt die Vermeidung von Laufwegen zu einer Steigerung der Produktivität.

7.5 Abweichungen zwischen der entwickelten Methode und den real gemessenen Interaktionszeiten (Eingabe-/Laufwegzeiten)

In den vorangegangenen **Kapiteln 7.2** und **7.3** werden die Kongruenz der einzelnen aus den drei Haupt-Einflussfaktoren **Anzahl der simultan zu bedienenden Maschinen**, **Komplexität der Menüstrukturen** und **Art und Anzahl von Laufwegen** resultierenden Effizienzpotenziale mit der Realität verglichen. Dazu sind jeweils die in der Probandenstudie zur Validierung gemessenen Interaktionszeiten (Eingabe-/Laufwegzeiten) gegenüber den mithilfe der entwickelten Methode prognostizierten Interaktionszeiten (Eingabe-/Laufwegzeiten) aufgetragen. Zusätzlich werden in **Kapitel 7.4** HMI-abhängig, unterschiedlich stark ausgeprägte Lerneffekte untersucht, um Aussagen über das Effizienzpotenzial einer Sprachsteuerung im Vergleich zu einer Touchsteuerung auch für längere Bedienungszeiträume treffen zu können.

Auf Basis der Evaluation der unterschiedlichen Einflussfaktoren anhand von abgeleiteten Parametern kann am Ende eine Einsparung von Nebenzeiten im Produktionsablauf prognostiziert werden. Im Rahmen der Validierung wird überprüft, inwieweit diese Prognose von den real gemessenen Interaktionszeiten (Eingabe-/Laufwegzeiten) abweicht. Dafür sind in **Tabelle 8** die minimale Abweichung, maximale Abweichung und mittlere Abweichung der mithilfe der Methode berechneten Differenz zwischen dem Median der Interaktionszeiten für alle Probanden, aller Durchläufe, jeweils für die 14 Befehle ohne und die sechs Befehle mit Laufweg, im Verhältnis zu den berechneten Interaktionszeiten aufgetragen.

Tendenziell ist die mittlere Abweichung für die Betrachtung der **Anzahl an Menüebenen** größer als für die Betrachtung der **Anzahl an Interaktionen**, da die Anzahl an Interaktionen die Komplexität der Menüstruktur noch präziser widerspiegelt. In dieser Validierung werden jedoch zum Teil für eine bestimmte Anzahl an Menüebenen und eine bestimmte zugehörige Anzahl an Interaktionen nur ein Befehl in 36-facher Wiederholung untersucht (vgl. **Kapitel 7.1**). Tritt in diesem untersuchten Befehl die maximale oder minimale Abweichung zwischen berechneter und gemessener Interaktionsdauer auf, sind die Abweichungen für die Anzahl an Menüebenen und Interaktionen gleich. Dies ist bei den untersuchten Befehlen der Laufwege in den rechten beiden Spalten der Fall.

Tabelle 8: Minimale, maximale und durchschnittliche Abweichung der Methode von der Realität

	Ohne Laufwege		Mit Laufwegen	
	Menüebenen	Interaktionen	Menüebenen	Interaktionen
Minimale Abweichung	2,04 % (0,06 s)	0,68 % (0,04 s)	0,68 % (0,03 s)	0,68 % (0,03 s)
Maximale Abweichung	28,66 % (1,10 s)	28,66 % (1,09 s)	8,92 % (0,33 s)	8,92 % (0,03 s)
Mittlere Abweichung	12,31 % (0,50 s)	8,51 % (0,33 s)	4,80 % (0,17 s)	4,80 % (0,17 s)

Insgesamt ist die Dauer der Interaktion zwischen Probanden und HMI länger, wenn zusätzlich ein Laufweg zurückgelegt werden muss. Daraus folgt, dass bereits kleine absolute Abweichungen der Laufwegdauer zu einer großen relativen Abweichung der Methode im Verhältnis zur Realität führen. Die maximale Abweichung der Methode liegt für Befehle ohne Laufwege bei der Betrachtung der **Anzahl an Menüebenen** bei 28,66 % (1,10 s), während die minimale Abweichung für Befehle ohne Laufwege bei der Betrachtung der Anzahl an Interaktionen bei 0,68 % (0,04 s) liegt. Für Befehle mit Laufwegen liegt die maximale Abweichung bei der Betrachtung der Menüebenen bei 8,92 % (0,33 s) und die minimale Abweichung bei 0,68 % (0,03 s). Wie oben beschrieben gilt dies analog auch für die Betrachtung der **Anzahl an Interaktionen**. Die mittlere Abweichung der Prognose zur Reduzierung von Nebenzeiten liegt mit 4,80 % (0,17 s) bis 12,31 % (0,50 s) unter der gemittelten und interindividuellen Abweichung der Probanden, die in **Abbildung 35**, **Abbildung 36** und **Abbildung 37** abgebildet sind. Damit wird die fünfte und letzten Anforderung aus dem Lastenheft, die höhere Präzision der Prognose als interindividuelle Unterschiede in Interaktionszeiten von Bedienern zur Gewährleistung der praktischen Nutzbarkeit, erfüllt (vgl. **Kapitel 1.4**).

8 Diskussion der Ergebnisse im Kontext der Forschungsfragen

Im Folgenden werden die Ergebnisse dieser Arbeit der Kapitel **3, 4, 5, 6 & 7** im Kontext der im **Kapitel 1.3** aufgestellten Forschungsfragen diskutiert. Zuerst werden die Hauptforschungsfrage und anschließend die Teilforschungsfragen diskutiert und in einen größeren Kontext gesetzt.

8.1 Methode zur Evaluation der Effizienz einer Sprachsteuerung

Die Hauptforschungsfrage dieser Arbeit lautet (vgl. **Kapitel 1.3**):

»Wie muss eine Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion gestaltet sein?«

Das übergeordnete Ziel der Methode ist die Identifikation von Effizienzpotenzial zur Reduzierung von Nebenzeiten als nicht wertschöpfende Tätigkeiten zur Steigerung der Produktivität von Produktionsabläufen. Ist eine Methode zu aufwändig oder kompliziert in der Anwendung, wird sie in der Praxis nicht genutzt. Das bedeutet, dass das Optimierungsziel für die Entwicklung einer Methode die Identifikation aller Einflussfaktoren ist, deren Parameter einen so großen Einfluss auf die Effizienz haben, dass deren Effekt nicht im Rauschen der Abweichung der Methode von der Realität untergeht. Dies kann passieren, wenn der Einfluss des jeweiligen Einflussfaktors kleiner als die statistische Unsicherheit der Ergebnisse der Methode auf Basis der empirischen Daten ist. Da das Optimierungsziel nicht berechenbar ist, muss eine solche Methode iterativ entwickelt werden. Das heißt, es müssen Schritt für Schritt weitere Einflussfaktoren hinzugefügt werden und die Präzision der Evaluation stets neu bewertet werden.

Die in dieser Arbeit ermittelten Zeitäquivalente zur Darstellung des Effizienzverhältnisses zwischen Sprach- und Touchbefehlen basieren genau wie die TMUs der MTM und Indexwerte der MOST zur Berechnung von Laufwegzeiten auf empirischen Daten. Während die MTM und MOST bereits seit langen Jahren etablierte Methoden sind, die in dieser Arbeit nochmal in einer eigenen Probandenstudie validiert und ins Verhältnis zu real gemessenen Werten gesetzt werden, ist die entwickelte *»Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion«* mit den ermittelten Zeitäquivalenten neu. Weitere Probandenstudien können daher in Zukunft die empirisch ermittelten Zeitäquivalente im Detail

noch verbessern. Wesentlicher Einflussfaktor ist jedoch eher die Länge und Komplexität des verwendeten Hotwords und Sprachbefehls im Vergleich zum Aufbau und der Komplexität des untersuchten GUI. [NER23]

8.2 Industrielle Produktionsabläufe mit Effizienzpotenzial durch Sprachsteuerung

Ergänzend zur Hauptforschungsfrage ergeben sich die folgenden Teilforschungsfragen. Die zweite Forschungsfrage lautet (vgl. **Kapitel 1.3**):

»Welche Produktionsabläufe lassen sich durch eine Sprachsteuerung im Vergleich zu einem konventionellen HMI effizienter gestalten?«

Sowohl die Industrie- als auch die Probandenstudien haben in denen verschiedenste Bediensituationen in unterschiedlichen Branchen gezeigt, dass ausschließlich teilautomatisierte Produktionsabläufe von einem gebrauchstauglichen und damit effizienten HMI profitieren. Das liegt daran, dass nur innerhalb teilautomatisierter Produktionsprozesse eine Kollaboration zwischen Mensch und Maschine über ein menübasiertes HMI stattfindet (vgl. **Kapitel 2.1**).

Im Besonderen profitieren teilautomatisierte Produktionsabläufe von einer Sprachsteuerung, wenn Mensch und Maschine wirklich kollaborieren und der Bediener nicht nur den eigentlichen wertschöpfenden vollautomatisierten Produktionsprozess vorbereitet oder Störungen beseitigt. Grund dafür ist der größtmögliche Anteil an Interaktion in kollaborativen, teilautomatisierten Produktionsabläufen. Dabei können nicht nur der eigentliche wertschöpfende Produktionsprozess, sondern auch Wartungs-, Reparatur und Rüstprozesse teilautomatisiert sein. Auch dann, wenn der eigentliche wertschöpfende Produktionsprozess manuell oder vollautomatisiert abläuft.

Die Effizienz in der Interaktion mit dem HMI lässt sich unabhängig von der zu bedienenden Produktionsmaschine anhand übergeordneter Einflussfaktoren und davon abgeleiteten Parametern quantifizieren (vgl. **Kapitel 5 & 6**). Typische Produktionsmaschinen, die hinsichtlich der Effizienz von HMIs in der Literatur untersucht werden, sind Produktionsautomaten, Roboter, Werkzeugmaschinen und Produktionsanlagen (vgl. **Kapitel 4**). Die Nutzung von übergeordneten Einflussfaktoren und Parametern innerhalb der entwickelten Methode gewährleistet die Allgemeingültigkeit und vom Produktionsablauf unabhängige Nutzbarkeit dieser (vgl. **Kapitel 1.4**).

8.3 Einflussfaktoren auf die Effizienz einer Sprachsteuerung

Die dritte Forschungsfrage lautet (vgl. **Kapitel 1.3**):

»Welche Einflussfaktoren sind für den effizienten Einsatz einer Sprachsteuerung als industrielles HMI einer Produktionsmaschine relevant?«

Die durchgeführten Probanden- und Industriestudien haben gezeigt, dass viele verschiedene Faktoren die Effizienz eines industriellen HMI beeinflussen (vgl. **Kapitel 5, 6 und 7**). Nach einem Überblick über alle Einflussfaktoren werden für die Entwicklung einer Methode zur quantitativen Evaluation der Effizienz einer Sprachsteuerung die Einflussfaktoren identifiziert, die einen besonders großen Einfluss auf die betrachtete Ressource, in dieser Arbeit die aufgewendete Zeit, haben (vgl. **Kapitel 4.2**).

Auf Basis der in den Probanden- und Industriestudien gewonnenen Erkenntnissen sowie begleitenden Experteninterviews und einer Literaturrecherche werden in dieser Arbeit die **Komplexität von Menüstrukturen, Art und Anzahl von Laufwegen** sowie die **Anzahl der simultan zu bedienenden Maschinen** als die drei entscheidenden Einflussfaktoren identifiziert. Diese und ergänzende weitere Einflussfaktoren, die unter Umständen die Abweichung zwischen der Methode und der Realität noch weiter verringern, werden in **Kapitel 4.2** beschrieben. In **Kapitel 7.5** wird diese Abweichung zwischen der entwickelten Methode und der Realität quantifiziert und die Abhängigkeit der relativen Abweichung von der Länge der Interaktionszeit beschrieben.

In der Industriestudie in **Kapitel 6** wird gezeigt welche Produktionsabläufe, mit welcher **Art von Laufwegen** in der Mehrmaschinenbedienung durch die Ergänzung oder den Austausch eines konventionellen, haptisch-visuellen HMIs mit einer Sprachsteuerung hinsichtlich ihrer Effizienz profitieren. Dabei hat der Einflussfaktor **Art und Anzahl der Laufwege** mit seinen Parametern **Häufigkeit** und **Länge** einen direkten Einfluss auf das relative und absolute Effizienzpotenzial. Die Abweichung des realen vom berechneten Effizienzpotenzial hängt im Wesentlichen von der Gewissenhaftigkeit ab, mit der der Produktionsablauf in seine einzelnen Bediensituationen zerlegt wird und der anschließenden Kategorisierung der Laufwege ab (vgl. **Kapitel 6.2**).

Die Ergebnisse der Probandenstudien in **Kapitel 5 und 7** zeigen, dass einfache Maschinenbefehle auf der Root-Ebene eines konventionellen, haptisch-visuellen HMIs, unabhängig vom Geschlecht, Alter oder Vorerfahrungen der Probanden immer schneller ausgeführt werden. Doch bereits ab nur einer zusätzlichen Menüebene oder drei Interaktionen benötigen die Probanden

mit der Sprachsteuerung eine geringere Eingabezeit. Laufwege zwischen unterschiedlichen HMIs oder zwischen HMI und eigentlichem Arbeitsplatz des Bedieners an der Maschine sowie eine Mehrmaschinenbedienung verstärken diesen Effekt zusätzlich. [NER23, NVE23]

Intuitivität und Lerneffekte einer Sprachsteuerung

Die Industrie- und Probandenstudien haben gezeigt, dass die Sprachsteuerung von Produktionsmaschinen die Möglichkeit bietet, komplexe Menüstrukturen zu durchbrechen, Laufwege zu vermeiden und eine Mehrmaschinenbedienung zu ermöglichen. Gleichzeitig ist die Steuerung von Produktionsmaschinen per Sprache eine für die Probanden intuitive Form der Interaktion. Obwohl keiner der Probanden aus den Probandenstudien vorher eine Produktionsmaschine per Sprache bedient hat, ist der Median der Eingabedauer bereits im ersten Durchlauf sehr gering und gleichzeitig auch immer niedriger als bei einer konventionellen Touchsteuerung. Zusätzlich bleibt dieser sehr konstant über mehrere Durchläufe. [NER23]

Alle durchgeführten Probandenstudien zeigen jedoch, dass ein Bediener mit ein wenig Übung durchaus auch in der Bedienung mit einer konventionellen Touchsteuerung Eingabezeiten von einer Sprachsteuerung erreichen kann. In den Probandenstudien zeigte sich dies, in Abhängigkeit des Anteils an komplexen Befehlen mit vielen Menüebenen und Interaktionen, bereits nach drei bis sechs Durchläufen. [NER23]

Verarbeitungs- und Wartezeit der eingesetzten Sprachsteuerung

Die in dieser Arbeit an den VCP-Technologieplattformen I & II eingesetzte Sprachsteuerung benötigt eine gewisse Zeit, um den eingehenden Audiostream zu verarbeiten. Gleichzeitig ist sie so programmiert, dass sie auf ergänzende Parameter wie Koordinaten oder boolsche Werte wartet. Dies führt dazu, dass die Sprachsteuerung besonders robust und zuverlässig eingegebene Sprachbefehle erkennt und erst eine Aktion der Maschine ausgelöst wird, wenn der vollständige Sprachbefehl vorliegt. Auf der anderen Seite führt dieses Vorgehen dazu, dass die Sprachsteuerung langsamer ist als sie sein müsste. Für erfahrene Maschinenbediener, die mit den zur Verfügung stehenden Sprachbefehlen vertraut sind, könnte diese Zeit noch um bis zu 1,3 s reduziert werden. [NER23]

Zur weiteren Gewährleistung einer ausreichenden Robustheit (in den Probandenstudien lag die Fehlerquote unter 2 %) ist es dann aber wichtig, dass die eingesetzten Sprachbefehle weiterhin eindeutig sind und kein Sprachbefehl Bestandteil eines anderen komplexeren Sprachbefehls ist. So ist sichergestellt, dass, wenn ein vollständiger Sprachbefehl erkannt wird, dies auch der Rich-

tige ist. Auch wenn weder die Verarbeitungs- noch die Wartezeit für zusätzliche Parameter vollständig eliminiert werden können, können sie reduziert werden, um den Effizienzvorteil einer Sprachsteuerung im Vergleich zu einer Touchsteuerung weiter zu steigern. [NER23]

Die Frage wie viele Einflussfaktoren verwendet werden sollen ist stets eine Abwägung zwischen der Präzision der Evaluation und der Nutzbarkeit der Methode in der Praxis. Eine Methode zur quantitativen Evaluation der Effizienz eines HMI ist ein Werkzeug für die Praxis, um Produktionsabläufe zu bewerten und auf Basis der Ergebnisse eine Entscheidung für oder gegen eine Sprachsteuerung zu treffen. Gleichzeitig basiert die entwickelte Methode auf empirischen Daten, die zum Teil aus anderen bereits etablierten Methoden (wie MTM/MOST) und zum anderen aus experimentell erhobenen Daten der Probanden- und Industriestudien stammen.

Dies führt zwangsläufig dazu, dass so gewissenhaft und präzise die Daten auch erhoben wurden, die Berechnung immer von der Realität abweichen wird. Das führt dazu, dass jeder weitere Einflussfaktor die Anwendung der Methode zwar aufwändiger gestaltet, die Präzision des Ergebnisses jedoch nicht zwangsläufig verbessert.

8.4 Parameter zur Quantifizierung des Effizienzpotenzials

Die vierte Forschungsfrage lautet (vgl. **Kapitel 1.3**):

»Welche Parameter können genutzt werden, um das Effizienzpotenzial einer industriellen Sprachsteuerung im Vergleich zu einem konventionellen HMI zu quantifizieren?«

Im ersten Schritt der Entwicklung einer solchen Methode ist es notwendig, die zu evaluierende Effizienz im Kontext der Betrachtung exakt zu definieren. Gemäß DIN EN ISO 9241-11 beschreibt die Effizienz das Verhältnis zwischen eingesetzten Ressourcen (verwendete Zeit, menschlicher Aufwand, finanzielle Ressourcen und verbrauchte Materialien) und den erzielten Ergebnissen (vgl. **Kapitel 2.4.1**) [DIN18].

In dieser Arbeit wird für die Entwicklung einer solchen Methode die verwendete Zeit als genutzte Ressource betrachtet. Neben der beschriebenen Interaktionszeiten bestehend aus Eingabe- und Laufwegzeiten als Hauptparameter werden die folgenden Parameter in der Methode genutzt, um das Effizienzpotenzial einer Sprachsteuerung im Vergleich zu einem konventionellen HMI zu quantifizieren.

- Anzahl simultan zu bedienender Maschinen
- Eingabezeit des konventionellen HMI
- Laufwegzeit des konventionellen HMI
- Art des Laufwegs
- Häufigkeit des Laufwegs
- Anzahl an Menüebenen
- Anzahl an Interaktionen
- Eingabezeit der Sprachsteuerung
- Laufwegzeit der Sprachsteuerung

Für jeden weiteren Einflussfaktor, um den man die Methode ergänzt muss mindestens ein weiterer Parameter eingeführt werden, um das Effizienzpotenzial der Sprachsteuerung weiterhin quantifiziert evaluieren zu können.

8.5 Entwicklung von Produktionsabläufen zur Nutzung des Effizienzpotenzials von Sprachsteuerung

Die fünfte Forschungsfrage lautet (vgl. **Kapitel 1.3**):

»Lässt sich die Effizienz von bestehenden Produktionsabläufen durch den Einsatz einer industriellen Sprachsteuerung signifikant steigern oder ist eine Anpassung von Produktionsabläufen an diese nutzerzentrierte HMI-Modalität notwendig, um das Effizienzpotenzial zu nutzen?«

Kurz gesagt: Beides! Aktuelle Produktionsabläufe sind so gestaltet, dass Sie mit aktuell etablierten und damit zur Verfügung stehenden HMI-Modalitäten möglichst effizient betrieben werden können. Mithilfe der entwickelten *»Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion«* können Produktionsabläufe, ohne die vorausgehende Integration einer industriellen Sprachsteuerung, identifiziert werden, die hinsichtlich ihrer zeitlichen Effizienz von einer Sprachsteuerung profitieren würden. Sowohl die Probanden- und Industriestudien als auch die Entwicklung und Validierung der Methode haben gezeigt, dass dies vornehmlich durch das Durchbrechen komplexer Menüstrukturen, die Vermeidung von Laufwegen oder eine Mehrmaschinenbedienung erfolgen kann (vgl. **Kapitel 5, 6 und 7**).

Da die Steuerung von Produktionsmaschinen per Sprache jedoch noch nicht flächendeckend verbreitet ist, sind aktuelle Produktionsabläufe nicht optimal auf diese HMI-Modalität ausgelegt.

Das heißt, dass ein zusätzliches Effizienzpotenzial von Sprachsteuerung darin liegt, dass Produktionsabläufe explizit auf den Einsatz eine Sprachsteuerung als ergänzende oder ersetzende HMI-Modalität zu einem konventionellen HMI ausgelegt werden.

Besonders in der Industriestudie, aber auch in der Literaturrecherche und in Gesprächen mit Fachexperten aus der Industrie hat sich gezeigt, dass deutsche Produktionsunternehmen Produktionsabläufe oft vollständig manuell oder aber vollautomatisiert auslegen. Eine Teilautomatisierung, in der eine Interaktion zwischen Mensch- und Maschine über ein HMI stattfindet, ist in der Auslegung vieler Produktionsabläufe nicht vorgesehen oder nur in rudimentären Zügen, wie dem Starten/Stoppen von Maschinen oder der Anpassung weniger oder sogar nur eines einzelnen Parameters. Hier liegt noch ein großes Effizienzpotenzial im Verborgenen, dass durch die in dieser Arbeit entwickelte Methode gewonnenen Erkenntnisse nutzbar gemacht wird.

9 Fazit und Ausblick

In dieser Arbeit wird eine Methode entwickelt, mit der Unternehmen in die Lage versetzt werden, ihre Produktionsabläufe hinsichtlich des Effizienzpotenzials einer Sprachsteuerung als Ergänzung oder Austausch zu einem konventionellen, haptisch-visuellen HMI systematisch zu bewerten. Zusätzlich werden Handlungsempfehlungen zur Gestaltung und zum Einsatz einer industriellen Sprachsteuerung zur Reduzierung von Nebenzeiten in Produktionsabläufen und der daraus resultierenden Steigerung der Produktivität gegeben. Mithilfe der entwickelten Methode wird HMI-, Steuerungs- und Maschinenherstellern sowie Betreibern von Produktionsmaschinen die Möglichkeit eröffnet, die Reduzierung von Nebenzeiten in Form von Interaktionszeiten (Eingabe-/Laufwegzeiten) zu prognostizieren, ohne die tatsächliche zeit- und kostenintensive Integration einer Sprachsteuerung. Diese Prognose dient Unternehmen anschließend als Entscheidungsbasis für oder gegen den Einsatz einer Sprachsteuerung. So wird die Nutzung der in dieser Arbeit beschriebenen Vorteile einer Sprachsteuerung systematisiert.

In Industriestudie, Expertengesprächen und einer ergänzenden Literaturrecherche werden dazu Anwendungsszenarien identifiziert, die potenziell hinsichtlich ihrer zeitlichen Effizienz von einer Sprachsteuerung profitieren. Aus diesen teilautomatisierten Produktionsabläufen, in denen Mensch und Maschine kollaborativ zusammenarbeiten, werden die Einflussfaktoren, die einen signifikanten Effekt auf die Effizienz der Interaktion mit dem HMI haben, extrahiert. Dabei werden die **Komplexität von Menüstrukturen, Art und Anzahl von Laufwegen** sowie die **Anzahl der simultan zu bedienenden Maschinen** als die drei Haupt-Einflussfaktoren identifiziert. Diese werden im Folgenden in unterschiedlichen Probandenstudien in verschiedenen Produktionsabläufen und zugehörigen Bediensituationen, die auf den Erkenntnissen der Industriestudie und Literaturrecherche basieren, quantifiziert.

Einflussfaktoren auf das Effizienzpotenzial von Sprachsteuerung

Für die **Komplexität von Menüstrukturen** ergibt der Einsatz einer Sprachsteuerung im Vergleich zu einer konventionellen Touchsteuerung ein Effizienzpotenzial von bis zu 67 %, pro Befehl bei fünf Menüebenen. Der Break-Even-Point für den Einsatz einer Sprachsteuerung wird dabei bereits bei Maschinenbefehlen mit nur einer zusätzlichen Menüebene (18 % Effizienzpotenzial) oder drei Interaktionen (21 % Effizienzpotenzial) erreicht. Die Analyse der **Art und Anzahl von Laufwegen** in neun Unternehmen unterschiedlicher Branchen und Maschinen mit unterschiedlichen Anwendungsszenarien und Produktionsabläufen sowie zugehörigen Bediensituationen

ergab ein jährliches relatives Effizienzpotenzial in der Interaktion von bis zu 84 %, durch die Vermeidung von Laufwegen per Sprachsteuerung. Die beiden Einflussfaktoren **Komplexität von Menüstrukturen** und **Art und Anzahl von Laufwegen** addieren sich in Bediensituationen mit komplexen Menüstrukturen, die durch eine Sprachsteuerung durchbrochen werden und Laufwegen, die durch eine Sprachsteuerung vermieden werden können. Zusätzlich skalieren beide Effekte über die **Anzahl der simultan zu bedienenden Maschinen** in einer Mehrmaschinenbedienung.

Neben den Haupt-Einflussfaktoren validieren und quantifizieren sowohl die Industrie- als auch die Probandenstudien unterschiedliche Thesen für Neben-Einflussfaktoren auf die Effizienz einer Sprachsteuerung aus der Literatur. Allen voran zeigten Lerneffekte in der Bedienung eines HMIs einen signifikanten Einfluss auf die zeitliche Effizienz. Während die Eingabedauer für die genutzte Sprachsteuerung in allen Probandenstudien sowohl für simple als auch komplexe Befehle von der ersten bis zur letzten Iteration nahezu konstant niedrig ist, sieht man für die haptische Bedienung des Touchscreens einen klaren Abwärtstrend in den Eingabezeiten und damit einen Aufwärtstrend in der Effizienz (37 % Effizienzpotenzial). Dies geht so weit, dass die Touchsteuerung sich für wenige Menüebenen und Anzahl an Interaktionen stark an die Eingabezeiten einer Sprachsteuerung annähert, sodass das Effizienzpotenzial einer Sprachsteuerung für einen erfahrenen Maschinenbediener deutlich geringer als der Durchschnitt ist.

Anwendung der entwickelten Methode in der Praxis

Die Identifikation von Einflussfaktoren sowie die Bündelung dieser in einer »*Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles HMI*« ermöglicht Unternehmen erstmalig die systematische quantitative Analyse ihrer Produktionsabläufe hinsichtlich des Effizienzpotenzials einer Sprachsteuerung. Dabei wird kein Eingriff in bestehende Hard- oder Software benötigt, da zum einen auf bereits im Produktionsablauf vorhandene Daten zurückgegriffen wird und zum anderen alle weiteren Daten mit einfachen Messwerkzeugen mit überschaubarem Aufwand erhoben werden können. Die Alternative, die tatsächliche Integration einer Sprachsteuerung als Austausch oder Ergänzung zu einem konventionellen HMI und der anschließende Vergleich von Eingabe- und Laufwegzeiten, bedeutet erheblich mehr Aufwand. Durch den Einsatz der entwickelten Methode wird für Unternehmen die Hürde für die Implementierung neuartiger, nutzerzentrierter HMIs, wie zum Beispiel einer Sprachsteuerung gesenkt, da bereits im Vorhinein eine Reduzierung von Nebenzeiten prognostiziert wird.

Validierung der entwickelten Methode

Die Abschätzung des Effizienzpotenzials auf Basis empirisch ermittelter Zeitäquivalente für die Eingabezeiten und mittels MTM errechneter Laufwegzeiten wird nie die Präzision des tatsächlichen Vergleichs desselben Produktionsablaufs mit zwei HMI-Modalitäten erreichen. Trotzdem zeigt die Validierung der Methode, bei der in einer zusätzlichen Probandenstudie zuerst das rechnerische Effizienzpotenzial ermittelt und dann die reale Interaktionsdauer gemessen wird, dass das rechnerische Effizienzpotenzial bereits für eine Entscheidung für oder gegen Sprachsteuerung ausreichend präzise Werte liefert. Die mittlere Abweichung der berechneten Interaktionsdauern der Probanden beträgt in der Validierung, bei der reinen Betrachtung der **Komplexität von Menüstrukturen** (ohne Laufwege) ca. 12,31 % (0,50 s). Für die Untersuchung der **Art und Anzahl von Laufwegen** sowie die damit verbundene **Mehrmaschinenbedienung** in der Probandenstudie zur Validierung sank die mittlere Abweichung der Methode von den real gemessenen Interaktionszeiten auf ca. 4,80 % (ca. 0,17 s). Die Identifikation der effizientesten HMI-Modalität wird gleichzeitig immer weniger Aufwand als die tatsächliche Integration einer Sprachsteuerung sein, da alle Messungen im Rahmen der Methode auch hier benötigt werden. Die Abschätzung in der Methode ersetzt also eins zu eins den Integrationsaufwand.

Die systematische Auseinandersetzung von Unternehmen mit den HMIs ihrer Maschinen führt gleichzeitig auch zu Synergieeffekten bei der Identifizierung verborgener Optimierungspotenziale vorhandener Produktionsabläufe und zugehöriger konventioneller HMIs. Besonders stark kommt dies im Bereich der Komplexität von historisch gewachsenen Menüstrukturen in GUIs und Laufwegen, die durch eine optimierte Arbeitsorganisation oder ein optimiertes HMI-Konzept vermieden werden können, zum Tragen.

Erweiterung der entwickelten Methode

Die entwickelte Methode bietet die Möglichkeit, das Effizienzpotenzial einer Sprachsteuerung im Vergleich zu einer konventionellen, haptisch-visuellen Touchsteuerung abzuschätzen. Die Touchsteuerung wurde als Ausgangsbasis für die Methode gewählt, da diese den aktuellen Stand der Technik von HMIs an modernen Produktionsmaschinen abbildet. Das bedeutet, dass ein Großteil der Produktionsabläufe, in denen HMIs mit der Methode bewertet werden können, eine entsprechende Steuerung aufweist. Nichtsdestotrotz kann die Methode in Zukunft noch flexibler für andere konventionelle HMIs und gegebenenfalls für diese auch präziser gestaltet werden, indem in weiteren Probandenstudien Zeitäquivalente für die **Anzahl an Menüebenen** und **Anzahl an Interaktionen** für die Navigation durch GUIs über Tasten neben dem Bildschirm oder eine traditionelle Maus und Tastatursteuerung ermittelt werden.

Neben unterschiedlichen konventionellen HMI-Modalitäten als Ausgangsbasis kann das Prinzip der Methode zur Berechnung des Effizienzpotenzials auch auf andere neuartige, nutzerzentrierte HMI-Modalitäten wie Gestensteuerungen, Eye Tracking oder in der Zukunft gar BCIs erweitert werden. Sowohl das Durchbrechen von **komplexen Menüstrukturen** als auch die **Vermeidung von Laufwegen** und die **Mehrmaschinenbedienung** sollten in diesem Kontext untersucht werden.

Ganzheitliche Betrachtung von Einflussfaktoren auf die Effizienz von HMI

Besonders für neuartige HMI-Lösungen, mit denen in der industriellen Praxis noch nicht so viele Erfahrungen gesammelt wurden, ist es wichtig, neben der Effizienz als Teil der Gebrauchstauglichkeit auch andere Effekte auf den Produktionsablauf und im Besonderen die Bediener zu untersuchen. Über besonders effiziente HMI-Systeme werden eine Vielzahl an Informationen zwischen dem Bediener und der Maschine innerhalb kürzester Zeit ausgetauscht. Dies kann bei Bedienern vor allem in der Anfangszeit innerhalb einer Eingewöhnungsphase zu einer kognitiven Überlastung (engl.: mental fatigue) und einer damit verbundenen kurzfristigen Reduzierung der Effizienz oder gar einer mittelfristigen vollständigen Ablehnung neuer Technologien führen. Gleichzeitig zeigten die strukturierten Befragungen begleitend zu den Probandenstudien in dieser Arbeit, dass einige Probanden sich bei der Nutzung einer Sprachsteuerung in Gegenwart von weiteren Personen weniger wohl fühlen als bei der Nutzung einer Touchsteuerung. Aus diesen Gründen sollten neben Probandenstudien zur Evaluation der Effizienz, Methoden zur Erhebung von subjektiven Maßen wie strukturierte Befragungen oder der NASA-TLX, aber auch objektive Maße wie die Herzfrequenzvariabilität (HFV), EEG oder mobile Elektrodermale Aktivität (EDA) genutzt werden. Dadurch kann ein ganzheitliches Bild des Einsatzes neuartiger HMI-Modalitäten in der Produktion gewonnen werden. Je neuartiger eine HMI-Modalität ist und gleichzeitig geringer die Erfahrungen mit dieser in der industriellen Praxis sind, desto größer ist die Gefahr, dass eine isolierte Betrachtung der Effizienz ohne die Berücksichtigung des Bedienerempfindens und -verhaltens zu verfälschten Ergebnissen in den Effizienzpotenzialen führt.

Notwendigkeit von neuen Produktionsabläufen für neuartige Human-Machine Interfaces

Aktuell etablierte Produktionsabläufe sind im besten Fall auf aktuell verbreitete und zur Verfügung stehende HMIs ausgelegt. Im schlechtesten Fall sind Produktionsabläufe bei der Einführung neuartiger HMIs weder überprüft noch speziell an diese angepasst worden. Neben der Nut-

zung des Effizienzpotenzials von Sprachsteuerung als Ergänzung oder Austausch zu konventionellen HMIs in aktuellen Produktionsabläufen sollten diese, auf Basis der Erkenntnisse dieser und ähnlicher Arbeiten an die eingesetzte HMI-Modalität angepasst werden, um das volle Effizienzpotenzial auszuschöpfen. Die Industriestudien haben gezeigt, dass branchenübergreifend viele Produktionsabläufe manuell oder voll automatisiert ablaufen. Gründe dafür sind unter anderem HMIs, die in teilautomatisierten Produktionsabläufen nicht ausreichend praktikabel sind. Während ortsfeste HMI-Terminals zu zusätzlichen Laufwegen führen, blockieren mobile HMIs wie Smartphones, Tablets oder andere mobile haptische Steuerungseinheiten die Hände von Maschinenbedienern. Neuartige nutzerzentrierte HMIs wie Gestensteuerung, Eye Tracking, BCIs oder aber Sprachsteuerung, als die HMI-Modalität, die von allen genannten der Schwelle zur flächendeckenden Industrialisierung am nächsten steht, eröffnen ganz neue Möglichkeiten für die Kollaboration von Mensch und Maschine in teilautomatisierten Prozessen. Damit diese Möglichkeiten aber auch genutzt werden, müssen die Effizienzpotenziale der jeweiligen HMI-Modalitäten bei für die Produktionsabläufe ausreichender Robustheit systematisch nachgewiesen werden.

Literatur

- [AFA19] Afanasev, Maxim Ya., Yuri V. Fedosov, Yuri S. Andreev, Anastasiya A. Krylova, Sergey A. Shorokhov, Kseniia V. Zimenko und Mikhail V. Kolesnikov. 2019. „A Concept for Integration of Voice Assistant and Modular Cyber-Physical Production System.“, 27–32. <https://doi.org/10.1109/INDIN41052.2019.8972015>.
- [AhZA17] Ahamed, Mohammed Mithun und Bakar ZA. 2017. „Analysis of Human Machine Interaction Design Perspective: A Comprehensive Literature Review.“ *International Journal of Contemporary Computer Research (IJCCR)* (Vol. 1 Issue. 1).
- [AiHe22] Aichele, Christian und Herrmann, Jörg, Hrsg. 2022 *Betriebswirtschaftliche KI-Anwendungen: Digitale Geschäftsmodelle auf Basis Künstlicher Intelligenz*. 2. Auflage. Wiesbaden: Springer Vieweg. 978-3-658-40098-9.
- [AmMa23] Ambadar, Fiolisya Faustine und Jude Joseph Lamug Martinez. 2023. „Improving the performance of speech-gesture multimodal interface in non-ideal environments.“ *Procedia Computer Science* 216:587–96. <https://doi.org/10.1016/j.procs.2022.12.173>.
- [Bal15] Balsliemke, Frank. 2015 *Kostenorientierte Wertstromplanung: Prozessoptimierung in Produktion und Logistik*. Essentials. Wiesbaden: Springer Gabler. 978-3-658-08698-5.
- [BMB15] Bates, Douglas, Martin Mächler, Ben Bolker und Steve Walker. 2015. „Fitting Linear Mixed-Effects Models Using lme4.“ *J. Stat. Soft.* 67 (1). <https://doi.org/10.18637/jss.v067.i01>.
- [Bec05] Becker, Torsten. 2005 *Prozesse in Produktion und Supply Chain optimieren*. Berlin, Heidelberg: Springer. 978-3-540-25841-4.
- [BiAy20] Bingol, Mustafa Can und Omur Aydogmus. 2020. „Performing predefined tasks using the human–robot interaction on speech recognition for an industrial robot.“ *Engineering Applications of Artificial Intelligence* 95. <https://doi.org/10.1016/j.engappai.2020.103903>.
- [Bin10] Binner, Hartmut F. 2010 *Handbuch der prozessorientierten Arbeitsorganisation: Methoden und Werkzeuge zur Umsetzung*. 4. Aufl. REFA-Fachbuchreihe Unternehmensentwicklung. München: Hanser. 9783446426412.
- [BBA21] Böge, Alfred; Böge, Wolfgang; Arndt, Klaus-Dieter; Bahmann, Werner; Barfels, Lutz; Bauer, Jürgen; Böge, Gert et al., Hrsg. 2021 *Handbuch Maschinenbau: Grundlagen und Anwendungen der Maschinenbau-Technik*. 24., überarbeitete und erweiterte Auflage. Wiesbaden, Heidelberg: Springer Vieweg. 978-3-658-30272-6.
- [BNP21] Breque, Maija, Lars de Nul und Athanasios Petridis. 2021 *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*. R&I Paper Series, policy brief. Luxembourg: Publications Office of the European Union. 978-92-76-25308-2.
- [BrLa11] Bressemer, Jana und Silva H. Ladewig. 2011. „Rethinking gesture phases: Articulatory features of gestural movement?“. *Semiotica* 2011 (184). <https://doi.org/10.1515/semi.2011.022>.
- [Bul95] Bullinger, Hans-Jörg. 1995 *Arbeitsgestaltung: Personalorientierte Gestaltung marktge-rechter Arbeitssysteme*. Technologiemanagement. Stuttgart: Teubner. 9783663077985.
- [BuSc21] Burggräf, Peter und Schuh, Günther, Hrsg. 2021 *Fabrikplanung*. 2., vollständig neu bearbeitete und erweiterte Auflage. Handbuch Produktion und Management / Günther Schuh (Hrsg.) 4. Berlin, Heidelberg: Springer Vieweg. 978-3-662-61968-1.

- [CZL23] Chen, Haodong, Niloofar Zendehelel, Ming C. Leu und Zhaozheng Yin. 2023. „Real-time human-computer interaction using eye gazes.“ *Manufacturing Letters* 35:883–94. <https://doi.org/10.1016/j.mfglet.2023.07.024>.
- [DIN22] DIN Deutsches Institut für Normung e.V. 2022 *Fertigungsverfahren: Begriffe, Einteilung*. 2022. Aufl. ICS 01.040.25, 25.020, Nr. 8580. Berlin: DIN Media GmbH. Zugriff am 21. Dezember 2024. <https://www.dinmedia.de/de/norm/din-8580/331759492>.
- [DIN20a] DIN Deutsches Institut für Normung e.V. / DIN-Normenausschuss Ergonomie. 2020 *Ergonomie der Mensch-System-Interaktion: Teil 220: Prozesse zur Ermöglichung, Durchführung und Bewertung (ISO 9241-220:2019)*. 2019. Aufl. ICS 13.180; 35.180, Nr. 9241-220. Berlin: Beuth Verlag GmbH.
- [DIN20b] DIN Deutsches Institut für Normung e.V. / DIN-Normenausschuss Ergonomie. 2020 *Ergonomie der Mensch-System-Interaktion: Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019)*. 2019. Aufl. ICS 13.180; 35.180, Nr. 9241-210. Berlin: Beuth Verlag GmbH.
- [DIN18] DIN Deutsches Institut für Normung e.V. / DIN-Normenausschuss Ergonomie. 2018 *Ergonomie der Mensch-System-Interaktion: Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018)*. 2018. Aufl. ICS 13.180; 35.080; 35.180, Nr. 9241-11. Berlin: Beuth Verlag GmbH.
- [DIN13] DIN Deutsches Institut für Normung e.V. / DIN-Normenausschuss Ergonomie. 2013 *Ergonomie der Mensch-System-Interaktion: Teil 154: Sprachdialogsysteme (ISO 9241-154:2013)*. 2013. Aufl. ICS 13.180; 35.240.99, Nr. 9241-154. Berlin: Beuth Verlag GmbH.
- [DIN08] DIN Deutsches Institut für Normung e.V. / DIN-Normenausschuss Ergonomie. 2008 *Sicherheit von Maschinen - Ergonomische Gestaltungsgrundsätze: Teil 2: Wechselwirkungen zwischen der Gestaltung von Maschinen und den Arbeitsaufgaben*. 2000/2008 ICS 13.110; 13.180, Nr. 614-2. Berlin: Beuth Verlag GmbH.
- [DIN00] DIN Deutsches Institut für Normung e.V. / DIN-Normenausschuss Ergonomie. 2000 *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten: Teil 14: Dialogführung mittels Menüs (ISO 9241-14:1997)*. 1999. Aufl. ICS 13.180; 35.080; 35.240.20, Nr. 9241-14. Berlin: Beuth Verlag GmbH.
- [EKM20] Ehrlenspiel, Klaus, Alfons Kiewert, Markus Mörtl und Udo Lindemann. 2020 *Kostengünstig Entwickeln und Konstruieren: Kostenmanagement bei der integrierten Produktentwicklung*. 8. Auflage. Berlin, Heidelberg: Springer Vieweg. 978-3-662-62590-3.
- [Eng25] Engel, Christoph. 2025. „Effizienz von Sprachsteuerung als HMI-System in der Mehrmaschinenbedienung: Kollaboration von Mensch und Maschine in der digitalen Fabrikhalle von morgen.“ Unveröffentlichtes Manuskript. Masterarbeit.
- [FMP23] Fazio, Roberto de, Vincenzo Mastronardi, Matteo Petrucci, Massimo de Vittorio und Paolo Visconti. 2023. „Human–Machine Interaction through Advanced Haptic Sensors: A Piezoelectric Sensory Glove with Edge Machine Learning for Gesture and Object Recognition.“ *Future Internet* 15 (1): 14. <https://doi.org/10.3390/fi15010014>.
- [FeSe10] Feng, Jinjuan und Andrew Sears. 2010. „Beyond errors: measuring reliability for error-prone interaction devices.“ *Behaviour & Information Technology* 29 (2): 149–63. <https://doi.org/10.1080/01449290802196596>.
- [FjKa22] Fjodorovs, Nikita und Marvin Kalla. 2022. „Einsatz eines Sprachassistenzsystems in der Produktion: Was gibt es zu beachten? Diskussion der Ergebnisse einer Laborstudie am Beispiel eines Montageprozesses.“ 117:634–38. <https://doi.org/10.1515/zwf-2022-1135>.

- [Fri20] Fritz, Jessica. 2020. „Gestaltungsregeln für digitale Technologien an der Mensch-Maschine-Schnittstelle in der Fabrikhalle: Konsolidierung, Weiterentwicklung und Standardisierungsmöglichkeiten.“ Zugriff am 6. Juni 2025. https://www.smartelectronicfactory.de/fileadmin/user_upload/bilder/digitaltalk/Zukunftsszenarien_Anforderungsgestaltung_2020-08-12.pdf.
- [GaPe18] Ganapathy, Sriram und Vijayaditya Peddinti. (2018) 2018. „3-D CNN Models for Far-Field Multi-Channel Speech Recognition.“ In *2018 IEEE International Conference on Acoustics, Speech, and Signal Processing: Proceedings : April 15-20, 2018, Calgary Telus Convention Center, Calgary, Alberta, Canada*, 5499–5503. Piscataway, NJ: IEEE.
- [Gei21] Geisler, Stefan. 2021. „Von Fahrinformation über Fahrassistenten zum autonomen Fahren.“ In *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement*, hrsg. von Christian Reuter. 2., aktualisierte Auflage, 383–403. Wiesbaden, Heidelberg: Springer Vieweg.
- [GhRe19] Ghofrani, Javad und Dirk Reichelt. 2019. „Using Voice Assistants as HMI for Robots in Smart Production Systems.“ Unveröffentlichtes Manuskript, zuletzt geprüft am 6. Juni 2025. <http://ceur-ws.org/Vol-2339/paper13.pdf>.
- [GSL14] Gorecky, Dominic; Schmitt, Mathias; Loskyll, Matthias und Zuhlke, Detlef, Hrsg. 2014 *Human-machine-interaction in the industry 4.0 era*.
- [Gra20] Graichen, Lisa. 2020 *Gestures for Human-Machine Interaction: Design Aspects, User Experience and Impact on Driving Safety*. Chemnitz. Zugriff am 27. November 2024. <https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-718279>.
- [GFS20] Gross, Erwin, Sven Finkbeiner, Jörg Siegert und Thomas Bauernhansl. 2020. „Sprachsteuerung für die Mensch-Roboter-Kollaboration in der Montage/Design of voice commands for human-robot collaboration in assembly.“ *wt* 110 (01-02): 73–79. <https://doi.org/10.37544/1436-4980-2020-01-02-75>.
- [GRC19] Guerin, C., P. Rauffet, C. Chauvin und E. Martin. 2019. „Toward production operator 4.0: modelling Human-Machine Cooperation in Industry 4.0 with Cognitive Work Analysis.“ *IFAC-PapersOnLine* 52 (19): 73–78. <https://doi.org/10.1016/j.ifacol.2019.12.111>.
- [Gün22] Günther, Birgit. 2022. „Künstliche Intelligenz in der Automobilindustrie: Von den ersten Visionen bis zu selbst denkenden Autos.“ In *Betriebswirtschaftliche KI-Anwendungen: Digitale Geschäftsmodelle auf Basis Künstlicher Intelligenz*, hrsg. von Christian Aichele und Jörg Herrmann. 2. Auflage, 243–68. Wiesbaden: Springer Vieweg.
- [Har06] Hart, Sandra G. 2006. „Nasa-Task Load Index (NASA-TLX); 20 Years Later.“ *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (9): 904–8. <https://doi.org/10.1177/154193120605000909>.
- [Har86] Hart, Sandra G. 1986. „NASA TASK LOAD INDEX.“ *NASA Ames Research Center*, 1–26. <https://ntrs.nasa.gov/api/citations/20000021488/downloads/20000021488.pdf>. Zugriff am 11. Februar 2025.
- [Hei24] Heidenhain, Johannes. 2024. „Heidenhain TNC7: Die Bahnsteuerung für Bearbeitungszentren und Fräs-Dreh-Maschinen.“ 1384156-12. Unveröffentlichtes Manuskript, zuletzt geprüft am 6. Juni 2025. https://www.heidenhain.de/fileadmin/pdf/de/01_Produnkte/Prospekte/PR_TNC7_Endkunde_ID1384156_de.pdf.
- [Hig13] Higgins, James M. 2013. „The Fourth Singularity and the Future of Jobs.“ *World Futures Review* 5 (1): 11–23. <https://doi.org/10.1177/1946756712473437>.
- [Inn16] Innovation Post. (2016) 2016. „Rockwell Automation potenzia Studio 5000 per uno sviluppo più efficiente del sistema di automazione.“ Zugriff am 6. Juni 2025. <https://>

www.innovationpost.it/tecnologie/industrial-it/rockwell-automation-potenzia-studio-5000-uno-sviluppo-piu-efficiente-del-sistema-automazione/.

- [IoSc21] Ionescu, Tudor B. und Sebastian Schlund. 2021. „Programming cobots by voice: A human-centered, web-based approach.“ *Procedia CIRP* 97:123–29. <https://doi.org/10.1016/j.procir.2020.05.213>.
- [ISIE12] ISO International Organization for Standardization und IEC International Electrotechnical Commission. 2012 *ISO/IEC 25041:2012-10 System und Software-Engineering - Qualitätskriterien und Bewertung von System- und Softwareprodukten (SQuaRE): Evaluierungshandbuch für Entwickler, Ankäufer und unabhängige Gutachter*. 2012. Aufl. ICS 35.080, Nr. 25041. Berlin: VDE Verlag. <https://www.vde-verlag.de/iec-normen/219286/iso-iec-25041-2012.html>.
- [JHR21] Janíček, M., R. Holubek, R. Ružarovský und K. Velíšek. 2021. „Voice frequency impact for voice control of collaborative robots.“ *IOP Conf. Ser. Mater. Sci. Eng.* 1009. <https://doi.org/10.1088/1757-899X/1009/1/012025>.
- [JRV21] Janíček, M., R. Ružarovský, K. Velíšek und R. Holubek. 2021. „Analysis of voice control of a collaborative robot.“ *J. Phys. Conf. Ser.* 1781 (1): 12025. <https://doi.org/10.1088/1742-6596/1781/1/012025>.
- [JuMa08] Jurafsky, Daniel und James H. Martin. 2008 *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2. ed. Prentice Hall series in artificial intelligence. Upper Saddle River, NJ: Prentice Hall. 9780131873216.
- [KaMe20] Kahle, Tim und Dominik Meißner. 2020 *All about voice*. 1. Auflage. Freiburg, München, Stuttgart: Haufe Group. <https://www.haufe.de/>. 978-3-648-13466-5.
- [KaBr03] Kamiske, Gerd F. und Jörg-Peter Brauer. 2003 *Qualitätsmanagement von A bis Z: Erläuterungen moderner Begriffe des Qualitätsmanagements*. 4., aktualisierte und erg. Aufl. München, Wien: Hanser. 9783446224582.
- [Kel83] Kelley, J. F. (1983) 1983. „An Empirical Methodology for Writing User-Friendly Natural Language Computer Applications.“ In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, hrsg. von Raoul N. Smith, 193–96. ACM Conferences. New York, NY: ACM.
- [KHJ23] Khan, Moin, Abid Haleem und Mohd Javaid. 2023. „Changes and improvements in Industry 5.0: A strategic approach to overcome the challenges of Industry 4.0.“ *Green Technologies and Sustainability* 1 (2): 100020. <https://doi.org/10.1016/j.grets.2023.100020>.
- [Koe23] Koerner, Konstantin. 2023. „The wage effects of offshoring to the East and West: evidence from the German labor market.“ *Rev World Econ* 159 (2): 399–435. <https://doi.org/10.1007/s10290-022-00471-4>.
- [KUK24] KUKA Deutschland GmbH. 2024. „Cobots: Ihr Einstieg in die Automatisierung mit kollaborierenden Robotern.“ Zugriff am 21. Dezember 2024. <https://www.kuka.com/de-de/future-production/mensch-roboter-kollaboration/cobots>.
- [KuKu16] Kumar, Naveen und Kumar, Jyoti, Hrsg. 2016 *Measurement of Efficiency of Auditory Vs Visual Communication in HMI : A Cognitive Load Approach // 2016 International Conference on Advances in Human Machine Interaction (HMI): 3-5 March 2016*. Unter Mitarbeit von Kumar, Naveen und Kumar, Jyoti. Piscataway, NJ: IEEE. 978-1-4673-8810-8.

- [LFK14] Lasi, Heiner, Peter Fettke, Hans-Georg Kemper, Thomas Feld und Michael Hoffmann. 2014. „Industrie 4.0.“ *Wirtschaftsinf* 56 (4): 261–64. <https://doi.org/10.1007/s11576-014-0424-4>.
- [Lau18] Laubheimer, Page. 2018. „Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests.“ Zugriff am 6. Juni 2025. <https://www.nngroup.com/articles/measuring-perceived-usability/>.
- [LFH17] Lazar, Jonathan, Jinjuan Heidi Feng und Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Second edition. Cambridge, MA: Elsevier Morgan Kaufmann Publishers. 978-0-12-805390-4.
- [LWL06] Li, Yang, Evan Welbourne und James A. Landay. 2006. „Design and experimental analysis of continuous location tracking techniques for Wizard of Oz testing.“ In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, hrsg. von Rebecca Grinter, Thomas Rodden, Paul Aoki, Ed Cutrell, Robin Jeffries und Gary Olson, 1019–22. New York, NY, USA: ACM.
- [LRP22] Linares-Garcia, Daniel Antonio, Nazila Roofigari-Esfahan, Kristina Pratt und Myounghoon Jeon. 2022. „Voice-Based Intelligent Virtual Agents (VIVA) to Support Construction Worker Productivity.“ *Automation in Construction* 143:104554. <https://doi.org/10.1016/j.autcon.2022.104554>.
- [LWL21] Liu, Yan, Wenqing Wang und Yanwei Li. 2021 - 2021. „Realization of Contactless Elevator Control Panel System Based on Voice Interaction Technology.“ In *2021 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, 591–94: IEEE.
- [LoDr20] Lodgaard, Eirin und Sebastian Dransfeld. 2020. „Organizational aspects for successful integration of human-machine interaction in the industry 4.0 era.“ *Procedia CIRP* 88:218–22. <https://doi.org/10.1016/j.procir.2020.05.039>.
- [LoPa20] Longo, Francesco und Antonio Padovano. 2020. „Voice-enabled Assistants of the Operator 4.0 in the Social Smart Factory: Prospective role and challenges for an advanced human-machine interaction.“ *Manufacturing Letters* 26:12–16. <https://doi.org/10.1016/j.mfglet.2020.09.001>.
- [Lor20] Lorenz, Sebastian. 2020. „Neue Dimensionen von Mensch-Maschine-Interfaces: Entwicklung eines Scoring-Systems zur Beschreibung und Evaluation von Mensch-Maschine-Interfaces für digitalisierte industrielle Anwendungen.“ *Design Research 2020 Kolloquium Technisches Design*. <https://d-nb.info/123423906X/34>. Zugriff am 6. Juni 2025.
- [Lot12a] Lotter, Bruno. 2012. „Einführung.“ In *Montage in der industriellen Produktion: Ein Handbuch für die Praxis ; mit 18 Tabellen*, hrsg. von Bruno Lotter. 2. Aufl., 1–8. VDI-/Buch]. Berlin, Heidelberg: Springer Vieweg.
- [Lot12b] Lotter, Bruno, Hrsg. 2012. *Montage in der industriellen Produktion: Ein Handbuch für die Praxis ; mit 18 Tabellen*. 2. Aufl. VDI-/Buch]. Berlin, Heidelberg: Springer Vieweg. 978-3-642-29060-2.
- [LBP21] Lüdecke, Daniel, Mattan Ben-Shachar, Indrajeet Patil, Philip Waggoner und Dominique Makowski. 2021. „performance: An R Package for Assessment, Comparison and Testing of Statistical Models.“ *JOSS* 6 (60): 3139. <https://doi.org/10.21105/joss.03139>.
- [Mäh95] Mählick, Heiner. 1995. *Herausforderung Lean Production: Möglichkeiten zur wettbewerbsgerechten Erneuerung von Unternehmen*. 2., überarb. Aufl. Düsseldorf: VDI-Verl. 3184014398.

- [MLK20] Mahmud, Saifuddin, Xiangxu Lin und Jong-Hoon Kim. 2020. „Interface for Human Machine Interaction for assistant devices: A Review.“ In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 768–73: IEEE.
- [MBP20] Makowski, Dominique, Mattan Ben-Shachar, Indrajeet Patil und Daniel Lüdecke. 2020. „Estimation of Model-Based Predictions, Contrasts and Means.“ Zugriff am 6. Juni 2025. <https://github.com/easystats/modelbased>.
- [MIK17] Mautua, Iñaki, Aitor Ibarguren, Johan Kildal, Loreto Susperregi und Basilio Sierra. 2017. „Human–robot collaboration in industrial applications.“ *International Journal of Advanced Robotic Systems* 14 (4): 1-10. <https://doi.org/10.1177/1729881417716010>.
- [MLA22] Mertes, Jan, Daniel Lindenschmitt, Masoud Amirrezai, Nima Tashakor, Moritz Glatt, Christian Schellenberger, Swati Matwankar Shah et al. 2022. „Evaluation of 5G-capable framework for highly mobile, scalable human-machine interfaces in cyber-physical production systems.“ *Journal of Manufacturing Systems* 64:578–93. <https://doi.org/10.1016/j.jmsy.2022.08.009>.
- [MTM19] MTM Association e. V. 2019. „MTM-UAS Basic Operations.“ Zugriff am 6. Juni 2025. <https://mtm.org/ueber-mtm/datenkarten>.
- [MVG15] Müller, R., M. Vette und A. Geenen. 2015. „Potentiale der Mensch-Roboter-Kooperation in der Flugzeugproduktion*/Potential of the human-robot-cooperation in aircraft assembly.“ *wt* 105 (09): 614–18. <https://doi.org/10.37544/1436-4980-2015-09-56>.
- [NMR22] N, Gayathri, Sundar M, Sargurunathan R, Sudharsan R und Sajith A. (2022) 2022. „Design of Voice Controlled Multifunctional Computer Numerical Control (CNC) Machine.“ In *2022 International Conference on Inventive Computation Technologies (ICICT)*, 657–63: IEEE.
- [Nat22] National Aeronautics and Space Administration. 2022. „NASA TLX: Paper and Pencil Version.“ Zugriff am 18. Mai 2025. <https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf>.
- [NiGo12] Nicolas-Alonso, Luis Fernando und Jaime Gomez-Gil. 2012. „Brain Computer Interfaces, a Review.“ *Sensors (Basel, Switzerland)* 12 (2): 1211–79. <https://doi.org/10.3390/s120201211>.
- [Nie05] Nielsen, Jakob. 2005. „Time Budgets for Usability Sessions.“ Zugriff am 26. Februar 2023. <https://www.nngroup.com/articles/time-budgets-for-usability-sessions/>.
- [NAL24] Norda, Marvin, Jens-E. Appell, Carsten Sven Lange und Axel Hahn. 2024. „Methodik zum effizienten Einsatz von Sprachsteuerung/Methodology for the efficient use of voice control.“ *wt* 114 (03): 66–70. <https://doi.org/10.37544/1436-4980-2024-03-38>.
- [NER23] Norda, Marvin, Christoph Engel, Jan Rennies, Jens-E. Appell, Sven Carsten Lange und Axel Hahn. 2023. „Evaluating the Efficiency of Voice Control as Human Machine Interface in Production.“ *IEEE Trans. Automat. Sci. Eng.*, Nr. 3: 4817-4828. <https://doi.org/10.1109/TASE.2023.3302951>.
- [NVE23] Norda, Marvin, Torben Veenhuis, Christoph Engel, Jens-E. Appell und Axel Hahn. 2023. „Effizienzsteigerung: Laufwegreduzierung durch Sprachsteuerung.“ *ZWF* 118 (10): 693–98. <https://doi.org/10.1515/zwf-2023-1132>.
- [Nzu21] Nzuva M., Silas. 2021. „Investigating Human-Machine Interfaces’ Efficiency in Industrial Machinery and Equipment: A Kenyan Context.“ *IJITCA* 11 (4): 1–8. <https://doi.org/10.5121/ijitca.2021.11401>.

- [Oeh00] Oehlert, Gary W. 2000, pp. 13-30 *A First Course in Design and Analysis of Experiments*. 1st print. New York, NY: Freeman. <http://www.loc.gov/catdir/enhancements/fy0811/99059934-d.html>. 978-0716735106.
- [PLK24] Panter, Lars, Rieke Leder, Dennis Keiser und Michael Freitag. 2024. „Requirements for Human-Machine-Interaction Applications in Production and Logistics within Industry 5.0 – A Case Study Approach.“ *Procedia Computer Science* 232:1164–71. <https://doi.org/10.1016/j.procs.2024.01.114>.
- [PPK15] Peddinti, Vijayaditya, Daniel Povey und Sanjeev Khudanpur. (2015) 2015. „A time delay neural network architecture for efficient modeling of long temporal contexts.“ In *Interspeech 2015*, 3214–18. ISCA: ISCA.
- [PeHi13a] Peissner, Matthias und Cornelia Hipp. (2013) 2013. „Herausforderungen und Lösungsansätze der Mensch-Technik Interaktion für die Produktion.“ In *Potenziale der Mensch-Technik-Interaktion für die effiziente und vernetzte Produktion von morgen*, hrsg. von Dieter Spath, Anette Weisbeckerr, Matthias Peissner und Cornelia Hipp, 38–65. Stuttgart: Fraunhofer-Verl.
- [PeHi13b] Peissner, Matthias und Cornelia Hipp. (2013) 2013. „Zusammenfassung.“ In *Potenziale der Mensch-Technik-Interaktion für die effiziente und vernetzte Produktion von morgen*, hrsg. von Dieter Spath, Anette Weisbeckerr, Matthias Peissner und Cornelia Hipp, 5–6. Stuttgart: Fraunhofer-Verl.
- [PfKa17] Pfister, Beat und Tobias Kaufmann. 2017 *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. 2., aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer Vieweg. 978-3-662-52838-9.
- [PKO21] Pizzagalli, S. L., V. Kuts und T. Otto. 2021. „User-centered design for Human-Robot Collaboration systems.“ *IOP Conf. Ser. Mater. Sci. Eng.* 1140 (1): 12011. <https://doi.org/10.1088/1757-899X/1140/1/012011>.
- [PPG16] Povey, Daniel, Vijayaditya Peddinti, Daniel Galvez, Pegah Ghahremani, Vimal Manohar, Xingyu Na, Yiming Wang und Sanjeev Khudanpur. (2016) 2016. „Purely Sequence-Trained Neural Networks for ASR Based on Lattice-Free MMI.“ In *Interspeech 2016*, 2751–55. ISCA: ISCA.
- [PYJ22] Priyadarshana, V.P.S.C, Y. M.P.G.R.R.Y. Yaparathne, D. B. S. Jayasooriya, T.M.T.S.D.B. Thennakoon, A.G.B.P. Jayasekara und D. P. Chandima. 2022. „Voice Controlled Robot Manipulator for Industrial Applications.“, 160–65. <https://doi.org/10.1109/IEMCON56893.2022.9946545>.
- [RCC21] R, Santhoshikka, Laranya C R und Harshavarthini C. 2021. „Eye Tracking and Its Applications.“ *International Advanced Research Journal in Science, Engineering and Technology* 8 (8). <https://doi.org/10.17148/IARJSET.2021.8824>.
- [REF25] REFA AG. 2025. „Produktionsprozesse: Definition.“ Zugriff am 6. Juni 2025. <https://refa.de/service/refa-lexikon/produktionsprozesse>.
- [REF21] *REFA-Grundausbildung 4.0: Begriffe und Formeln*. 2021. 1. Auflage. Hanser eLibrary Band 3. München: Carl Hanser Verlag. 9783446471788.
- [RPS22] Rendyansyah, Rendyansyah, Aditya P. P. Prasetyo und Sarmayanta Sembiring. 2022. „Voice Command Recognition for Movement Control of a 4-DoF Robot Arm.“ *ELKHA* 14 (2): 118. <https://doi.org/10.26418/elkha.v14i2.57556>.
- [RBR20] Rogowski, Adam, Krzysztof Bieliszczyk und Jerzy Rapcewicz. 2020. „Integration of Industrially-Oriented Human-Robot Speech Communication and Vision-Based Object Recognition.“ *Sensors (Basel, Switzerland)* 20 (24): 1–23. <https://doi.org/10.3390/s20247287>.

- [Saf08] Saffer, Dan. 2008 *Designing Gestural Interfaces: Touchscreens and Interactive Devices*. 1. ed. Beijing, Kön: O'Reilly. <http://swb.ebibli.com/patron/FullRecord.aspx?p=443463.9780596554224>.
- [SRA24] Sami, Ali, Khosro Rezaee, Mojtaba Ansari, Mohamadreza Khosravi und Vahid Karimi. 2024. „Review of Brain-Computer Interface Applications in Neurological Disorders.“ In *Proceedings of the Second International Conference on Computing, Communication, Security and Intelligent Systems*, hrsg. von Shahid Mumtaz, Danda B. Rawat und Varun G. Menon, 391–407. Algorithms for Intelligent Systems. Singapore: Springer Nature Singapore.
- [ScSc14] Schuh, Günther und Carsten Schmidt. 2014 *Produktionsmanagement*. Berlin, Heidelberg: Springer Berlin Heidelberg. 978-3-642-54287-9.
- [SSK21] Schulte, Daniel, Martin Sudhoff und Bernd Kuhlenkötter. 2021. „Spracherkennung in der manuellen Montage/Speech recognition in manual assembly.“ *wt* 111 (09): 579–82. <https://doi.org/10.37544/1436-4980-2021-09-5>.
- [SBS94] Someren van, Maarten W., Yvonne F. Barnard und Jacobijn A. C. Sandberg. 1994 *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. London: Academic Press. 0-12-714270-3.
- [SKM18] Spille, Constantin, Birger Kollmeier und Bernd T. Meyer. 2018. „Comparing human and automatic speech recognition in simple and complex acoustic scenes.“ *Computer Speech & Language* 52:123–40. <https://doi.org/10.1016/j.csl.2018.04.003>.
- [SSS16] Suarez Fernandez, Ramon A., Jose Luis Sanchez-Lopez, Carlos Sampedro, Hriday Bavle, Martin Molina und Pascual Campoy. 2016. „Natural user interfaces for human-drone multi-modal interaction.“, 1013–22. <https://doi.org/10.1109/ICUAS.2016.7502665>.
- [Sys06] Syska, Andreas. 2006 *Produktionsmanagement: Das A - Z wichtiger Methoden und Konzepte für die Produktion von heute*. 1. Aufl. Wiesbaden: Gabler. 978-3-8349-0235-1.
- [TAG17] Thommen, Jean-Paul, Ann-Kristin Achleitner, Dirk Ulrich Gilbert, Dirk Hachmeister und Gernot Kaiser. 2017 *Allgemeine Betriebswirtschaftslehre: Umfassende Einführung aus managementorientierter Sicht*. 8., vollständig überarbeitete Auflage. Lehrbuch. Wiesbaden: Springer Gabler. <http://www.springer.com/>. 978-3-658-07767-9.
- [TDK21] Tropschuh, Barbara, Fabian Dillinger, Svenja Korder, Maria Maier, Quirin Gärtner und Susanne Vernim. 2021. „Industrie 5.0 – ein menschenzentrierter Ansatz.“ *Zeitschrift für wirtschaftlichen Fabrikbetrieb* 116 (6): 387–92. <https://doi.org/10.1515/zwf-2021-0091>.
- [Tsc83] Tschätsch, Heinz. 1983. „Bestimmung der Fertigungszeit.“ In *Praktische Betriebslehre*, hrsg. von Heinz Tschätsch, 36–55. Wiesbaden: Vieweg+Teubner Verlag.
- [VaBo16] Vajpai, Jayashri und Avnish Bora. (2016) 2016. „Industrial Applications of Automatic Speech Recognition Systems.“ In *Industrial Applications of Automatic Speech Recognition Systems*, hrsg. von Int. Journal of Engineering Research and Applications, 88–95 6.
- [Vee23] Veenhuis, Torben. 2023. „Entwicklung einer Methodik zur systematischen Optimierung Entwicklung einer Methodik zur systematischen Optimierung von Laufwegen in der Produktion mithilfe von Sprachsteuerung.“ Unveröffentlichtes Manuskript. Bachelorarbeit.
- [Vid73] Vidal, J. J. 1973. „Toward Direct Brain-Computer Communication.“ *Annual review of biophysics and bioengineering* 2:157–80. <https://doi.org/10.1146/annurev.bb.02.060173.001105>.

- [VSL21] Villani, Valeria, Lorenzo Sabattini, Frieder Loch, Birgit Vogel-Heuser und Cesare Fantuzzi. 2021. „A General Methodology for Adapting Industrial HMIs to Human Operators.“ *IEEE Trans. Automat. Sci. Eng.* 18 (1): 164–75. <https://doi.org/10.1109/TASE.2019.2941541>.
- [VoBu08] Vorley, Geoff und Mary-Clare Bushell. 2008. „Mini Guide to: Root Cause Analysis.“ Zugriff am 9. Oktober 2024. <https://www.root-cause-analysis.co.uk/images/Green%20RCA%20mini%20guide%20v5%20small.pdf>.
- [WWY21] Wang, Zhimin, Haofei Wang, Huangyue Yu und Feng Lu. 2021. „Interaction With Gaze, Gesture, and Speech in a Flexibly Configurable Augmented Reality System.“ *IEEE Trans. Human-Mach. Syst.* 51 (5): 524–34. <https://doi.org/10.1109/THMS.2021.3097973>.
- [YKK21] Yadav, Sudeept Singh, Praveen Kumar, Sanil Kumar und Sanjeet Singh. 2021 - 2021. „Google Assistant Controlled Home Automation with Voice Recognition.“ In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 1039–44: IEEE.

Hinweise zu bereits veröffentlichten Inhalten

Die Inhalte dieser Arbeit, die mit den folgenden Literaturangaben versehen sind, wurden im Vorhinein bereits peer-reviewed in den jeweiligen wissenschaftlich-technischen Journalen veröffentlicht.

- [NAL24] Norda, Marvin, Jens-E. Appell, Carsten Sven Lange und Axel Hahn. 2024. „Methodik zum effizienten Einsatz von Sprachsteuerung/Methodology for the efficient use of voice control.“ *wt* 114 (03): 66–70. <https://doi.org/10.37544/1436-4980-2024-03-38>.
- [NER23] Norda, Marvin, Christoph Engel, Jan Rennies, Jens-E. Appell, Sven Carsten Lange und Axel Hahn. 2023. „Evaluating the Efficiency of Voice Control as Human Machine Interface in Production.“ *IEEE Trans. Automat. Sci. Eng.*, Nr. 3: 4817-4828. <https://doi.org/10.1109/TASE.2023.3302951>.
- [NVE23] Norda, Marvin, Torben Veenhuis, Christoph Engel, Jens-E. Appell und Axel Hahn. 2023. „Effizienzsteigerung: Laufwegreduzierung durch Sprachsteuerung.“ *ZWF* 118 (10): 693–98. <https://doi.org/10.1515/zwf-2023-1132>.

© 2023 IEEE. Reprinted, with permission, from Marvin Norda, Christoph Engel, Jan Rennies, Jens-E. Appell, Sven Carsten Lange und Axel Hahn. 2023. „Evaluating the Efficiency of Voice Control as Human Machine Interface in Production.“ *IEEE Trans. Automat. Sci. Eng.*, Nr. 3: 4817-4828. <https://doi.org/10.1109/TASE.2023.3302951>.

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Carl von Ossietzky Universität Oldenburg’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Abschließende Erklärung

Ich versichere hiermit, dass ich die Dissertation »*Entwicklung einer Methode zur Evaluation der Effizienz von Sprachsteuerung als industrielles Human-Machine Interface in der Produktion*« selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt sowie die allgemeinen Prinzipien guter wissenschaftlicher Praxis der Carl-von-Ossietzky Universität befolgt habe.

Oldenburg, den 23.09.2025

Marvin Leon Norda