

Technologies and Education for a Smart World

**Proceedings of the 8th FDIBA Conference
Sofia, Bulgaria, 30 November 2018**

FDIBA Conference Proceedings

Volume 2, November 2018

ISSN 2535–132X (Print)

ISSN 2535–1338 (Online)

Technologie und Ausbildung für Smart World

**Tagungsband der 8. FDIBA Konferenz
Sofia, Bulgarien, 30. November 2018**

Tagungsbände der FDIBA Konferenz

Volumen 2, November 2018

ISSN 2535–132X (Gedruckt)

ISSN 2535–1338 (Online)



Technische Universität Sofia
Karlsruher Institut für Technologie
Hochschule für Ökonomie und Management Essen
Technische Universität Braunschweig
Otto von Guericke Universität Magdeburg

With the financial support of the German Academic Exchange Service



Mit der Förderung des Deutschen Akademischen Austauschdienstes

Technical Program Committee

Chair: Alexandar Tsenov

Thomas Heupel
Klaus Tönnies
Peter Lang
Jivka Ovcharova
Hristomir Yordanov

Organisation Committee

Chair: Hristomir Yordanov

Alexandar Tsenov
Yordan Yordanov
Evegniy Petrov
Marin Marinov

Reviewers

Dimo Chotrov
Vasil Galabov
Vladislav Ivanov
Sanal Purgic
Nikola Rantchev
Uwe Siart
Irina Topalova
Aleksandar Tsenov
Magdalina Uzunova
Hristomir Yordanov
Sebastian Zug

Technologies and Education for a Smart World
Proceedings of the 8th FDIBA Conference, Sofia, Bulgaria, 30 November, 2018
FDIBA Conference Proceedings, 2018, Volume 2
Alexandar Tsenov, Hristomir Yordanov (Eds.)
Book size A4-ISO (210 × 297 mm), 72 pages
Paperback
Published by TU-Sofia Publishing House
Printed by Accent 96
Sofia, Bulgaria
April 2019
ISSN 2535-132X (Printed)
ISSN 2535-1338 (Online)

FDIBA Conference Program

30. 11. 2018

Venue

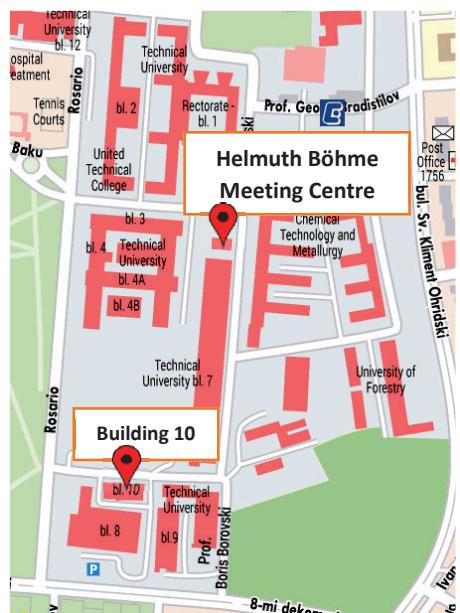
The conference will be held in two locations in the campus of the Technical University of Sofia. The first is the conference room of the Helmut Böhme Meeting Centre (IBZ). The opening ceremony, the PhD Students session and the Session on Mechanical Engineering, Mechatronics and Economics will take place there. The sessions on Computer Science will be held in room 10100 in Building 10.

The Banquet will take place in the V Parka hall (vparka.com), located in Bulgaria square 1 (The Small Palace of Culture).

Wireless Internet Access

Free Wi-Fi will be provided to all conference participants during the event at both locations. The name of the network is FDIBA Conference 2018. Please take into consideration that data is not encrypted over this network and using only secure protocols is recommended. You can also use the eduroam network if your institution participates in eduroam (not available in Helmut Böhme Meeting Centre).

Friday, 30. 11. 2018



The Helmut Böhme International Meeting Centre (IBZ)	
9:00	Official Opening Ceremony
9:30	Plenary Talk Prof. Zwetelina Gankova - Ivanova Erneuerbare Energiequellen und Beschäftigung

Room	10100	IBZ
10:30	Session INF1 Chair: Velko Ilchev	Svetla Ivanova-Vassileva Hygiene Factors of Usability (Precondition for GUI Acceptability)
10:50		Svetla Ivanova-Vassileva Associativity of Visual Metaphors and User Performance (Study of a Banking Application)
11:10		Vanya Markova, Ventseslav Shopov Comparative Analysis of Trajectory Controllers for Noncholoniomic Mobile Robot
11:30		Ventseslav Shopov, Vanya Markova Identification of Non-linear dynamic system
11:50		Alek Borisov, Nikolay Belobradov Graph Laplacian
12:10		Lunch break
13:30	Session INF2 Chair: Ventseslav Shopov	Angel Garabitov, Aleksandar Tsenov Using Formant as an indication of speech intelligibility in VOIP
13:50		Daniel Hristov Application of Hilbert Transform for Phased Antenna Array Calibration
14:10		Ina Nikolova-Jahn, Emil Enchev Principles and consistency in automation of logistic information processes
14:30		Emil Enchev Moderne Logistikinformationsysteme
14:50		Blagovest Kirilov Utilizing Blockchain technology for data analysis and interaction through The Formal Method of Things
15:30		Hashem Badra, Jivka Ovtcharova Navigation System for Virtual Manufacturing Process
18:30		Georgi Yanachkov, Simona Hesapchieva Integral Brake System for Electric Motorcycle Using Maximum Energy Regeneration
		Stefan Hensel, Marin B. Marinov, Christoh Kehret Implementation of a system for testing of mapping algorithms
		Marin Marinov, Ivan Topalov Location based calibration of low cost particulate matter sensors
		Anna-Maria Staiger, Malte Fiedler, Wolfgang Fritz, Dietrich von der Oelsnitz Competence in the Context of Digitalization
		Official closing ceremony
		BANQUET

THIS PAGE IS INTENTIONALLY LEFT BLANK

Contents

Plenary talk	1
Renewable Energy Sources and Employment Zwetelina Gankova-Ivanova	1
Session Informatics	3
“Hygiene Factors” of Usability (Precondition for GUI Acceptability) Svetla Ivanova-Vassileva	5
Associativity of Visual Metaphors and User Performance (Study of a Banking Application) Svetla Ivanova-Vassileva	9
Informational aspects of modelling the Health Regions in Bulgaria Velko Iltchev, Aleksandar Tsenov	13
Methods for Eventual Consistency in Collaborative Editing Systems Boyan Katsarski, Velko Iltshev	17
Mixed-Reality Spatial Configuration with a ZED Mini Stereoscopic Camera Dimo Chotrov, Zlatka Uzunova, Yordan Yordanov, Stoyan Maleshkov	21
Prediction of EIGRP Traffic Parameters by 3D Approximation Using Neural Networks Irina Topalova	25
Session Mechatronics	29
Navigator System for Virtual Manufacturing Process Hashem Badra, Jivka Ovtcharova	31
Integral Brake System for Electric Motorcycle Using Maximum Energy Regeneration Georgi Yanachkov, Simona Hesapchieva	35
Design of a mobile platform for the evaluation of Localization and Mapping Algorithms Stefan Hensel, Marin B. Marinov, Christoh Kehret	39
Location based calibration of low cost particulate matter sensors Ivan P. Topalov	43
Session Economics	47
Competence of Employees in the Context of Digitalization Anna-Maria Staiger, Malte Fiedler, Wolfgang Fritz, Dietrich von der Oelsnitz	49
Session Student Papers	53
Using Formant as an indication of speech intelligibility in VOIP Angel Garabitov, Aleksandar Tsenov	55
Application of Hilbert Transform for Phased Antenna Array Calibration Daniel Hristov	59
Utilizing Blockchain technology for data analysis and interaction through The Formal Method of Things Blagovest Kirilov	63

THIS PAGE IS INTENTIONALLY LEFT BLANK

Renewable Energy Sources and Employment

Erneuerbare Energiequellen und Beschäftigung

Plenary talk

Zwetelina Gankova – Ivanova

Faculty of Economics, Technical University of Gabrovo
Stefan Karadja Str. 9/24, 5300 Gabrovo, Bulgaria, e-mail: zwetelina_gankova_7@yahoo.de

ABSTRACT

In the 21st century, energy supply is undergoing significant structural change, which is a major challenge. This is the sustainable transformation of the energy system, which means reduced energy consumption, greater energy efficiency and increased use of renewable energies. In this context, clear targets have been set in Bulgaria, covering at least 10

This talk deals with this topic with the aim of answering the question of whether the increased use of renewable energies and the creation of new jobs can be permanently harmonized and whether a positive employment impulse exists.

From an energy and environmental perspective, both the current and future development of renewable energy sources is no longer a matter of dispute not only at national but also at international level. In view of the fact that financial support in this area is needed, as has already been mentioned, certain conflict of objectives could arise with other policy and economic sectors. After all, it should be emphasized that the increased use of renewable energy sources can have a positive effect on high unemployment by reducing it through job creation.

At present, the impact of developing renewable energy on the Bulgarian labour market is of great interest. In this sense, it is necessary to answer the question of whether the use of renewable energy in an overall economic perspective ultimately leads to an increase in employment, or to losing jobs. This debate is due, on the one hand, to the complexity of the issue and, on the other hand, to the lack of data in some areas. In addition, various assumptions about future development play a crucial role, because different interpretations of the macroeconomic effects that could increase the use of renewable energies are allowed.

Employment effects can be defined by two terms. A distinction must be made between direct and indirect employment. Direct employment results from investment in

equipment and its operation with manufacturers, operators and service companies. In order to fulfill orders in the form of goods, goods and services from other economic sectors, indirect employment is created in subcontracting and subcontracting companies. The result of both types of employment is the so-called gross employment, which must always be considered positive. However, any negative employment effects due to macroeconomic analysis are taken into account. In this respect, the so-called net employment effect should also be analyzed as a balance of all effects because the result could be not only positive but also negative.

The gross employment can be determined within a given framework, but the net employment effect can be calculated as the difference between two realistic, future scenarios. If the effect is positive, actual additional employment will result in the increased expansion of renewable energies.

Renewable energy products—electricity, heat and fuel—will generally replace fossil fuels. In this context, it is necessary to establish whether the exploitation of biomass, hydro, wind, solar and geothermal resources will lead to a negative or positive impact on investment and therefore on employment not only in the energy sector but also in other sectors. It is not negligible that the public or private budgets are burdened, because the renewable energy is not yet cost-efficient and competitive compared to the conventional. The resulting differential costs are borne by the energy consumers. However, these consumers have no capacity to ask for more goods and this budget effect will lead to negative employment effects in other sectors, which should also be taken into account in a net balance.

It should also be emphasized here that the impact on employment in the area of foreign trade is not easy to forecast and therefore needs to be analyzed, because the volume of imported and exported goods, goods and services has a significant impact on domestic employment. In the future, this aspect will also be given a high priority.

KURZFASSUNG

Im 21. Jahrhundert unterliegt die Energieversorgung einem bedeutenden Strukturwandel, was eine große Herausforderung darstellt. Dabei handelt es sich um die nachhaltige Umgestaltung des Energiesystems, worunter reduzierter Energieverbrauch, höhere Energieeffizienz und verstärkter Einsatz erneuerbarer Energien zu verstehen sind. In diesem Zusammenhang sind klare Ziele in Bulgarien gesetzt worden, wobei mindestens 10

Der Vortrag beschäftigt sich mit diesem Thema und dabei wird zum Ziel gesetzt, die Frage zu beantworten, inwieweit die verstärkte Nutzung erneuerbarer Energien und die Schaffung neuer Arbeitsplätze dauerhaft aufeinander abgestimmt werden können und ob ein positiver Beschäftigungsimpuls nachhaltig existieren wird.

Aus Energie- und umweltpolitischer Sicht ist sowohl der bisherige als auch der künftige Ausbau erneuerbarer Energiequellen keine Streitfrage mehr nicht nur auf nationaler, sondern auch auf internationaler Ebene. Da finanzielle Förderung in diesem Bereich erforderlich wird, wie dies bereits erwähnt wurde, könnten gewisse Zielkonflikte mit anderen Politik- und Wirtschaftssektoren entstehen. Immerhin ist es zu betonen, dass die verstärkte Nutzung erneuerbarer Energieträger positive Effekte auf die hohe Arbeitslosigkeit haben kann, indem sie durch die Schaffung von Arbeitsplätzen reduziert wird.

Derzeit sind die durch den Ausbau erneuerbarer Energien entstehenden Auswirkungen auf den bulgarischen Arbeitsmarkt von großem Interesse. In diesem Sinne ist es erforderlich, die Frage zu beantworten, ob die Nutzung erneuerbarer Energien bei einer gesamtwirtschaftlichen Betrachtung letztendlich zu einer Zunahme von Beschäftigung führt, oder ob dadurch Arbeitsplätze verloren gehen. Diese Debatte ist einerseits auf die Komplexität des Sachverhalts, andererseits auf die unzureichenden Daten in einigen Bereichen zurückzuführen. Ferner spielen diverse Annahmen über die künftige Entwicklung eine entscheidende Rolle, weil unterschiedliche Interpretationen in Bezug auf die gesamtwirtschaftlichen Wirkungen zulässig sind, welche die verstärkte Nutzung erneuerbarer Energien hervorrufen könnten.

Die Beschäftigungseffekte lassen sich durch zwei Begriffe definieren. Dabei ist zwischen einer direkten und einer indirekten Beschäftigung zu unterscheiden. Die *direkte Beschäftigung* ergibt sich aus Investitionen in Anlagen und

deren Betrieb bei Herstellern, Betreibern und Dienstleistungsunternehmen. Damit Aufträge in Form von Waren, Gütern und Dienstleistungen aus anderen Wirtschaftssektoren erfüllt werden können, wird eine *indirekte Beschäftigung* in Vorleistungs- und Zulieferunternehmen geschaffen. Als Ergebnis beider Beschäftigungsarten resultiert die sog. *Bruttobeschäftigung*, die immer als positiv zu beurteilen ist. Jedoch sind eventuelle negative Beschäftigungseffekte aufgrund einer gesamtwirtschaftlichen Analyse zu berücksichtigen. In dieser Hinsicht ist auch der so genannte *Nettobeschäftigungseffekt* als eine Bilanz aller Effekte zu analysieren, weil das Ergebnis nicht nur positiv, sondern auch negativ ausfallen könnte.

Die Bruttobeschäftigung kann innerhalb eines vorgegebenen Rahmens bestimmt werden, jedoch lässt sich der Nettobeschäftigungseffekt als Differenz zweier realistischer, zukünftiger Szenarien ermitteln. Ist der Effekt positiv, entsteht somit die tatsächliche Mehrbeschäftigung im verstärkten Ausbau erneuerbarer Energien.

Die Produkte der erneuerbaren Energien – Elektrizität, Wärme und Kraftstoffe – werden in der Regel die fossilen Brennstoffe ersetzen. In diesem Zusammenhang ist es notwendig festzustellen, ob die Erschließung der Ressourcen von Biomasse, Wasser-, Wind- und Sonnenkraft sowie von Geothermie zu negativen oder zu positiven Auswirkungen auf die Investitionen und dadurch auf die Beschäftigung nicht nur im Energiesektor, sondern auch in anderen Wirtschaftszweigen führen wird. Es ist nicht zu vernachlässigen, dass die öffentlichen bzw. privaten Budgets belastet werden, weil die erneuerbaren Energien im Vergleich zu den konventionellen bislang noch nicht kosteneffizient und konkurrenzfähig sind. Die daraus resultierenden Differenzkosten sind von den Energieverbrauchern zu tragen. Diese Verbraucher verfügen aber über keine Kapazitäten, weitere Güter nachzufragen und dieser Budgeteffekt wird zu negativen Beschäftigungswirkungen in anderen Branchen führen, was bei einer Nettobilanz ebenfalls zu berücksichtigen ist.

Hier ist es auch hervorzuheben, dass der Einfluss auf die Beschäftigung im Bereich des Außenhandels nicht leicht zu prognostizieren und daher auch zu analysieren ist, weil das Volumen der ein- bzw. ausgeführten Waren, Güter und Dienstleistungen die Beschäftigung im Inland deutlich beeinflusst. Künftig wird es diesem Aspekt auch einen großen Stellenwert beigemessen.

Session Informatics

Chair: Velko Iltchev, Ventseslav Shopov

THIS PAGE IS INTENTIONALLY LEFT BLANK

“Hygiene Factors” of Usability (Precondition for GUI Acceptability)

Hygienefaktoren der Gebrauchstauglichkeit (Voraussetzung für Akzeptabilität der Schnittstellen von graphischen Oberflächen)

Svetla Ivanova-Vassileva

Engineering Design Department, Technical University of Sofia
Sofia, Bulgaria, vassileva@fdiba.tu-sofia.bg

Abstract — Present paper describes an approach to increase GUI acceptability. The main goal is to rank such GUI characteristics related to Ergonomics and Usability as metaphors' associativity, aesthetics, consistency, mapping, etc. according to their importance. Specific "hygiene factors of usability" and user motivators based on Herzberg's two-factor theory are summarized. In the investigation take part Master students in Graphic and Web Design and Engineering Design at the Technical University – Sofia.

Zusammenfassung — Der vorliegende Artikel beschreibt einen Ansatz zur Erhöhung der GUI-Akzeptabilität. Das Hauptziel besteht darin, solche GUI-Merkmale, die sich auf Ergonomie und Gebrauchstauglichkeit beziehen, wie Assoziativität der Metaphern, Ästhetik, Konsistenz, Mapping usw. entsprechend ihrer Wichtigkeit zu rangieren. Spezifische „Hygienefaktoren der Gebrauchstauglichkeit“ und Benutzermotivatoren wurden aufgrund der Zwei-Faktoren-Theorie von Herzberg zusammengefasst. An der Untersuchung nehmen Master-Studenten in Graphischem und Web-Design und Engineering Design an der Technischen Universität - Sofia teil.

I. EINFÜHRUNG

Im Mittelpunkt jeder Strategie zur Steigerung der Benutzerzufriedenheit steht die Untersuchung menschlicher Wünschen und Erwartungen. Es wird ein Ansatz zur Erhöhung der GUI-Akzeptabilität vorgestellt.

II. DEFINITION DER HYGIEINEFAKTOREN DER GEBRAUCHSTAUGLICHKEIT

A. Zwei-Faktoren-Theorie von Herzberg

Die vorliegende Untersuchung basiert auf Zwei-Faktoren-Theorie von Herzberg mit der Annahme, dass es zwei voneinander unabhängige Dimensionen der Arbeitszufriedenheit gibt: (1) Unzufriedenheit / Nicht-Unzufriedenheit und (2) Zufriedenheit / Nicht-Zufriedenheit. Bedingungen für eine Senkung der Unzufriedenheit als Übergang zur Nicht-Unzufriedenheit liegen v.a. im Arbeitsumfeld (Hygienefaktoren). Bedingungen, die einen Übergang von Nicht-Zufriedenheit zur Zufriedenheit auslösen können, liegen schwerpunktmäßig im Arbeitsinhalt selbst (Motivatoren) [1, 2]. Weil die Arbeitszufriedenheit vielmehr von intrinsischen Motivation abhängig ist und Faktoren wie Arbeitsinhalt relevanter sind, können zwei Ziele erreicht werden: die Unzufriedenheit bekämpfen und durch Steigerung der Zufriedenheit die Leistung fördern. Das kann aber nicht automatisch auf die digitale Welt übertragen werden. Auch hier hängen die *Hygienefaktoren* mit der Unzufriedenheit der Benutzer zusammen und die *Motivatoren* beeinflussen die Zufriedenheit. Hier kann man statt situative Bedingungen die wesentlichen, aus der Sicht der Softwareergonomie und der Gebrauchstauglichkeit, GUI-Charakteristiken bzw. deren Abwesenheit/Mängel als Hygienefaktoren betrachten. In den

weiteren Untersuchungen steht die Aufgabe, aufgrund des Wissens von Usability Engineering herauszufinden, die Mängel welcher GUI-Eigenschaften (z. B. Konsistenz, Mapping, Assoziativität der Sinnbilder u.a.) zur Missbilligung seitens Nutzer führt. So können solche Merkmale der graphischen Schnittstellen als „Hygienefaktoren der Gebrauchstauglichkeit“ bezeichnet werden, weil sie nicht unbedingt die positive Einstellung der Benutzer gegenüber dem System-GUI verursachen, ihre Abwesenheit aber führt unbestreitbar zur Unzufriedenheit der Benutzer.

B. Fragebogen Hygienefaktoren der Gebrauchstauglichkeit

Teilnehmer an der Untersuchung sind 20 Master-Studenten in Graphischem und Web-Design an der Technischen Universität – Sofia. Bei der Zusammenstellung der Liste der GUI-Merkmale sind im Zeitraum 2012-2018 folgende Objekte untersucht worden: die Web-Seite eines bulgarischen Unternehmens für Computerreparaturen, On-Line-Anwendung für Fragebogenverteilung und -auswertung und die interne Anwendung eines Unternehmens für Computerreparaturen.

Die Web-Seite (Abb. 1) ist von einem bulgarischen Unternehmen für Computerreparaturen [3]. Die Seite ist nach einer Suche in Google zufällig gewählt – Suchbegriff “Computer-Reparatur“ (Bemerkung: Suche in bulgarischer Sprache).

Die On-Line-Anwendung LimeSurvey (Abb. 2) ist ein Open-Source-Software für Fragebogenverteilung und -auswertung. Die Firmensoftware (Abb. 3) ist eine interne Anwendung eines Unternehmens für Computerreparaturen.

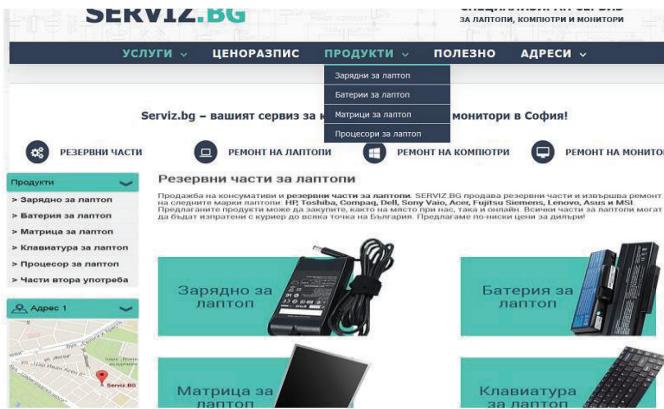


Abb. 1. GUI von Serviz.Bg

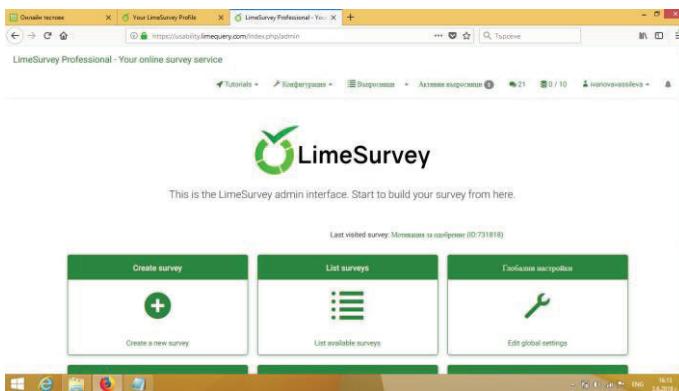


Abb. 2. GUI von LimeSurvey



Abb. 3. GUI von LAN Service.bg

Die gewählten GUI-Merkmale, die sich auf Ergonomie und Gebrauchstauglichkeit beziehen [5, 6, 7, 8, 9, 10, 11, 12, 13, 14], sind: Assoziativität der Metaphern, Ästhetik, Konsistenz, Mapping (Präsenz eines klaren Zusammenhangs zwischen Aktion und Folge), Individualisierbarkeit, klare Struktur, klare Begriffe, Funktionalität, Lesbarkeit und Farbgestaltung. Sie sind entsprechend ihrer Wichtigkeit auf die folgende Weise rangiert (Tabelle I). Die Studenten erfüllen im Rahmen des Semesters Aufgaben, gerichtet an die GUI-Gestaltung und -Evaluierung, und aufgrund dieser sammeln sie Erfahrung. Die letzte Semesteraufgabe ist als schon vorbereitete Usability-Spezialisten, die GUI-Merkmale nach ihrer Relevanz einzuschätzen.

Die in Tabelle I aufgelisteten Merkmale von gebrauchstauglichen GUIs können als Basis der so genannten "Hygienefaktoren der Gebrauchstauglichkeit" dienen. Ursprünglich ist die Liste aufgrund der Erfahrung, gesammelt während der Betreuung von praxisbezogenen studentischen Aufgaben in Softwareergonomie und Usability Engineering im Rahmen des Ingenieur-Design-Studiums für einen Zeitraum

von mehr als zehn Jahren, zusammengestellt. Diese Liste ist schon mehrmals revidiert und ergänzt worden.

TABELLE I. RANGIERTE MERKMALE VON GEBRAUCHSTAUGLICHEN GUIs

Benutzer-Bewertung (Rangstufe)	Expertengewichtung	Merkmale von gebrauchstauglichen GUIs
2,33	hoch	Assoziativität der Metaphern
3,28	hoch	klare Struktur
3,33	mittelmäßig	Lesbarkeit inkl. (passende Farbgestaltung)
3,69	hoch	Mapping
3,88	mittelmäßig	klare Begriffe
4,28	niedrig	Zugänglichkeit
4,50	hoch	Konsistenz
7,18	niedrig	Ästhetik
7,65	mittelmäßig	Funktionalitätsbreite
8,31	niedrig	Individualisierbarkeit

Im nächsten Schritt werden die wichtigsten Faktoren anhand einer Pareto-Analyse rangiert. Die Werte ergeben sich folgenderweise: es spielen Rolle die Wichtigkeit (entspricht dem Gewicht) und die Rangierung durch Benutzer (Testpersonen), d.h. Assoziativität der Metaphern bedeutet – vorwiegend an 1. bis 3. Stelle (2,33 - Mittelwert) und Merkmal mit hohem Gewichtung (hohe Bewertung der Experten). Dagegen ist Individualisierbarkeit mit an 8. bis 10. Stelle (8,31 - Mittelwert) und mit niedrigem Gewichtungswert der Experten. Unabhängig von seiner Stelle in der Rangliste ist jeder Faktor eine Voraussetzung für Akzeptabilität eines Systems.

Die ausgewählten Charakteristiken sind als grundlegende Empfehlungen betrachtet, wobei das Überschreiten einer bestimmten Schwelle zur Nicht-Akzeptanz der Benutzerschnittstelle führen könnte. Obwohl nach Herzberger Theorie als eigentliche Motivatoren bei der Nutzung eines Systems manche organisatorische Aspekte wie Zeiterparung, 24/7-Zugänglichkeit, finanzielle Vorteile u.a. genannt werden können, spielen eigentlich auch die Merkmale von gebrauchstauglichen GUIs auch wichtige Rolle für die Benutzerzufriedenheit.

TABELLE II. HYGIENEFAKTOREN DER GEBRAUCHSTAUGLICHKEIT

Hygienefaktoren	HF Bewertung
1 mangelnde Assoziativität der Metaphern	2,33
2 nicht klare Struktur	3,28
5 Unlesbarkeit inkl. (nicht passende Farbgestaltung)	6,66
3 Mangelndes Mapping	3,69
6 nicht klare Begriffe	7,76
7 mangelnde Zugänglichkeit	12,84
4 mangelnde Konsistenz	4,50
9 mangelnde Ästhetik	21,54
8 mangelnde Funktionalität	15,30
10 mangelnde Individualisierbarkeit	24,93

Die in Tabelle II angegebene Liste der GUI-Defizite, die entweder mit nicht vorhandenen Charakteristiken oder mit

mangelhafter Gestaltung dieser verbunden sind, wird als Hygienefaktoren der Gebrauchstauglichkeit bezeichnet. Die Gewichtung ist aufgrund der Rangierung in Tabelle I automatisch übertragen. Benutzerbewertung ergibt sich aus dem Mittelwert der Rangierung der GUI-Faktoren durch den Benutzer. Gewichtungskoeffiziente entsprechen der Expertenbewertung: 1 bei hoch, 2 bei mittelmäßig und 3 bei niedrig (Tabelle I). Hygiene-Faktoren-Bewertung (Tabelle II) wird nach der Formel (1) berechnet.

$$\text{Benutzerbewertung} * \text{Gewichtungskoeff.} = \text{HF Bewertung} \quad (1)$$

Andere mögliche Merkmale bzw. Vorschläge für Hygienefaktoren schließen ein: ob die GUI interaktiv ist und die Konformität der üblichen Funktionselemente gegenüber den allgemeinen Normen vorhanden ist.

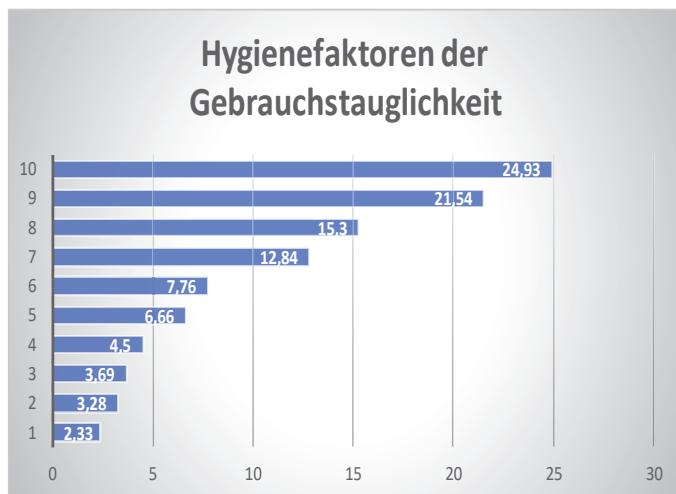


Abb. 4. „Hygienefaktoren“ der Gebrauchstauglichkeit

Diese Forschung basiert auf die momentane Aktualität jeder der GUI-Merkmale. In der Entwicklungsdynamik der Technologien, kann eins von denen ausfallen, anderes an Bedeutung verlieren, der überwiegende Anteil bleibt zumindest in den nächsten Jahren relevant.

III. ANFRAGE ÜBER DIE GUI-AKZEPTABILITÄT

Bei der Erforschung der „Hygienefaktoren der Gebrauchstauglichkeit“ aus der Sicht der Akzeptanz durch die Benutzer sind Master-Studenten in Graphischem und Web-Design befragt worden. Die Teilnehmer sind Leute mit unterschiedlichem Hintergrund aus der Bachelor-Stufe: die Hälfte kommt von der Ingenieur-Design-Fachrichtung und die anderen – aus der Informatik-Fachrichtung. Die Befragung wurde am Ende des Masterstudiums durchgeführt, als die Studenten schon verschiedene Fächer aus dem Bereich der Informatik und des Designs bzw. der Kunst behandelt haben. Die Fragebögen sind nicht anonym und bei der Bearbeitung der Ergebnisse wurde festgestellt, daß keine Tendenzen wegen des unterschiedlichen Hintergrundes beobachtet werden.

TABELLE III. TEILNEHMER AN DER FORSCHUNG

Anzahl	20
Geschlecht	10 Frauen 10 Männer
Alter	23,1
Ausbildung	Universitätsabschluß
Bachelor-Fachrichtung	10 Designer 10 Informatiker
GUI Erfahrung	Min. 5 Jahre

In Tabelle III sind Informationen über die befragten Personen angegeben. Sie sind nach Geschlecht und Fachrichtung gleich verteilt: die 50% Frauen und 50% Männer, die Hälfte Informatiker, die andere Hälfte – Designer mit Bachelor-Abschluß und mindestens fünfjährige GUI-Erfahrung.

Die Frageliste ist aufgrund der zusammengefassten Hygienefaktoren der Gebrauchstauglichkeit erstellt, wobei die Fragen auf verständliche Weise für den Benutzer formuliert sind. Die Fragen, die Hygienefaktoren (in Klammern und kursiv) und die Ergebnisse der Rangierung sind in Tabelle IV vorgestellt. An den ersten Stellen in Bezug der Benutzerakzeptanz sind klare Struktur, Assoziativität, Mapping und Konsistenz geordnet.

Die Gewichtung der Experten in Tabelle I entspricht der Ergebnisse dieser Anfrage.

TABELLE IV. UNTERSUCHUNG DER HYGIENEFAKTOREN DER GEBRAUCHSTAUGLICHKEIT (GUI-MERKMALE)

Fragen über GUI-Merkmale - Voraussetzung für die Definition der „Hygienefaktoren der Gebrauchstauglichkeit“	Rangierung
1. Erkennen Sie die Bedeutung der Sinnbilder und Symbole? (Assoziativität)	2
2. Sind der Bildschirm und seine Elemente ästhetisch und angenehm? (Ästhetik)	9
3. Lässt das System eigene Einstellungen zu? (Individualisierbarkeit)	8
4. Reagiert das System und sieht es gleich in allen Bestandteilen? (Konsistenz)	4
5. Erkennen Sie die Folgen Ihrer Aktion und die Systemreaktion? (Mapping)	3
6. Ist die Information klar verteilt und arrangiert? (klare Struktur)	1
7. Sind die Titel und Meldungen klar formuliert? (klare Begriffe)	5
8. Sind genug Funktionen vorhanden? (Funktionalität)	7
9. Ist die Information lesbar auch bezüglich der Farbgestaltung? (Lesbarkeit)	6
10. Sind visuelle Metaphern für die schnelle und sprachunabhängige Informationssuche vorhanden? (Zugänglichkeit)	10

Als Zusammenfassung ist wichtig zu betonen, daß Ergonomen inkl. Gebrauchstauglichkeitsexperten und Gestalter permanent streben, durch Evaluierungen und Entwicklung neuer Ansätze den Stress [15] zu minimieren, und demzufolge die Motivation und Leistung zu steigern. Weiterhinaus ist das Ziel, die Benutzerzufriedenheit zu erhöhen oder mindestens die Benutzerunzufriedenheit zu bekämpfen und dadurch GUI-Akzeptabilität zu erreichen. Deswegen ist die Definition spezifischer, wichtige für die Benutzer, GUI-Merkmale bzw. Hygienefaktoren der Usability ein Schritt vorwärts.

IV. AUSBLICK

In weiteren Forschungen werden die Hygienefaktoren der Gebrauchstauglichkeit weiter in Bezug zu konkreten Benutzerschnittstellen untersucht. Ein weiteres Ziel ist festzustellen, wie diese Faktoren als Grundlage minimaler Anforderungen bei der GUI-Gestaltung dienen können.

LITERATURVERZEICHNISS

- [1] Gabler *Wirtschaftslexikon*, <https://wirtschaftslexikon.gabler.de/definition/zweifaktorentheorie-48072/version-271330>, Maier, G. W., „Zweifaktorentheorie, Definition“, Oktober 2018
- [2] Schlick, Christopher M., Bruder, R., Lutzak, H., *Arbeitswissenschaft*, Springer Verlag, Berlin Heidelberg, 2010

- [3] Web-Seite von der Firma Serviz.bg <https://serviz.bg/>, Januar-März 2018
- [4] LimeSurvey®, Official web site of LimeSurvey, <http://limesurvey.org>, Januar-März 2018
- [5] Boehringer J., P. Buehler, P. Schlaich, *Kompendium der Mediengestaltung für Digital- und Printmedien*, Springer-Verlag Berlin Heidelberg, 2008
- [6] Galitz, W. O., *The Essential Guide to User Interface Design*. An Introduction to GUI Design Principles and Techniques, Wiley Publishing Inc., Indianapolis, 2007
- [7] Evtimova M. *Color interpretations*, (in Bulgarian language), Color Group – Bulgaria, Sofia, 2009, ISBN 978-954-92092-5-9
- [8] Nielsen, J. *Usability Engineering*. San Francisco, Morgan Kaufmann Publishers Inc., 1993
- [9] Dimitrova V., "Framework for implementing, evaluating and analyzing some algorithms for non-photorealistic image processing", Proceedings of the 7-th FDIBA Conference "The Digital Transformation: Challenges in Technological, Scientific and Social Development", 30 November - 1 December 2017, Sofia, Bulgaria, Vol. 1, ISSN 2535-1338
- [10] Preece, J., Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, *Human Computer Interaction*. Harlow, England: Addison-Wesley, 1994.
- [11] Richter M., M. Flückiger, *Usability Engineering kompakt*, Benutzbare Software gezielt entwickeln ISBN 978-3-8274-2328-3 2. Auflage 2010 Spektrum Akademischer Verlag Heidelberg (Spektrum Akademischer Verlag ist ein Imprint von Springer), 2010
- [12] Stapelkamp, Torsten, *Interaction- und Interfacedesign, Web-, Game-, Produkt- und Servicedesign*. Usability und Interface als Corporate Identity, Reihe: X.media.press, Springerverlag, 2011, ISBN 978-3-642-02074-2, Print version ISBN 978-3-642-02073-5
- [13] Zühlke, Detlef: *Nutzergerechte Entwicklung von Mensch-Maschine-Systemen: Useware-Engineering für technische Systeme*, 2. Auflage, Berlin 2012
- [14] International Standard Organization, ISO 9241-11 (2018) Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts
- [15] Vladkova, B.: "Stress Impact at the Working Place – Comparative Study for Bulgaria, Macedonia and Germany" - IV International Scientific and Technical Conference Engineering. Technologies. Education. Security, Veliko Tarnovo, 01-03 June 2016, ISSN: 1310 – 3946.

Associativity of Visual Metaphors and User Performance (Study of a Banking Application)

Assoziativität von GUI-Metaphern und Benutzerperformanz

(Studie über eine Bank-Anwendung)

Svetla Ivanova-Vassileva

Engineering Design Department, Technical University of Sofia
Sofia, Bulgaria, vassileva@fdiba.tu-sofia.bg

Abstract — The research includes two experiments with students in Master Courses Engineering Design and Graphical and web design at the Technical University of Sofia. First, the associativity of the icons for a bank application is evaluated. Students try to connect icons with their meanings. Next step is to investigate the influence of understandable or confusing visual metaphors on user performance. Test participants complete three tasks with different complexity.

Zusammenfassung — Die Forschung umfaßt zwei Experimente mit Studenten in Master-Studiengängen in Engineering Design und Graphischer und Web Design an der Technischen Universität - Sofia. Zunächst wird die Assoziativität der Icons für eine Bankanwendung ausgewertet. Studenten versuchen, Symbole mit ihrer Bedeutung zu verbinden. Der nächste Schritt besteht darin, den Einfluss verständlicher oder verwirrender visueller Metaphern auf die Benutzerleistung zu untersuchen. Testteilnehmer lösen drei Aufgaben mit unterschiedlicher Komplexität.

I. EINFÜHRUNG

In unserem Alltag mit steigendem Anteil der elektronischen Geräte sind die Benutzer durch den Umgang mit diversen GUIs beansprucht. Die Gestalter versuchen ständig die Herausforderungen der neuen digitalisierten Umgebung zu treffen. Das Ziel ist Verwirrungen und Desorientierung explicit bei älteren und unerfahrenen Benutzer von neuen Systemen vorzubeugen. Unabhängig davon, ob typische Webseiten, Selbstbedienungsautomaten, mobile Anwendungen oder Smart-Systeme in Frage kommen sind die HCI und Usability Empfehlungen [1, 2] und die Konformität mit den Normen [3, 4] zu erreichen.

Metaphern hätte einen negativen Effekt auf solche GUI-Merkmale wie die Assoziativität, Konsistenz und Konformität mit den Richtlinien und allgemeinen Benutzererwartungen.

Im ersten Schritt wurden die GUI-Sinnbilder von Studenten im Fach Softwareergonomie des Masterstudiums Graphischer und Web-Design bewertet (Tabelle 1).

TABELLE I. TEILNEHMER AM EXPERIMENT SINNBLIDERASSOZIATIVITÄT

Anzahl	20
Geschlecht	10 Frauen 10 Männer
Alter	23,1
Ausbildung	Universitätsabschluß
GUI Erfahrung	Min. 5 Jahre

Den Testpersonen wurden keine weitere Informationen außer den GUI-Metaphern und der zugehörigen Liste der wahrscheinlichen Sinnbilderetiketten zur Verfügung gestellt. Die Liste ist folgenderweise zusammengestellt: zuerst stehen die vier Hauptkategorien auf der Startseite (Positionen 1 bis 4), weiter sind die Dienstleistungen (5 bis 14). Alle den Benutzern eingereichten Abbildungen sind in zufälliger Reihenfolge:

Liste der wahrscheinlichen Sinnbilderbedeutungen:

1. Einzelpersonen, 2. Freie Berufe, 3. Klein- und Mittelstand, 4. Firmen und Investitionen;
5. Unternehmensfinanzierung, 6. Spareinlagen, 7. POS-Terminal, 8. Payroll, 9. Alltägliches Banking, 10. Kredite, 11. Jugendliche Pakete, 12. Versicherungen 13. Bankkontos, 14. Sonstige Dienstleistungen

B. Evaluierungsergebnisse

Nach der Durchführung der Untersuchung sind die ausführlichen Daten über die Assoziativität der Sinnbilder zusammengefaßt (Tabelle 1). Die Kommentare von Testpersonen sind relevant für die Benutzerüberlegungen währnd der Interaktion mit dem Prototypen.

TABELLE II. INTERPRETATION DER METAPHERN BEI BANKANWENDUNG

Metapher	Bedeutung	Komentare
	5% Assoziativität, 22% Vermutung für Freiberufe	Nur einer der Benutzer hat Vermutung für Bankdatei des Kunden und assoziiert mit einem Bankkonto
	100 % Assoziativität	„Das beste Sinnbild, trotzdem nicht eindeutig. Interessant wäre die Assoziativität bei Benutzer anderer Nationalitäten.“
	10% Assoziativität 50% Vermutungen für Versicherungen	Bei Vermutungen für Versicherungen: „Hinzufügen von Schutzschild“, „Regenschirm“, „Vorhangeschloss“
	10% Assoziativität, heterogene Vermutungen	Bei Vermutungen für Sonstige Dienstleistungen: „Hinzufügen von Papierblatt mit Bookmarking“
	33% Assoziativität, je 22% Vermutungen für Kredite und Peyroll	„Für offensichtlichen Geldübergabeprozess noch eine Hand hinzufügen – erste Hand reicht Geld der zweiten ein“, „offene Geldbörse und hineinfallende Münzen“
	78% Assoziativität	„Hinzufügen von Papierblatt mit Bookmarking“
	22% Assoziativität je 17% Vermutungen für Sonstige Dienstleistungen und Alltägliches Banking	„Statistische Daten“.
	45% Assoziativität 22% Vermutungen für POS Terminal je 17% Vermutungen für Alltägliches Banking und Firmen und Investitionen	
	50% Assoziativität je 17% Vermutungen für Einzelpersonen und für Firmen und Investitionen	

	50% Assoziativität 17% Vermutung für Freiberufe	„Jugendansicht durch Visier“, „Verständlichkeitsverbesserung durch Wegschaffen der Krawatte und andere Frisur“
	0% Assoziativität je 28% Vermutungen für Klein- und Mittelstand und für Firmen und Investitionen 17% Vermutungen für Kredite	„für Klein- und Mittelstand: statt Pfeilen Business-Gebäuden“
	67% Assoziativität 17% Vermutungen für Bankkontos	„Zahlungskarte und/oder Geheimzahleingabe“
	17% Assoziativität 22% Vermutungen für Unternehmensfinanzierung	„für Unternehmensfinanzierung Business-Gebäude oder Aktentasche hinzufügen“
	0% Assoziativität 33% Vermutungen Spareinlagen und 17% für Peyroll	„Im Alltag gebraucht man eine Geldbörse, deswegen assoziiert man sie mit Alltäglichem Banking“

Die Ergebnisse der Evaluierung der 14 Metaphern zeigen, daß bei bulgarischen jüngeren Benutzern nur eine von diesen mit Einsparen assoziiert wird. Nur bei zwei anderen (Private Personen mit 78% und POS Terminal - 67%) ist die Assoziativität annehmbar. Daraus kann man die eindeutige Schlußfolgerung ziehen, dass das formelle Hinzufügen von unerkennbaren Sinnbildern, würde kaum die Zugänglichkeit, auch nicht die Benutzerleistung- und/oder GUI-Akzeptabilität steigern. Maßgeblich ist auch, dass die Kreativität bei zwei der Sinnbilder zu 0% Assoziativität führte und bei weiteren drei zwischen 5 und 10% liegt (Abbildung 1).

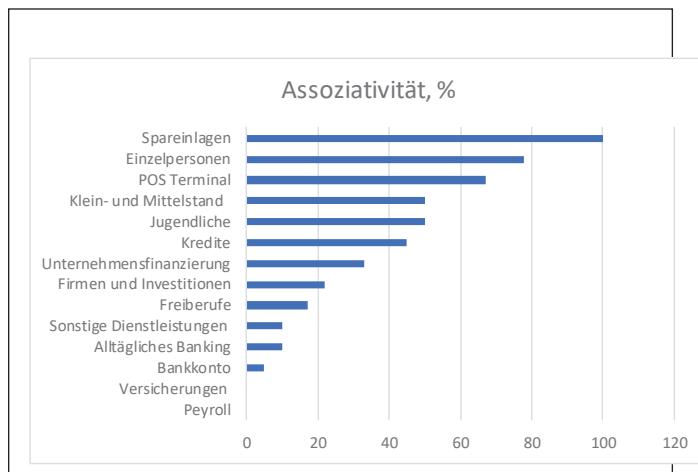


Abb. 1. Assoziativität der Metaphern

III. PERFORMANZTEST

A. Experimentvorbereitung

Um festzustellen, ob eine Beziehung zwischen der Performanz und der Gebrauchstauglichkeit am Beispiel der Erkennbarkeit der GUI-Sinnbilder wurde eine Untersuchung mit Studenten in Usability Engineering des Master-Studiums in Ingenieurdesign und im Fach Softwareergonomie des Masterstudiums Graphischer und Web-Design (Tabelle 3).

TABELLE III. TEILNEHMER AM EXPERIMENTBENUTZERPERFORMANZ

Anzahl	10
Geschlecht	6 Frauen 4 Männer
Alter	23,4
Ausbildung	Universitätsabschluß
GUI Erfahrung	Min. 5 Jahre

Die Benutzerleistung kann man durch die Erfolgsquote (Fehleranzahl, Aufgabenerfüllung), Zeit einschätzen. In der vorliegenden Untersuchung werden die mittlere Anzahl der Benutzeraktionen und die Summe der Fehler als Experimentvariablen ausgewählt.

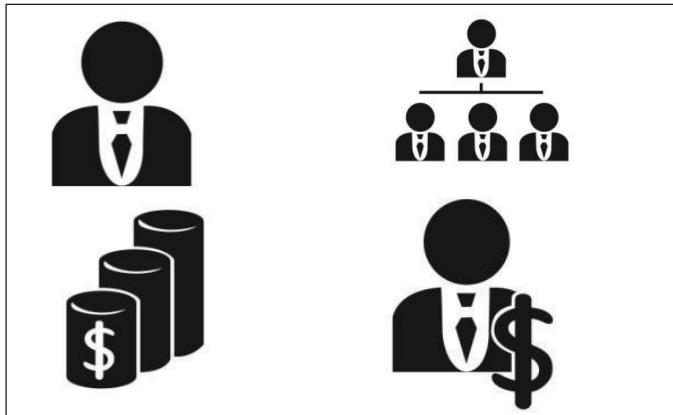


Abb. 2. Startbildschirm

Zur Erfüllung der Aufgaben werden den Benutzern die Bildschirme auf Abbildung 2 und 3 zur Verfügung gestellt.

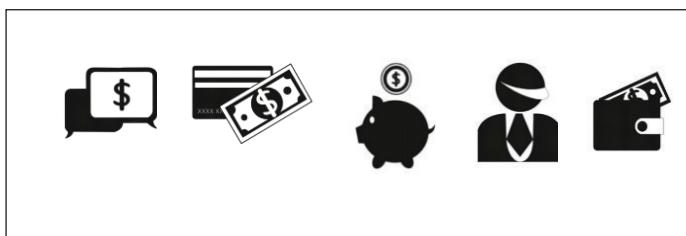


Abb. 3. Bildschirm "Dienstleistungen für Einzelpersonen"

Bei der Untersuchung der Beziehung zwischen Sinnbilder-Assoziativität und Benutzerperformanz wird die Performanz nicht nur durch die Anzahl der erfolgreich bzw. nicht erfolgreich erfüllten Aufgaben geschätzt, aber wird auch die Komplexität der Aufgaben in Bezug genommen. Bei diesem Experiment ergibt sich die Komplexität aufgrund der Metaphernverständlichkeit infolge der o.g. Untersuchung.

Die Aufgaben sind nach ihrer Komplexität als:

- leicht – in zwei Schritten und mit assoziativen Sinnbildern (Finden Sie „Spareinlagen“ für Einzelpersonen)
- mittlerer Schwierigkeit - in zwei Schritten und mit einem assoziativen und einem mittelmäßig

assoziativen Sinnbild (Finden Sie „Jugendliche“ für Einzelpersonen)

- und schwer in zwei Schritten und mit einem assoziativen und einem nicht assoziativen Sinnbild (Finden Sie „Versicherungen“ für Einzelpersonen) bezeichnet.

TABELLE IV. VARIABLEN DER PERFORMANZBEURTEILUNG

Variable	Bezeichnung	Werte
Aktionenanzahl	NoA NumberOfActions	Mittelwert
Fehleranzahl	NoE NumberOfErrors	Summe

Die Performanz wird auf der Basis der Variablen in Tabelle IV: Anzahl der Benutzeraktionen (Mittelwert) und -fehler (Summe) gemessen.

B. Experimentergebnisse

Nach der Erfüllung der Aufgaben durch die wahrscheinlichen Benutzer sind die Experimentdaten in Tabelle V zusammengefasst.

TABELLE V. DATEN VON DER PERFORMANZUNTERSUCHUNG

Aufgaben-Komplexität	NoA (MW)	NoE	Aufgabenbeschreibung
niedrige ET Easy	2,2	1	 Einzelpersonen
mittlere MT Average	5,2	13	 Einzelpersonen
hohe CT Complex	6,0	20	 Einzelpersonen

Die Anzahl der Fehler aller beteiligten an die Untersuchung Benutzer beträgt bei den assoziativen Sinnbildern 1 - ein sehr niedriger Wert im Vergleich zu den Werten bei mangelnder oder auch fehlender Assoziativität: 13 bzw. 20. Ähnlich sieht es bei den Werten der mittleren Aktionenanzahl: 2,2 gegenüber 5,2 bzw. 6,0.

Die Experimentergebnisse bestätigen eindeutig, daß zwischen assoziativen Metaphern und Nutzerperformanz ein positiver Zusammenhang besteht.

IV. AUSBLICK

In weiteren Forschungen ist vorgesehen, relevante Usability-Merkmale von GUI wie Mapping, Hierarchie, Konsistenz, u.a. zu untersuchen. Infolge einer Serie Experimente kann man eine vergleichende Analyse unterschiedlicher Benutzerleistungsfaktoren durchführen, wichtig für die Design-Empfehlungen bei verschiedenen Benutzerschnittstellen und Projekten [5, 6, 7, 8].

LITERATURVERZEICHNISS

- [1] Wiedenbeck S, The use of icons and labels in an end user application program: an empirical study of learning and retention, *Behaviour & Information Technology*, 18(2), 1999, pp. 68-82
- [2] Leporini, B., F. Paterno, Criteria for Usability of Accessible Web Sites. Lecture Notes in Computer Science, Volume 2615/2003, 2003, pp. 43-55
- [3] International Standard Organization, ISO 9241-125 Ergonomics of Human-System Interaction - Part 125: Guidance on Visual Presentation of Information, 2017
- [4] International Standard Organisation, ISO 9241-171 Ergonomics of Human-System Interaction - Part 171: Guidance on Software Accessibility, 2008
- [5] Boehringer J., P. Buehler, P. Schlaich, *Kompendium der Mediengestaltung für Digital- und Printmedien*, Springer-Verlag Berlin Heidelberg, 2008
- [6] Galitz, W. O. *The Essential Guide to User Interface Design An Introduction to GUI Design Principles and Techniques*, Wiley Publishing Inc., Indianapolis, 2007
- [7] Stapelkamp, Torsten, *Interaction- und Interfacedesign, Web-, Game-, Produkt- und Servicedesign* Usability und Interface als Corporate Identity, Reihe: X.media.press, Springerverlag, 2011, ISBN 978-3-642-02074-2, Print version ISBN 978-3-642-02073-5
- [8] Pearrow, M., *Web Site Usability Handbook*, Charls River Media Inc., Rockland, Massachusetts, 2000

Informational aspects of modelling the Health Regions in Bulgaria

Informationsaspekte bei der Modellierung der Gesundheitsregionen in Bulgarien

Velko Iltchev, Aleksandar Tsenov

FDIBA, Technical University
Sofia, Bulgaria

Abstract — The main objective of the work is, based on the terms "Regional Health Care" and "Health Regions", to create a suitable data management and data processing concept, which ensures complete information treatment of a health region. A simple model of a health region is shown. The components forming the model are divided according to its main properties. These properties are defined as key indicators and appropriately quantized to determine the required data types and the appropriate data structure.

Zusammenfassung — Hauptziel der Arbeit ist, aufgrund der Vorstellung der Begriffe „Regionale Gesundheitsversorgung“ und „Gesundheitsregionen“ ein passendes Konzept zur Datenhaltung und Datenbearbeitung zu erstellen, die eine vollständige Informationsbehandlung einer Gesundheitsregion gewährleistet. Ein einfaches Modell einer Gesundheitsregion ist dargestellt. Die, das Modell bildende Komponente sind nach deren Haupteigenschaften aufgeteilt. Diese Eigenschaften werden als Schlüsselindikatoren definiert und passend quantisiert, um die benötigten Datentypen und die geeignete Datenstruktur bestimmen zu können.

I. EINFÜHRUNG

Es gibt keinen einheitlichen Begriffsgebrauch und Begriffsverwendung „Gesundheitsregion“. Ein eingeführtes oder doch zumindest allgemein zugrunde liegendes Verständnis, was eine Gesundheitsregion ist, fehlt, und auch die verschiedenen vorgetragenen Initiativen lassen dieses in aller Regel offen [1].

Definition Gesundheitsregion: Als Gesundheitsregion kann eine Region bezeichnet werden, in welcher es zu einem reflektierten Zusammenschluss von sowohl in der Patientenversorgung als auch in der Gesundheitswirtschaft, in daran angrenzenden Bildungs- und Forschungseinrichtungen als auch im Tourismus handelnde Organisationen und Personen gekommen ist, die zugunsten eines neuen, abgestimmten Leistungs-, Entwicklungs- und Qualitätsgeschehens miteinander kooperieren. Neben den Zielen der solchermaßen wirtschaftlich Involvierten (a) und der Region (b) sollen auch die Bürger (potentielle Patienten) und Gäste (c) von den Leistungen der Gesundheitsregion profitieren [2].

Anhand mehrerer Untersuchungen im Rahmen des Wettbewerbs „Gesundheitsregionen der Zukunft“, initiiert von dem Bundesministerium für Bildung und Forschung (BMBF) erstellt und veröffentlicht W. George ein einfaches Modell der Struktur und der Verbindungen zwischen den Hauptkomponenten einer Gesundheitsregion /s. Abb. 1/.

II. MODELBESCHREIBUNG

Definition Regionale Gesundheitsversorgung (RGV): Unter RGV werden all diejenigen koordiniert vorgetragenen Aktivitäten einer Region (oder auch Stadt) verstanden, die dem Ziel dienen, die gesundheitliche Versorgungsqualität (a), die Lebensqualität der Betroffenen (b) und die Attraktivität einer

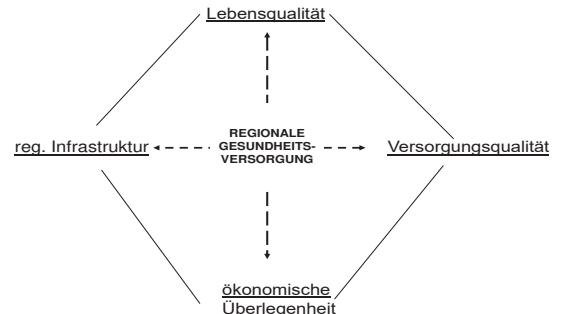


Abb. 1. Die 4 Leitkriterien einer regional vorgetragenen Gesundheitsversorgung: Lebensqualität der Betroffenen, medizinische Ergebnisqualität, Ausprägung einer regionalen Infrastruktur (Standortmerkmal) und ökonomische Effizienz [2]

Region (c) bestmöglich zu organisieren. Dabei wird eine Praxis gewählt, die sich zusätzlich auch an dem Kriterium ökonomischer Überlegenheit gegenüber zentral vorgetragenen Versorgungsstrategien orientiert (d).

A. Gesundheitsgebundene Lebensqualität

Zu den wichtigsten gesundheitsgebundenen Lebensqualitätsfaktoren kann man unter allem gezählt werden:

- Gesundheit und Alter;
- Gesundheit und Geschlecht;
- Gesundheit und Familienstand (inkl. Kinderanzahl);
- Gesundheit und Ausbildungsniveau;
- Gesundheit und Einkommen;
- Gesundheit und Beschäftigung;

- Gesundheit und Sport;
- Gesundheit und ungesunde Gewohnheiten (Alkohol, Rauchen).

Alle o.a. Faktoren lassen sich quantitativ bestimmen (z. B. als Prozentwert) und die Daten können relativ leicht und regelmäßig gesammelt werden – z.B. bei Volkszählung. Das erlaube die Erstellung einer Sammlung von Daten gleicher Art, die als eine geeignete Datenstruktur implementiert werden kann.

Jeder Faktor kann weiter dekomponiert werden um höhere Genauigkeit und Umfangreichbarkeit zu erreichen [3].

B. Regionale Versorgungsqualität

Die Regionale Versorgung erfolgt aufgrund mehrerer Strategien für Prävention, Behandlung und Rehabilitation: Die Entwicklung diesen Strategien ist gewährleistet durch die Analyse mehreren Daten über die Ausbreitung von Krankheiten, Krankenhäuser, Rehabilitationszentren, Objekte der Heilungstourismus usw.,

- Anzahl der Bevölkerung in den administrativen Regionen des Landes;
- Anzahl der Kranken von unterschiedlichen Krankheiten;
- Anteil der Mortalität und Inzidenz verschiedener Krankheiten;
- Grad der Prävention, Behandlung und Rehabilitation;
- Verfügbarkeit von Gesundheitseinrichtungen;

Dazu kann man auch die laufenden Projekte auf dem Gebiet der Gesundheitsversorgung. Die hohe Erfolgsrate dieser Projekte ist vor allem durch die guten Ergebnisse im Sozialwesen, und nicht so viel durch das Gesundheitssystem zu kennzeichnen.

C. Regionale Infrastrukturmerkmale

Um geeignete Merkmale zur Modelbildung der Gesundheitsregionen einzuführen zu können und diese als Schlüsselindikatoren zu bestimmen, wird hier den Vorschlag eingegangen, folgende Punkte zu beachten:

- Leichte Zugänglichkeit: diese wird dadurch gesichert, dass die Patienten in geeigneter Form und Anspruchsauswahl die Möglichkeit erhalten ihre Beschwerden vorzutragen.
- Zügige Bearbeitung: die zeitnahe Bearbeitung ist wichtig. Der Eingang wird bestätigt, bei komplexem Sachverhalt wird ein Zwischenbericht gegeben.
- Unabhängigkeit: dies bedeutet neben einem klaren Vorgehensprozess unabhängige und qualifizierte Bearbeitung durch geeignete und geschulte Mitarbeiter.
- Transparenz: die für Dritte durchsichtigen Wege des Beschwerdemanagements von der Annahme einer Beschwerde bis zur Interventionsempfehlung sowie ein jährlich veröffentlichter Bericht.
- Verantwortung: durch Sicherung einer finalen Bearbeitung, in der die Verpflichtungen der Beteiligten definiert sind.

Zuletzt soll die Legitimität regionaler Versorgung durch ihr potentiell enges Verhältnis zu den integrierten Versorgungsverträgen begründen.

D. Kostengünstige Auswirkungen und Kooperative Gesundheitsökonomie

Das Ziel wäre, die Systeme der Integrierten Versorgung stärken und eine bessere Abstimmung zwischen stationären und ambulanten Bereich ermöglichen. Zur Erfüllung dieses Ziels benötigt man bestimmten Maßnahmen.

- Gründung von medizinischen Versorgungszentren (MVZ) zur Teilnahme an der vertragsärztlichen Versorgung;

- Institutionelle Öffnung der Krankenhäuser bei Unterversorgung im vertragsärztlichen Bereich;
- Einbeziehung der Krankenhäuser in die ambulante Teilnahme an Diseasesmanagement Programmen;
- Hochspezialisierte Leistungen und die Behandlung seltener Erkrankungen können durch Krankenhäuser ambulant erbracht werden;
- Regelung der Verträge über integrierte Versorgungsformen im Bereich der Behandlung, Prävention und Rehabilitation, Diseasesmanagement-Programme u.a. Auch Krankenhäuser sollen daran partizipieren können. Es soll Budgetverantwortlichkeit herrschen.

Potentiell möglich werden durch diese sozialrechtlichen Veränderungen:

- ein weitgehend freies Verhandeln zwischen den Kostenträgern und den Leistungserbringern;
- das Recht der gemeinsamen Berufsausübung;
- die Überwindung sektoraler Budgets durch Einführung eines Prozess- oder Case- oder Diseasesmanagement Honorierungssystems,
- Entwicklung bedarfsgerechter und kostengünstiger Betreuung der Bevölkerung.

Eine große Herausforderung in der Zukunft sei die Art und Weise der Quantisierung und der Einschätzung der Ergebnisse von der Erfüllung der o.a. Maßnahmen.

III. INFORMATIONS- UND DATENMODEL

Wichtiger Bestandteil einer optimalen Versorgungsstruktur ist gegenwärtig die Implementierung einer geeigneter Informationsinfrastruktur. Der Erste Schritt ist die Bildung der Regionen landesweit. Dies erfolgt aufgrund Bestimmung einzelnen Bereichen mit ähnlichen Werten, der bis zu diesem Punkt definierten Schlüsselindikatoren.

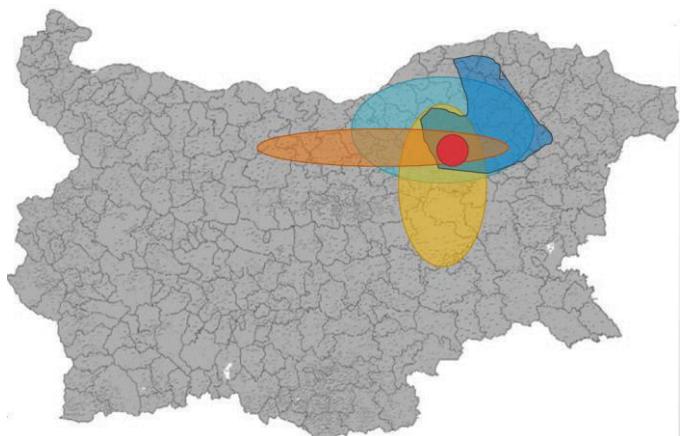


Abb. 2. Geografische Bildung einer Gesundheitsregion aufgrund des beschriebenen Models und Bestimmung des Regionszentrums

Das Zusammenbinden diesen Bereichen findet man das Zentrum einer Gesundheitsregion. Die genauere Lokation des Zentrums erfolgt je nach Gegebenheiten der untersuchten Landesgebiet. Die so gebildete Region kann nicht von der administrativen Aufteilung des Landes (Abb. 2)

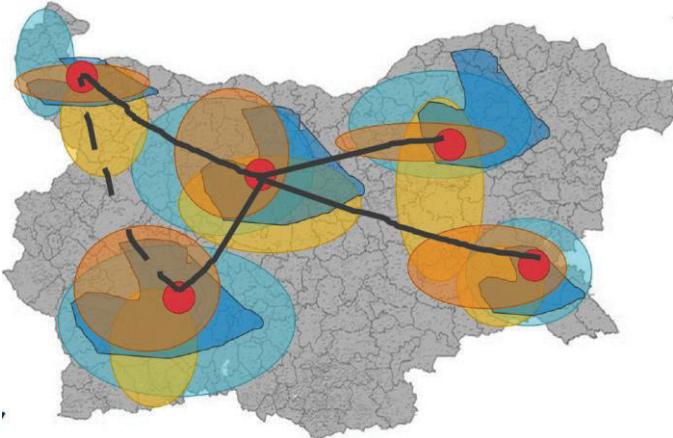


Abb. 3. Vernetzung der Regionszentren

Nach der Bestimmung des Regionszentrums folgt die Errichtung des Informationszentrums der Region. Die Struktur des Zentrums wird hier nicht behandelt. Es wird angenommen, dass die Rechnerarchitektur und die Anwendungsarchitektur des Informationszentrums vorhanden sind.

Es besteht die Möglichkeit, ein von den Zentren als Mittelpunkt des Netzes zu benutzen, oder ein separates Datenhaltung und Datenbearbeitungszentrum zu errichten (Abb.3).

Die Hauptherausforderung der Vernetzung der Regionszentrums ist die Heterogenität der Daten, erstellt durch die verschiedenen Anwendungen, verschiedenen Datentypen und die unterschiedlichen Methoden der Datenübertragung.

Eine Struktur eines geeigneten Datenmodell ist auf der Abb. 4. dargestellt. Man behauptet eine Vorbearbeitung der Daten je nach dem Typ der Schlüsselindikatoren. Einige Vorbearbeitungen können vorgenommen werden, bevor die Daten zum Datenzentrum zu schicken. So werden die Daten entsprechend filtriert und deren Volumen kann wesentlich reduziert werden.

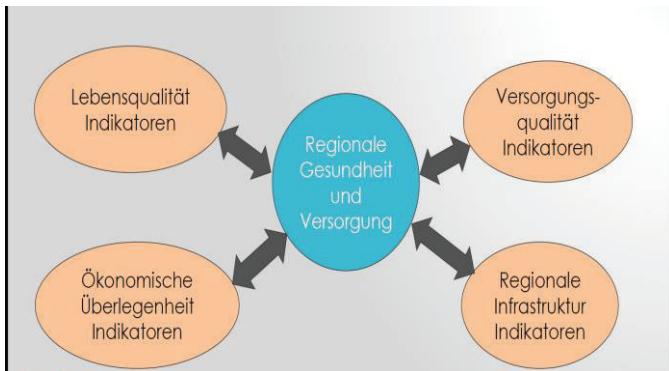


Abb. 4. Funktionales Model der zu erstellenden Daten

IV. DATENHALTUNG UND DATENBEARBEITUNG

Die Daten über die einzelnen Indikatoren kommen aus unterschiedlichen Quellen, wie z.B.:

- Nationales Institut für soziale Sicherheit - Pro-Kopf-Einkommen, Arbeitslosigkeit u.a.
- Krankenkasse - Lebensdauer, Häufigkeit bestimmter Krankheiten, sowie auch Zugänglichkeit, Rechtzeitigkeit und Angemessenheit der medizinischen Versorgung u.a.
- Wirtschaftsministerium - Inflationsrate/Wachstumsrate, Lebenshaltungskosten, Immobilienpreise, Aktienkurse u.a.

- Ministerium für Ausbildung und Wissenschaft - Bildung, Forschung und Innovation

Die Struktur dieser Daten ist heterogen, denn jede dieser Institutionen besitzt ihr eigenes Datenbanksystem. Trotzdem müssen diese Daten in eine regionale Datenbank und danach in eine zentrale Datenbank gespeichert werden /s. Abb. 2/. Die regionalen und die zentrale Datenbank werden eine und dieselbe Datenstruktur besitzen.

Deswegen muss /als erster Schritt/ ein einheitliches Format für die Datenübertragung entwickelt werden. Dieses Format soll die Übertragung von semi-strukturierten Daten ermöglichen. Am bestens für diesen Fall geeignet sind Sprachen wie XML [7] und JSON [4], mit den entsprechenden Definitionen der Struktur der Daten DTD [6], XSD [8] und JSON-Schema [5]. Der Vorteil bei diesem Vorgehen ist, dass Ergänzungen und Korrekturen der Struktur der Daten leicht machbar sind.

Für die Generierung der Datensätze aus den entsprechenden Institutionen, aufgrund der o.g. Definitionen der Datenstruktur, muss ein System von Mikroservices entwickelt werden. Diese werden in jede Organisation implementiert. So werden sie zu jedem Zeitpunkt aktuelle Informationen für die regionalen und für die zentrale Datenbank liefern.

Nachdem die Daten über die Indikatoren in den regionalen und in der zentralen Datenbank schon gespeichert sind, müssen diese vor deren weiterer Bearbeitung normiert werden. Nach dieser Normierung sollen die Werte im Intervall zwischen 0 und 1 liegen. Das kann man mithilfe der folgenden Formel erreichen:

$$x_{norm}[i] = \frac{x[i] - x_{min}}{x_{max} - x_{min}}$$

Die schon gespeicherten Daten über die Indikatoren werden als Fakten für ein Expertensystem dienen. Es wird genutzt, um Ratschläge zu generieren, in Bezug auf die zukünftige Entwicklung der Regionen, wie z.B.:

- welche Einrichtungen müssen noch gebaut werden
- welche Maßnahmen muss man zusätzlich ergreifen
- usw.

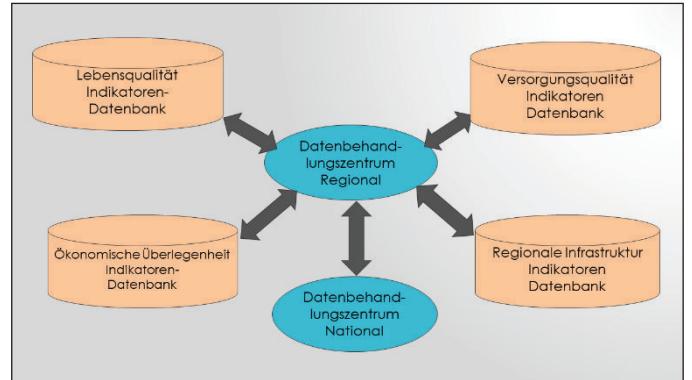


Abb. 5. Beispiel eines Systems zur Datenhaltung und Datenbearbeitung von Daten aus den Gesundheitsregionen und aus den entsprechenden Institutionen.

Expertensysteme, die auf Neuronale Netzwerke basiert sind, sind zu diesem Zweck nicht gut geeignet, denn es ist schwer zu definieren, welches Wissen das System erlernen muss. Damit ist es auch schwer die Muster zu bestimmen, die dem neuronalen Netzwerk dieses Wissen beibringen werden.

Am bestens zu diesem Zweck geeignet sind regelbasierte Expertensysteme mit Regeln, die mithilfe propositionaler Logik definiert sind. Durch Einfügen von Gewichtungs-faktoren vor den Fakten und vor den Regeln im Regelrumpf, kann man die Wichtigkeit der einzelnen Indikatoren variieren.

Der Prozessor, der diese Regel bearbeiten wird, soll eine *top-down* Strategie benutzen, bei der er seinen eigenen Stack verwalten soll, was bedeutet - nicht den Programmstack benutzen. Die Vorteile dieses Vorgehens sind:

- zyklische Regeln sind leichter zu erkennen und damit leichter zu vermeiden, sonst kann der Prozessor in eine Endlosschleife geraten;

- es ist leicht festzulegen ob eine Regel schon bearbeitet ist oder nicht. Wenn - ja, dann kann man das berechnete Wert aufspeichern. Falls dieselbe Regel wieder in irgendeinem Regelrumpf auftritt, dann braucht man nicht sie erneut zu berechnen. So kann man Berechnungen, die sich wiederholen, vermeiden, was die Geschwindigkeit wesentlich steigert.

DANKSAGUNG

Diese Publikation wurde innerhalb des Anbahnungsprojekt „Arteria Danubia“ im Themenfeld „Soziale Lage und Gesundheit“, gefördert vom BMBF, Deutschland, erstellt.

LITERATURVERZEICHNIS

- [1] Hilbert, J.: "Das Gesundheitswesen als Jobmaschine? Vor den Erfolg haben die Götter den Schweiß gesetzt!", *Jahrbuch Gesundheitswirtschaft 2008: Prozessoptimierung, eHealth und Vernetzung*, Berlin Wegweiser, S. 144-145
- [2] George W., "Gesundheitsregionen und Regionale Gesundheitsversorgung", *Gesundheitsnetzwerke initiieren: Kooperationen erfolgreich planen*, Berlin: Med. Wiss. Verl.- Ges., ISBN 978-3-939069-72-0, - 2009, pp. 293-307
- [3] Vankova D., N. Usheva, N. Feschieva, "Quality of Life Related to Health in the Community", *Social Medicine*, 2/3 2012, ISSN 1310-1757, pp. 26-29
- [4] European Computer Manufacturers Association, *Standard ECMA-404, The JSON Data Interchange Syntax*, <http://www.ecma-international.org/publications/standards/Ecma-404.htm>, 2017
- [5] JSON-Schema Organization, *JSON Schema: A Media Type for Describing JSON Documents*, <http://json-schema.org/latest/json-schema-core.html>, 2018
- [6] World Wide Web Consortium, *DTD Standard*, <http://www.w3.org/standards/dtd/core>, 2017
- [7] World Wide Web Consortium, *XML Standard*, <http://www.w3.org/standards/xml/core>, 2018
- [8] World Wide Web Consortium, *XSD Standard*, <http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/>, 2018

Methods for Eventual Consistency in Collaborative Editing Systems

Methoden für eventuelle Konsistenz in kollaborativen Bearbeitungssystemen

Boyan Katsarski, Velko Iltschev

Department of Computer Science: Technical University Sofia, branch Plovdiv
Plovdiv, Bulgaria, Email: katzarsky@gmail.com, iltchev@tu-plovdiv.bg

Abstract — The paper analyzes methods for Operational Transformations and models for Conflict-free Replicated Data Types in the light of Collaborative Editing Systems and preservation of semantics. Benefits and disadvantages of each method or model are presented. A proposal for semantic improvement and usability has been made, which includes: switching atoms from characters to words and adding inverse indexing to the documents with words stored private dictionaries.

Zusammenfassung — Der Artikel analysiert Methoden für operative Transformationen und Modelle für konfliktfreie replizierte Datentypen in Hinblick auf kollaborativen Textverarbeitungssystemen und auf Erhaltung der Semantik. Vor- und Nachteile jeder Methode oder jedes Modells werden vorgestellt. Ein Vorschlag für semantische Verbesserung und Nutzbarkeit wird gemacht, der Folgendes umfasst: Anwendung ganzer Wörter anstelle einzelner Zeichen als Atome und Einfügen inverser Indexierung zu den Dokumenten mit den Wörtern, die in den privaten Wörterbüchern gespeichert sind.

I. INTRODUCTION

Users of mobile applications expect them to continue working even when the device is offline or has bad network connectivity and when the network becomes available again to synchronize with other devices. Such applications are: calendars, address-books, password-managers, task-organizers, etc. Likewise, collaborative work requires several users to edit the same text document (text, spreadsheet, etc.) at the same time. Each user's concurrent edits must be reflected on the others' replicas with minimal delay.

The key requirement here is that the state of the distributed application needs to be replicated on many devices (nodes). Each of them modifies the state locally and propagates the changes optimistically, while continuing to work locally.

The traditional approach to concurrency control - serializability (with locking), causes the application to become unusable when network connectivity is poor [1].

For applications, which tolerate temporal network outages, we must assume that users can make updates concurrently on different nodes and the resulting conflicts must be resolved.

The naive way to resolve conflicts is to discard some of the updates when a conflict arises. This approach leads to loss of updates and potentially data. Another approach is to make the user manually resolve the conflict using some logic or common-sense. This is tedious and prone to errors and should be avoided if possible. Nowadays applications solve this problem with a range of ad-hoc and application-specific means.

In this paper an overview of the benefits and disadvantages of the current general-purpose optimistic replication methods and models has been made. The focus is on Operational Transformation (OT) methods and Conflict-free Replicated Data Types (CRDTs) models and their derivatives in the context of collaborative editing systems. Finally, a proposal for semantic improvement and usability has been made, which

includes: switching atoms from characters to words and adding inverse indexing to the documents with words stored private dictionaries.

II. STRONG EVENTUAL CONSISTENCY IN PARTITIONED NETWORKS. COLLABORATION.

A. Partitioned network

In the real networks, devices (especially mobile ones) go offline and then online all the time in a random manner, with periods of no connectivity in-between. Under these conditions it is impossible to achieve immediate strong consistency of the replicated state across all devices. Moreover, due to poor networks, updates made in the originating node may be received by the other nodes in a different order. While some of the updates are in transit and some of the replicas are only converging to the *eventual* consistent state.

B. Strong Eventual Consistency (SEC)

Eventual consistency requires that at the end – when all updates from all the nodes are sent and received – all replicated states will become the same.

More formally *eventual consistency* is defined [2] as:

- **Eventual Delivery:** Every node will see the updates made by the other nodes *eventually*. These updates may come out-of-order due to network partitioning.
- **Convergence:** Same operations should lead to the same state. Even when received out-of-order.
- **No updates are lost due to concurrent modifications:** A good example for a violation is the “Last Writer Wins” (LWW) policy.

C. Collaborative Editing Systems (CES)

In the constraints imposed by partitioned networks and eventual consistency, collaborative editing becomes a non-

trivial task.

Nowadays, collaborative editing systems are based on the old Operational Transformation principles. There is however a new approach – CRDT, which is still in research and development.

III. APPROACHES TO OPTIMISTIC REPLICATION, COLLABORATIVE EDITING AND CONFLICT RESOLUTION

A. Operational Transformation (OT)

The algorithms for Operational Transformation appear first in the works of Ellis & Gibbs [3] in 1989. Later improvements are made by M. Ressel et al [4] and C. Sun et al [5]. Most of them treat a document as a single list of characters.

First OT system is GROVE based on the “dOPT” algorithm [3]. Currently Google Docs uses a modified version of it.

The essence of OT is that operations made in one node are *transformed* on the fly when applied to another node in order to get the same resulting state. A transformation function is used to change the position of the incoming operations in order to be adequately applied to the local text.

For example: if Alice deletes a character at position 5 while concurrently Bob inserts a character at position 10. Bob’s operation should be transformed to position 9 = (10-1) when applied to Alice’s text.

The transformation function must take into account all prior operations made locally. Moreover, if there are many concurrent editors, the transformations become exponentially more complex as the operations come in different order for each of the users. Complexity is aggravated when an *offline* user has produced a big batch of operations and now becomes *online*. This makes OT unsuitable for many users working in poor networks.

Notable implementations are Grove (1989), Jupiter (1995) and its descendants: Google Docs, Apache Wave (formerly Google Wave), Etherpad and ShareJs.

To alleviate the problem with the complexity mentioned above Google Docs applies restrictions like:

- Using a single centralized server to sequence the order of updates.
- The centralized server makes some or all (details are proprietary) of the transformations of the operations.
- Restricted or disabled offline editing capabilities to limit the transformations produced by stale updates. (Stale updates lead to recalculation of all the updates made in the meantime.)

There are a couple of problems with these limitations:

- A centralized server introduces a single point of failure.
- A centralized server makes transformations and it needs to have the operations’ content, which means end-to-end encryption is impossible.
- Offline work is limited severely.

B. Conflict-free Replicated Data Types (CRDT)

Later in 2011 a notion for special replication-ready data model appears in the works of Shapiro et al [6].

CRDT considers that transforming operation is too complex and establishes a new model to handle real-time collaboration. While OT attempts to make non-commuting operations commute after the fact (via transformation). A better approach is to design operations to commute in the first place. This avoids the complexities of OT.

Operations that update the model must follow 3 constraints:

- Commutative
- Associative: these two ensure that operations can be applied in any order
- Idempotent: an operation can be applied many times and it yields the same result as if applied once.

The first two (commutative & associative) deal with the out-of-order problem due to poor networks. The last one (idempotent) remedies the problem that arises when an operation is sent to a remote node and no acknowledgement is returned. At this point it is unclear whether the node has *not received* the operation or it has received it but *not acknowledged* it or the acknowledgement has been *lost* in the network. Hence, CRDT operations need idempotency.

Model with these restrictions makes conflicts impossible [6], but as a result, the model is monotonically growing (while leaving tombstones) and designing operations with the above constraints becomes difficult quickly.

The essence of CRDT is that operations should be designed following the 3 rules above. A trivial example is an increment-only counter, which in CRDT terms must be designed as an array of integers [6]. Each index is owned only by one user. Write operations are done only to the user’s element but incoming updates can be received on all elements of the array. The actual value of the counter is the sum of all elements.

The benefits of the CRDT model include:

- Built-in strong eventual consistency
- Possible server-less implementation, providing end-to-end encryption and no single point of failure
- Offline capability with no extra complexity.

The disadvantages are:

- Ever-growing state due to more users added or data being “deleted” under tombstones [7]. Distributed garbage collection methods are non-trivial and require some sort of locking.
- Difficult to design operations (API) for the CRDT type (map, set, counter, text document, etc.)

Some of the trivial CRDTs include [8]:

- Increment-Only Counter
- PN (increment/decrement) Counter
- Add-Only Set
- Directed Graph CRDT
- Last-Writer-Wins Register

Notable implementations of trivial CRDTs are made by: Riak (library), Bet365 (counters) and League of Legends (chat).

Some of the non-trivial CRDTs (plus algorithms) for collaborative text editing are:

- **WOOT** by Oster et al [9]
- **Treedoc** by Nuno Preguiça et al [10]
- **RGA** by Roh et al [11]
- **Logoot** by Weiss et al [12]

They share some common characteristics. A character is considered an *atom* in the model (recently a UTF-8 character which is represented by more than one byte). The structure varies (linked lists, trees and semi-lattices) but to *preserve the intent* of the original editor tombstones [7] are used. Tombstone is an atom marked as “deleted” and skipped when presenting to the user. Tombstones are needed when an atom is deleted but it serves as a reference point for concurrent insert.

IV. CRDTs FOR COLLABORATIVE EDITING

A. WOOT

The “WithOut Operational Transformation” method is developed by Oster et al [9] in 2005. The essence of it is that it treats characters as atoms in a *linked list*. Each element has id indicating precedence. Inserts contain `<character, id, preceding-id>`. Deletes just mark the id as a tombstone. There is no update operation but delete and insert. Tombstones are needed when one user deletes a character while another concurrently adds a character after it. Generation of IDs is special (free from vector clocks) so that a sorting function can linearize the resulting semi-lattice. A filtering function skips the tombstones when presenting the content to the user.

The method works well even with stale updates and does not explode in complexity when more users join the editing, unlike any OT method.

Operations Complexities [11]:

- Local: Insert $O(N^2)$, Delete $O(1)$
- Remote: Insert $O(N^3)$, Delete $O(N)$

Main problem is the unbound growth of the document when lots of edits are submitted over time. The garbage collection of the tombstones is non-trivial and requires all nodes to be online to reach a consensus on which tombstones are not needed anymore and will be physically deleted from the document. This is aggravated by the fact that the atom is a character and update of a character is translated to delete and insert, producing a tombstone in the process. Algorithm of purging the tombstones is not yet presented.

B. Treedoc

Treedoc is presented by Nuno Preguiça et al [10] in 2009. It stores atoms in a balanced binary-tree structure (with the extension of *mini-nodes* to handle more than two children stemming from the same parent). Treedoc is a binary tree whose paths to nodes are unique indices and ordered totally in infix order. It tries to keep the tree balanced and executes *explode* and *flatten* routines over a subtree to minimize tombstones and tree structure overhead. Flattening uses voting commit, similar to ACID databases’ two-phase commit. Still uses tombstones.

It works but has issues:

- if the user appends at the end of the document it unbalances the tree
- optimizations for tree flattening require locking which makes it less usable in poor networks
- a lesser problem is the overhead of maintaining the tree structure balanced

C. RGA

“Replicated Growable Array” is proposed by Roh et al [11] in 2011 as a Replicated Abstract Data Type (RADT) aiming for text and data modelling. It treats the text as a linked-list of linked-lists (text blocks). It uses hash-tables to accelerate the access and improves garbage collection of tombstones by adding them to a *cemetery*.

Operations Complexities [11]:

- Local: $O(N)$
- Remote: $O(1)$

Complexity is lower than earlier models due to hashes. Moving text fragments (copy/paste) is simpler because a text-block is represented by a sub-linked-list. It uses tombstones to preserve intention as well.

Negatives include the overhead of supporting tombstones and the problem with inserting new text at the same coordinates concurrently.

D. LOGOOT

A CRDT proposed by Weiss [12] in 2009. A Logoot document is composed by lines defined by: `<id, content>` where *content* is a text line and *id* - a unique position identifier. Its structure supports easy deletion and insertion of new line(s). It does NOT rely on tombstones for consistency. Complexity is logarithmic. It preserves the meaning of a text line better because a user can only add/remove a line, not a single character. Atom ownership is not over a character but a line.

V. COMMON PROBLEMS PLAGUING THE CRDTs FOR COLLABORATIVE EDITING

A. Atom Ownership is over characters

Most of the CRDT for text editing, *except Logoot* are based on atoms that are characters. This is a problem because meaning is contained not within a single character but rather within words, phrases and sentences. Assigning blame for badly formed phrases that result in a collaborative editing is difficult since one word may be edited by couple of users. Only Logoot hits the mark in that area with having lines of text as atoms, which is not ideal but at least addresses the issue. A better solution may be using words or phrases as atoms.

B. Moving Text Blocks

Moving big blocks of text is a common practice when editing. Most of the methods above *except Logoot* suffer from structural issues when this is done: unbalanced trees and massive tombstone production. Since current tombstone garbage collection algorithms require locking commits they compromise the P2P-ready nature of CRDTs. A better approach for tombstone purging is needed or a method that does not even require them.

C. Search

Yielding the resulting text in CRDTs that are based on characters and tombstones is another issue because the text is calculated rather than used “as-is” and searching through documents based on these models require an extra step before the actual search.

D. Concurrently Inserting at the Same Point

It is “solved” in current models by inserting both fragments while *stabilizing* the order of the inserts. But this approach will produce semantically questionable results and is still unsolved semantically by any of the models.

VI. CONCLUSION AND A WAY FORWARD

A step towards a more meaningful policy of semantic ownership is assigning words or phrases as atoms. Since phrases are difficult to isolate even using grammatical analysis for different natural languages – words will be a wiser choice. An editor can be built on top to color-code the ownership of words, hence assigning blame for bad phrasing can be done easily.

Isolating words that are capitalized or punctuated may be represented as the non-punctuated word with added attributes to the atom. This will keep the dictionary smaller and more meaningful, since the words will be in cleaner form.

These word-atoms can form a dictionary that will enable inverse document storage right from the start. The inverse indexing will enable *built-in search* of the edited text. The document structure will use dictionary-pointers instead of the words themselves.

The dictionary can be partitioned in sections owned by the users. Since a section is owned only by one user – it can be

garbage collected only by the owner-user and not requiring a consensus or a lock protocol.

The editor which can be built on top of this proposed data structure will have to observe user actions such as:

1. typing or deleting a character
2. copy-pasting or moving a block of text
3. deleting a block of text

Adding or deleting a character (1) in a word will introduce temporary noise in the dictionary since the currently edited word will change. But this will affect only the private section of the user-dictionary. If the word was not owned by the editing user the new word is created in the user-owned partition of the dictionary.

Copy-pasting or cut-pasting (2) a big block of text will not make a lot of noise. The dictionary will stay mostly unchanged, except in the case when the text block cuts words in the middle. The structure made of dictionary-pointer atoms will reflect the change as in WOOT or LOGOOT while keeping the atoms count low.

Deleting a block of text (3) will leave fewer tombstones, since they are word-pointers not characters.

REFERENCES

- [1] S. B. Davidson, H. Garcia-Molina, and D. Skeen, “Consistency in partitioned networks,” *ACM Computing Surveys*, vol. 17, no. 3, pp. 341–370, Sep. 1985.
- [2] Kleppmann, Martin & R. Beresford, Alastair. (2016). “A Conflict-Free Replicated JSON Datatype.” *IEEE Transactions on Parallel and Distributed Systems*. Volume 28, Issue 10, Oct 2017, pp. 2733-2746. 10.1109/TPDS.2017.2697382.
- [3] C. Ellis and S. J. Gibbs, “Concurrency control in groupware systems,” in *ACM International Conference on Management of Data* (SIGMOD), May 1989, pp. 399-407.
- [4] M. Ressel, D. Nitsche-Ruhland, and R. Gunzenhauer, “An integrating, transformation-oriented approach to concurrency control and undo in group editors,” in *ACM Conference on Computer Supported Cooperative Work (CSCW)*, Nov. 1996, pp. 288-297.
- [5] C. Sun and C. Ellis, “Operational transformation in real-time group editors: Issues, algorithms, and achievements,” in *ACM Conference on Computer Supported Cooperative Work (CSCW)*, Nov. 1998, pp. 59-68.
- [6] M. Shapiro, N. Preguiça, C. Baquero, and M. Zawirski, “A comprehensive study of convergent and commutative replicated data types,” *[Research Report] RR-7506, INRIA*. 2011, pp.50.
- [7] G. Oster, P. Urso, P. Molli, and A. Imine, “Tombstone transformation functions for ensuring consistency in collaborative editing systems,” *IEEE Conference on Collaborative Computing: Networking, Applications and Worksharing - CollaborateCom 2006*, Nov 2006, Atlanta, Georgia, USA, pp.1-10. 10.1109/COLCOM.2006.361867
- [8] Marc Shapiro, Nuno Preguiça, Carlos Baquero, Marek Zawirski. “Conflict-free Replicated Data Types.”, *Stabilization, Safety, and Security of Distributed Systems*, Springer Berlin Heidelberg 2011, pp. 386-400.
- [9] Gérald Oster, Pascal Urso, Pascal Molli, Abdessamad Imine. “Real time group editors without Operational transformation”, *[Research Report] RR-5580, INRIA*. 2005, pp. 24.
- [10] Nuno Preguiça, Joan Manuel Marquès, Marc Shapiro, Mihai Letia. “A commutative replicated data type for cooperative editing”. *29th IEEE International Conference on Distributed Computing Systems (ICDCS 2009)*, Jun 2009, Montreal, Québec, Canada. IEEE Computer Society, pp. 395-403. 10.1109/ICDCS.2009.20.
- [11] Hyun-Gul Roh, Myeongjae Jeon, Jin-Soo Kim, Joonwon Lee. “Replicated abstract data types: Building blocks for collaborative applications”, *Journal of Parallel and Distributed Computing*, Volume 71, Issue 3, March 2011, pp. 354-368.
- [12] Stéphane Weiss, Pascal Urso, Pascal Molli. “Logoot: A Scalable Optimistic Replication Algorithm for Collaborative Editing on P2P Networks”. *29th IEEE International Conference on Distributed Computing Systems - ICDCS 2009*, Jun 2009, Montreal, Canada. IEEE, pp. 404-412.

Mixed-Reality Spatial Configuration with a ZED Mini Stereoscopic Camera

Räumliche Mixed Reality Konfiguration mit einer ZED Mini Stereoskopischen Kamera

Dimo Chotrov, Zlatka Uzunova, Yordan Yordanov, Stoyan Maleshkov

Virtual and Augmented Reality Laboratory, Research and Development and Innovation Consortium
Sofia, Bulgaria, vr-lab@sofiatech.bg

Abstract — This paper starts with a short overview of how a depth map can be calculated from images captured with stereoscopic cameras, and in particular - the work principal of the ZED Mini stereoscopic camera. After that we present three approaches we have identified that can be used to create a spatial configuration based on a real-world environment. The result of this spatial configuration is a mixed environment containing virtual objects aligned with the physical objects from the real environment. At the end we provide the results from experiments we have conducted using developed prototypes to validate the proposed approaches.

Zusammenfassung — Am Anfang wird einen kurzen Überblick gegeben, wie eine Tiefenkarte aus mit stereoskopischen Kameras aufgenommenen Bildern berechnet werden kann, und insbesondere das Arbeitsprinzip der stereoskopischen Kamera ZED Mini. Anschließend stellen wir drei Ansätze vor, mit denen eine räumliche Konfiguration basierend auf einer reellen Umgebung erstellt werden kann. Das Ergebnis dieser räumlichen Konfiguration ist eine gemischte Umgebung mit virtuellen Objekten, die mit den physischen Objekten aus der reellen Umgebung ausgerichtet sind. Am Ende stellen wir die Ergebnisse von Experimenten vor, die wir mit entwickelten Prototypen durchgeführt haben, um die vorgeschlagenen Ansätze zu validieren.

I. INTRODUCTION

Creating a virtual configuration of some space to see how it will look like before placing the actual physical objects in it allows trying out different combinations of objects and layouts. Through the application of mixed reality this can be done in the actual space for which the virtual configuration is intended combining the real environment with the virtual models to be placed in it. Using a stereoscopic camera capable of generating a depth map and calculating measurements of the real environment, a mixed reality application can preserve the depth information from the real environment (when presenting it to the user) and seamlessly combine it with the virtual objects.

II. DEPTH CALCULATION USING STEREOSCOPIC CAMERA

A. General Principle for Estimating Depth from Stereoscopic Images

The distance from a selected point in the 3-dimentional scene to the image plane is estimated using two images of the scene, taken from different points of view at the same time. The relations between the 3D points of objects in the scene and their projections onto the 2D image planes are described by the rules of the epipolar geometry which define constraints between corresponding points in the image planes. The general principle is illustrated on Fig. 1 [1], [2].

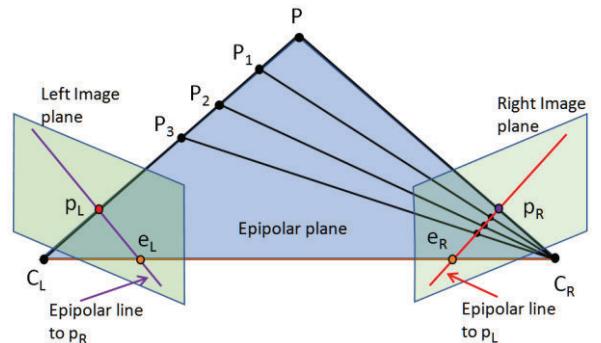


Fig. 1. Epipolar geometry of stereo vision.

The scene is captured simultaneously from two cameras (Left and Right), which are located at different positions and their optical centers are coded C_L and C_R respectively. The line between C_L and C_R forms the baseline while the three points: P , C_L and C_R build the epipolar plane. Each optical center projects into a point in the image plane of the other camera: the optical center C_L projects into e_R , while C_R projects into e_L . These two image points are called epipoles. Let's consider the 3D point P located in the scene at unknown distance from the baseline. This point is captured by both cameras and projected as 2D points: on the Left image plane as p_L and on the Right image plane as p_R . If we move the 3D point P from its initial position towards the optical center of the left camera at indicative positions P_1 , P_2 and P_3 , the corresponding projections of these points into the image plane of the other (Right) camera build a line passing through e_R called epipolar line which corresponds to the projected point p_L in the other image plane. The same is valid for the projected point p_R in the Right image plane - it is related to the epipolar line created in the other (Left) image

plane. The epipolar lines are formed at the intersection of the epipolar plane with both image planes and pass through the corresponding epipoles. If the 3D point P is moving in the scene the epipolar plane rotates around the baseline. An important feature known as epipolar constraint is the condition that if the relative position of the two cameras and the projection p_L of the 3D point P in the Left image plane are known then the projection p_R of the same 3D point P in the Right image plane must lie on the epipolar line (e_R, p_R). This constraint is valid for all points (e.g. P_1, P_2, P_3) which lie on the 3D line P, C_L . By analogy we can observe the same constraint for points which lie on the line P, C_R , their projection in the Left image plane must lie on the epipolar line (e_L, p_L). Another interpretation of the epipolar constraint is that it defines a mapping between points in the left image plane and lines in the right image plane and vice versa. This projective mapping from points to lines can be represented by the fundamental matrix F, which gives algebraic description of epipolar geometry.

The epipolar constraint allows us to calculate the position of the 3D point P from known positions of the projections p_L and p_R using the condition that if these projections correspond to the same 3D point, then the projection lines must intersect at P. The described principle called triangulation forms the basis of 3D reconstruction e.g. estimating the position of the 3D point from the known positions of the projections of this point in the image planes [3].

Several numerical methods have been developed for solving the problem for estimating the position of 3D points in the scene based on measurements of their projections on the two image planes. One common approach for estimating the position of the 3D points is to minimize the sum of the squared errors between the measured image positions and the re-projected positions applying pseudo-inverses or singular values decomposition numerical methods [1], [3], [4]. A major problem of this approach is that it fails to deliver consistent and stable results when the measurements are noisy. In such situation the RANSAC (RANdom SAmple Consensus) method is more appropriate [1], [3], [4]. It delivers a suitable result with a certain probability, which increases with the number of iterations. The big number of calculations which have to be performed in this case can be executed in parallel on GPUs [5].

B. The ZED Mini Stereoscopic Camera

To date, the technological achievement of 3D sensors has not allowed them to offer depth perception at longer ranges and outdoors. The ZED Stereo camera manufactured by Stereolabs is the first sensor to introduce longer range depth perception (up to 20m) indoors and outdoors [7]. This enables a variety of new applications in areas like robotics, augmented reality, security, etc. The ZED Mini camera allows for a smaller maximum supported range of 12-15m and due to the smaller distance between its two lenses which approximates the average human's interpupillary distance, it is better suited for seeing nearby objects.

The ZED camera reproduces the way humans' natural vision works. Human eyes are separated horizontally by about 65mm which means that each eye has a slightly different view of the scene. These different perspectives are merged by the brain to create the feeling of depth and 3D motion in space. The device uses its two lenses and triangulation to "understand" its surroundings and produce a 3D model of the observed scene. The output data consists of a high-resolution side-by-side colour video through a USB 3.0 interface. This video contains two synchronised left and right video streams which are used by the ZED application on the host computer to calculate a depth map of the scene. In a depth map the

camera stores a distance value (Z, in metric units) for each pixel (X, Y) in the image which is calculated from the back of the left lens of the camera to the scene object [7](Fig. 2).

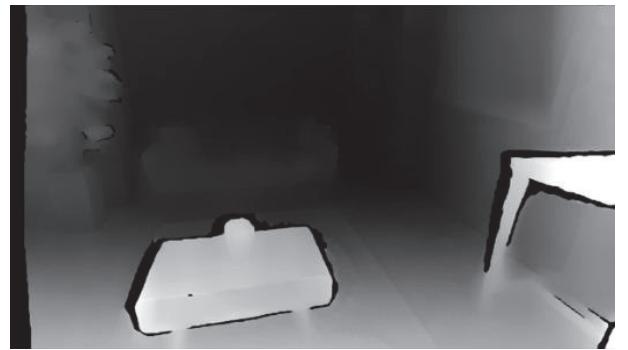


Fig. 2. A depth map created by ZED Mini [7]

The depth map is a 32-bit encoded image which cannot be displayed directly without converting it first to a monochrome image. This grayscale representation (8-bit) has values between 0 and 255 with 0 meaning the most distant possible depth value and 255 – the closest possible. A significant advantage of the device is its capability to create the depth map in real-time thus avoiding the need of a pre-scan of the environment [7].

III. APPROACHES FOR SPATIAL CONFIGURATION

In this paper we present three possible approaches for creating a virtual spatial configuration based on a real environment using the SDK of the ZED Mini stereoscopic camera. The purpose is to be able to create a virtual scene with objects that are aligned to the real environment and the physical objects in it. The three approaches we have identified are:

- Using hit tests against the calculated depth map;
- Using hit tests against reference planes generated according to surfaces identified in the depth map;
- Using hit test against a generated spatial map of the environment.

In all approaches a specialized tracked controller is used as a pointing device to generate a ray with which to specify the desired position where a virtual object should be placed.

A. Hit Testing Against the Depth Map

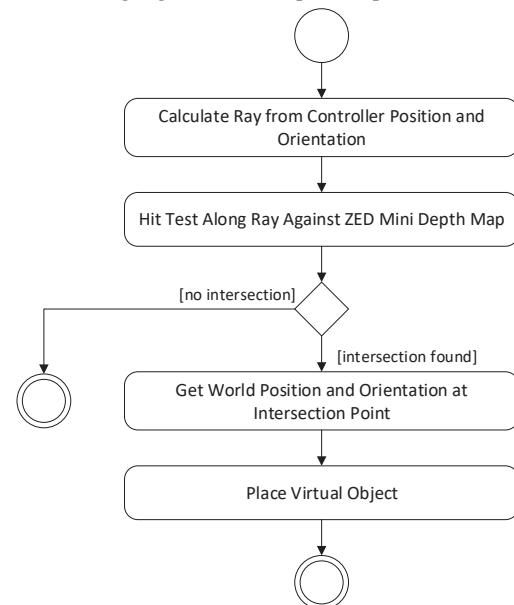


Fig. 3. Algorithm for placing a virtual object by hit testing against the ZED Mini generated depth map.

This approach follows the algorithm shown on Fig. 3. First a ray is generated starting from the tracked controller's position and in the direction of its orientation. This ray is passed to a method in the ZED Mini SDK which generates points with increasing distance (with a specified step) along the ray and for each point checks whether its distance to the camera is greater than the distance calculated in the depth map. If so a hit is detected, and the position of the point is returned. Then the normal at that position in the real environment is queried and an object is placed with the calculated (virtual) position and orientation.

B. Hit Testing Against Reference Planes

This approach is composed of two stages (see Fig. 4):

- First one or more “reference planes” are generated. The algorithm is similar to the hit test algorithm from the first approach but this time if a hit is found its position is passed to a method in the ZED Mini SDK which generates a plane containing all points from the depth map identified as belonging to the same surface.
- In the second step a hit test is performed against the generated reference planes and the virtual object is placed at the intersection point on the reference plane.

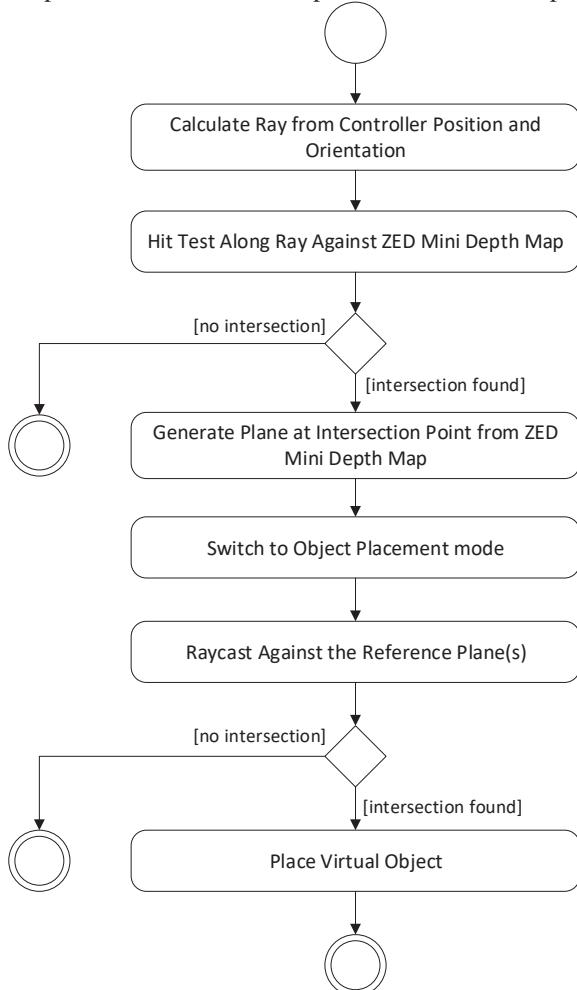


Fig. 4. Algorithm for placing a virtual object using reference planes.

C. Hit Testing Against a Generated Spatial Map

The third approach is similar to the second one – it also consists of two stages, but instead of reference planes during the first stage the ZED Mini stereoscopic camera is used to generate a spatial map of the environment. When the spatial map is complete the virtual objects are placed by hit testing against the spatial map.

IV. EXPERIMENTS

Series of Experiments were performed to validate the three implemented approaches. The experiments were conducted using a ZED Mini stereoscopic camera attached to the front of an HTC Vive Headset where HTC Vive's controllers were used for interaction.

The developed prototypes use 3D models of furniture for the creation of a virtual spatial configuration. Fig. 5 shows a sample virtual spatial configuration in a real environment.



Fig. 5. Example of creating a virtual spatial configuration.

A. Using hit tests against the calculated depth map

Pressing the trigger on the Vive Controller generates a ray that detects where the point of the environment is and places the object oriented by the normal vector of the intersection point (Fig. 6).

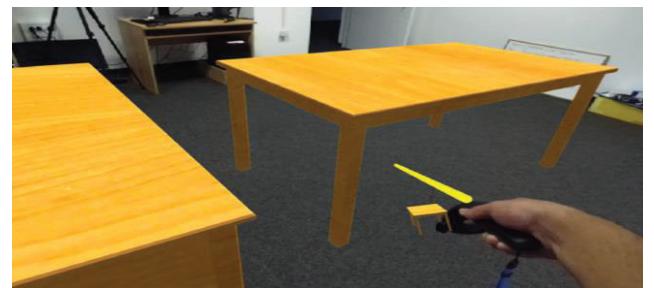


Fig. 6. Example of using hit tests against the calculated depth map.

B. Using hit tests against reference planes generated according to surfaces identified in the depth map;

In this approach when the trigger on the controller is pressed a reference plane is generated (Fig. 7). After that, virtual objects are placed on the desired position using the reference plane. Sometimes the hit point is created behind the real object and the plane is not positioned correctly to the surrounding environment. On the other hand, when the planes are generated properly several objects can be placed on one level (Fig. 8) which is more difficult to achieve with the previous method.

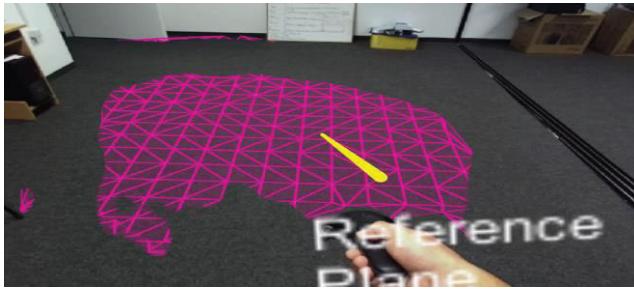


Fig. 7. Using real world hit test to generate a reference plane from surfaces identified in the depth map.

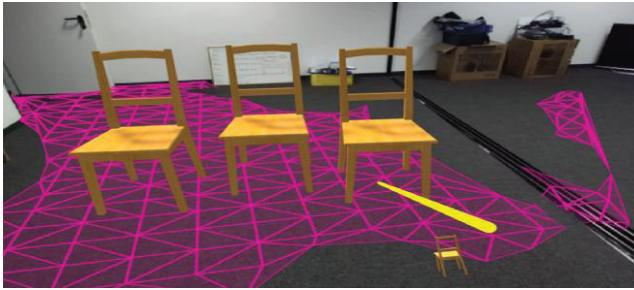


Fig. 8. Example of placing several objects on one plane.

C. Using hit test against a generated spatial map of the environment.

During the initial extraction the spatial map is created accurately, and the virtual spatial configuration can be created correctly. Once the user moves horizontally into space, the spatial map no longer coincides with the real environment, and the virtual scene with placed objects “swims” in the air at a slight distance from the ground. If the user moves vertically, the displacement from the ground in height can become significant.

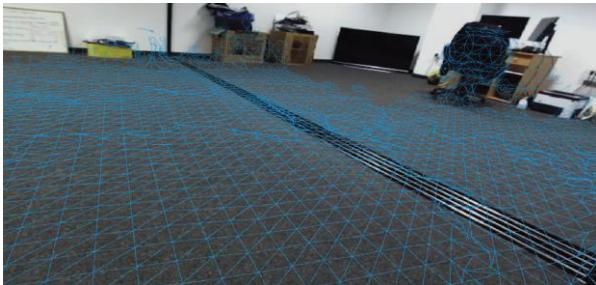


Fig. 9. Example of generating spatial map of the environment.

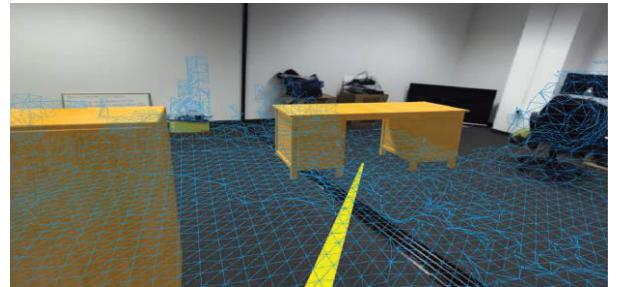


Fig. 10. Example of using hit test against a generated spatial map of the environment.

The conducted experiments showed that it is possible to use the ZED Mini stereoscopic camera in combination with an HTC Vive Headset to create a spatial configuration, but the results are not always satisfactory.

ACKNOWLEDGMENT

The results presented in this paper are obtained in the framework of research activities at the Virtual and Augmented Reality Laboratory, funded by the Research and Development and Innovation Consortium, Sofia Tech Park, project "Research and development of methods and algorithms for interaction in mixed reality environments performing tasks for spatial configuration and collocation of objects".

REFERENCES

- [1] R. Hartley, A Zisserman, *A Multiple View Geometry in Computer Vision*, 2-nd ed., Cambridge University Press, 2003.
- [2] M. Aladem, *Robust Real-Time Visual Odometry for Autonomous Ground Vehicles*, Master Thesis, University of Michigan-Dearborn, 2016.
- [3] R. Hartley, P. Sturm, "Triangulation", *Computer vision and image understanding*, Vol. 68, Nr. 2, pp. 146-157, 1997.
- [4] Y. Gao, *A user-oriented markerless augmenter reality framework based on 3D reconstruction and loop closure detection*, Ph.D. thesis, University of Birmingham, 2016.
- [5] S. Ouerghi, R. Boutteau, X. Savatier, F. Tlili. "CUDA Accelerated Visual Egomotion Estimation for Robotic Navigation". *12th International Conference on Computer Vision Theory and Applications*, Feb 2017, Porto, Portugal. SCITEPRESS - Science and Technology Publications, Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp.107-114, <10.5220/0006171501070114>.
- [6] M. Senthilvel, R. Soman, K. Varghese, "Comparison of Handheld devices for 3D Reconstruction in Construction", *34-th International Symposium on Automation and Robotics in Construction (ISARC)*, 2017.
- [7] Depth Sensing, <https://www.stereolabs.com/docs/depth-sensing/>, as seen on 18.11.2018

Prediction of EIGRP traffic parameters by 3D approximation using neural networks

Vorhersage von EIGRP-Parametern durch 3D-Approximation mittels neuronaler Netze

Irina Topalova

FDIBA, Technical University Sofia
Sofia, Bulgaria, itopalova@abv.bg

Information Technology, University of Telecommunication and Post
Sofia, Bulgaria, itopalova@abv.bg

Abstract — In the dynamic development of modern communication technologies, the use of intelligent methods for automatic adaptation of communication traffic parameters is increasingly necessary. The problems of self-decision and predictable adaptation of traffic parameters have to be solved within the autonomous system. The article presents a study of the ability of neural networks of type Multi-Layer-Perceptron and Kohonen, to adaptively predict the necessary change of the bandwidth, depending on the dynamic changes in other parameters of the dynamic routing protocol metric. The experiment was conducted for the Enhanced Interior Gateway Routing Protocol, observing different ratios between two of its basic metrics parameters - load and reliability. The input and output vectors of the neural networks are represented by a 3D approximation surface, where different approaches to approximation were used and compared. The precision of the prediction is tested when the metric parameters change. The accuracy of the proposed method is analyzed. The achieved results are represented and discussed, as well feature work is proposed.

Zusammenfassung — Bei der dynamischen Entwicklung moderner Kommunikationstechnologien wird zunehmend der Einsatz intelligenter Verfahren zur automatischen Anpassung von Kommunikationsverkehrsparametern notwendig. Die Probleme der Selbstentscheidung und der vorhersehbaren Anpassung von Verkehrsparametern müssen innerhalb des autonomen Systems gelöst werden. Der Artikel präsentiert eine Studie über die Fähigkeit neuronaler Netze vom Typ Multi-Layer-Perceptron und Kohonen, die notwendige Änderung der Bandbreite abhängig von den dynamischen Änderungen anderer Parameter der dynamischen Routing-Protokoll-Metrik adaptiv vorherzusagen. Das Experiment wurde für das Enhanced Interior Gateway Routing Protocol durchgeführt, wobei unterschiedliche Verhältnisse zwischen zwei seiner grundlegenden Parameter - Last und Zuverlässigkeit - beobachtet wurden. Die Eingangs- und Ausgangsvektoren der neuronalen Netze werden durch eine 3D-Approximationsfläche dargestellt, wobei verschiedene Approximationsansätze verwendet und verglichen wurden. Die Genauigkeit der Vorhersage wird getestet, wenn sich die metrischen Parameter ändern. Die Genauigkeit der vorgeschlagenen Methode wird analysiert. Die erzielten Ergebnisse werden dargestellt und diskutiert, sowie zukünftige Arbeit wurde vorgeschlagen.

I. INTRODUCTION

In the dynamic development of modern communication technologies, the use of intelligent methods for automatic adaptation of communication traffic parameters is increasingly necessary. The problems of self-decision and predictable adaptation of traffic parameters have to be solved within the autonomous system. At present, there are various studies and attempts for practical realizations of dynamic adjustment of the parameters of the routing protocols as well as of the routes themselves. In most modern studies, this is accomplished by applying adaptive decision-making methods such as applying different types of neural networks. This article presents a study of the ability of neural networks of type Multi-Layer-Perceptron (MLP) and Kohonen, to adaptively predict the necessary change of the bandwidth, depending on the dynamic changes in other parameters of the EIGRP (Enhanced Interior Gateway Routing Protocol) dynamic routing protocol metric. The experiment was conducted, observing different ratios between two of its basic metrics parameters - load and reliability. The input and output vectors of the neural networks

are represented by a 3D approximation surface, where different approaches to approximation were used and compared. The precision of the prediction is tested when the metric parameters change and the accuracy of the proposed method is analyzed. The next section II presents some related works, the next one section III explores the proposed method. Section IV shows the experiment and the obtained 3D surface results. Section V includes the conclusion.

II. RELATED WORKS

Many modern researchers are trying to find effective adaptive methods to optimize the parameters of IP/TCP traffic in terms of predicting the best route, by providing the appropriate bandwidth. Methods using neural networks are increasingly being used, due to their good adaptive capabilities.

The routing algorithm proposed in [1], is based on link state routing protocols, considering the number of hops, bandwidth, load and delay. The authors have created a new metric, that uses artificial neural network logic, in order to optimize route

selection, optimize network resources and reduce link occupancy. They use two different Hopfield neural networks: the first - in order to ensure that every change of network topology (or network parameters) is distributed to the network as fast as possible and the second one is dedicated to the route selection problem, based on previously collected information. The authors investigate the relation between bandwidth and load, but instead for the EIGRP - for the OSPF, because EIGRP is a proprietary Cisco routing protocol. Search for the shortest path with a Genetic algorithm (GA) is used for routing in packet switched data networks in [2]. The authors explore solution space in multiple directions at once. They claim that GA is well suited for routing problem as it explores solution space in multiple routing directions at once. But to improve the results, the authors are looking for other options for applying intelligent approaches for populating the routing table and using mutation probabilities, enhancing it to support for load balancing. The authors of [3], determine the final route through two phases. First, a Kohonen [4] neural network creates link clusters, with formulating the Cost parameter (based on Bandwidth, Density, Delay and Reliability, in the same manner as metric within EIGRP protocol, and added it to every link. With costs reduced in such a manner, the second phase of the algorithm is entered. This phase demands the use of Hopfield's neural network, which has exhibited good results while searching for the optimal route. Further, on all four parameters are being analyzed, but only the links with higher reliability are stimulated. The second phase demands the use of Hopfield's neural network, which has exhibited good results while searching for the optimal route. All of the above-mentioned methods aim to optimize routing parameters but are based on complex algorithms and calculations without examining the achieved accuracy and performance of the result.

The simple method proposed in this work presents a study of the ability of neural networks of type MLP [5] and Kohonen, to adaptively predict the necessary change of the bandwidth, depending on the dynamic changes in other parameters of the EIGRP dynamic routing protocol metric. The experiment was conducted, observing different ratios between two of its basic metrics parameters - load and reliability. The output vectors of the neural networks are represented by two different 3D approximation surface approaches when applying Kohonen SOM (Self Organized Map), compared with the results when MLP neural network is used.

III. THE PROPOSED METHOD

The proposed approach for dynamically tuning one of the parameters forming the EIGRP metric, namely the bandwidth, is based on defining a relative factor. This factor adapts dynamically, according to the neural network decision, which is taught by the other two main parameters defining the EIGRP metric - load and reliability.

A. Selecting the parameters

The parameters load, reliability and bandwidth are represented by relative units whose high values (maximum = 1) represent a strong participation in the EIGRP metric of the relevant parameter, reciprocal to the low values. They serve as recommendation about giving relatively high, low or middle bandwidth values for the selected route, depending on the combination and the ratio between load and reliability. That is the reason to train the neural network with these two parameters and teach it to give appropriate output bandwidth. The corresponding values are given in Table I.

TABLE I. RELATIVE PARAMETER VALUES

reliability	load	bandwidth requirements
0.5	0.5	0
1	0	1
1	1	1
0	1	1
0	0	1
0.5	0	1
1	0.5	1
0.5	1	1
0	0.5	1
0.5	0.25	0.5
0.75	0.5	0.5
0.5	0.75	0.5
0.25	0.5	0.5
0.25	0	1
0.75	0	1
1	0.25	1
1	0.75	1
0.75	1	1
0.25	1	1
0	0.75	1
0	0.25	1

B. Maintaining the neural network types

To examine the accuracy of the approximation, two types of neural networks - of the MLP type and the self-organizing Kohonen Map – SOM, were trained, aiming to compare the results and recommend the better method. First two different topologies of SOM were trained with structure 3-inputs and 21 outputs in the feature map. Then, the number of neurons in the map was incrementally increased to allow for greater accuracy in structures 3-inputs and 36 outputs in the map – shown in Fig. 1. The two parameters *load* (*L*) and *reliability* (*R*) were given also to the input of a MLP neural network with structure 2-5-1. The idea of the combination of the set relative values of the parameters in Table 1 consists of the following: when at least one of the two *L* or *R* parameters has a very low or very high value - the *bandwidth* (*B*) parameter can be defined by the network as a prescription/need for dynamic increase of the bandwidth. Accordingly, in all other cases where the same two parameters have average values, the requirements for bandwidth should not be changed, the process is considered to be constant and stable. The difference between the MLP and the SOM neural network is that the MLP approximates its output (*B*) depending on the values of *L* and *P*, while SOM simultaneously adapts the three variables *L*, *R* and *B*, assigning to each winner-neuron, three approximated values. However, in both cases, the three variables will approximate a 3D surface that will set *B* on its Z axis.

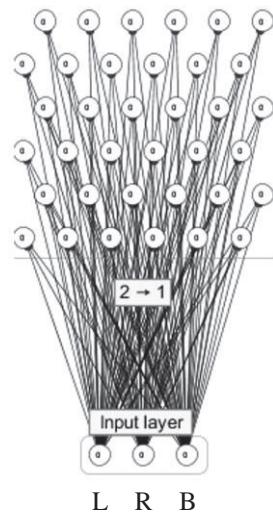


Fig. 1. Trained SOM structure with 3 inputs and 36 neurons in the map

IV. EXPERIMENTAL RESULTS

To determine the structure of the SOM that provides a good ratio between the number of neurons in the map and accuracy, two different methods to create/ approximate a 3D surface of the weight values resulting in the "winning" neurons were applied. These are the methods *Inverse Distance to a Power (IDP)* and *Minimum Curvature (MC)*.

A. IDP Method

The IDP method is fast but has the tendency to generate "bull's-eye" patterns of concentric contours around the data points. Inverse Distance to a Power does not extrapolate Z values beyond the range of data [6]. It uses the approximation calculations given in (1), where

$$\hat{Z}_j = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n h_{ij}^b} \quad (1)$$

$$h_{ij} = \sqrt{d_{ij}^2 + \delta^2}$$

is the effective separation distance between grid node j and the neighboring point i ;

\hat{Z}_j is the interpolated value for grid node j ; Z_i are the neighboring points; d_{ij} is the distance between the grid node j and the neighboring point i ; b is the weighting power parameter; δ is the smoothing parameter.

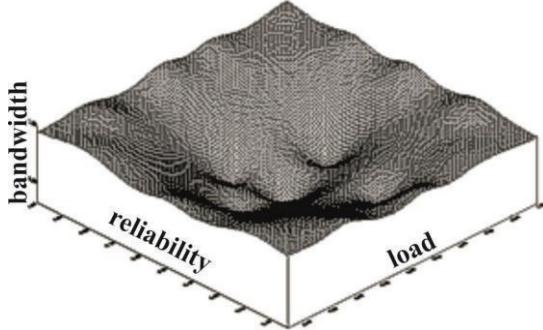


Fig.2 3D approximation results with IDP approximation method and SOM 3-inputs and 21 neurons in the map

Fig. 2 represents the obtained surface when the SOM network with 21 neurons in the map, is trained with all combinations of 3 values shown in Table 1 and fed to the three input neurons. Fig. 3 represents the obtained surface when the SOM network with 36 neurons in the map, is trained with all combinations of 3 values shown in Table 1 and fed to the three input neurons.

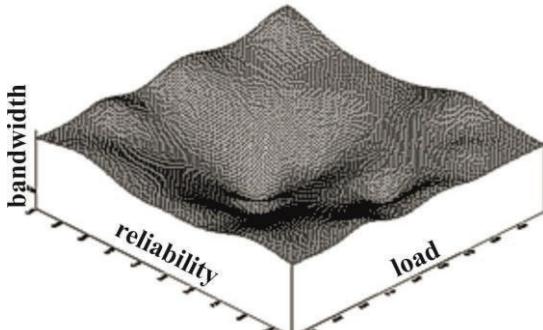


Fig.3 3D approximation results with IDP approximation method and SOM 3-inputs and 36 neurons in the map

B. MC Method

Minimum Curvature generates smooth surfaces and is fast for most data sets. The internal tension and boundary tension allow controlling over the amount of smoothing. Minimum Curvature can extrapolate values beyond data's Z range [6]. Fig. 4 and Fig. 5 represent the obtained surface, applying MC method, when the SOM network with respectively 21 and 36 neurons in the map, is trained with all combinations of the 3 values shown in Table 1 and fed to the three input neurons.

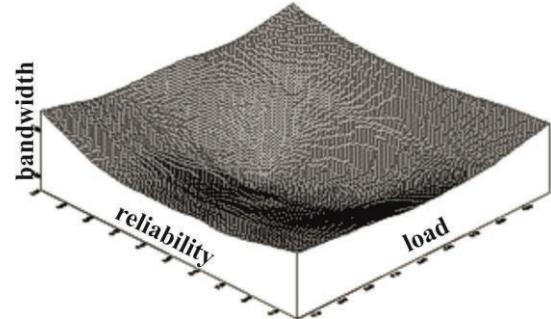


Fig.4 3D approximation results with MC approximation method and SOM 3-inputs and 21 neurons in the map

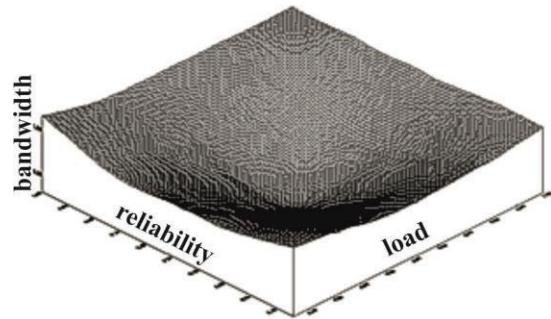


Fig.5 3D approximation results with MC approximation method and SOM 3-inputs and 36 neurons in the map

C. MLP Method

The next experiment was conducted by training the MLP neural network of structure 2-5-1. It has been trained with the input data *load* and *reliability* of Table 1, with a set output neuron value corresponding to the *bandwidth* column. Both networks SOM and MLP were trained with the same number of 1250 iterations and train error of 0.01. In this case, the graph of the 3D surface shows steeper slopes, which would reflect in more dynamic changes of the Z-Axis, i.e. of the bandwidth parameter.

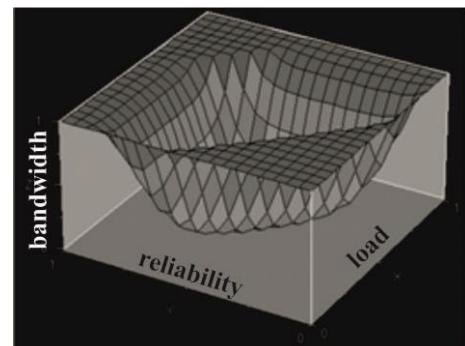


Fig.6 3D approximation results with MLP approximation method and structure 2-5-1

D. Methods comparison

To compare the accuracy of approximation using SOM and MLP for the same parameter training, the obtained graphics are shown in Fig. 7 and Fig. 8.

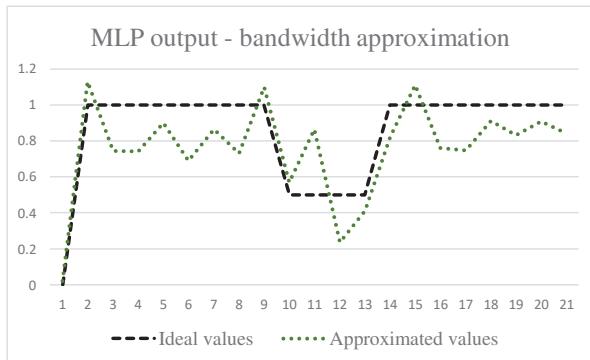


Fig.7 Approximation of MLP 2-5-1 output “bandwidth” with ideal (desired) and the obtained values

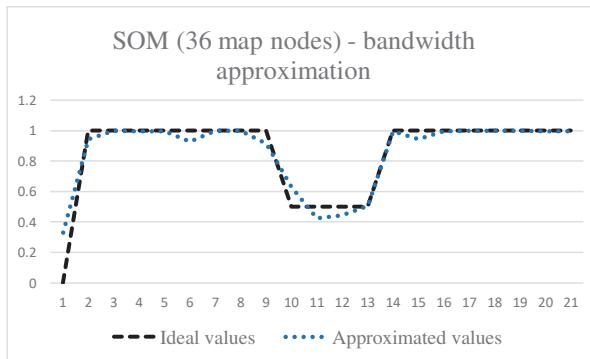


Fig.8 Approximation of MLP 2-5-1 output “bandwidth” with ideal (desired) and the obtained values

The error $Approx_{err}$ of approximating the output neuron of the MLP and B neuron of the SOM networks is calculated using the equation given in (2),

$$Approx_{err} = \sqrt{\frac{\sum_{i=1}^N (BW_{ideal} - BW_{app})^2}{N}} \quad (2)$$

where BW_{ideal} is the desired bandwidth and BW_{app} represents the approximated value. Fig. 9 shows the obtained results for $Approx_{err}$ for SOM-map-21; SOM-map-25; SOM-map-36 and MLP 2-5-1. It is obvious, that with increasing the number of neurons in the SOM map, the accuracy of approximation

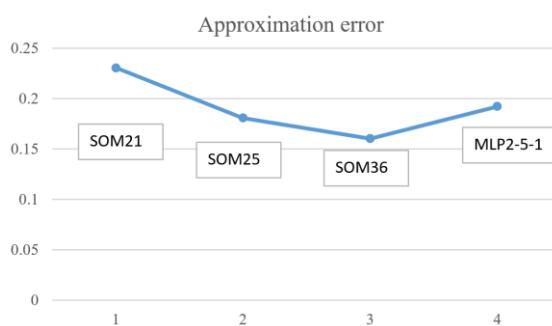


Fig.9 Approx_{err} for different SOM maps and for MLP 2-5-1

improves. The calculation according to equation (2) is done for the 3D approximation, according to the MC method, since according to the graphs of Fig.2-5 it produces better results.

V. CONCLUSION

The represented paper is a study of the ability of neural networks of type Multi-Layer-Perceptron (MLP) and Kohonen, to adaptively predict the necessary change of the bandwidth, depending on the dynamic changes in other parameters of the EIGRP dynamic routing protocol metric. The experiment was conducted, observing different ratios between two of its basic metrics parameters - load and reliability. The 3D approximation surface was obtained, where different approaches to approximation were used and compared. The precision of the prediction is tested when the metric parameters change and the accuracy of the proposed method is analysed.

The results obtained show, that compared to the MLP, the better method is that with the application of the SOM neural network, which according to Fig. 9 gives the lowest value of the error of approximation. In MLP case, the graph of the 3D surface shows steeper slopes, which would reflect in more sharp changes of the Z-Axis - of the bandwidth parameter. It is obvious, that with increasing the number of neurons in the SOM map, the accuracy of approximation improves. Furthermore, it is evident that the MC method is superior to the IDP method for 3D approximation, as the comparison of the graphs of Figs. 2-5 shows a more uniform surface representation, which would result in a smoother change of the bandwidth parameter.

This pre-comparative study is necessary to precede the implementation of the real-time method in communication in topology working with the specified routing protocol, which is also the subject of further research development.

REFERENCES

- [1] N. Kojić, I. Reljin and B. Reljin, “A Neural Networks-Based Hybrid Routing Protocol for Wireless Mesh Networks,” *Sensors* 2012, 12(6), 7548-7575; doi:10.3390/s120607548, June 2012.
- [2] G. Krishna, N. K. Pandey and S. Bajpai, “Optimal Path Routing in Variable Data Network using Genetic Algorithm,” *International Journal of Engineering Science & Humanities*, ISSN 2250-3552, pp.1-15, 2012.
- [3] N. Kojić, M. Pavlović and I. Reljin, “Dynamic Routing Algorithm based on Artificial Intelligence,” *ICIST 2012 - 2nd International Conference on Information Society Technology*, 29.2. - 3.3, Kopaonik, Serbia, 2012.
- [4] T. Kohonen, *Self-Organizing Maps*, New York, Springer-Verlag, 1997.
- [5] NeuroSystems V4.0 Manuall. Germany: Siemens GmbH, 2016.
- [6] Surfer, *User’s Guide, Contouring and 3D Surface Mapping for Scientists and Engineers*, Golden Software, Inc., 2006.

Session Mechatronics

Chair: Vasil Galabov

THIS PAGE IS INTENTIONALLY LEFT BLANK

Navigator System for Virtual Manufacturing Process

Navigator-System zum virtuellen Fertigungsprozess

Dr.-Ing. Hashem Badra*, Prof. Dr. Dr.-Ing. Dr. h. c. Jivka Ovtcharova†

*Institute for Information Management in Engineering, Karlsruhe Institute of Technology
Karlsruhe, Germany, hashem.badra@kit.edu

†Institute for Information Management in Engineering, Karlsruhe Institute of Technology
Karlsruhe, Germany, Jivka.ovtcharova@kit.edu

Abstract — Due to the constantly growing complexity of products and their manufacturing processes, product manufacturing processes should be effectively controlled and managed. In this article, a navigator system is presented and briefly described in which the virtual manufacturing processes are effectively and thoughtfully controlled and managed. It is also referred to as a system network in which systems or software are integrated that make a major contribution to the execution of the virtual manufacturing processes of a product.

Zusammenfassung — Aufgrund der permanent wachsenden Komplexität der Produkte und deren Fertigungsprozesse sollten die Produktherstellungsprozesse effektiv gesteuert und gemanagt werden. In dem Beitrag wird ein Navigator-System dargestellt und kurz beschrieben, in dem die virtuellen Fertigungsprozesse effektiv, durchdacht gesteuert und gemanagt werden. Es wird auch als System-Verbund, in dem Systeme oder auch Software integriert werden, bezeichnet. Diese Systeme leisten einen großen Beitrag zur Ausführung der virtuellen Fertigungsprozesse eines Produkts.

I. EINFÜHRUNG

Die aktuellen Anforderungen des Marktes, wie Varianten, Flexibilität, kurzer Produktlebenszyklus usw. stellen die produzierenden Unternehmen ständig vor neuen und wachsenden Herausforderungen, damit sie sich weiter erfolgreich behaupten und um wettbewerbsfähig bleiben zu können. Dies kann durch eine wissenschaftliche und praxisnahe sowie durchdachte Fabrik-/Fertigungsplanung entstehen. Die Steuerung und Regelung der Fertigungsprozesse leisten in diesem Fall einen bedeutenden Beitrag. In diesem Artikel wird eine Methodik zur Steuerung und Regelung der virtuellen Fertigungsprozesse mittels eines sog. Navigator-Systems dargestellt, das im Rahmen der Fabrikplanung verwendet werden kann.

II. MOTIVATION

Die jetzigen und zukünftigen mechatronischen Produkte beruhen vielfach auf dem Zusammenwirken von Mechanik, Elektronik, Regelungstechnik und Softwaretechnik, die durch komplexe Bauteile und Funktionalitäten gekennzeichnet sind und die von immer komplexer werdenden Produktentstehungs- und Produktionsprozessen begleitet werden [1]. Die Herstellung solcher Produkte ist mit den klassischen Vorgehensweisen wirtschaftlich nicht beherrschbar. Für die Planung und die Entstehung solcher Produkte werden zunehmend neue Technologien bspw. der Ansatz der digitalen Fabrik mit deren Werkzeugen eingesetzt. Somit ist die Digitale Fabrik auch als Muster des Virtuellen Engineerings zu betrachten [6], [7].

III. PROMBLEMSTELLUNG

Der virtuelle Fertigungsprozess eines Produkts, der mittels mehrerer und unterschiedlicher Tools von verschiedenen Softwareherstellern ausgeführt werden können, kann in das Konzept „digitale Fabrik“ eingebettet bzw. integriert werden. Die meisten existierenden Tools bzw. Werkzeuge für die Fabrik-/Fertigungssimulation wurden als Einzellösungen (sog. Insellösungen) entwickelt und sind in der Praxis im Einsatz. Diese Einzellösungen sind für bestimmte Einzelaufgaben bspw. die Montagesimulation (vgl. Abb. 1) geeignet [3].



Abb. 1. Einzellösungen/Insellösungen zur virtuellen Fertigung

Beim Einsatz der Insellösungen für die virtuellen Fertigungsprozesse entstehen mehrere Probleme, z. B. ungünstiger Datenaustausch, häufiges Fehlerauftreten etc. Hieraus resultiert, dass Produkte zu späteren Zeitpunkten auf

dem Markt eingeführt werden. Somit verfügen die heutigen Lösungen für CAD und digitale Fabrikplanung über ungenügendene Durchgängigkeit [9]. Dies hat zur Folge, dass die Konkurrenzfähigkeit leidet. Daher entsteht der Bedarf nach Verbesserungen einerseits zum Datenaustausch und andererseits zum Navigieren unter den eingesetzten Tools unterschiedlicher Herkunft, um die Steuerung und Regelung in dem virtuellen Fertigungsprozess zu verbessern bzw. zu optimieren.

IV. ZIELSETZUNG

Als Lösungsansatz der dargestellten Problematik bietet sich ein Tool-Verbund (auch Navigator-System genannt) basierend auf einer graphischen Modellierungsmethode an, in dem die Durchführung der virtuellen Fertigungsprozesse (Teilefertigung, Montage und Endkontrolle) und Modellierung der Produkte sowie Produktionslinien stattfindet.

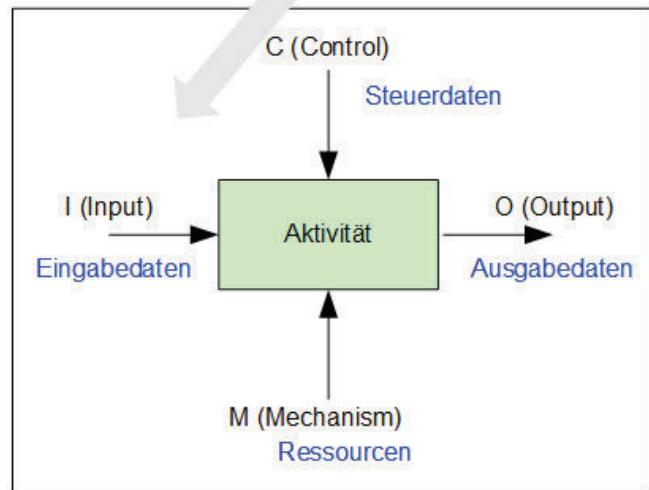
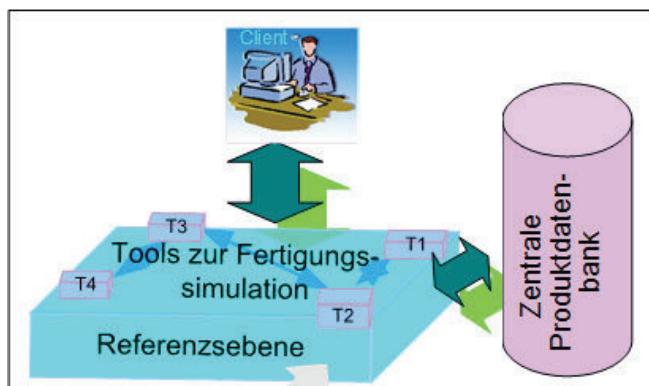


Abb. 2. Ansatz zum Toolverbund/ Navigator-System basiert auf der SADT-Methode

V. METHODEN ZUR MODELLIERUNG

Da die Fertigungssysteme eigentlich als Subsystemen der Fabrik zu betrachten sind, wird ihre Planung im Rahmen der Fabrikplanung durchgeführt. Für solche Planung existieren diverse Methoden, die aufgrund der Planungsdauer und –häufigkeit den heutigen Anforderungen nicht entsprechen. Deshalb ist eine Rechnerunterstützung im gesamten Planungsprozess erforderlich, um eine möglichst schnelle Umsetzung der Kundenwünsche und Produktideen zu erzielen und konkurrenzfähig bleiben zu können.

Hierfür werden Methoden zur Prozessmodellierung eingesetzt, die z. B. Fertigungssysteme realitätsnah beschrieben und darstellen können. Graphikbasierte Methoden sind zur Modellierung

von Prozessen wie „Produkt fertigen“ und „Produkt montieren“ sehr gut geeignet. Mit der Prozessmodellierung können mehrere Ziele erreicht werden, u. a. die Vermeidung von Prozessschleifen, Medienbrüchen und die Bestimmung von Abhängigkeiten der Prozesse untereinander [2]. Ein Vergleich mehrerer graphikbasierten Methoden, die zur Anwendung für die Modellierung von virtuellen Fertigungsprozessen in Frage kommen, weist darauf hin, dass die SADT-Methode „Structured Analysis and Design Technique“ aus mehreren Gründen u. a. der Übersichtlichkeit, Rückführung, Navigation und Verfeinerung bis zu einem definierten Detaillierungsgrad geeignet ist [4].

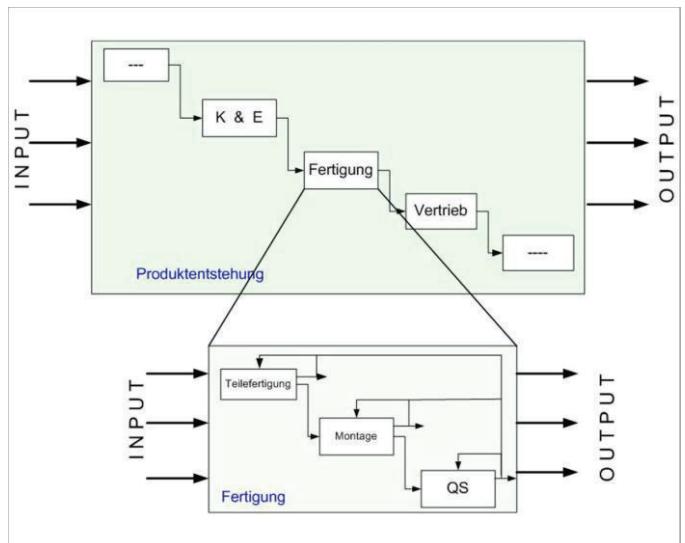


Abb. 3. SADT-Diagramm - virtuelle Fertigung

VI. SADT-METHODE UND VIRTUELLER FERTIGUNGSPROZESS

Abb. 3 zeigt die Prozesskette „virtuelle Fertigung / virtuelle Produktentstehung“ in einem SADT-Diagramm. Hierbei sind alle Hauptbereichen bzw. Teilprozesse Teilefertigung, Montage und Endkontrolle/ Prüfung vorhanden, die sequentiell, parallel oder auch zeitlich überlappend anzutragen sind [4]. Für die virtuelle Fertigung sind die Daten aus der Konstruktion und Entwicklung als INPUT zu betrachten, die innerhalb des virtuellen Fertigungsprozesses mittels bestimmten Tools verarbeitet und als OUTPUT ausgegeben werden. Weitere Konkretisierung bzw. Detaillierungen der virtuellen Fertigung lassen sich in Abb. 4 darstellen. Dabei lässt sich erkennen, dass innerhalb der Aktivitäten z.B. virtuelle Teilefertigung die Eingabedaten von den Ressourcen (Software, Methoden, Personal) verarbeitet und in Ausgabedaten ausgegeben werden.

3D-Konstruktionen und technische Zeichnungen sind als Eingangsdaten und auch als Basis für die virtuelle Fertigung zu bezeichnen, die für die virtuelle Teilefertigung der Einzelteile von den Ressourcen verarbeitet werden. In dem Teilprozess virtuelle Montage werden die schon virtuell gefertigten Einzelteile montiert. Dies erfolgt entweder manuell mittels eines Menschmodells oder automatisch mittels Roboter-Modells. Eine hybride virtuelle Montage kann auch mittels bestimmter Software verwendet werden.

Als Ergebnis der virtuellen Montage sind digital montierte Produkte oder auch Baugruppen zu betrachten, welche in dem letzten Schritt der virtuellen Fertigung auf ihre Funktionsfähigkeit virtuell geprüft werden.

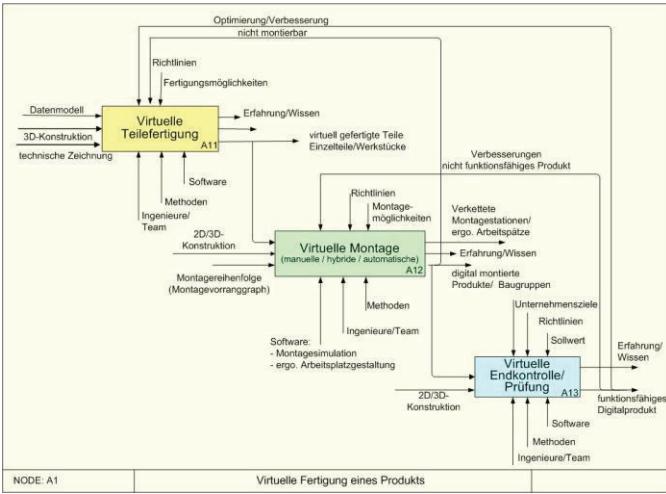


Abb. 4. Detaillierung der virtuellen Fertigung basierend auf die SADT-Methode

VII. METHODIK DER STEUERUNG UND DER NAVIGATION

Für die Steuerung und die Navigation der virtuellen Fertigungsteilprozesse sowie die durchgängig kontinuierliche Verarbeitung der innerhalb des Prozesses „virtuelle Fertigung“ generierten Daten und Dateien ist ein Leitsystem bzw. ein Navigator-System (NS) erforderlich. Dieses System soll eine gemeinsame Infrastruktur mit einer grafischen Benutzeroberfläche besitzen. Beim Entwurf des NS sind einige Anforderungen u.a. die Navigation unter den einzelnen Tools ohne Medienbruch zu beachten [4]. Innerhalb des NS sind die virtuellen Fertigungsprozesse nach der SADT-Methode verbunden. Diese Prozesse können sequenziell, parallel, zeitversetzt oder auch bei Möglichkeit gleichzeitig mittels externer Anwendungen ausgeführt werden.

VIII. AUFBAU DES NAVIGATOR-SYSTEMS

Abb. 5 vermittelt einen groben Überblick über den Datenfluss des Navigator-Systems (NS) und dessen Bestandteile. Es besteht aus einer gemeinsamen und graphischen Benutzeroberfläche (BO), einem Prozess-management-System (PMS) sowie einer zentralen Produktdatenbank (ZPDB).

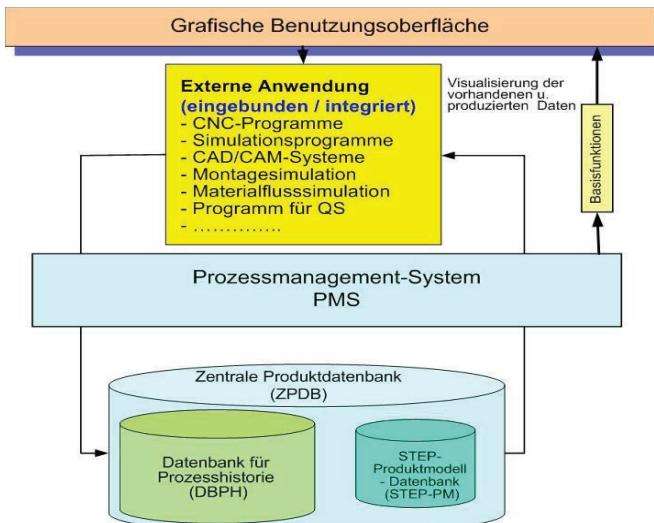


Abb. 5. Überblick über die Bestandteile / Datenfluss im NS [4]

Auf der BO können Produkt- und Prozessdaten visualisiert und externe Anwendungen (CAD/ CAM-Systeme, Simulationsprogramme) für den virtuellen Fertigungsprozess ausgeführt werden. Das PMS ist die Kommandozentrale zur Steuerung des Datenflusses innerhalb des NS. Die ZPDB wird zur Abspeicherung und Archivierung der Produkt- und Prozessdaten verwendet. Während des virtuellen Fertigungsprozesses findet ein Zusammenspiel der erwähnten Systemkomponenten statt [3].

IX. DARSTELLUNG DES NS

Das NS lässt sich in Abb. 6 darstellen. In der grafischen BO werden die virtuellen Teilprozesse der virtuellen Fertigung in einem Hauptmenü platziert, die auf Wunsch graphisch und basierend auf der SADT-Methode [4], [5], [6], auf der Arbeitsfläche dargestellt werden können. Die Verknüpfungen zu den externen Anwendungen sind ebenfalls in der BO platziert und können bei Bedarf verwendet werden. Zur Ausführung eines Teilprozesses der virtuellen Fertigung eines Produkts werden die entsprechenden externen Anwendungen in der Benutzungsoberfläche durch den Programmstart aktiviert, die in einem separaten Fenster in der Arbeitsfläche ausgeführt werden. Im Hintergrund werden den aktivierte externen Anwendungen benötigte Daten und Dateien aus der ZPDB über das PMS gesendet.

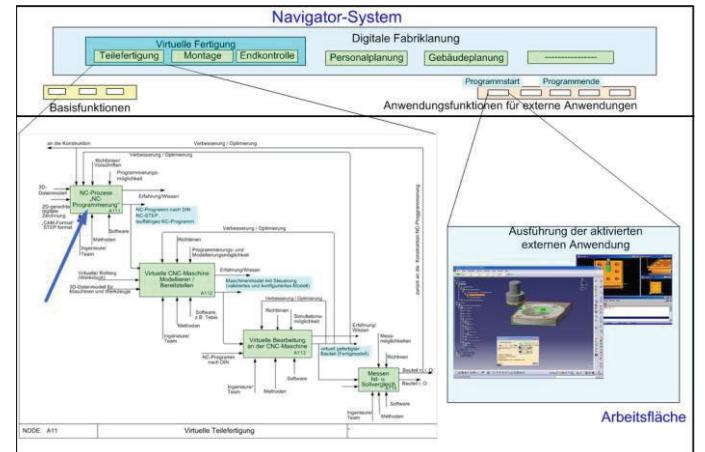


Abb. 6. Virtuelle Teilefertigung im Navigator-System

X. ANWENDUNGSBEISPIEL

Zur Verifizierung der entwickelten Methode soll das Produkt Hydropumpe virtuell gefertigt werden (vgl. Abb. 7). Hierfür werden mehrere Eingangsdaten dem NS über das PMS von einer externen Datenstelle importiert, die in der ZPDB entsprechend zur Weiterverwendung abgelegt und archiviert werden. Hier gilt wie oben erwähnt, dass die Navigation unter den Prozessen/Teilprozessen basierend auf der SADT-Methode auf der gemeinsamen Ebene BO des NS erfolgt. Diese Prozesse/Teilprozesse werden mittels externer Anwendungen ohne Unterbrechungen (Medienbrüche) ausgeführt. Hierfür ist der Anstoß der *virtuellen Fertigung* des Produkts „Hydropumpe“ vorzunehmen. Dabei sind externe Anwendungen für die Teilprozesse „virtuelle Teilefertigung“, „virtuelle Montage“ und „virtuelle Prüfung“ nötig. Die digitalen Bauteile, Baugruppen sowie Kauf- und Normteile sind aus mehreren CAD-Systemen ins NS über eine externe Datenschnittstelle zu importieren.

Beispielhaft wird lediglich die virtuelle Teilefertigung für das Teil „Gehäuse“ innerhalb des NS durchgeführt. Zur Durchführung dieses Teilprozesses „Teile-fertigung“ werden geeignete und im NS integrierte externe Anwendungen in der grafischen BO des NS aktiviert bzw. gestartet, die autonom und unbeeinflusst vom NS arbeiten. Diese externen Anwendungen öffnen sich in einem neuen Fenster mit eigenem Menü und führen bspw. den NC-Prozess für das Gehäuse, den Prozess zur Modellierung virtueller CNC-Maschinen usw. aus.

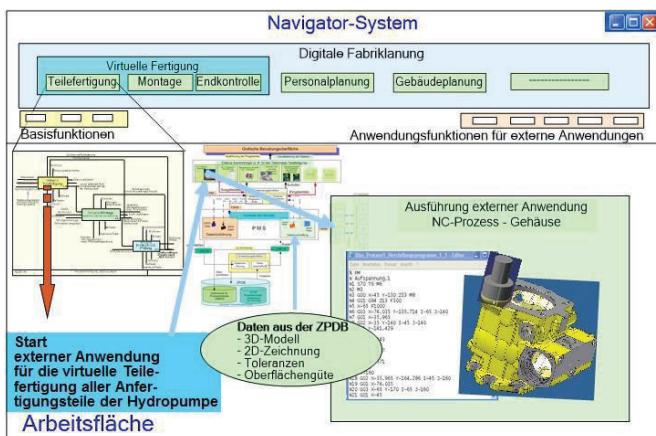


Abb. 7. Hydropumpe-Gehäuse im Navigator-System

Hierbei werden die benötigten Daten des Gehäuses, wie 2D-Zeichnung, 3D-Modell des Gehäuses usw. durch das PMS aufbereitet und der aktivierte externen Anwendung gesendet. Die durch die externe Anwendung modifizierten Daten (hier das Modell für das bearbeitete Gehäuse, NC-Programm) werden über das PMS an die ZPDB zurückgesendet und dort sortiert sowie passend abgelegt. Diese Daten stehen für die weitere Verwendung in der virtuellen Montage und Prüfung zur Verfügung.

LITERATURVERZEICHNIS

- [1] T. Günther, R. Brandis, D. Nordsiek, S. Peters, „Frühzeitige und qualifizierte Kostenprognose für Produktions2systeme“. In: BMF-Kongress 10. Karlsruher Arbeits-gespräche Produktionsforschung 2010, KIT Karlsruhe.
- [2] R. Anderl, „Produktdatentechnologie -Produkt- u. Prozessmodellierung, Skript“ SS2011, Fachgebiet Datenverarbeitung TU Darmstadt.
- [3] B. Awiszus, A. Hoffmann, et al, „Durchgängige Virtualisierung der Entwicklung u. Produktion von Fahrzeugen“. Gemeinsamer FuE-Abschlussbericht, BMBF 2012.
- [4] H. Badra, „Konzeptionelle Entwicklung einer Methodik zur Navigation in der Fertigungssimulation im Rahmen der Fabrikplanung“, Dissertation, TU Chemnitz, 2012.
- [5] H. Badra, E. Müller, „Prozessmanagement in den Fertigungssystemen - Navigator System“, Konferenz VPP2013 – Trend and Strategien für die Produktion von morgen – 24.10.2013, Chemnitz 2013.
- [6] H. Badra, „Virtual Manufacturing- Navigator System“ In: 5th International Conference on Engineering and Sustainability (ICES5), Dec. 2014 Islamic University of Gaza, Palestine.
- [7] J. Ovtcharova, „Virtuelles Abbild - neue Ingenieurmethoden für Industrie 4.0“. In: 3.Fachkonferenz zu VR/AR-Technologien in Anwendung und Forschung an der Professur Werkzeugmaschinen und Umformtechnik. Technische Universität Chemnitz, 2015.
- [8] J. Ovtcharova, „Virtual Engineering I, Vorlesungskript 2007“, Institut für Informationsmanagement im Ingeneurwesen (IMI) Karlsruhe Institut für Technologie (KIT)
- [9] N.N., „Virtuelle Inbetriebnahme“, IT&Production, https://www.it-production.com/produktentwicklung/virtuelle_inbetriebnahme/- Nov. 2018

Integral Brake System for Electric Motorcycle Using Maximum Energy Regeneration

Integrales Bremssystem für elektrisches Motorrad mit maximaler Energieregeneration

Georgi Yanachkov*, Simona Hesapchieva†

* Department of internal combustion engines, automobiles and transport, Technical University of Sofia
Sofia, Bulgaria, *gyanachkov@tu-sofia.bg, †Simona_Hesapchieva@abv.bg

Abstract — This publication discusses the possibilities of regenerative braking of an electric motorcycle. The peculiarities of the construction of the motorcycle and the braking forces acting on it are discussed. The braking performance of the rear wheel has been examined and the brake force in which the maximum energy recovery rate is found. An integrated braking system model has been proposed to allow maximum energy recovery during braking.

Zusammenfassung — In der Artikel werden die Möglichkeiten des regenerativen Bremsens eines elektrischen Motorrads diskutiert. Die Besonderheiten der Konstruktion des Motorrads und der darauf wirkenden Bremskräfte werden diskutiert. Die Bremsleistung des Hinterrads wurde untersucht und die Bremskraft, bei der die maximale Energierückgewinnungsrate gefunden wird, ermittelt. Ein integriertes Bremssystemmodell wurde vorgeschlagen, um eine maximale Energierückgewinnung während des Bremsens zu ermöglichen.

I. INTRODUCTION

Electric vehicles are increasingly spreading. Their advantages over those equipped with ICE are that they do not pollute the environment, the electric motor has a higher efficiency, there is the possibility of reverse and stop-mode operation, higher acceleration, noise, lower cost per kilometer road.

The disadvantage is the low capacity of the batteries and their slow recharging. For this reason, energy regeneration systems have been developed in which the electric motor operates in a generator mode and the resulting electricity is used to power the on-board systems and to charge the battery. This extends the range and efficiency of the vehicle.

Motorcycles are characterized by a relatively high mass center and a short wheelbase, resulting in greater redistribution of the braking forces. As is known from the theory of the vehicle in braking mode, a torque is obtained that unloads the rear wheel. The shorter the base and the higher the center of the mass center, the greater the moment. In most motorcycles, there is a real danger of an intense braking momentum increasing to the extent that the rear wheel loses contact with the road before it reaches the adhesion limit. Exceptions are the heavy motorcycles with a long base.

Another feature of the motorcycle is that it is always with a rear wheel drive and only the braking force in it can be used to drive the electric motor in a generating mode for energy regeneration. Options for the deployment of a generator on the front wheel are not considered, this would increase the weight of the construction and lead to an increase in non-suspended masses in the suspension. This publication will examine how much of the total braking force can be achieved only with the rear wheel and, respectively, what part of the braking energy can be regenerated.

II. FORCES ACTING ON THE MOTORCYCLE IN BRAKING MODE AND BASIC GEOMETRIC PARAMETERS

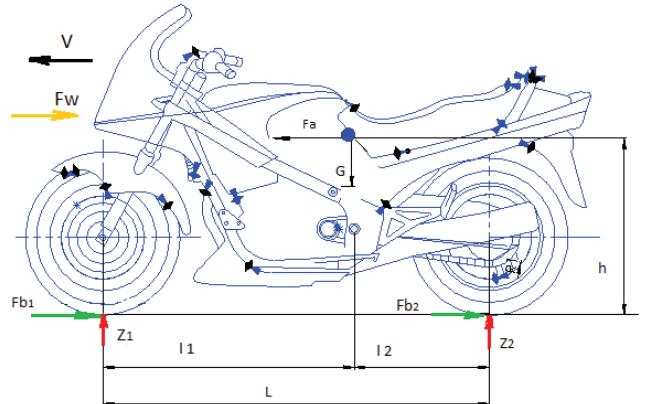


Fig. 1. Forces acting on the motorcycle in braking mode.

On the scheme on Fig. 1 are made the following indications [2]:

- Z_1 - normal front wheel reaction, N;
- Z_2 - normal rear wheel reaction, N;
- F_w – aerodynamic resistance force, N;
- F_{b1} - front wheel braking force, N;
- F_{b2} - rear wheel brake force, N;
- F_a - Inertial force, N;
- L - wheelbase of the motorcycle, mm;
- h - height of the center of gravity, mm;
- l_1, l_2 - coordinates of the mass center respectively by the front and rear axes, m
- G - mass of the motorcycle, N

III. BRAKING FORCE IN THE REAR WHEEL

To obtain how much of the braking energy can be regenerated, only the braking force applied to the rear wheel of the motorcycle is considered because it is connected to the

electric motor. While braking the electric motor works as a generator.

The maximum braking deceleration achieved only with the rear brake is[1]:

$$a_{br.opt.} = \frac{\mu g \cdot l_1}{\delta(L + \mu h)} \quad (1)$$

For greater braking deceleration, it is also necessary to activate the front brake. This results in an additional unloading of the rear wheel and lowering the corresponding braking force. This leads to a reduction of the percentage of regenerated energy relative to the total braking energy.

IV. OPPORTUNITIES TO RECOVER THE ENERGY CONSUMED WHEN THE REAR WHEEL BRAKES. INTEGRAL BRAKE SYSTEM

In the case of low-intensity braking, it is possible to stop only with the rear wheel, which means that all braking energy can be restored as is shown in Fig.2.

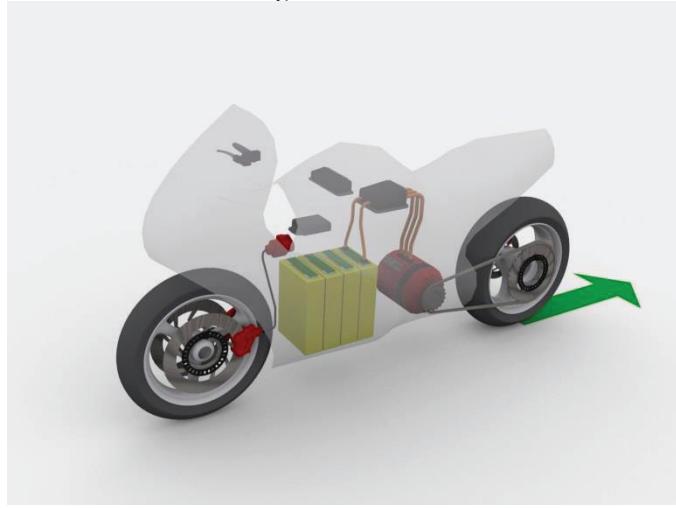


Fig. 2. Motorcycle braking with braking force applied only on the rear wheel.

For higher braking deceleration, it is also necessary to activate the front brake. Due to the progressive unloading of the rear axle, the braking effort on the rear wheel needs to be reduced Fig.3, otherwise the rear wheel will lose traction with the road, will slip and the motorcycle will lose stability.

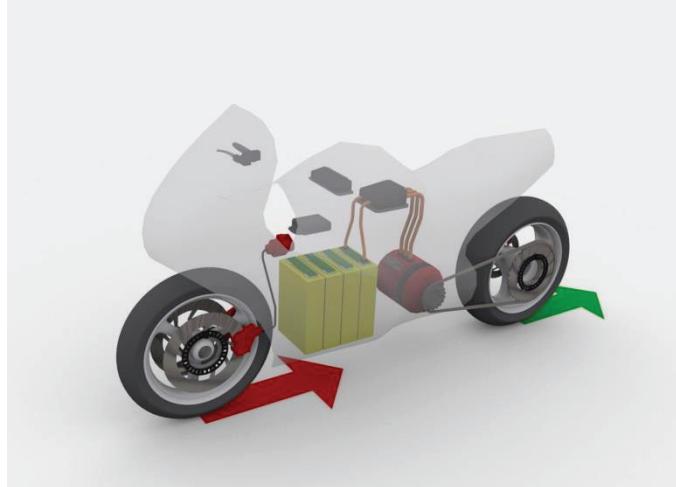


Fig. 3. Motorcycle braking with both wheels.

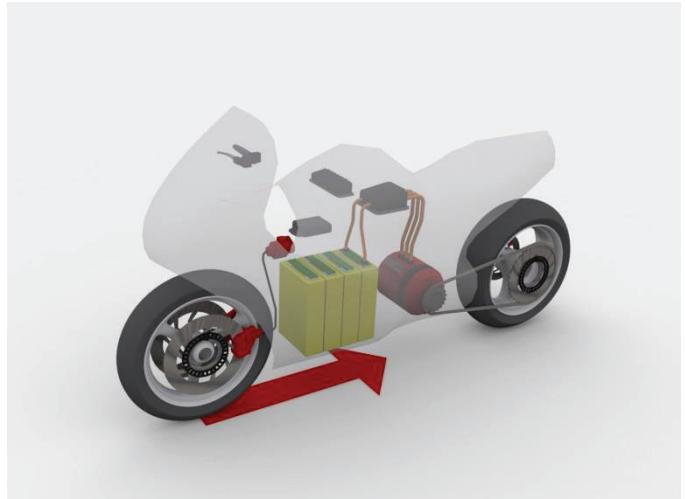


Fig. 4. Motorcycle braking with braking force applied only on the front wheel.

There is an extreme case of braking when the braking force in the front wheel is so great that lifts the rear wheel. In these case the motorcycle will lose contact with the road with his rear wheel and will flip over. That depends on the dimensions of the motorcycle – wheelbase and center of masses, and the traction coefficient [1]. The anti – flipover condition is:

$$\mu > \frac{l_1}{\delta h} \quad (2)$$

The higher the center of masses and the shorter the wheelbase, the more prone to flip is the motorcycle. Especially vulnerable to that effect are the super-sport bikes and city scooters.

The exact law on the distribution of braking forces depends on the geometric and mass parameters of the particular motorcycle. To make a simple example the dimensions of Honda CBR1100 XX are used. This is a sport-touring motorcycle with widespread proportions. The dimensions are shown in Table I. The traction coefficient is $\mu = 0.8$ – something common for an asphalt road surface. The exact distribution law is shown in Fig. 2

TABLE I.

Designation	Honda CBR 1100 XX		
	Parameter	Value	Dimention
L	Wheelbase	1450	mm
l_1	Coordinates of the mass center respectively from the front wheel.	725	mm
l_2	Coordinates of the mass center respectively from the rear wheel.	725	mm
h	Height of the center of masses.	800	mm
G	Mass of motorcycle	330	Kg

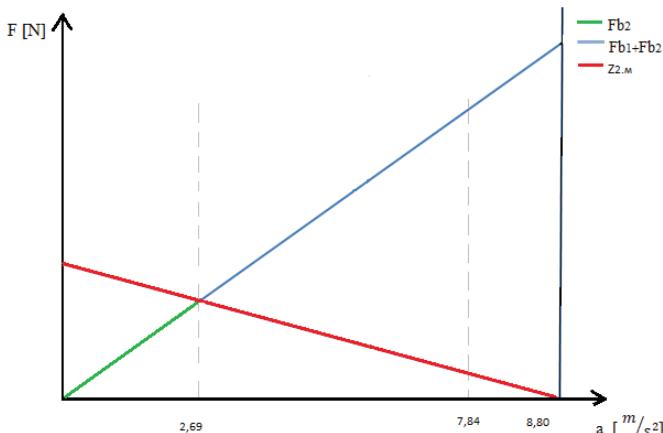


Fig. 5. Proportional distribution of braking forces between front and rear wheels depending on braking deceleration.

As seen on the figure in the area of brake decelerations below $2,69 \text{ m/s}^2$ it is possible to stop only with the rear wheel. In that case all of the braking force can be transferred to the motor/generator and be converted to energy. The maximum deceleration is $7,84 \text{ m/s}^2$. If the traction coefficient was higher, then the motorcycle can brake more intense until the deceleration is $8,80 \text{ m/s}^2$. By that time the braking force is so high that the rear wheel loses traction and the motorcycle flips over. For the current calculations, the coefficient of the rotating masses, the rolling resistance and the air resistance are not taken into account.

To achieve the optimum distribution of braking forces between the wheels an Integral Brake System is needed.

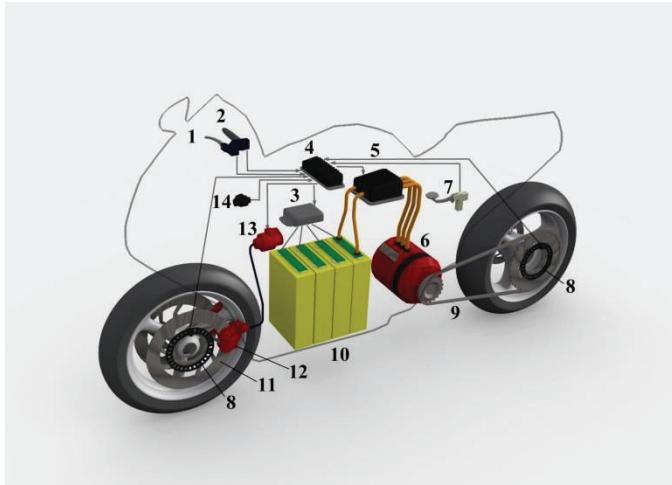


Fig. 6. Basic schematic view of an Integral Brake System optimized for maximum energy regeneration.

The Integral Brake System consists of following elements as shown on Fig. 6:

- 1 - Brake lever;
- 2 - Throttle;
- 3 - Battery Management System;
- 4 - Boardcomputer;
- 5 - Controller;
- 6 - Electric motor;
- 7 - Brake pedal;
- 8 - ABS sensor;
- 9 - Final drive;
- 10 - Battery pack;
- 11 - Front brake disk;

- 12 - Front brake caliper;
- 13 - Hydrocylinder;
- 14 - Braking deceleration sensor;

The main control instrument of the system is the hand brake lever. The input value is the brake handle position and its corresponding brake deceleration. The relation between them can be linear or nonlinear, depending on the rider preferences. Other source of input value can be the foot brake pedal. The brake handle sends signal to the boardcomputer and commands to the motor controller. The controller is connected to the battery pack and its battery management system. For applying brake force to the front wheel is used the standard brake disk and caliper. The only difference is that the brake is not directly controlled by the handle, but through an electrohydraulic cylinder, controlled by the boardcomputer. The ABS sensors for the front and the rear wheel are also connected to the boardcomputer. Information for the intensity of the braking is given by a braking deceleration sensor. The system can be further complicated by other sensors – for leaning in the corners, tyre condition, weather condition, etc.

The simplified diagram of the functioning system is shown on Fig. 7.

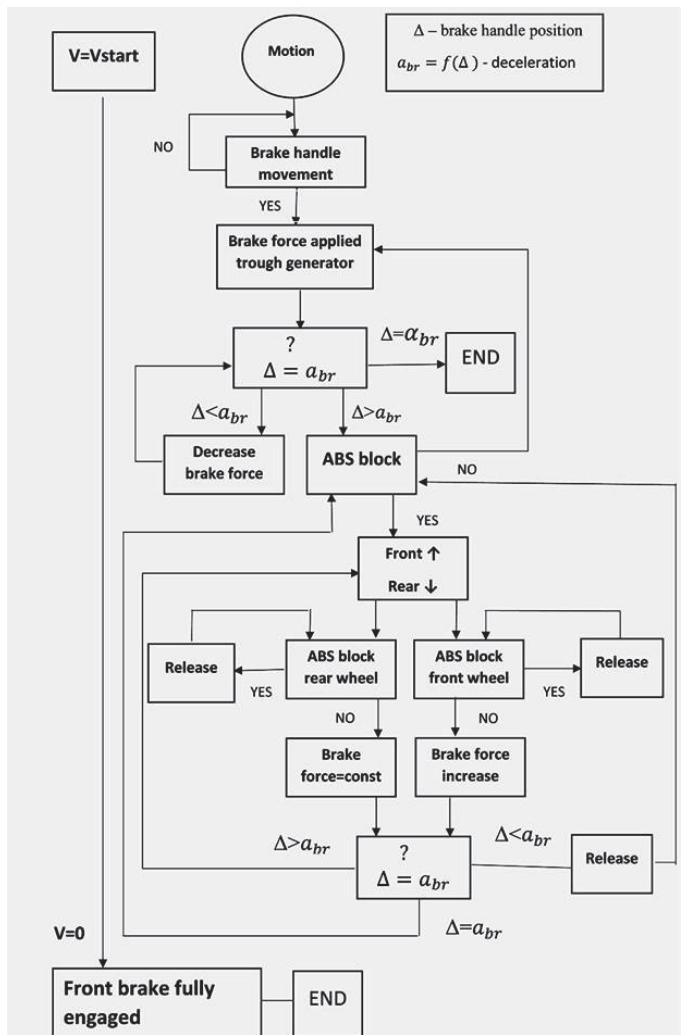


Fig. 7. Block diagram of the function of an Integral Brake System optimized for maximum energy regeneration.

The main parameters of the control logic are the brake handle position Δ and the brake deceleration a_{br} . The braking process begins in certain speed. Then the brake lever is actuated. That gives a signal to the controller to switch the electric motor in generator regime at minimum resistance. The computer checks

if the brake deceleration corresponds with the brake handle position and if there are differences increases or decreases the braking force (generators resistance). There are constant ABS-checks for locking the rear wheel and if a lock is detected then the front wheel brake is engaged and the brake force in the rear wheel is decreased. From that point on there is a big logic cycle for balancing the forces so the rear wheel is always on the edge of blocking and the rest of the needed braking force is achieved with the front brake. When the speed decreases to a stop the front brake is fully engaged the vehicle stands still and the whole braking process ends.

When the rider releases the brake lever the front brake disengages, the electric motor is back to traction regime and the motorcycle is ready to accelerate again.

CONCLUSIONS

It is possible to use the rear brake only and therefore all the energy consumed during braking to regenerate until a certain braking delay.

When the braking delay increases, the percentage of recovered energy decreases.

Greater opportunity for regeneration of energy is found in motorcycles with a longer base and a lower center of gravity.

The above dependencies can be used to program an integral braking system with maximum use of the braking power to charge the battery of an electric motorcycle

REFERENCES

- [1] G. Yanachkov, S. Hesapchieva, "Regenerative braking possibilities of an electric motorcycle," *Bultrans 2018*, Sozopol, October 2018.
- [2] D. Semov, N. Ivanov, D. Lozanov, *Automobiles, tractors and forklift trucks*, Tehnika, Sofia, 1992.
- [3] S. Dimitrov, *Theory of the automobile*, Technical University - Sofia, Sofia, 2016.
- [4] K. Kosev, L. Kunchev, *Motorcycles, sport- and racecars*, Technical University - Sofia, Sofia, 2010.
- [5] N. Djagarov, *Electric drive*, Technical University - Varna, 2011.
- [6] M. Minchev, J. Shopov, E. Rac, *Conversion technology*, Avangard Prima, Sofia, 2006.

Design of a mobile platform for the evaluation of Localization and Mapping Algorithms

Systementwurf für die Lokalisierung und Kartierung mobiler Plattformen

Stefan Hensel*, Marin B. Marinov**, Christoh Kehret*

* University of Applied Sciences Offenburg, Department for Electrical Engineering,
Badstraße 24, D-77652 Offenburg, Germany, stefan.hensel@hs-offenburg.de

** Technical University of Sofia, Faculty of Electronic Engineering and Technologies, Department of Electronics,
8, Kliment Ohridski Blvd., BG-1756 Sofia, Bulgaria, mbm@tu-sofia.bg

Abstract — In this thesis various algorithms of simultaneous localization and mapping (SLAM) are evaluated. The methods use different sensor combinations to determine the current position and to create a map. Two approaches are compared. Firstly, the position is determined exclusively on the basis of consecutive scans of a LiDAR system and the resulting point cloud of the environment is made available as a three-dimensional map. On the other hand, the combination of laser scan data, odometry and inertial measurements is carried out on the basis of an online SLAM procedure in which the position relative to the map is determined and the map is simultaneously created and updated taking into account the current position estimate. In this thesis, an experimental platform has been developed which is suitable for evaluation under real conditions.

Zusammenfassung — In dieser Arbeit werden diverse Algorithmen der simultanen Lokalisierung und Kartierung (SLAM) evaluiert. Die Verfahren nutzen für die Aufgabe unterschiedliche Sensorkombinationen um die aktuelle Position zu bestimmen und eine Karte zu erstellen. Es werden zwei Ansätze verglichen. Zum einen wird die Position ausschließlich auf Basis konsekutiver Scans eines LiDAR-Systems bestimmt und die resultierende Punktwolke der Umgebung als dreidimensionale Karte zur Verfügung gestellt. Zum anderen erfolgt die Kombination von Laserscandaten, Odometrie und Inertialmessungen auf Basis eines online SLAM-Verfahrens, in dem die Position relativ zur Karte bestimmt und die Karte gleichzeitig unter Berücksichtigung der aktuellen Positionsschätzung erstellt und aktualisiert wird. In der Arbeit wurde eine experimentelle Plattform aufgebaut, die für eine Evaluation unter realen Bedingungen geeignet ist.

I. EINFÜHRUNG

Die Navigation und Ortsbestimmung in einer unbekannten Umgebung stellt seit Beginn der mobilen Robotik die fundamentale Herausforderung für den Einsatz autonomer Systeme im praktischen Einsatz dar. Während die ersten Ansätze militärischen Systemen und kostenintensiven Fluganwendungen vorbehalten waren, erlaubt die kontinuierliche Weiterentwicklung der Algorithmen und eingesetzten Sensorsystemen mittlerweile die Anwendung in der zivilen Nutzung. Durch den stetig steigenden Grad der Komplexität für die detaillierte Umwelterfassung, die beispielsweise im innerstädtischen autonomen Fahren benötigt wird, und die geforderte Flexibilität der Einsatzmöglichkeiten mobiler Plattformen haben sich bestimmte Sensorsysteme und deren Kombination als außerordentlich erfolgreich herausgestellt [1]. Die hohe Informationsdichte optischer Systeme und die geforderte Güte für die genutzten Karten stellen zudem die Mindestanforderungen an die genutzten Kartierungs- und Lokalisierungsalgorithmen. Die Familie der simultanen Lokalisierungs- und Kartierungsalgorithmen (SLAM) haben sich hierbei als besonders erfolgreich erwiesen [2].

II. MATERIALEN UND METHODEN

Dieses Kapitel gibt einen Überblick über die eingesetzten Sensortypen und -systeme. Das als Sensorplattform verwendete mobile System und die durchgeführten Anpassungen werden danach erläutert, bevor detailliert auf die verwendete Software eingegangen wird, die sich aus der Middleware des Robot Operating Systems (ROS) und den letztendlich evaluierten Algorithmen zusammensetzt.

A. Hardwarekomponenten

1) 3D-Lasersensor

In der präzisen und ortsgenauen Vermessung von Objekten mit einer hohen Messrate haben sich zum aktuellen Zeitpunkt Laserscanner gegenüber optischen und radarbasierten Systemen durchgesetzt [3]. Die in anpruchsvollen Aufgaben eingesetzten 3D-Laserscanner werden hierbei als LiDAR (engl. Light Detection And Ranging) bezeichnet.

2) LiDAR Velodyne VLP-16

In der Arbeit wird der VLP-16 der Firma Velodyne eingesetzt. Er besteht aus 16 vertikal angeordneten Lasern, die sich in X° Intervallen auffächern und einen Öffnungswinkel von Y° aufweisen. Durch die Rotation mit 10 Hz werden 300 000 Laserpunkte gewonnen, die Objekte bis zu einer Distanz von 100 Metern und einem horizontalen Sichtfeld von 360° detektieren können. Das System arbeitet mit dem Pulslaufzeitverfahren und bestimmt die Entfernung des reflektierenden Objekts mit einer Abweichung von 10 cm. Durch die Miniaturisierung des Systems konnte der Preis in den vierstelligen Bereich verringert werden, was den Sensor zum de facto Standard im autonomen Fahren macht [3].

3) Inertiale Messeinheit (IMU)

Während optische Sensorsysteme wie Laser und Kameras die Positionsbestimmung bezüglich der Umwelt erlauben, werden sogenannte inertiale Sensoren dazu genutzt, die relative Bewegung einer Plattform aus deren Bewegung zu bestimmen.

Eine inertiale Messeinheit (IMU) besteht aus drei Beschleunigungssensoren, drei Drehratensensoren (Gyros) und für gewöhnlich drei Magnetfeldsensoren, die jeweils als Triaden orthogonal zueinander angeordnet werden und auf Basis der Massenträgheit die aktuelle Orientierung im Raum bestimmen.

Die Sensordatenfusion der Einzelsensoren erlaubt hierbei die Kompensation vorhandener Drift, welche durch die Integration der Drehraten und Beschleunigungen unweigerlich auftritt [4]. Die Lage im Raum wird durch die Drehratensensor-Triade bestimmt, deren Drift um die Nick- und Wankachse durch die mit Hilfe der Beschleunigungs-sensortriade gemessenen Erdanziehung kompensiert wird. Die Drift des Gierwinkels wird durch die Bestimmung der Nordrichtung kompensiert, die wiederum mit der Hallsensortriade aus dem Erdmagnetfeld bestimmt wird. Die letztendliche Schätzung der Lage wird entweder mit Hilfe nichtlinearer Zustandsbeobachter, z.B. dem extended Kalman Filter [5], oder orthogonalen Filtern, beispielsweise in [6] beschrieben, durchgeführt.

4) XSens IMU

In dieser Arbeit wurde die IMU MTi-300 von XSens verwendet. Diese besteht aus den beschriebenen Sensortriaden und schätzt die Lage mit Hilfe eines Kalman Filters. Die Unsicherheit des Gierwinkels liegt unter 1°, die Nick- und Rollwinkel werden mit einer Genauigkeit von 0,2° bestimmt. Vorteile des Systems sind die IP67 Schutzklasse und die Verfügbarkeit der Treiber, die eine Integration auf der Roboterplattform erleichtern und einen Einsatz im Außenbereich erlauben.

5) Mobile Plattform – Clearpath Robotics Husky

Als Sensorträger kommt die mobile Roboterplattform Husky von Clearpath Robotics zum Einsatz. Es handelt sich um eine mobile Basis, welche durch die Traglast von 75 kg mit einer Vielzahl von Sensoren, Rechnern und Batterien für die Energieversorgung ausgerüstet werden kann. Durch die kompakten Abmaße, den Radstand, gegebene Radhöhe sowie den Skid-Steer-Antrieb kann ein Einsatz im Innen- und Außenbereich erfolgen, was die Evaluation der Kartierung für strukturierte und natürliche Umgebungen erlaubt. Die Eigenschaften des Systems sind zur Übersicht in Tabelle I aufgeführt, in Abbildung 1 ist die Plattform im Rahmen des Gesamtsystems zu sehen.

TABELLE I. KENNDATEN DES HUSKY ROBOTERS

Eigenschaften	Husky
Größe [mm]	990x670x390
Gewicht [kg]	50
Max. Traglast [kg]	75
Max. Geschwindigkeit [m/s]	1.0
Laufzeit [h]	3
Treiber und APIS	ROS, C++, Mathworks

Die Steuerung der Einheit erfolgt über einen Rechner mit x64 Architektur, der mit einer Intel Core-i5 CPU, 8 GByte RAM und dem Betriebssystem Linux Ubuntu sowie ROS ausgestattet ist.

B. Softwarekomponenten

1) ROS – Robot operating system

Für die Implementierung der Algorithmen und Sensoren wurde die Middleware ROS [7] verwendet. Unter einer Middleware versteht man eine Verteilungsplattform, welche im abstrakten Sinn die Kommunikation zwischen Prozessen

erlaubt und regelt. Die vorgestellten sensoren verfügen ebenso wie der Husky Roboter über Treibersoftware für ROS, was die Integration im Gesamtsystem deutlich vereinfacht. ROS bietet zudem eine Vielzahl von existierenden Programmiersprachen, die die Bereiche Visualisierung, Datenaufzeichnung, Simulation und Systemüberwachung beinhalten. Selbst erstellte Module können in den Programmiersprachen C++, Python oder mit Matlab/Simulink erstellt werden.

III. VEFahren ZUR KARTENERSTELLUNG

A. Laser Odometry and Mapping – LOAM

Laser Odometry and Mapping ist eine echtzeitfähige Methode, um mit Hilfe von 3D LiDAR-Daten die aktuelle Position zu schätzen. Die gefahrene Distanz zwischen zwei konsekutiven Laserscans wird mit Hilfe eines Point-Matching-Verfahrens bestimmt und die aktuelle Position aus der Integration bestimmt. Das Verfahren ist hierbei zweistufig und besteht aus einer schnellen lokalen Positionsschätzung und einem langsameren Korrekturschritt, der aus den aufgezeichneten Messungen eine Punktwolke mit Hilfe des Blockbündelausgleichs erstellt. Die Punktwolke fungiert als Karte, welche wiederum als Bezugssystem für die erneute Positionsbestimmung genutzt werden kann. Diese Parallele Struktur erlaubt eine hohe Berechnungsgeschwindigkeit bei gleichzeitig hoher Präzision. Detaillierte Informationen finden sich in [8] und [9].

Das System nutzt explizit keine zusätzlichen Sensoren, so dass das Verfahren in die Klasse der Scan-Matching basierten SLAM-Methoden zu zählen ist. Für die Evaluation ausgewählt wurde es aufgrund seiner geringen Rechenlast, was die Echtzeitfähigkeit auf handelsüblichen CPUs erlaubt, zudem sind Varianten des Algorithmus auf den vordersten Rängen neutraler Leistungsbenchmarks, z.B. der KITTI-Datenbank [10], zu finden.

B. Cartographer SLAM

Methoden der simultanen Lokalisierung und Kartierung haben eine lange Historie in der Robotik [1] und können grob in zwei relevante Klassen eingeteilt werden: Probabilistische Verfahren basierend auf Zustandsschätzern (Extended-Kalman-Filter-SLAM) und Verfahren, die einen Graphen aus den verhandenen Posen (Position und Orientierung) und den Sensorbeobachtungen der Umgebung erstellen und diesen optimieren (engl. Pose-Graph-SLAM).

Bei Cartographer handelt es sich um ein Pose-Graph-SLAM-Verfahren, basierend auf LiDAR-Sensoren [11]. Eine Erweiterung zu den klassischen SLAM-Verfahren stellt die Unterteilung in einen lokalen, echtzeitfähigen, SLAM-Prozess, der mit einer globalen Optimierungsstufe verknüpft wird.

Der lokale Prozess weist eine starke Ähnlichkeit zu LOAM auf und erzeugt aus konsekutiven Laserscandaten eine konsistente Darstellung der Umgebung die ebenfalls für die Berechnung der zurückgelegten Strecke genutzt werden kann.

Die globale Optimierung beruht auf einem Blockbündelausgleichsverfahren. Diese Optimierung von dünnbesetzten Matrizen beruht auf photogrammetrischen Algorithmen und korrigiert die Ausrichtung der Ursprungspunkte der lokalen Karten, was sich über diese wiederum auf die geschätzte Position des Roboters auswirkt.

Die verwendete Implementierung von Cartographer-SLAM stellt zwei Optionen zur Verfügung, die eine nutzt den 3D-Laserscan und erzeugt eine zweidimensionale Karte, die auf den horizontalen Laserstrahlen beruht, die andere erzeugt eine interne 3D-Karte mit allen Laserstrahlen, die allerdings auf eine 2D-Darstellung zurückprojiziert wird.

C. Aufnahmearten des Systems

In Abbildung 2 ist der Aufbau des Roboters dargestellt. Die Kommunikation der einzelnen Komponenten erfolgt über ROS, die Bewegung des Roboters erfolgt über einen Controller, so dass keine autonome Exploration erfolgt, sondern ausschließlich die Kartierung und Lokalisierung erprobt wird.

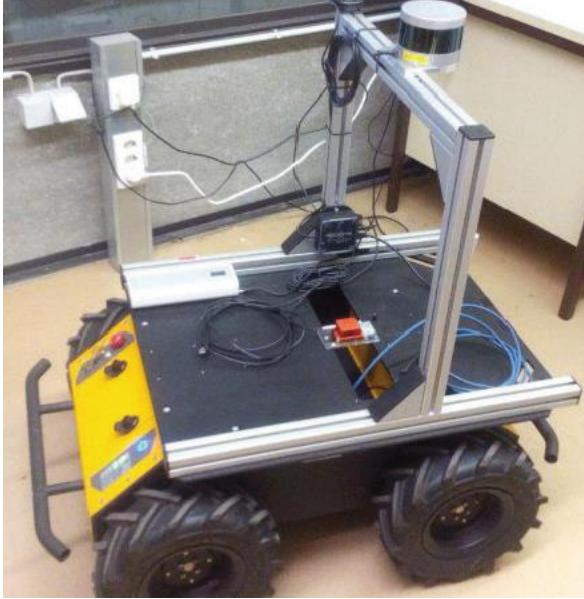
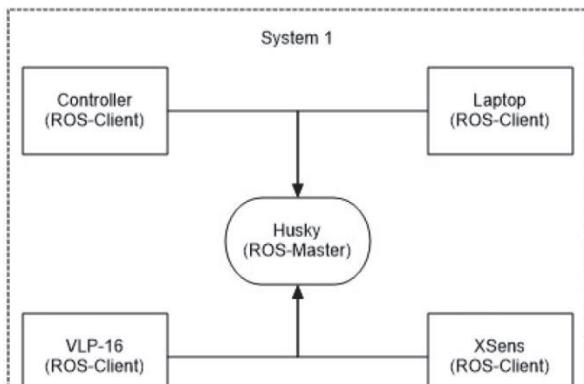


Abb. 1. Aufbau des vollen Systems

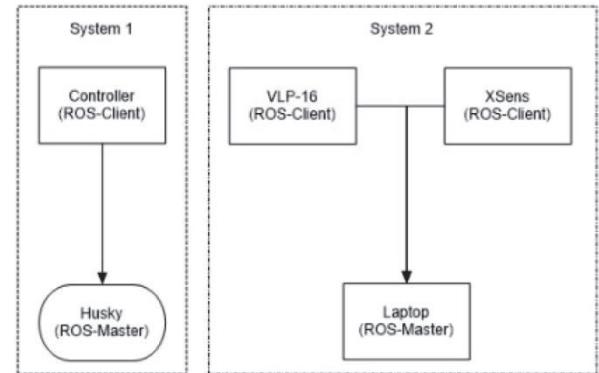
Für die Evaluation wurden zwei Arten der Datenakquise realisiert, welche in Abbildung 3 dargestellt sind. Die erste Möglichkeit erlaubt es, die Sensoren über den eingebauten Rechner zu steuern und mit Hilfe eines externen Laptops die Daten über eine SSH-Schnittstelle (LAN oder WLAN) aufzuzeichnen. Dieser Ansatz bietet den Vorteil, neben den Sensordaten auch direkt die fusionierte Odometrie des Huskys sowie die Steuerdaten des Controllers aufnehmen zu können.

Die zweite Methode trennt die Datenaufnahm, die mit Hilfe eines zweiten Laptops erfolgt, von der Steuerung des Roboters.

Obwohl die erste Methode den direkten Einsatz und online-SLAM-Berechnung ermöglicht, wurde sich im Rahmen der Veröffentlichung für den zweiten Ansatz entschieden, da hier eine höhere Datenrate und lückenlose Aufzeichnung, im Speziellen die Laserscan-Daten betreffend erreicht werden konnte. Die Erstellung der Karte und Berechnung der Position kann danach offline mit Hilfe des ROS-Frameworks simuliert werden, um eine Reproduzierbarkeit der Parametereinstellungen sicherzustellen.



(a) Indirekte Aufnahme



(b) Direkte Aufnahme

Abb. 2. Aufnahmearten

IV. ERGEBNISSE UND AUSWERTUNG

Mit dem beschriebenen System wurden zuerst die qualitativen und danach die quantitativen Aspekte untersucht und beurteilt.

A. Qualitative und quantitative Aspekte bei der Untersuchung

1) Qualitative Auswertung

In der qualitativen Auswertung werden folgende Punkte beleuchtet:

- Kartentyp: Hierbei handelt sich um den Typ der Karte, entweder die Karte wird in 2D oder 3D dargestellt.
- Ressourcen: Ressourcen bezeichnet die Hardwareanforderungen. Dabei bedeutet schlecht, dass eine hohe Rechenleistung benötigt wird.
- Ansicht: Hiermit wird die Auflösung der Karte bewertet. Je höher die Auflösung ist, desto besser können Menschen Informationen aus der Karte entnehmen.
- Navigation: Unter dem Aspekt Navigation wird angegeben, ob sich eine spätere Navigation mithilfe der erzeugten Karte bewerkstelligen lässt.
- Robustheit: Mit der Robustheit wird eine qualitative Aussage darüber geschlossen, wie robust der Algorithmus arbeitet.
- Zubehör: Hierdurch wird der benötigte Aufwand beschrieben. In dieses Kriterium geht vor allem die benötigte Zahl an Sensoren ein.

2) Quantitative Auswertung

Die quantitative Auswertung erfolgt auf Basis der Start- und End-Pose des Roboters, hierbei gelten folgende Definitionen:

- Absoluter Fehler: Beim absoluten Fehler handelt es sich, um die gemessenen Abweichungen der Endposition mit der Anfangsposition. Die Abweichung wird einerseits in X-Richtung und andererseits in Y-Richtung ermittelt. Die Einheit ist Meter.
- Relativer Fehler: Der relative Fehler bezeichnet den prozentualen Fehler. Dieser gibt an, um wie viel Prozent der Algorithmus falsch liegt in Abhängigkeit von der gefahrenen Gesamtstrecken. Die Formel für den relativen Fehler lautet:

$$\text{Rel. Fehler} = \frac{\text{abs. Fehler}}{\text{Strecke}} \cdot 100\%. \quad (1)$$

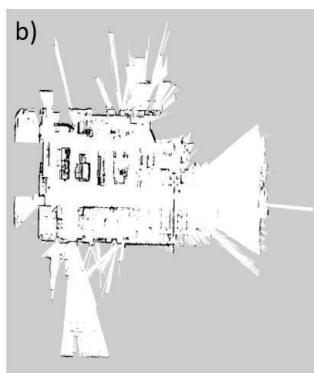
Der Fehler gibt an, wie genau und robust die Berechnungen sind. Dies ist die wichtigste Eigenschaft und deswegen liegt hierbei das Augenmerk und die Relevanz der Arbeit.

B. Exemplarische Experimentelle Ergebnisse

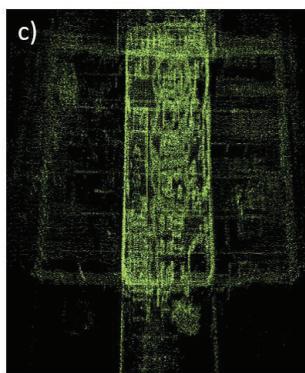
Die Evaluation der Verfahren erfolgte anhand mehrere typischen Sequenzen für mobile Roboter. Als Beispiel ist In Abb. 3 eine industrielle Anlage mit hohem Strukturanteil dargestellt. Es ist jeweils ein Beispieldvideobild, die erstellte LOAM-Punktwolkenkarte und die in Cartographer erstellte 2D Karte gezeigt. Zusätzlich erfolgte die Auswertung mit unterschiedlichen Streckenlängen und zusätzlichen Szenarien, einmal in Bürogebäuden mit Korridoren und einmal in Außenbereichen mit Sträuchern und Bäumen.



a)



b)



c)

Abb. 3. Exemplarisches Ergebnis der Auswertung mit a) Videobild, b) Cartographer "D-Karte und c) LOAM 3D-Punktwolkenkarte

Das quantitative Ergebnis der Testfahrten (Abb. 3 findet sich in Industry Double wieder) für den Cartographer-SLAM ist in Tabelle II dargestellt, welche auch für die anderen Sequenzen die Ergebnisse als relativen Fehler der Lokalisierungs-genauigkeit in x- und y-Koordinatenrichtung auflistet.

TABELLE II. QUANTITATIVE AUSWERTUNG DES CARTOGRAPHER-2D-VERFAHRENS

Umgebung	Gesamtlänge [m]	rel. Fehler X []	rel. Fehler Y []
IndoorShort	61	0,033	0,016
IndoorShort3	183	0	0
IndoorMiddle	152	1,7	0,2
IndoorMiddle3	456	0,4	0,02
IndoorBig	475	0,48	0,18
IndoorBigLoop	500	0,41	0,06
IndustryLoop	50	0,6	0,04
IndustryReverse	100	0,06	0,02
IndustryDouble	115	0,18	0,02
OutdoorShort	75	0,64	0,65
OutdoorShort3	225	0,64	0,65
OutdoorMiddle	150	0,28	0,12
OutdoorBig	500	0,24	0,14

Zum Vergleich sind in Tabelle III die Ergebnisse für den 3D-LOAM-SLAM aufgeführt. Die Ergebnisse zeigen deutlich, dass Cartographer-SLAM dem LOAM-Verfahren in puncto Genauigkeit überlegen ist. Dies war durch den zusätzlichen Optimierungsschritt auch zu erwarten. Der Vorteil des LOAM-Verfahrens zeigt sich eher in der hohen Bearbeitungsgeschwindigkeit und der möglichkeit auf einfache Weise eine relative Lokalisierung zu erhalten.

Der systematisch größte Vorteil von Cartographger-SLAM besteht in der Erstellung der Rasterkarte, welche als Eingang für die spätere Roboternavigation dient. Die 3D-Punktwolke des LOAM-Verfahrens stellt dagegen eine visuell ansprechende Darstellung des befahrenen Gebietes dar, welche auch von Menschen sehr gut interpretiert werden kann.

TABELLE III. QUANTITATIVE AUSWERTUNG DES LOAM-ALGORITHMUS

Umgebung	Gesamtlänge [m]	rel. Fehler X []	rel. Fehler Y []
IndoorShort	61	0,08	0,43
IndoorShort3	183	0,47	0,01
IndoorMiddle	152	1,66	0,68
IndoorMiddle3	456	-	-
IndoorBig	475	0,18	0,24
IndoorBigLoop	500	0,07	0,27
IndustryLoop	50	1,22	0,14
IndustryReverse	100	0,54	0,04
IndustryDouble	115	0,24	0,04
OutdoorShort	75	0,31	0,11
OutdoorShort3	225	0,35	0,03
OutdoorMiddle	150	0,46	0,15
OutdoorBig	500	0,03	0,10

V. ZUSMMENFASSUNG UND AUSBLICK

In dem vorliegenden Aufsatz konnte das System für die Implementierung und Evaluation komplexer Lokalisierungs- und Kartierungsverfahren vorgestellt werden. Beide Verfahren eignen sich grundlegend für eine robuste Lokalisierung in fordernden Szenarien wie Pick and Place Anwendungen in der Industrie 4.0, Erkundung und Überwachung in der Feldrobotik oder Anwendungen der Servicerobotik. Komplette SLAM-Ansätze zeigen sich hier systemisch überlegen, da alle Informationen einer Vielzahl an Sensoren kombiniert und genutzt werden kann.

Nächste Schritte bestehen nun in der Nutzung der erstellten Karten, was sich im Falle der präzisen 3D-Punktwolke als komplexe Aufgabe erweist.

LITERATURVERZEICHNISS

- [1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid and J. Leonard, "Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age," vol. 32, no. 6, pp. 1309-1332, December 2016.
- [2] Sebastian Thrun, Wolfram Burgard, Dieter Fox, *Probabilistic Robotics*, ISBN:978-0-262-20162-9., Cambridge : MIT Press, 2005.
- [3] Andreas Nüchter, *3D Robotic Mapping - The Simultaneous Localization and Mapping Problem with Six Degrees of Freedom*, Berlin, Heidelberg: Springer Science & Business Media,, 2009.
- [4] J. Wendel, *Integrierte Navigationssysteme: Sensordatenfusion, GPS und Inertiale Navigation*, 2009.
- [5] P. Corke, *Robotics, Vision and Control*, 2nd. Edition, Springer, 2017.
- [6] R. Mahony, T. Hamel and J.-M. Pflimlin, "Nonlinear Complementary Filters on the Special Orthogonal Group," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203-1217, 2008.
- [7] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler and A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA Workshop on Open Source Software*, 2009.
- [8] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," *Robotics Science and Systems*, 2014.
- [9] J. Zhang and S. Singh, "Low-drift and real-time lidar odometry and mapping," *Autonomous Robots*, vol. 41, no. 2, pp. 401-416, 2017.
- [10] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [11] W. Hess, D. Kohler, H. Rapp and D. Andor, "Real-Time Loop Closure in 2D LIDAR SLAM," in *Robotics and Automation (ICRA)*, pp. 1271-1278, 2016.

Location based calibration of low cost particulate matter sensors

Standortbasierte Kalibrierung von kostengünstigen Partikelsensoren

Ivan P. Topalov *

* Department of Electronics, Technical University of Sofia
Sofia, Bulgaria, vantop@hotmail.com

Abstract — Air pollution nowadays is reaching levels beyond the human body can live with and health issues caused by bad air quality are increasing and in particular cases leading to death. Analyzing air quality has no meaning if not backed up by real measurements. The biggest problem with measuring air quality is the price of the particulate matter equipment which sometimes goes over tens of thousands of dollars. Very often such a big investment can be afforded only by big companies or government/municipal agencies which leaves ordinary people “blind” regarding what they breathe in their residential area. This paper provides a device for measuring air particulate matter using off the shelf, low cost sensors. To improve the accuracy of the sensor data, the system localizes itself and performs calibration procedures using data from the nearest government or municipal air quality station.

Zusammenfassung — Die Luftverschmutzung erreicht heutzutage ein Niveau, mit dem der menschliche Körper nicht leben kann, und Gesundheitsprobleme, die durch schlechte Luftqualität verursacht werden, nehmen zu und führen in bestimmten Fällen zum Tod. Luftqualität hat keine Bedeutung, wenn sie nicht durch reale Messungen unterstützt wird. Das größte Problem bei der Messung der Luftqualität ist der Preis der Feinstaubmesssysteme, der oft mehrere Tausende Euro übersteigt. Sehr oft kann eine so große Investition nur von großen Unternehmen oder Behörden / Stadtverwaltungen geleistet werden. Diese Arbeit präsentiert ein mobiles System zur Messung von Luftpartikelzentrationen mit handelsüblichen, kostengünstigen Sensoren. Um den erhaltenen Ergebnissen zu vertrauen, lokaliert das System sich selbst und vergleicht die Daten mit der nächstgelegenen staatlichen oder kommunalen Luftqualitätsstation.

I. INTRODUCTION

One of the main components in the air pollution is the particulate matter (PM). PM is made of a mixture of solid particles and liquid droplets. Some particles, such as dust, dirt, smoke, are big and dark enough to be seen with a naked eye while others are so small that can be seen only through a microscope. When talking about particle sizes, PM is divided into size groups – PM_{2.5} and PM₁₀ are of a main importance. PM₁₀ are inhalable particles which can pass through a size-sensitive inlet with diameters that are equal or less of 10 micrometers while PM_{2.5} are fine inhalable particles like PM₁₀ but with diameters that are equal or less of 2.5 micrometers.

The sources of the PM are diverse and the particles are made up of hundreds of different chemicals. Some of them are coming straight from a source, such as constructions areas, fields, unpaved roads and fires and other particles are formed in the atmosphere as a result of complex chemical reactions which produce sulfur dioxide, nitrogen oxides and others which are all pollutants emitted from power plants, industries and combustion vehicles [1].

The smaller the size of particles, the bigger is the impact on human's health. Small particles PM₁₀ are the source of the greatest problems because they can penetrate deep into the lungs and PM_{2.5} can even get into the bloodstream. Extreme and prolong exposure to such pollutants can impact human's lungs and heart and thus lead to variety of problems like premature death in people with heart or lung diseases [2], heart attacks, irregular heartbeat, asthma, respiratory symptoms like coughing and others.

Poor air quality is a growing global health problem which attacks millions of people worldwide, mainly in big cities. Big

industry, which is increasingly situated outside of metropolitan regions and urban areas, is no longer considered the main cause of air quality problems. Recent studies indicate that road traffic is the primary source of air pollution. Road transport is responsible for an average of 25% of all harmful emissions in Europe. In many EU countries this value is higher than 30%. Thus, poor environmental quality, especially in urban areas, is one of the greatest environmental concerns of this century as it affects both health and welfare [2, 3]. Many studies show that today's increase in respiratory diseases, as compared to other related diseases and allergies, is primarily due to air pollution. According to official figures of the EU more than 225 000 people die every year from diseases caused by car emissions in Europe. To combat this threat, the European Union has introduced stricter laws and regulations and intends to reduce car emissions by 20% by 2020 [4].

The official monitoring networks for particulate matter are usually concentrated on a small number of locations where the measurements are made high accuracy and thus with high costs. This approach is very convenient for monitoring of long term trends of timely averaged particulate matter indexes mainly if the concentrations of the pollutants are poorly influenced by local sources of pollution. In such cases, the accuracy of the measurement equipment allows for detecting even the lowest changes of the tendencies of the main pollutants, in our case particulate matter. On the other hand, the above mentioned approach doesn't allow a complete spatial visualization of minor changes in particulate matter concentration and short term variations caused by local sources.

The use of cheap sensors is still questionable because of their error and reliability. But if somehow the output of the

particulate matter sensors can be calibrated they can be accepted as a trustful solution for air quality monitoring which is the idea in the presented paper.

II. PARTICULATE MATTER MEASUREMENTS – PARAMETERS AND STANDARDS

Nowadays there is no specific standard or algorithm describing the measurement process of particulate matter because the health impact was not considered a big threat. Usually, aerosols and particles are side effect product of natural and human engaged processes and are found in different shapes, densities, chemical structures and biological properties.

A. Basic parameters

There are a number of significant parameters of PM in ambient air that should be taken into account:

- particle number concentration
- total mass concentration of selected fractions of particulate matter,
- particle size distribution,
- 24-hour variations of concentrations with peak values,
- chemical composition, etc.

Currently, the most common health applicable metric is mass related to particle size, which is represented as PM_x .

In most sources, the PM_x abbreviation is incorrectly defined as "all fine particles with size smaller than $x \mu m$ ". The correct definition is "particulate matter which passes through a size-sensitive inlet with a 50% efficiency at $x \mu m$ aerodynamic diameter" [5].

PM limits and targets for 24 hours and annual averages differ significantly from country to country. This is shown in Table 1 which provides examples of PM standards and objectives of some countries and the World Health Organization (WHO).

TABLE I. STANDARDS AND OBJECTIVES FOR PM MONITORING IN URBAN AREAS IN SOME COUNTIES [6]

Country	PM fraction	Averaging period	Limit /Standard $\mu g/m^3$
EU	PM_{10}	Annual	40
		24 h	50
	$PM_{2.5}$	Annual	25/20
US	PM_{10}	24 h	150
	$PM_{2.5}$	Annual	12
		24 h	35
China	PM_{10}	Annual	70
		24 h	150
	$PM_{2.5}$	Annual	35
India	PM_{10}	24 h	75
		Annual	60
		24 h	100
WHO	PM_{10}	Annual	20
		24 h	50
	$PM_{2.5}$	Annual	10
		24 h	25

Apart from diverse national and international standards, there is the challenge that air quality varies non-linearly by

locations and a by a great number of factors, such as the weather conditions, traffic, land use, etc., which affect it and make it very complicated to be modeled. In fact, unless there is a monitoring station, we do not really know what the air quality of a location is.

B. Fractions of particulate matter

The size of the particles is directly related to the potential they have to cause health problems - the finer the particles, the more difficult they are to disperse and the deeper they can penetrate into the lungs and even into the blood stream thus causing more harm. PM_{10} enters the respiratory tract, and has been associated with health risks such as bronchitis, asthma, and upper respiratory tract infections. PM_{10} magnifies symptoms of existing diseases rather than triggers new conditions.

Fine particles in the $PM_{2.5}$ size range get into the respiratory tract and can reach the lungs and the blood stream causing cardiovascular problems.

The statistics show that 7 % of the urban population in the EU-28 was exposed to $PM_{2.5}$ levels above the EU limit value, and approximately 82 % was exposed to concentrations exceeding the stricter WHO AQG value for $PM_{2.5}$ in 2015 [7].

Advances in embedded systems and new sensors technologies have made it possible for a new generation of low-cost PM monitoring systems to emerge. Portable and autonomous sensors have the potential to take measurements with sufficient accuracy and in this way to capture effectively the spatial variability of the air pollutants. The number of those commercially available devices has increased considerably over the last five years although the quality of the data which they provide is still questionable [8]. The main goal of our study is to test the quality of the data obtained by off-the-shelf cheap sensors and compare these results with those reported by the official authority stations. We want to find out whether such cost-effective systems can provide reliable results and indications about air quality and can be used in practice.

III. ARCHITECTURE OF THE PROPOSED MEASUREMENT

The overall system architecture and the design of hardware and software components are presented in details in this section.

- High reliability and availability of the device for long-term measurements.
- Ability to measure PM_{10} and $PM_{2.5}$.
- Use of off-the-shelf cost-effective components for Wi-Fi implementation.
- Computing power to perform on-board calculations, scalable architecture that supports easy expansions with peripherals and environmental parameter sensors.
- Capability for remote status monitoring, GPS localization and software updates.
- Battery powered operation

Fig. 1 shows the overall system architecture of an environmental PM monitoring system that we have developed. The system consists of a development board based on the microcontroller ESP32 with integrated Wi-Fi, a GPS module with integrated RTC, a temperature/humidity/pressure sensor and an Optical Particle Counter SDS011 for measuring $PM_{2.5}$ and PM_{10} .

The heart of the system is the System-on-Chip (SoC) ESP32 from the company Espressif [9]. Main advantage of this SoC is its price tag – between €5 and €10 depending on the peripherals included on the board. The SoC is an Xtensa LX6 dual core 32bit microprocessor with performance up to 600

DMIPS. It has an integrated Wi-Fi 802.11 b/g/n and Bluetooth connectivity. The SoC has a lot of peripherals implemented but in our case of interest are the I2C, the UART and SPI interfaces. In our particular case the board used is Wemos Lolin32 operating at 80MHz and has 4MB of FLASH memory.

For temperature, humidity and pressure is used the BOSCH sensor BME280 packed in 2.5 mm x 2.5 mm x 0.93 mm metal lid LGA [10]. This is an I2C device with a temperature range -40...+85 °C, relative humidity range 0...100 % and pressure range 300...1100 hPa. The accuracy of the temperature is ±1.0°C, the accuracy of the relative humidity is ±3 % and the accuracy of the pressure is ±1.0hPa.

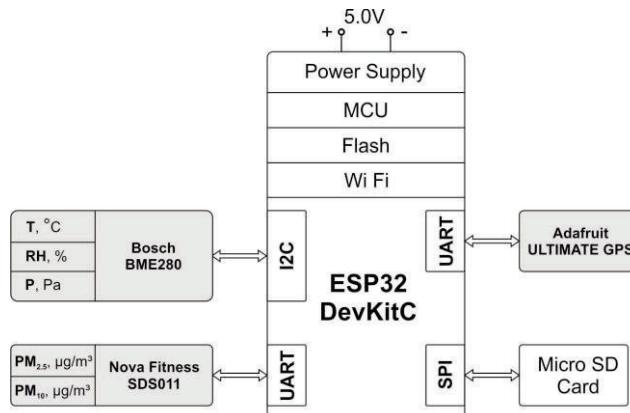


Fig. 1. Block diagram of the developed system

Regarding the particulate matter measurements we have used a low cost sensor SDS011 from the company Nova Fitness [11]. It is a counting particle device using laser scattering method. It can output via a UART the concentration of the PM particles in two sizes – 2.5µm and 10µm. The output range of both sizes is 0.0-999.9 µg/m³ with a maximum error of +/- 15%. The price tag of this sensor is around €20.

As for the location, a GPS from Adafruit is used [12]. Ultimate GPS is a very compact, with built in antenna and has a maximum update rate of 10Hz.

IV. SOFTWARE AND PROGRAM WORKFLOW

The software is developed using the Arduino IDE since the ESP32 fully supports the environment.

At the start of the program, the microcontroller initializes the Wi-Fi, the sensors, the GPS and also the SD Card. Then it sets its own real time clock using either time server on internet or time provided by the GPS. After all these procedures the software goes in measurement mode. Every minute the software measures PM_{2.5}, PM₁₀, temperature, humidity and pressure and records the location of the current measurement. Every hour, an average of all of the above readings is calculated and is saved to a CSV file on the SD Card together with the same readings from the closest government or municipal air quality station. The problem of finding the closest government or municipal air quality station and getting the data from it was solved using a service from the aqicn.org website. This is an information website on internet where most of the government and municipal air quality station submit their data. The website aqicn.org provides Application Programming Interface (API) through which data from any air station can be downloaded locally. One of the API functions can be used to search for the nearest station from a given latitude/longitude. Using this method the system downloads air quality data from the nearest government or municipal station using the coordinates from system's GPS. This data is saved together with the data measured by system's sensors.

V. EXPERIMENTAL RESULTS

The experimental results were conducted on 14.11.2018 at the location of the Technical University of Sofia, Bulgaria with coordinates for latitude 42.655642 and for longitude 23.356051. Using the above mentioned coordinates and with the help of aqicn.org the closest municipal air quality station was reported to be in the area of Mladost, which is roughly 2km away from our testing point. The measurements were made continuously for 24 hours, saving each hour the values for PM_{2.5} and PM₁₀ which are an average for 1 hour and saving them together with the data from the Mladost air quality station. As can be seen on Fig. 2 the samples of the sensor SDS011 are very close to the public environmental air quality station Mladost. The linearity of both samples (from SDS011 and from Mladost station) is equal.

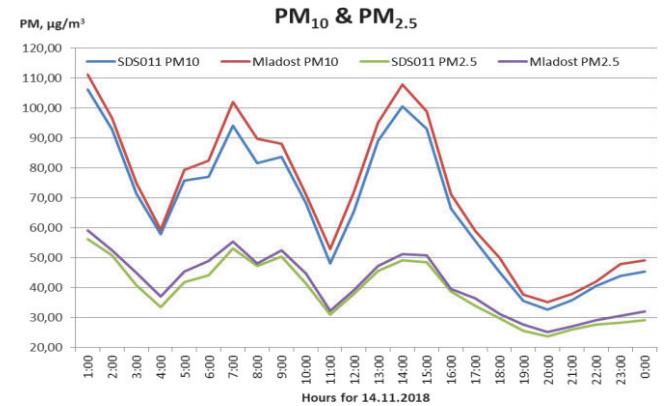


Fig. 2. Measurement results for PM_{2.5} and PM₁₀ taken on 14.11.2018 by the proposed system using the particulate matter sensor SDS011 and the public environmental air quality station Mladost in Sofia, Bulgaria

On Fig. 3 can be seen the error in percentage between the samples from SDS011 and the values the public environmental air quality station Mladost which can be used as references.

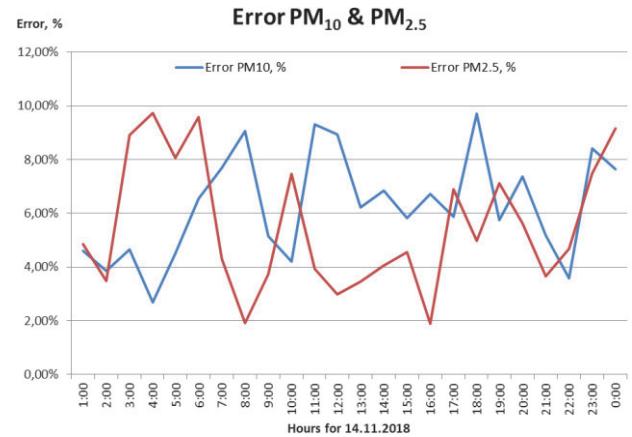


Fig. 3. Error in percentage of the measurement results for PM_{2.5} and PM₁₀ taken on 14.11.2018 by the proposed system using the particulate matter sensor SDS011 and the public environmental air quality station Mladost in Sofia, Bulgaria

Since the suggested particulate matter is a low cost sensor using a laser scattering method, its technical parameters need to be verified before the measurements can be trusted. As it is seen on Fig. 3 the error never goes beyond 10% for both PM_{2.5} and PM₁₀. The specification of the particulate matter sensor SDS011 says that maximum error of the measurements is +/- 15% and with the above mentioned results we can be assured that sensor's performance is according to the datasheet. The

error is always positive which means that area of the Mladost station is dirtier than the area of Technical University where the real measurements were made assuming that the samples from the sensor SDS011 are trustworthy. The chosen date 14.11.2018 was selected amongst several other days because the weather was stable, no wind, no high humidity and the results for particulate matter from other nearby government air quality stations were nearly equal presuming the same conditions. The equipment used as a reference for particulate matter measurements in Mladost air quality station is from the company "Thermo Scientific" Model "5030 SHARP" [13]. It is a synchronized hybrid ambient real-time particulate monitor with an accuracy of $\pm 5\%$. The results from the Mladost station are official and are collected and analyzed by the European Environment Agency. For calibrating our system we used the so called "Blind calibration" model whose idea is to reach high correspondence between all sensors in a given network. A basic condition is that nearby sensors should have almost equal readings, Fig. 4s. This method was developed at the beginning for general sensor networks and especially for temperature and relative humidity [14], [15]. In our case, we are adding an offset to our low cost sensor which offset is the difference between SDS011 particulate matter sensor and the government air quality station [14].

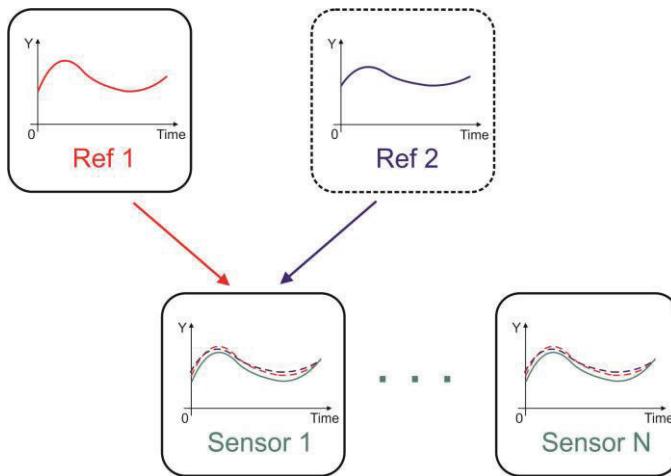


Fig. 4. Blind calibration scenario with urban sensors Sensor 1 - Sensor N and urban reference Ref. 1 and Ref. 2.

VI. CONCLUSION

From the experimental results is obvious that the use of low-cost off the shelf particulate matter sensors is reliable and after a short period of time calibration with a reliable source of samples like the public air quality stations the data can be used for further analyses of the quality of the air at the sampled

location. The proposed system automatically finds the closest reference station and saves its data together with the data from the SDS011 sensor. In case of errors from the reference values of more than 10% the data can be considered untrustworthy. In the experimental test the biggest calculated error was 9.7%.

REFERENCES

- [1] „Air Pollution Monitoring,“ [Online]. Available: <http://www.epa.gov/airquality/monitring.html>. [Accessed 9 November 2018], “ U.S. Environmental Protection Agency, 2018.
- [2] „7 million premature deaths annually linked to air pollution,“ World Health Organisation, 2012.
- [3] „Impact Analysis,“ European Commission, 2013.
- [4] M. Jovasevic-Stojanovic, „On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter,“ *Environmental Pollution*, tom 206, pp. 696-704, 2015.
- [5] R. Williams, R. Long and M. Beaver, „Sensor Evaluation Report,“ Environmental Protection Agency, Office of Research and Development, Washington, DC, May 2014.
- [6] „Air Quality in Europe report,“ European Environment Agency, ISSN 1977-8449, 2017.
- [7] *Smart City-Strategie Berlin*, Berlin: Senatsverwaltung für Stadtentwicklung und Umwelt, 2015.
- [8] E. M. Viana, „Urban Air Quality in Europe, Berlin Heidelberg: Springer-Verlag,“ 2013.
- [9] „ESP32 datasheet,“ [Онлайн]. Available: www.espressif.com.
- [10] „BME280 datasheet,“ [Online]. Available: www.bosch-sensortec.com.
- [11] „SDS011 datasheet,“ [Online]. Available: www.inovafitness.com/en.
- [12] „Ultimate GPS description,“ [Online]. Available: www.adafruit.com.
- [13] „Thermo Scientific Model 5030 SHARP,“ [Online]. Available: www.thermofisher.com.
- [14] Y. Wang, A. Yang, Z. Li, X. Chen, P. Wang and H. Yang, „Blind drift calibration of sensor networks using sparse bayesian learning,“ *IEEE Sensors Journal*, pp. vol. 16, no. 16, pp. 6249–6260, 2016.
- [15] L. Balzano and R. Nowak, „Blind calibration of sensor networks,“ in *Proc. of IPSN. IEEE/ACM*, pp. 79-88, 2007.
- [16] B. Maag, Z. Zhou and L. Thiele, „A Survey on Sensor Calibration in Air Pollution Monitoring Deployments,“ *IEEE Internet of Things Journal*, pp. Volume 5, Issue 6, December 2018.

Session Economics

Chair: Vasil Galabov

THIS PAGE IS INTENTIONALLY LEFT BLANK

Competence of Employees in the Context of Digitalization

Mitarbeiterkompetenz im Zeitalter von Digitalisierung

Anna-Maria Staiger^{*}, Malte Fiedler[†], Wolfgang Fritz[†], Dietrich von der Oelsnitz^{*}

^{*} Institut für Unternehmensführung, Technische Universität Braunschweig
Braunschweig, Deutschland

[†] Institut für Marketing, Technische Universität Braunschweig
Braunschweig, Deutschland

Abstract — The today's working environment is characterized by a disruptive transformation. Digitalization and automatization change the conditions of work, what forces reaction of involved actors. This paper intends to show the state-of-the-art of competence development and future trends of work. By conducting a systematic literature review the main focus of existing research regarding competencies is identified. Results show the increasing importance of technical and social skills in order to cope with future working trends.

Zusammenfassung — Die Arbeitswelt befindet sich im Umbruch. Digitalisierung und Automatisierung führen zu veränderten Rahmenbedingungen, auf die es sich einzustellen gilt. Dieses Paper arbeitet im Rahmen einer Literaturanalyse den Stand der Forschung systematisch auf. Dabei werden sowohl die Trends der Arbeitswelt, als auch die notwendigen zukünftigen Kompetenzen der Mitarbeiter betrachtet. Hier zeigt sich ein Schwerpunkt der Literatur insbesondere auf der zunehmenden Bedeutung von technischen und sozialen Kompetenzen.

I. BEZUGSRAHMEN UND FORSCHUNGSFRAGE

Die Arbeitswelt und alle beteiligten Akteure unterliegen einem stetigen Wandel, der durch den Einfluss verschiedener globaler Trends vorangetrieben wird. Das Akronym VUKA kann dabei als übergeordneter Rahmen interpretiert werden. VUKA steht für Volatilität, Ungewissheit, Komplexität und Ambiguität und beschreibt damit Faktoren, die zunehmend auf die moderne Arbeitswelt und damit auf alle (Management-)Entscheidungen einwirken. Volatilität meint dabei die schnelle und disruptive Dynamik von Veränderungen. Die Unsicherheit steigt an, da Zusammenhänge und grundlegende Muster aufgrund zunehmender Verflechtungen von Unternehmen an Transparenz verlieren. Konsequenzen unternehmerischer Entscheidungen bleiben ungewiss. Aufgrund der globalen Verflechtungen werden Handlungsmöglichkeiten diverser, was in einer Erhöhung der allgemeinen Komplexität, aber auch in einer Mehrdeutigkeit von Informationen und Rahmenbedingungen mündet. [1] Es wird deutlich, dass die genannten Entwicklungen dazu beitragen, dass die unternehmerische Umwelt nicht vollständig kontrolliert werden kann. Vielmehr wird die Notwendigkeit unterstrichen, sich mit den systemischen Verflechtungen auseinanderzusetzen und diese in strategische (Personal-)Entscheidungen zu integrieren. Diese Tendenzen werden durch globale und langfristige Megatrends verschärft und angetrieben. [2] Der vorliegende Beitrag baut auf den Megatrends auf, die die fortschreitende Technisierung und Informatisierung der Arbeitswelt adressieren. Exemplarisch lässt sich hier die Digitalisierung nennen. Diese und die fortschreitende Automatisierung bewirken neben der Transformation von traditionellen Geschäftsmodellen auch eine grundlegende Veränderung von Produktionsprozessen. In diesem Zusammenhang wird häufig der Begriff Industrie 4.0 verwendet. [3] Daneben werden aber auch die Art und Weise der Zusammenarbeit in und zwischen Unternehmen durch verschiedene Technologien, wie die Verbreitung von Internettechnologien oder Cloudsystemen, beeinflusst. [4] Es etablieren sich

zunehmend individualisierte und flexible Arbeitsformen, wie bspw. das virtuelle-, mobile oder plattformbasierte Arbeiten. [4] Die skizzierten Entwicklungen werden in der Literatur unter Begriffen wie New Work oder Arbeit 4.0 subsummiert. [5] Diese Arbeitsformen ermöglichen gerade im Bereich der Wissensarbeit zwar einerseits eine Steigerung der Performance [6], erfordern andererseits von Unternehmen und Mitarbeitern eine Weiterentwicklung vorhandener sowie eine Erschließung zukunftsweisender Kompetenzen. Es lässt sich ableiten, dass die Arbeitswelt technischer, vernetzter, informatisierter und komplexer wird. [1] Unternehmen sind daher gefordert, entsprechende Personalentwicklungsmaßnahmen anzustossen und die betriebliche Ausbildung hinsichtlich der vernetzten Arbeitswelt anzupassen. Die fortwährende Erhaltung der Handlungskompetenz des einzelnen Mitarbeiters sollte dabei im Fokus stehen. [7] Daraus erwachsen das Interesse und die Notwendigkeit, sich folgender Fragestellung zu widmen: Welche Kompetenzen und Fähigkeiten gewinnen in der technisierten Arbeitswelt an Bedeutung?

Das Ziel dieses Beitrages ist es, eine systematische Aufarbeitung des Literaturstandes im Rahmen der aufgestellten Forschungsfrage vorzunehmen. Untergeordnet ist es so möglich, erste theoriegeleitete Handlungsempfehlungen für die betriebliche Personalentwicklung abzuleiten. Dabei beziehen sich alle Ausführungen ausschließlich auf Wissensarbeiter. Diese Einschränkung basiert auf der informationsintensiven und nicht-routinierten (Kreativ-)Arbeit, die Wissensarbeiter unter Nutzung von Technologien ausüben. In der Regel verfügen sie über eine gute und umfassende Ausbildung, die durch ein breites Allgemeinwissen ergänzt wird. Systemisches und innovatives Denken sowie Kreativität sind weitere Attribute dieser *kreativen Klasse*. [8] Im Zuge der neuen Arbeitswelt nimmt die Zahl dieser Wissensarbeiter zu, da sich durch die Automatisierung und Technisierung der Arbeitsmarkt und Berufsprofile verändern. [9]

II. THEORETISCHE GRUNDLAGEN

Der berufsbezogene Kompetenzbegriff wird sowohl in der Wissenschaft als auch in der unternehmerischen Praxis unterschiedlich verwendet und aufgefasst. Das etablierte Verständnis des Kompetenzbegriffes umfasst die Ausstattung einer Person mit einer gewissen Befugnis. Daneben wird der Begriff synonym für den Sachverstand einer Person verwendet. Vertreter des OECD-Projektes DeSeCo (Defining and Selecting Key Competencies) erweitern das Verständnis, indem davon ausgegangen wird, dass Mitarbeiter nicht über einzelne Fähigkeiten verfügen, sondern vielmehr über Kompetenz-Sets. Diese kommen in komplexen Situationen zum Tragen und führen durch eine situative Skalierung zu einer erfolgreichen Bewältigung dieser. Dabei integriert kompetentes Handeln verschiedene Elemente, wie Fachwissen oder kognitive und praktische Fähigkeiten. Zudem ergänzen soziale und verhaltensorientierte Komponenten das Kompetenzportfolio. [10] Essenziell ist dabei, dass die verschiedenen Elemente der Kompetenz sowohl erlern- als auch durch eigene Erfahrungen entwickelbar sind. [11] Die durch individuelles Verhalten angepasste Kombination der einzelnen Fähigkeiten und Fertigkeiten mündet in der erfolgreichen Bewältigung beruflicher Herausforderungen. Das skizzierte Verständnis des Kompetenzbegriffs wird diesem Beitrag zugrunde gelegt. Um die übergeordnete Frage des Beitrages zu diskutieren, sollen folglich sämtliche Fertigkeiten und Fähigkeiten gesammelt und hinsichtlich der skizzierten Trends und Entwicklungen innerhalb der Arbeitswelt ausgewertet werden.

III. METHODIK

Die relevante Literatur sowie aussagekräftige Studien wurden durch eine systematische Datenbankauswertung identifiziert. Dazu wurde sich aus forschungswissenschaftlichen Gründen auf die wirtschaftswissenschaftliche Datenbank SCOPUS beschränkt. Das Vorgehen der systematischen Literatursuche folgt der Building-Blocks-Methode. Goodman et al. (2014) beschreiben diese als einen zentralen Ansatz zur elektronischen Literaturrecherche. Die Datenbankabfrage erfolgt hierbei unter Rückgriff auf eine Kombination aus Schlüsselwörtern, durch den Einsatz von Verknüpfungen (bspw. AND/OR) und mittels einer spezifischen Eingrenzung von Suchparametern. Bei der Entwicklung des Suchstrings liegt so ein iterativer Prozess zugrunde, der durch wiederholte Anpassung der Suchbegriffe und Kriterien eine Eingrenzung der Suchergebnisse anstrebt. [12] So wurden im Rahmen der Building-Blocks-Methode zunächst zentrale Schlüsselbegriffe auf Basis zuvor identifizierter Literatur festgelegt. Durch sukzessive Erweiterung dieser um zusätzliche Suchwörter, konnte die in Tabelle 1 dargestellte Schlüsselwortliste generiert werden. Unter Verwendung der Booleschen Operatoren „AND“ und „OR“ wurden die Suchworte zu einem Suchstring verknüpft, der als Grundlage für die Recherche in der Datenbank SCOPUS diente. Die eingesetzte Suchphrase setzte sich sowohl aus englischen als auch deutschen Begriffen zusammen.

Da das Kernthema des Beitrags zukünftige Kompetenzen von Mitarbeitern adressiert, sind alle Begriffe, die diese Kategorie abdecken, als Kernbegriffe anzusehen. Untereinander stehen die Begriffe innerhalb des Suchstrings in einer nicht-ausschließende disjunktiven Beziehung und werden durch den Operator „OR“ verknüpft. Weiter fokussiert sich der vorliegende Beitrag auf eine sowohl technisierte als auch informationsgetriebene Arbeitswelt. Dies hat für die weitere Recherche zur Folge, dass Aspekte der Technisierung und Digitalisierung von der Gestaltung, Ausführung und Art der Arbeit aufgenommen werden müssen. Der zweite Kernaspekt ist unmittelbar mit der ersten Kategorie verbunden, was eine Adjunktion

und somit eine „AND“-Verknüpfung erfordert. Weiterhin wirken allgemeine Trends auf die Veränderungen von Arbeit ein, sodass unter der dritten Kategorie erweiterte, in den Kontext passende Aspekte aufgenommen und analog in den Suchstring eingebunden wurden. Dabei wurden entsprechend der Building-Blocks-Methode alle Kernbegriffe um synonyme bzw. verwandte Begriffe ergänzt.

Eine iterative Anpassung des initialen Suchstrings und eine Eingrenzung der Ergebnisse auf die Veröffentlichungsjahre 2015-2018 und die Publikationsart auf Journalbeiträge, reduzierten die Suchtreffer in SCOPUS von 18.803 auf 124 Ergebnisse. Im Anschluss an ein Titel-Screening (Reduktion auf 22 Veröffentlichungen) wurde diese Auswahl durch ein Abstract-Screening auf elf Veröffentlichungen eingegrenzt. Aus dieser Eingrenzung wurden nach Volltextsuchung letztlich sieben Beiträge in die Analyse einbezogen. Ergänzend zu dieser Auswahl konnten durch manuelle Nachrecherche weitere Paper identifiziert und in der weiteren Untersuchung berücksichtigt werden.

TABELLE I. SUCHWORTLISTE

AND		
Kernaspekt 1 „Kompetenzen“	Kernaspekt 2 „Technisierung von Arbeit“	Aspekt 3 „umfassende Begriffe“
- (Job) Skills	- Collaborative	- VUCA
- Competence	- Technology	- Megatrends
(Development)	- Digitalisation	- Arbeitswelt 4.0
- Qualification	- ICT	- Employment
- Resources	- Industrial Revo-	Market
- Graduate	lution	- HRM
Employability	- Smart Work	- Education
	- Future of Work	

IV. ZENTRALE ERGEBNISSE UND DISKUSSION

Durch die Analyse der identifizierten Literatur ergibt sich eine logische Untergliederung der für die zukünftige Entwicklung der Arbeitswelt relevanten Kompetenzen. Abstrahiert ist in den vorliegenden Artikeln zunächst festzustellen, dass die Themen Innovation, Kreativität sowie das Entrepreneurship von elementarer zukünftiger Bedeutung für Unternehmen sind. [13] Die von Bruyne und Gerritse (2018) ermittelte zukünftige Wichtigkeit von Künstlicher Intelligenz (KI), Digitalisierung und Automatisierung [4] wird auch von Sousa und Wilks (2018) bestätigt. [14]

Aus der identifizierten Literatur können vorwiegend zwei grundlegende Kategorien von Kompetenzen abgeleitet werden: generische sowie fachliche Kompetenzen. Damit wird sich der Klassifizierung von Bozic et al. (2018) angeschlossen. [15] Bruyne und Gerritse (2018) beschreiben anlehnd das sogenannte T-Profil von Mitarbeitern. Dieses beschreibt ein Fähigkeitsprofil, was neben einer hohen fachlichen Spezialisierung auch die Entwicklung von Überblickswissen sowie Teilkompetenzen in anderen Fachgebieten vorsieht. [16]. Die zunehmend auf IT sowie KI ausgerichteten Qualifikationen bilden dabei die fundamentale Basis der zukünftigen Beschäftigungsfähigkeit. Visualisiert wird dieses durch die vertikale Ausrichtung des „T“. Zudem wird das horizontale Element des „T“ durch generisch-versatile Fähigkeiten bzw. Metawissen komplettiert. [4] An dieser Stelle findet sich die zuvor skizzierte Definition des Kompetenzbegriffes wieder. Gleichzeitig wird die Mehrdimensionalität von Kompetenzen verdeutlicht.

Die Präsenz und die Durchschlagkraft der Digitalisierung und ihrer Konsequenzen lassen sich in allen Beiträgen wiederfinden. Die klare Einordnung in eine der genannten Dimensio-

nen ist nicht abschließend möglich, sodass für eine Beurteilung die dritte Dimension *Digitale Kompetenz* als sinnvoll erachtet und somit ergänzt wird.

Im Folgenden werden die Dimensionen auf Grundlage der systematischen Literaturanalyse näher spezifiziert.

Nach Motyl et al. (2017) umfassen zukünftig relevante Fachkenntnisse die Softwareprogrammierung und das Wissen um Softwareprogramme sowie deren Steuerung. [17] Bruyne und Gerrits (2018) ergänzen dazu Qualifikationen im Bereich der Robotik und der angewandten Mathematik. [4] Es wird deutlich, dass gerade in diesem Ausbildungsbereich Schulen und Hochschulen eine deutliche Verantwortung zugeschrieben bekommen. [18] Derzeitige Studien prognostizieren, dass nur 15% der zukünftigen Arbeitnehmer über genannte wettbewerbsfähige IT-Fähigkeiten verfügen werden. [4]

Generische Kompetenzen stellen eine Kombination aus erlernbaren sowie analytischen und problemlösenden Fähigkeiten dar. Diese sind nicht direkt kontextbezogen und äußern sich bei unterschiedlichen beruflichen Situationen und Herausforderungen. [15] Llorens et al. (2017) gliedern an dieser Stelle auch Softskills ein, die zunehmend an Relevanz gewinnen. Exemplarisch lassen sich hier Proaktivität, Teamfähigkeit und Kooperationsfähigkeit sowie der Umgang mit Informations- und Kommunikationstechnologien (IKT) nennen. [13] Während der Trend darauf hindeutet, dass die genannten Technologien in naher Zukunft die Prozesse der Unternehmen bestimmen werden, bleibt der Faktor Mensch als Vermittler und Bediener unerlässlich. Dies rückt den versierten Umgang mit IKT in den Fokus der Kompetenzbetrachtung. Dabei ist anzumerken, dass IKT-Fähigkeiten nicht formal institutionell erlernt werden. Vielmehr werden diese on-the-job entwickelt und ausgebaut. [18] Ihre Wichtigkeit wird dadurch unterstrichen, dass höhere IKT-Fähigkeiten zu einem Lohnanstieg von 25% führen können. [18] Das Beherrschung der Fähigkeiten kritisch und lösungsorientiert zu denken, wird von Sousa und Wilks (2018) ergänzt. [14] Vor dem Hintergrund des Innovationsdrucks verweisen die Autoren auf die gesteigerte Wichtigkeit der Neugierde. [14] Kirchherr et al. (2018) titulieren die generischen Fähigkeiten auch als *nicht-digitale Schlüsselkompetenzen*. Zu diesen zählen sie auch das unternehmerische Denken sowie die Adoptionsfähigkeit. [19]

Die letzte Dimension umfasst die bereits erwähnten Digitalen Kompetenzen. Hierzu werden sämtliche Fähigkeiten gezählt, die im digitalen Kontext zum Tragen kommen. An dieser Stelle lassen sich die Fähigkeit des effektiven agilen Arbeitens sowie die virtuelle Kollaboration verorten. Zentrales Element der digitalen Kompetenzen ist die sogenannte Digital Literacy, die den informierten und sicheren Umgang mit persönlichen und unternehmerischen Daten im Internet meint. Ergänzend muss an dieser Stelle auf die digitale Ethik hingewiesen werden. Durch die Vernetzung entstehen große Datenmengen, deren Nutzung und Analyse sich zukünftig zu einem lukrativen Geschäftsfeld entwickelt und einen großen Beitrag zur Wissensgenerierung im Unternehmen stiftet. Mitarbeiter müssen zunehmend das eigene digitale Verhalten mit den vorliegenden Daten kritisch hinterfragen und ethische Entscheidungen treffen. Eine weitere Dimension der Digitalen Kompetenz ist die Fähigkeit der digitalen Interaktion und des digitalen Lernens. Die Nutzung von Online-Kanälen und das Wissen um ein angemessenes virtuelles Verhalten gewinnen an Bedeutung. Diese Fähigkeiten lassen sich als digitale Schlüsselkompetenzen verstehen. [19] In diesem Zusammenhang lässt sich auch der bewusste Einsatz von Technologie zu den digitalen Schlüsselkompetenzen zählen. [20] Es wird deutlich, dass der Forschungsstand einheitlich den Trend der Technisierung und Digitalisierung abbildet und als Kernaufgabe zukünftiger Organisationsentwicklung adressiert.

V. LIMITATIONEN UND IMPLIKATIONEN

A. Limitationen

Der vorliegende Beitrag weist einige Limitationen auf. Zunächst erlaubt es die gewählte Methode der systematischen Literaturanalyse nur vage Aussagen hinsichtlich der Verallgemeinerung der recherchierten Ergebnisse zu treffen. Neben verschiedenen Branchenbetrachtungen legen die Beiträge unterschiedliche Studiendesigns zur Erhebung der relevanten Daten zugrunde und vertreten somit unterschiedliche Evidenzklassen. Des Weiteren ergeben sich Limitationen aus der Datenbankrecherche, insbesondere durch die gewählten Filter und die Beschränkung der Analyse auf nur eine Datenbank. Auch die methodenbedingte deduktive Entwicklung der Suchbegriffe kann als Limitation aufgefasst werden. Gegebenenfalls konnten nicht alle relevanten Begriffe berücksichtigt werden. Es fällt zudem auf, dass in vielen identifizierten Artikeln und Veröffentlichungen sehr oberflächlich argumentiert wird und Begriffe nicht trennscharf voneinander abgegrenzt werden. Zum Teil werden von den Autoren identifizierte Kompetenzen nicht weiter erläutert oder hinreichend differenziert betrachtet. So bleibt es mitunter unklar, wie in einzelnen Publikationen beispielsweise der Begriff der digitalen Kompetenzen charakterisiert wird. Eine valide Messung von Kompetenzen wird durch unzureichende Indikatoren erschwert.

B. Implikationen

Die Analyse der Artikel legt die Ableitung einiger Handlungsempfehlungen nahe, die sich insbesondere für Personalverantwortliche ergeben. Übergeordnet lässt sich die Förderung der fortwährenden Beschäftigungsfähigkeit (employability) der Mitarbeiter im Zuge des Lebenslangen Lernens proklamieren. Starre Kompetenzen tragen langfristig nicht zum Geschäftserfolg bei und müssen stetig ausgebaut und trainiert werden. [14] So entwickelt sich der versierte Umgang und kompetente Einsatz von digitalen Technologien zur wesentlichen Schlüsselkompetenz im beruflichen Alltag. Dazu ist eine Auseinandersetzung mit den T-Profilen der Mitarbeiter unumgänglich. Reine Fachspezialisten ohne ergänzende soziale Teilkompetenzen können in der vernetzten Arbeitswelt von morgen nur einen Teil der erforderlichen Potenziale freisetzen. Nur durch ein adäquates Zusammenspiel innerhalb des Kompetenz-Sets ist es Mitarbeitern möglich, lösungsorientiert komplexen Herausforderungen zu begegnen und Innovationen zu generieren. Die Aufgabe der Personalentwickler wird es an dieser Stelle sein, den zukünftigen Anforderungen gerecht werdende Maßnahmen anzustossen. Dabei gilt es, das vorhandene Kompetenzportfolio des Mitarbeiters stetig zu erweitern und individuell zu fördern. Durch den zunehmenden Einfluss der VUCA-Umgebung werden Mitarbeiter zu Schlüsselakteuren im Wettbewerb um Innovationen innerhalb der digitalen und smarten Welt.

LITERATURVERZEICHNIS

- [1] Bennett N, Lemoine GJ. "What a difference a word makes: Understanding threats to performance in a VUCA world". *Business Horizons* 2014; 57(3):311–7. doi: 10.1016/j.bushor.2014.01.001.
- [2] Graf N, Gramß D, Edelkraut F. "Agiles Lernen": Neue Rollen, Kompetenzen und Methoden im Unternehmenskontext. 1. Auflage. Freiburg, München, Stuttgart: Haufe Gruppe; 2017.
- [3] Schuh G, Potente T, Wesch-Potente C, Weber AR, Prote J-P. "Collaboration Mechanisms to Increase Productivity in the Context of Industrie 4.0". *Procedia CIRP* 2014; 19:51–6. doi: 10.1016/j.procir.2014.05.016.
- [4] Bruyne E de, Gerrits D. "Exploring the future workplace: results of the futures forum study". *Journal of Corporate Real Estate* 2018; 20(3):196–213. doi: 10.1108/JCRE-09-2017-0030.
- [5] Hackl B, Wagner M, Attmer L, Baumann D. "New Work: Auf dem Weg zur neuen Arbeitswelt": Management-Impulse, Praxisbeispiele, Studien. Wiesbaden: Springer Gabler; 2017.

- [6] Rigby KD, Tager S. "Leading a Digital Transformation"; 2014 [Stand: 16.10.2018]. Verfügbar unter: https://www.bain.com/contentassets/43b1e275d2fe4eeb90e9e19ad588b55d/bain_brief_leading_a_digital_transformation_.pdf.
- [7] Heinze H, Kaßebaum B. "Kompetenzentwicklung als Gestaltungsaufgabe für eine „erweiterte moderne Beruflichkeit“". In: Ahrens D, Molzberger G, Hrsg. *Kompetenzentwicklung in analogen und digitalisierten Arbeitswelten: Gestaltung sozialer, organisationaler und technologischer Innovationen*. Berlin, Heidelberg: Springer; 2018. S. 173–86 [Kompetenzmanagement in Organisationen].
- [8] Florida R, Mellander C, Stolarick K. "Inside the black box of regional development—human capital, the creative class and tolerance". *J Econ Geogr* 2008; 8(5):615–49. doi: 10.1093/jeg/lbn023.
- [9] Ahmadi A, Abzari M, Nasr Isfahani A, Safari A. "High-performance, knowledge sharing and ICT skills". *HSM* 2018; 37(3):271–80. doi: 10.3233/HSM-17169.
- [10] Rychen DS, Salganik LH, Hrsg. "Key competencies for a successful life and a well-functioning society". Cambridge, Mass.: Hogrefe & Huber; 2003.
- [11] Krumm S, Mertin I, Dries C. "Kompetenzmodelle". Göttingen: Hogrefe; 2012, in: *Praxis der Personalpsychologie; Bd. 27*. Verfügbar unter: http://sub-hh.ciando.com/book/?bok_id=471885.
- [12] Goodman JS, Gary MS, Wood RE. "Bibliographic Search Training for Evidence-Based Management Education: A Review of Relevant Literatures". *AMLE* 2014; 13(3):322–53. doi: 10.5465/amle.2013.0188.
- [13] Llorens A, Berbegal-Mirabent J, Llinàs-Audet X. "Aligning professional skills and active learning methods: an application for information and communications technology engineering". *European Journal of Engineering Education* 2017; 42(4):382–95. doi: 10.1080/03043797.2016.1189880.
- [14] Sousa MJ, Wilks D. "Sustainable Skills for the World of Work in the Digital Age". *Syst. Res* 2018; 35(4):399–405.
- [15] Bozic M, Certic JD, Vukelic M, Cizmic S. "New Instructional Approach for Fostering Generic and Professional Competences: Case Study of the Project and Problem Based Learning Engineering Practice Course.". *International Journal of Engineering Education* 2018; (34):1581–91.
- [16] Busch MW, Oelsnitz Dvd. "Teammanagement": Grundlagen erfolgreichen Zusammenarbeitens. 1. Auflage. Stuttgart: Verlag W. Kohlhammer; 2018, in: *Organisation und Führung*. Verfügbar unter: <https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5431603>.
- [17] Motyl B, Baronio G, Uberti S, Speranza D, Filippi S. "How will Change the Future Engineers' Skills in the Industry 4.0 Framework? A Questionnaire Survey". *Procedia Manufacturing* 2017; 11:1501–9. doi: 10.1016/j.promfg.2017.07.282.
- [18] Falck O, Schüller S. "Querschnittstechnologie Internet – Universallösung für den Arbeitsmarkt der Zukunft?". *Wirtschaftsdienst* 2016; 96(8):609–13. doi: 10.1007/s10273-016-2023-z.
- [19] Kirchherr JW, Klier J, Lehmann-Brauns C, Winde M. "Future Skills: Welche Kompetenzen in Deutschland fehlen. Future Skills - Diskussionspapier 1"; 2018.
- [20] Krumm S, Kanthak J, Hartmann K, Hertel G. "What does it take to be a virtual team player? The knowledge, skills, abilities, and other characteristics required in virtual teams". *Human Performance* 2016; 29(2):123–42. doi: 10.1080/08959285.2016.1154061.

Session Student Papers

Chair: Marin Marinov

THIS PAGE IS INTENTIONALLY LEFT BLANK

Using Formant as an indication of speech intelligibility in VOIP

Verwenden der Formant als Indikation der Sprachverständlichkeit in VOIP

Angel Garabitov^{*}, Aleksandar Tsenov[†]

Fakultät für deutsche Ingenieur- und Betriebswirtschaftsausbildung, TU – Sofia
Sofia, Bulgaria, ^{*}angel.garabitov@fdiba.tu-sofia.bg, [†]aleksandar.tsenov@fdiba.tu-sofia.bg

Abstract — Voice over Internet Protocol (VoIP) is a revolutionary technology which is acting as a platform for the development of latest trends in modern communication world. The speech signal quality in VoIP is governed by the speech coding techniques employed. Conveying the message is one of the most important goals. This can be challenging since the intelligibility of the speech may be harmed at various stages before, during and after the transmission process from sender to receiver. Effective quality evaluation methodologies are important for system development and refinement, particularly by adopting user feedback-based measurement. The article researches an affect-based QoE (Quality of Experience or QoE-based) evaluation framework which attempts to capture users' perception while they are engaged in voice communication. The article makes a proposal to use a formant as criteria for the prediction of subjective voice quality assessment. In the work we propose a new measure which show good prediction results with the intelligibility of nonlinear processed speech. The newly proposed measure is of a low computational complexity and mathematically tractable which make them suitable for optimization of new signal processing solutions which aim for improving speech intelligibility.

Zusammenfassung — Voice over Internet Protocol (VoIP) ist eine revolutionäre Technologie, die als Plattform für die Entwicklung der neuesten Trends in der modernen Kommunikationswelt dient. Die Sprachsignalqualität in VoIP wird durch die verwendeten Sprachcodierungstechniken bestimmt. Die Übermittlung der Botschaft ist eines der wichtigsten Ziele. Dies kann eine Herausforderung darstellen, da die Verständlichkeit der Sprache in verschiedenen Stadien vor, während und nach dem Übertragungsprozess vom Sender zum Empfänger beeinträchtigt sein kann. Effektive Qualitätsbewertungsmethoden sind wichtig für die Systementwicklung und -verfeinerung, insbesondere durch die Verwendung von Benutzer-Feedback-basierten Messungen. Der Artikel untersucht ein affektives QoE (Quality of Experience oder QoE-basiertes) Evaluierungs-Framework, das versucht, die Wahrnehmung der Nutzer während der Sprachkommunikation zu erfassen. Der Artikel schlägt vor, einen Formanten als Kriterium für die Vorhersage der subjektiven Sprachqualitätsbewertung zu verwenden. In der Arbeit schlagen wir ein neues Maß vor, das gute Vorhersageergebnisse mit der Verständlichkeit von nichtlinear verarbeiteter Sprache zeigt. Die neue vorgeschlagene Maßnahme ist von geringer rechnerischer Komplexität und mathematisch handhabbar, was sie zur Optimierung neuer Signalverarbeitungslösungen geeignet macht, die auf eine Verbesserung der Sprachverständlichkeit abzielen.

I. EINFÜHRUNG

Sogar in einer reinen Klangsилбе kann es eine ganze Reihe von Klangfrequenzen sein. Die menschlichen Stimmorgane fügen dem Grundton, den die Stimmbänder bilden, zusätzliche harmonische Komponenten hinzu. Diese Komponenten verleihen der Stimme Farbe, wodurch wir insbesondere die Sprache der Personen, die wir kennen, erkennen können. Als Ergebnis der Forschungen wurde gefunden, dass vier Frequenzen, die sich aktiv in den Resonanzhohlräumen des Stimmtrakts bilden, aktiv an der Sprachbildung beteiligt sind. Diese Frequenzen werden Formanten genannt. Bei der Artikulation können sich sowohl die Amplitude als auch die Frequenz der Formatkomponenten des Klangs ändern. Gleichzeitig bleibt jedoch die Anzahl der Formanten in Sprachtönen konstant und beträgt immer vier. [1]

Bei Geräuschen ist es schwierig, Formantkomponenten darin zu identifizieren.

In Bezug auf die Wissenschaft, die sich mit der Erkennung und Synthese von Sprache befasst, stellen wir hier ein anderes Konzept der Einheit des Tonsystems vor – das Phonem.

Phoneme werden nicht direkt mit Wörtern oder Silben identifiziert. Sie spielen die Rolle unteilbarer Teilchen, Zungenatome und sind Tonfolgen. [2,3] Alle anderen Sprachkonstrukte wie Silben und Wörter setzen sich aus Phonemen zusammen. Phoneme verfügen über zahlreiche Funktionen, die zur Klassifizierung und Erkennung verwendet werden können.

Das Wissen, dass die Frequenzen der ersten Formanten der Phoneme (d. H. Die niedrigsten akustischen Resonanzen des Stimmtrakts) primäre Informationsträger in der menschlichen Sprache sind.

Dies hat viele Forscher veranlasst, Mittel zum Abschätzen von Formantfrequenzen aus dem akustischen Signal zu entwickeln. Ein Algorithmus zur automatischen Formantenschätzung kann Anwendungen finden, darunter: Klassifizierung von vokalähnlichen phonetischen Einheiten in einem sprecherunabhängigen Erkennungssystem [4], Erfassung von Parametern für Diphone-Text-zu-Sprache-Systeme [5] und Datenreduktion für die Forschung in der akustischen Phonetik.

Die Natur der Formant - Stimmfalten (Bündel) schwanken, wenn sie unter dem Bernoulli-Effekt durch sie geblasen werden.

Anregung einer Volumengeschwindigkeit mit einer einzigen Amplitude am Eingang des Wellenleiters (1).

$$u(t) = \sin(\omega t) \quad (1)$$

Die Übertragungsfunktion ist in der Form:

$$u(x, t) = \frac{\sin(\omega t) \cos\left(\frac{\omega(l-x)}{c}\right)}{\cos\left(\frac{\omega l}{c}\right)} \quad (2)$$

Maxima im Spektrum entsprechen stehenden Wellen mit einer Wellenlänge (3):

$$\lambda(n) = \frac{1}{1/4 + n/2}, n = 0, 1, \dots \quad (3)$$

Oder Frequenzen:

$$F(n) = \frac{c}{4l} (1 + 2n), n = 0, 1, \dots \quad (4)$$

Die Übertragungsfunktion des Vokaltrakts und folglich das Sprachsignal sind durch Maxima in dem um mehrere hundert Hertz getrennten Spektrum gekennzeichnet und hängen im Wesentlichen von der Form des Vokaltrakts ab. Für Vokale werden diese Maxima Formanten genannt.

Eine effiziente und kompakte Darstellung der zeitabhängigen Sprachmerkmale bietet potenzielle Vorteile für die Spracherkennung. Formant-Tracking-Methoden, die auf Linear Prediction Analysis (LPC) basieren, haben beträchtliche Aufmerksamkeit gefunden. Wurzelfindungsalgorithmen werden verwendet, um die Nullstellen des LPC-Polynoms zu finden, oder lokale Maxima der LPC-Hüllkurve [6] werden unter Verwendung von Spitzenspiking-Techniken gesucht. Das Problem bei Verfahren zur numerischen Lösung ist, dass die Bestimmung der Formantenfrequenzen und -bandbreiten nur für komplexe konjugierte Pole und nicht für reelle Pole erfolgreich ist.

Es kann jedoch bestimmte Aspekte geben, aufgrund der formantbasierten Parameter attraktiv sind:

- Formanten gelten als robust gegenüber Kanalverzerrungen und Rauschen.
- Formant-Parameter können ein Mittel sein, um das Problem einer Nichtübereinstimmung zwischen Trainings- und Testbedingungen zu lösen.
- Es gibt eine enge Beziehung zwischen Formant-Parametern und modellbasierten Ansätzen der Sprachwahrnehmung und -produktion.

II. LPC-FORMANT-SCHÄTZUNG

Das Sprachsignal wird durch die Faltung der Anregungsquelle und der zeitveränderlichen Komponenten des Vokaltrakt-systems erzeugt. [7] Diese Anregungs- und Vokaltraktkomponenten sind von dem verfügbaren Sprachsignal zu trennen, um diese Komponenten unabhängig zu untersuchen. Zur Dekonvolution der gegebenen Sprache in Anregungs- und Vokaltraktssystemkomponenten werden Methoden entwickelt, die auf homomorphen Analyse wie Cepstralanalyse basieren. Da die Cepstralanalyse die Dekonvolution von Sprache in Quell- und Systemkomponenten

durchläuft, indem sie durch den Frequenzbereich wandert, wird die Dekonvolutionsaufgabe ein rechenintensiver Prozess. Um eine solche Art von Berechnungskomplexität zu reduzieren und die Quellen- und Systemkomponenten aus der Zeitdomäne selbst zu finden, wird die lineare Vorhersageanalyse entwickelt.

Die isolierten Signale werden mit einem linearen prädiktiven Codierungsmodell analysiert, um die Formanten zu finden.

Beschreibung des Algorithmus:

1. Führen Sie ein autoregressives Autokorrelationsmodell (LPC) der Ordnung LPC_COEFF aus
2. Berechnen Sie die komplexen Wurzeln des LPC-Modells
3. Ignorieren Sie Wurzeln mit positivem Imaginärteil (beliebig, hätte die Negative auch ignorieren können)
4. Konvertieren Sie die Wurzeln in den Frequenzbereich basierend auf der Abtastfrequenz

Die Formantenfrequenzen (Abbildung 1) des normalen Schreibens von Säuglingen wurden unter Verwendung von drei verschiedenen Schätztechniken gemessen; Schallspektrographie, lineare prädiktive Codierung (LPC) und Leistungs-spektrumanalyse.

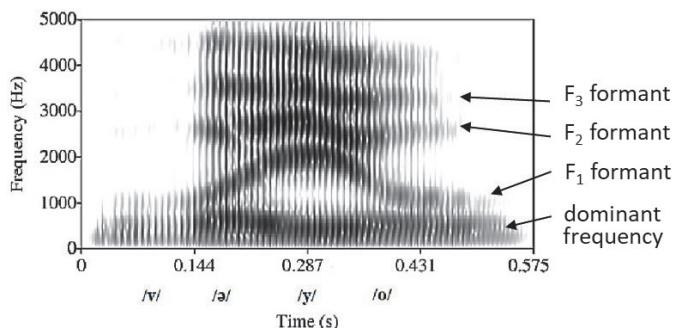


Abb. 1. Beispiel für Formanten

Resonanzen in Sprache Vocal Tract (Hals + Zunge + Lippen) agiert als variabler Resonator. Resonanzen sind Formanten. [8]

III. SPRACHERKENNUNG DER DEUTSCHEN LANGEN MONOPHTHONGE MIT FORMANTEN

Es gibt sieben lange Monophthonge in deutscher Sprache (Tabelle 1).

TABELLE I. LANGE MONOPHTHONGE IN DEUTSCHER SPRACHE

Lange Monophthonge	Geschrieben	Gesprochen
i:	viel	/fi:l/
e:	Beet	/be:t/
a:	Saat	/za:t/
o:	Boot	/bo:t/
u:	Hut	/hu:t/
y:	Rübe	/ry:bə/
ø:	Öl	/ø:l/

Die verwendeten Vokale stammen von den E-Learning-Einheiten auf dem Virtual Linguistics Campus. [9]

Auf Abbildung 2 sind sichtbar: die Amplitude, der Frequenzgang, erste und zweite Formant. Sie sind durch die Algorithmen der Praat Programme [10] berechnet.

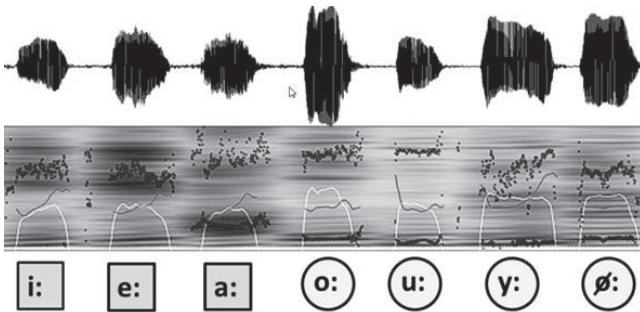


Abb. 2. Lange Monophthonge mit Formanten

IV. BERECHNUNGEN FÜR MONOPHTHONGE I

Bevor wir die Spracherkennung versuchen, müssen wir das Sprachsignal vorverarbeiten. Während dieser Verarbeitung sollten Rausch- und Fremdsignale entfernt werden, deren Frequenzspektrum außerhalb des menschlichen Sprachspektrums liegt.

Um das Problem der Erkennung zu lösen, müssen die Hauptmerkmale der Sprache identifiziert werden, die in den nachfolgenden Stufen des Erkennungsprozesses verwendet werden. Primärmerkmale werden durch Analyse der spektralen und dynamischen Eigenschaften des Sprachsignals unterschieden.

Vokale können phonetisch schwer zu beschreiben sein, da sie Punkte oder Bereiche innerhalb eines zusammenhängenden Raums sind. Die ersten beiden Formanten sind wichtig für die Bestimmung der Qualität von Vokalen. Zur Hervorhebung der informativen Merkmale des Sprachsignals wird die spektrale Darstellung von Sprache verwendet.

Wir berechnen die Frequenzantwort des Tons i: nach den klassischen Methoden. Das Ergebnis ist auf Abbildung 3 dargestellt. Im Frequenzgang sind die schwarzen Linien der Formate deutlich sichtbar.

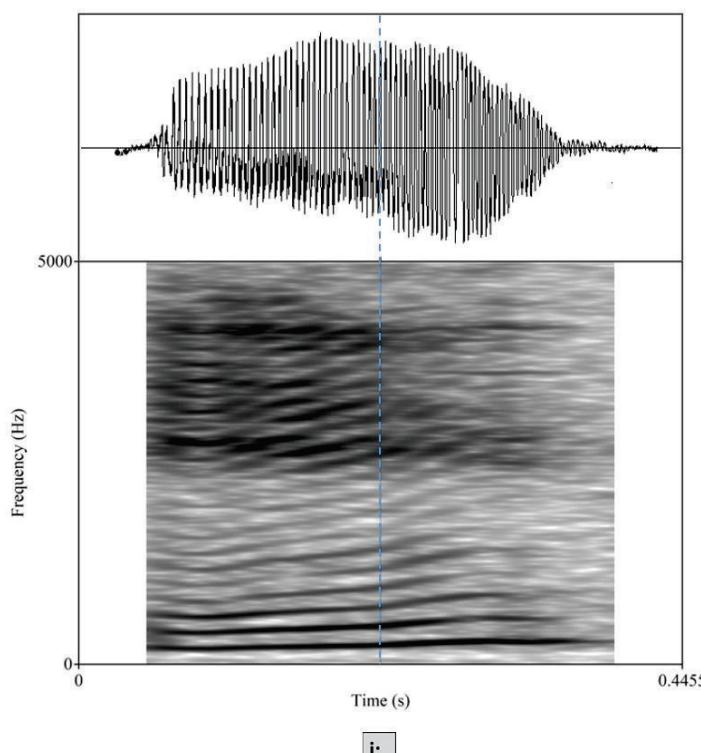


Abb. 3. Amplitude und Frequenzgang der Monophthong i:

Abbildung 4 zeigt eine Frequenzantwort mit der Basisfrequenz beider Formante für einen bestimmten Zeitpunkt.

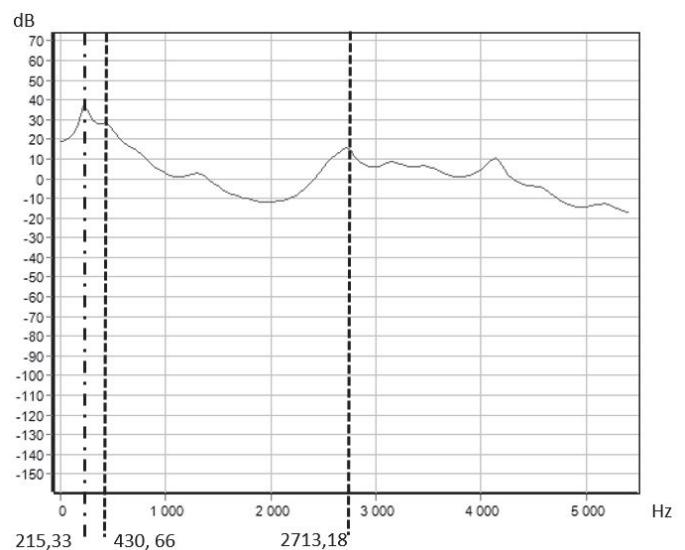


Abb. 4. Frequenzband der Monophthong i

Mit dem in Punkt 2 beschriebenen LPC-Formant-Schätzungs-Algorithmus werden die Werte der ersten beiden Formante für das gesamte Zeitintervall berechnet.

Wir haben MATLAB verwendet, um einen Algorithmus zum Schätzen der Positionen der Formanten von stimmhaften Sprachintervallen basierend auf den Positionen der Sprachpolynomwurzeln zu realisieren, wie sie aus einer rahmenbasierten Analyse eines Sprachsignals unter Verwendung des Verfahrens der linearen Vorhersage-Analyse erhalten werden. Ein Schlüsselaspekt dieses Algorithmus besteht darin, Regionen von quasi zusammenhängenden Sprachrahmen zu finden (wobei die ersten drei Formanten über die Dauer der stimmhaften Region eine zusammenhängende Region bilden) und dann diese Regionen (sowohl rückwärts als auch vorwärts) basierend auf einer etwas schwächeren Region zu erweitern Kriterium für das zusammenhängende Verhalten der Formanten.

Beide Formanten sind auf Abbildung 5 dargestellt.

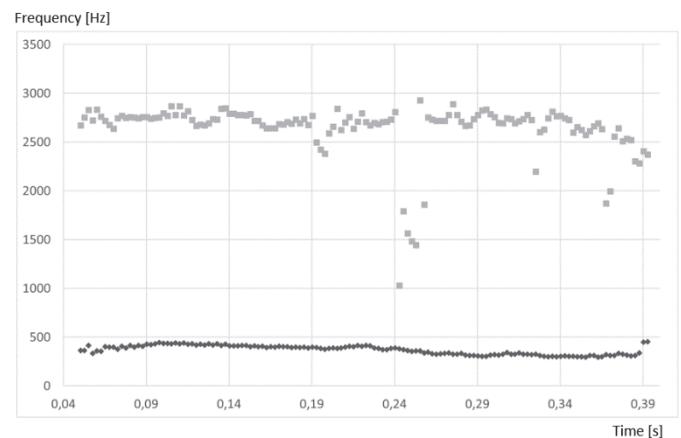


Abb. 5. Erste und zweite Formant der Monophthong i:

V. SCHLUSSFOLGERUNG

Hier haben wir versucht, den akustischen Raum in Tabelle 2 zu bringen. Die ersten beiden Formante sind für die Bestimmung der Vokalqualität am wichtigsten, und dies wird häufig in Form einer Auftragung der ersten Formante gegen die zweiten Formante angezeigt.

Die Erkennung kontinuierlicher Sprache ist ein mehrstufiger Prozess. Nach der Vorverarbeitung des Sprachsignals und dem

Extrahieren von informativen Merkmalen daraus werden die lexikalischen Elemente der Sprache extrahiert. Dies ist die erste Stufe der Anerkennung.

TABELLE II. VERHÄLTNIS DER FORMANTEN DER LANGE MONOPHTHONGE IN DEUTSCHER SPRACHE

Monophthonge	F2/F1
i:	9,47
e:	1,43
a:	3,62
o:	8,05
u:	11,72
y:	12,19
ø:	6,23

Wenn Sie mit einem Lehrer ein Netzwerk unterrichten, können Sie dem Netzwerk beibringen, Objekte zu erkennen, die zu einer vorgegebenen Gruppe von Klassen gehören. Wenn das Netzwerk ohne Lehrer trainiert wird, können Objekte nach ihren digitalen Parametern in Klassen eingeteilt werden.

Der Ansatz basiert auf der Auswahl lexikalischer Elemente (Phoneme, Allophone, Morpheme usw.) in der Sprache.

Das hier erwähnte Verfahren zur Isolierung der lexikalischen Elemente der Sprache basiert auf der Verwendung eines Verhältnisses des zweiten zu dem ersten Format.

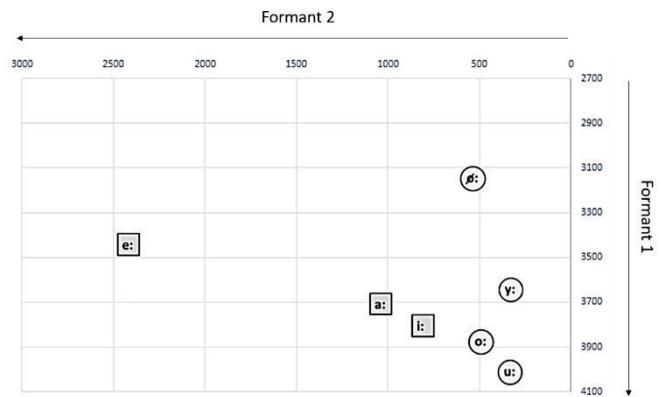


Abb. 6. Grafische Darstellung der Formanten der 7 langen Monophthonge

ANERKENNUNG

Diese Publikation wurde finanziert durch Vertrag zur Doktorandenunterstützung mit Nummer 172PD0010-07 der TU – Sofia.

LITERATURVERZEICHNISS

- [1] Rabiner L., R. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978
- [2] Rishma Rodrigo R., R. Radhika, M. Vanitha Lakshmi, „Speech Enhancement of Vowels Based on Pitch and Formant Frequency“, World Academy of Science, Engineering and Technology, *International Journal of Computer and Systems Engineering*, Vol:9, No:3, 2015
- [3] Madisetti V., *Video, Speech, and Audio Signal Processing and Associated Standards (The Digital Signal Processing Handbook, Second Edition)*, 2009
- [4] Prodeus A., „Formant-Modulation Method of Speech Intelligibility Evaluation“ *Measuring and Exactness*, MEMSTECH'2011
- [5] Gopi E. S., *Digital Speech Processing Using Matlab*, Springer, 2014
- [6] Furui S., *Digital Speech Processing, Synthesis and Recognition*, CRC Press, 2000
- [7] Zheng-Hua Tan, B. Lindberg, *Automatic speech recognition on mobile devices and over communication networks*, Springer, 2008
- [8] Dong Yu, Li Deng, *Automatic Speech Recognition A Deep Learning Approach*, Springer, 2015
- [9] Handke J., The Virtual Linguistics Campus, Online: <http://linguistics.online.uni-marburg.de/> Accesed 10.25.2018
- [10] Boersma P., D. Weenink, University of Amsterdam, Online: <http://www.fon.hum.uva.nl/praat/> Accesed 10.25.2018

Application of Hilbert Transform for Phased Antenna Array Calibration

Anwendung der Hilbert-Transfomation zum Kalibrieren von phasengesteuerten Gruppenantennen

Daniel Hristov

Technical University of Sofia, Faculty of German Engineering and Economics Education
Kliment Ohridski blvd. 8 1756 Sofia, Bulgaria, e-mail: danielh@gilat.com

Abstract — The paper presents a procedure for phase errors estimation across an aperture of phased antenna array applying Hilbert transform on the measured data. The measurement principle, signal processing and the error compensation are described. The procedure is validated on receiving phased antenna array for Ku band, composed of 16x16 V- and H-polarized circular elements. The pattern of calibrated antenna is presented and compared with the pattern of uncalibrated antenna.

Zusammenfassung — Die Arbeit stellt ein Verfahren zur Ermittlung von Phasenfehlern entlang der Apertur einer phasengesteuerten Gruppenantenne unter Anwendung der Hilbert-Transformation auf den gemessenen Daten vor. Das Messprinzip, die Signalverarbeitung und die Fehlerkompensation sind beschrieben. Das Prinzip wurde an einer Empfangsanntenne für Ku-Band validiert, die aus 16x16 V- und H- polarisierten Patchelementen besteht. Das Antennendiagramm der kalibrierten Antenne ist dargestellt und mit dessem unkalibrierter Antenne verglichen.

I. INTRODUCTION

Phased antenna arrays consist of multiple radiating elements, whose phase and amplitude can be individually controlled in order to dynamically modify the radiation pattern of the antenna. Due to technological limitations each phase and amplitude control element introduces an error, or a difference between the optimal and the obtained feeding signal, which leads to a radiation pattern, different from the desired one. The phased antenna array calibration is process of error correction. In order to be efficient, a calibration technique for serial antenna production has to be both fast and precise and must take into account both the features of measurement equipment and the behavior of antenna circuitry, as well as their dynamic characteristics.

The calibration procedure is divided into three steps - field measurement, data processing and error compensation. The measurement is arranged in accordance with the widely used Rotating Electric-Field Vector (REV-) Method.

The data processing relies on conversion of measured sequence to analytical signal using the Hilbert transform and subsequent calculation of correlation with a referent complex sine-function.

II. MEASUREMENT PRINCIPLE AND TEST SETUP

The measurement principle is based on Rotating Electric-Field Vector Method, described in [1, 2, 3] and illustrated on Fig. 1. The block diagram of the test setup is shown on Fig. 2.

The coordinates on Fig. 1 represent the orientation of incident V- or H- polarized waves, emitted by every single radiating element, towards a sensing antenna - with orientation corresponding to the polarization. A spectral analyzer detects the superposition of the incident waves into the sensing aperture.

The measurement is carried out inside of an anechoic chamber in order to minimize errors, caused by interference. The power level is chosen according to the maximal dynamic range of the measured antenna in order to obtain maximal possible signal-to-noise ratio (SNR).

If the feed signal of a particular/individual radiating element is sequentially phase shifted from 0 to 2π radians, the level received by the sensing antenna will describe a sine wave in euclidean plane or a circle in the polar plane. The center of this circle coincide with the level caused by the field of the remaining elements and the radius corresponds to the level, emitted by the analyzed element.

In the classical REV-Algorithm the information about the relative phase is retrieved correlating the measured sequence with a sine and cosine function. This approach delivers correct values only if the shape if the received power sequence is tightly close to an ideal circle and if the shape-deviation is normally distributed along all phase states [4, 5]. If these conditions are not met, the phase-estimation will be biased by the curve shape imperfections.

The phase control circuitry is a typical source of shape distortion. Another source of shape distortion in the case of microstrip antenna arrays is the mutual coupling between the radiating elements. If distortion occurs, the underly-

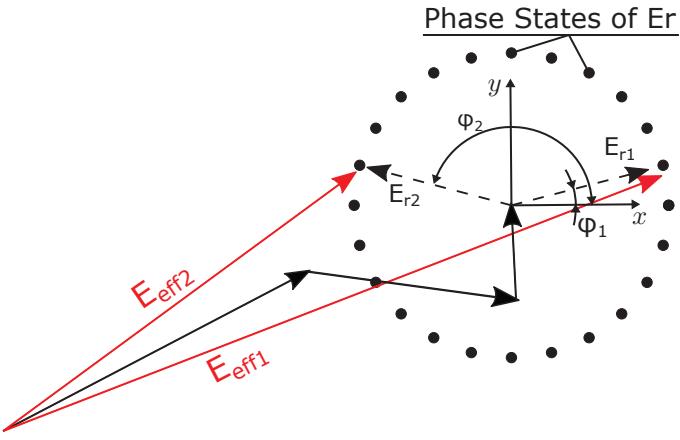


Fig. 1. REV- Principle of superposition.

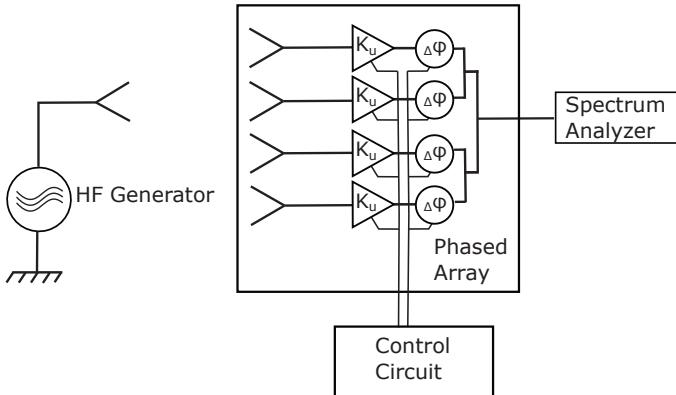


Fig. 2. Test Setup - Block Diagram.

ing waveform could be restored deploying different signal processing algorithms - e.g. point-to-point interpolation, low-pass filtering and least-square-error algorithms , but they require enough sample points around the fault one.

All this motivates the search for a phase estimation approach, providing sufficient accuracy with reduced count of samples. "Sufficient accuracy" is understood as less then a half phase step, if a digital phase shifters are used.

A. Experimental Setup Description

The test setup is arranged as shown on Fig. 2. The chosen low gain sensing antenna is placed in the boresight of the measured antenna and the distance between them meets the far-field criterion.

The measured antenna contains 256 patch elements, arranged into 16x16 regular planar array. The phase and amplitude of every element's feed power is controlled by 64 four-channel MMICs, as shown on Fig. 3. The gain of every channel is controlled by a 3-bit digital attenuator with 1 dB step. The phase shift is controlled by a 5-bit digital phase shifter, providing a phase step of $2\pi/32$ radians.

For the purpose specifically to verify the approach usability to extract phase information from measured sequence, the power difference ζ across all elements was reduced down to ± 1.5 dB, prior the main REV-measurement.

The feed-signal of every particular patch-element is sequentially shifted through all possible 32 phase states and the received signal—Fig. 4—is placed into an 256x32 matrix. One can recognize the underlying sinusoidal wave and the non-linear signal distortions.

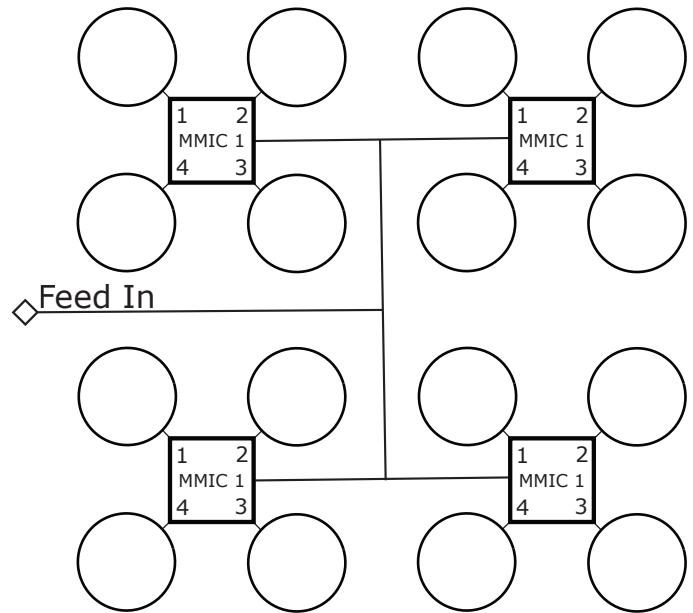


Fig. 3. Simplified antenna layout.

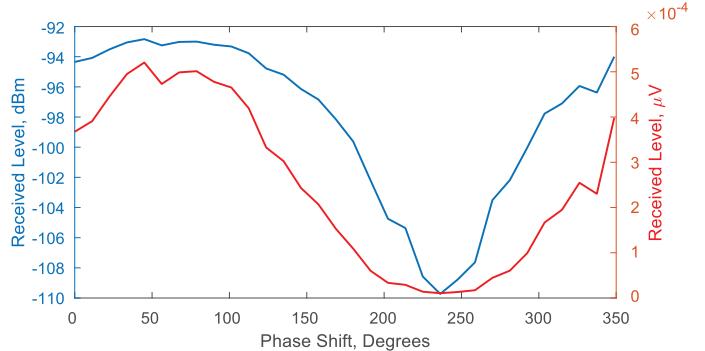


Fig. 4. Received Level as function of phase shift.

III. ANALYTICAL SIGNALS AND THE HILBERT TRANSFORM

The use of analytical signals is justifiable in situations, where the signal processing in the complex space is preferable, compared to the simple time-space processing - e.g. single side band modulation and demodulation, speech processing and direction-of-arrival measurements.

An analytical signal $S_a[t]$ is defined as a complex valued function with no negative spectral components:

$$S_a[t] = S_r[t] + j\hat{S}_r[t] = S_r[t] + H\{S_r[t]\}, \quad (1)$$

where $S_a[t]$ represents real valued original sequence and the $H\{S_a\}$ - it's Hilbert transform, also real valued but in quadrature with the original one [6, 7].

The analytical signals reveal their advantages in polar form:

$$\begin{aligned} S_a[t] &= S_m[t]e^{j\phi[t]} \\ S_m[t] &= |S_a[t]| \\ \phi[t] &= \arg[S_a[t]] \\ f[t] &= \frac{1}{2\pi} \frac{\delta\phi(t)}{\delta t}, \end{aligned} \quad (2)$$

where $S_m[t]$, $\phi[t]$ and $f(t)$ represent the so called "local features" - "instantaneous amplitude, phase" and "frequency" of the analytical signal.

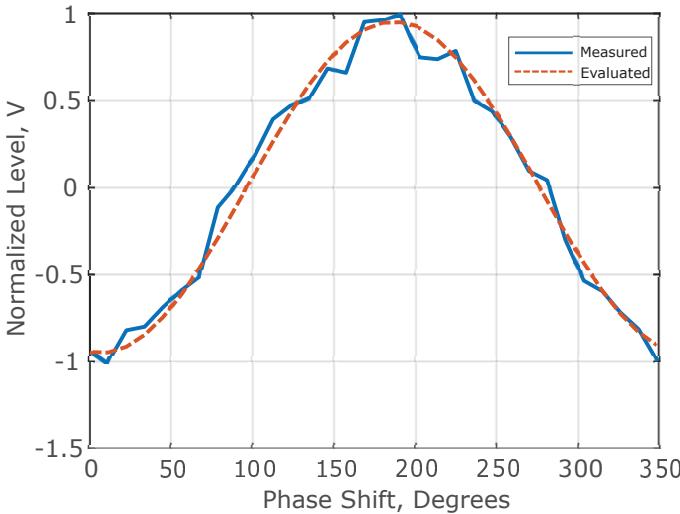


Fig. 5. $[A_j]_{\text{center}}$ and Approximated sine-wave

The work assumption is that the "local features" - can be directly implemented for deriving the parameters of the underlying harmonic wave.

IV. DATA PROCESSING

The data processing is split into five different steps:

- Conversion from dBm to μV
- Centering
- Conversion to analytical form
- Correlation calculation and Phase extraction
- Compensation.

A. Conversion to μV and Centering

The array $[A]$ consists 256 columns $[A_j]$, each with the 32 values of received power for every phase state. $[A]$ has to converted from dBm to μV and in order to avoid value-induced biasing, a column-wise normalization to value of 2 is performed in accordance with:

$$\begin{aligned} [A]_{\mu\text{V}} &= 10^{[A]_{\text{dBm}}/20} \\ \overline{[A_j]} &= 2 * \frac{[A_j] - \min([A_j])}{\max([A_j]) - \min([A_j])} \end{aligned} \quad (3)$$

The centering of every level vector A_j (columns in $[A]\mu\text{V}$) is performed calculating vector's zero-frequency component:

$$[A_j]_{\text{center}} = \overline{[A_j]} - |\text{FFT}_0[\overline{[A_j]}]| \quad (4)$$

B. Conversion to analytical form

The conversion of $[A]_{\text{center}}$ to analytical form $[A]_H$ is performed using the built in MATLAB function `hilbert()`.

In the same way an additional "reference" sine wave is generated and transformed to analytical form:

$$\begin{aligned} S_{\text{ref}} &= \sin([n\pi]), \text{ where } n = 0 : 0.1963 : 2 \\ S_H &= H\{S_{\text{ref}}\} \end{aligned} \quad (5)$$

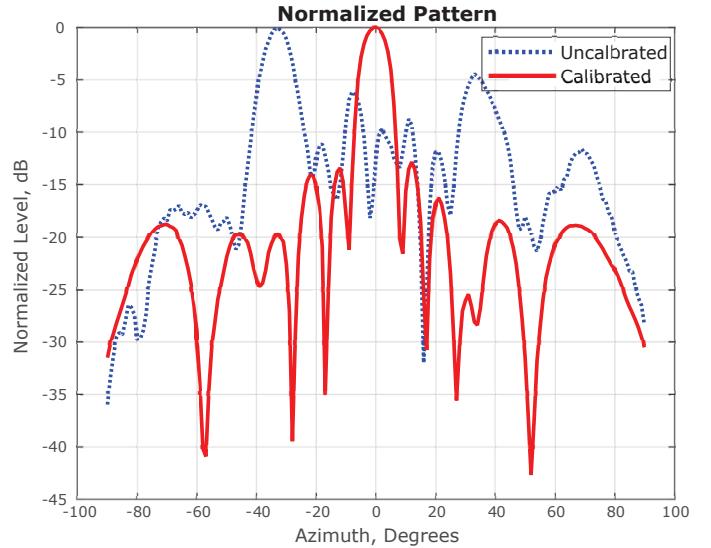


Fig. 6. Patterns before and after calibration

C. Correlation calculation and phase extraction

The correlation between the measured sequence $[A_j]_H$ and the reference signal S_H is calculated using the inner-product of both complex sequences: per-element multiplication of $[A_j]_H$ with the conjugate of S_H :

$$\Lambda_j = \frac{1}{32} \sum [A_j]_H \bullet S_H^* \quad (6)$$

The phase shift of $[A_j]$ in relation to the reference sinusoidal sequence $[S_H]$ can be derived as mean of the local phases of Λ_j

$$\Phi_j = \overline{\angle \{\Lambda_j\}} \quad (7)$$

D. Compensation

A binary configuration, corresponding to the correction phase shift Ω_j

$$\Omega_j = -\Phi_j \quad (8)$$

is loaded into the phase shifter for the particular MMIC's channel. An example of measured sequence and approximated sine-wave $\sin(\omega t + \Omega_j)$ is presented on Fig. 5.

The Antenna Pattern with the loaded correction phase shifts is presented in red on Fig. 6. For the sake of comparability, the antenna pattern before the phase correction is shown in blue on the same figure.

V. CONCLUSION

It was presented the conversion of measurement sequence to analytical signals provides a convenient way to make a correct estimation of phase errors from every single radiation element. The pattern of the calibrated antenna exhibits close to the theoretical maximum -13dB Peak to the first side lobe and well defined lobes and zeros.

Future work will be dedicated to reduction of measurement samples used for phase estimation and to its comparison with other phase estimation techniques like the Pisarenko Harmonic Decomposition, MUSIC, ESPRIT and others using the same count of input samples.

REFERENCES

- [1] Katagi,T., Mano, S., "A Method for Measuring Amplitude and Phase of Each Radiating Element of a PhasedArray Antenna", *Electronics and Communications in Japan*, 1982, Vol. 65-B, No. 5, 58 - 64
- [2] Yonezava, R., Konishi, Y., et. al. "Beam-Shape Correction in Deployable Phased Array", *IEEE Transactions on Antennas and Propagation*, 1999, MARCH, VOL. 47, NO. 3, 482 - 486
- [3] Sorace, R., "Phased Array Calibration", *IEEE Transactions on Antennas and Propagation*, 2001, Apr, Vol. 49 , Issue: 4, 517 - 525
- [4] Hazewijk, J., "Fast Determination of Element Excitation of Active Phased Array Antennas", *Antennas and Propagation Society International Symposium*, 1991, Digest, 1478 - 1481
- [5] Takahashi, T., Miyashita, H., et. al., "Theoretical Study on Measurement Accuracy of Rotating Element Electric Field Vector (REV) Method" *Electronics and Communications in Japan*, 2006, Part 1, Vol. 89, No. 1, 22 - 33
- [6] Oppenheim, A., Schafner, R., *Discrete-Time Signal Processing*, 3-th ed., 2010, Prentice Hall Press Upper Saddle River, NJ, USA
- [7] Hahn, S., *Hilbert Transforms in Signal Processing*, 1996, Artech House, Michigan, USA

Utilizing Blockchain technology for data analysis and interaction through The Formal Method of Things

Anwendung von Blockchain Technologie zur Datenanalyse und Interaktionen durch den Formal Method of Things

Blagovest Kirilov

Faculty of German Engineering and Economics Education,
Technical University Sofia, Sofia, Bulgaria

Abstract — This paper looks upon utilizing blockchain technology for data analysis and interaction through the Formal Method of Things (FMoT), using Process and Trust as its main building blocks and focusing on security, scalability and data input for different applications and usages by end users, devices and systems. The main focus is on the utilisation of large amounts of data whilst also keeping user profile security and interaction abilities at a high level.

Zusammenfassung — Dieses Schreiben betrachtet die Anwendung von Blockchain Technologie zur Datenanalyse und Interaktion durch den Formal Method of Things (FMoT), durch Prozesse und Vertrauen als Hauptbausteine des Systems. Der Fokus liegt auf Sicherheit, Skalierbarkeit sowie Dateneingabe für verschiedene Anwendungen durch Nutzer, Geräte und Systeme. Es wird grundsätzlich die Bearbeitung von großen Datenvolumen sowie Nutzerprofilsicherheit und Interaktionsmöglichkeiten mit dieser Methode beschrieben.

I. INTRODUCTION

A draft process flow model is an outline that gives the base process steps for the formalization of a data analysis and interaction process. Overlying on this, an interface should be created that will enable the rollout of the formalization for all types of data input, as well as all methods to be used. This model features end-to-end functionality, going from the pre-analysis and integrated descriptor checks of input data through applying data analysis methods through formalization and ending in user output. This method will be further described in this paper.

Given the development and importance of the topic for all industries and public areas of importance, this would give a solid foundation for a set-up of “formalization to various methods, so as to have scalability, cost-effectiveness and future-proofing going ahead” [1].

II. THE FORMAL METHOD OF THINGS

FMoT uses an entirely new approach to digital and realworld processes. Any and every process can be broken down into a series of tasks; a finite set of directional but variable, step-by-step interactions between people and objects in a particular location – and therein present a way to manage such processes on any device.

There are two main building blocks to understanding the ‘totality’ of FMoT. Each block builds upon and enables the next, the blocks are Process and Trust. Each building block has great social and commercial application and value in its own right, solving significant problems and limitations in several fields, but in combination they become exponentially more powerful and virtually limitless in scope and opportunity.

A. FMoT and Process

The first block is understanding that FMoT can manage any given process. Regardless if it is a process, whether it is technology, social, domestic, commercial, logistical, industrial, political, or environmental, FMoT can make this process easier, more efficient intuitive and secure, faster, safer, cheaper and far more seamlessly integrated with all the digital and information resources now available, such as the web and the Internet of Things.

With FMoT each successive task in a process has an easily coded microservice written for it, so FMoT can walk the user through all the variables, the decisions and actions needed to complete the task and move the process forward. Microservices are designed so they can easily be created with anyone with basic web programming skills. Most microservices or tasks have already been written, and will be part of an open source library so the ‘process provider’ only needs to fit them together in the particular order required for any particular process. Most microservices will be used many times in many different processes. Understanding an organisation’s processes, and compiling the open source microservices in order to allow FMoTs to manage those processes will likely become a marketable skill.

Processes are often aggregated (by organisations, groups, businesses and individuals) with many process-streams feeding into each other to become larger more complex processes. Even the most complex multi-stream, multi user processes can be easily and intuitively managed by FMoT

It is the user letting FMoT know their location that triggers and defines the process options for that location, according to the users trust and access status, their preferences, community and business affiliations and the location specific processes they choose to interact with. They might want to open a security door, check out tonight’s offers in preferred local venues, or select which fruit trees to harvest, it is the location

and personal settings and permissions that will determine what process options are relevant and available. The interface, menu options, buttons and even the entire appearance of the application will seamlessly change according to the user's location, process, task, services, subscriptions, permissions and the objects in the user's current location.

Objects are often produced, transformed or consumed in some way during a process or task. Pre-existing objects may change status or are transformed into something new; stock becomes a sale, raw materials become a product, an off-light switch becomes on, a stranger becomes a friend, an unoccupied restaurant table becomes occupied. With FMoT, objects can communicate their location, their present status, and what they need to further the process or task to the user in that location. They can even communicate their entire history to users, and authorized users in the object's vicinity can communicate with those objects and update their status and history in real time.

Using FMoT and blockchain technology it will be theoretically possible to publicly verify all the details of an object's supply chain, the entire history of a product/project; where and how raw materials were sourced, who handled/processed/transformed it, how where and how long it was stored and at what temperature, who bought and sold the object and at what price, how it was broken and how/who fixed it and so on. For example, a garden chair is an object that arises out of a number of unrelated process streams and transformations that just happen to converge in its production and are only related to each other by its production. The chair can reliably let merchants middlemen and consumers know the details of multiple process streams that are otherwise unrelated and unavailable.

The object itself can let us know what forest the timber came from, how that forest is managed and through what environmental protocols the trees were selected for felling. But also the previously separate details of the entire process regarding the preservative chemicals which have been applied can be added to the chair's database. The constantly updating and aggregating information belonging to that object moves and travels with it through time and space in its own distributed (non-centralised) database, to be retrieved or added to by the chain of producers and consumers along the way. When this information is committed to a blockchain it cannot be altered or erased and it can be verified. However, some (or all) information regarding the chair and its production streams may not be recorded by FMoT, or it may be encrypted or privatized, requiring private keys to unlock portions of the information now belonging to the chair. So for example a sawmill may entirely privatize the name of the particular sawmill operator that day (only available to the company in case of complaint), make the price paid for the timber available to particular users (their accountants), whilst choosing to make entirely public their company name and location for retrieval by anyone that interacts with the object's database further down the line. Ultimately the extent to which information is made publically available will be subject to consumer demand and uptake of FMoT along the process of production but FMoT provides the means by which whole new realms of communication and accountability become possible as demanded. In reality there will likely be gaps in information and a mixture of private and public information, but this will still massively enhance efficiencies, as producers, consumers and middlemen learn to converse with each other publically and privately, directly and

indirectly in all kinds of ways generating new demands that have not been conceivable up until now.

It is important to re-iterate that use of FMoT does not ever require, store or broadcast personal data unless specifically authorized by users. Hard anonymity and encryption is ensured by personalized public/private keys.

With FMoT an object's needs and status, it's own history and process, can continually be updated in real time by the object itself, by other users, by the location, or by other objects. This can enhance efficiencies and prevent wastage.

An object can accumulate information through feeds of data from many different sources and communicate them to the process in hand, updating the task and the user. In this way the huge FMoTential of internet of things will start to become realized. For example a bottle of milk could update its use-by date and communicate that to a domestic consumer via their fridge, according to the temperature data previously logged in a transport container.

B. FMoT and Trust

When the Process functionality of FMoT is combined with further blockchain innovations it becomes possible to solve a very significant and longstanding problem in the digital and virtual fields, which is the issue of Trust. Trust is the second building block to understanding the application.

The internet of things has so far been very vulnerable to hacking and misuse of data collection. Furthermore, people's precious identities and data have also been widely exposed to fraud, theft and manipulation in both the 'real' and virtual worlds.

The Trust, and cross-blockchain application that of FMoT will make it inherently anonymous and secure, yet provide the opportunity for users to maintain a unique Trust profile. To do this FMoT does not require biographical or biometric data, and so does not store it. FMoT is location aware but cannot be used to track or trace a user, nor can it be used to transfer real world biographical identity to malicious actors. Nonetheless it will be possible for users to establish a self-asserting Trust identity.

If Trust can be digitally and reliably established, it enables an entire new realm for further technical innovations.

Trust is an emergent property of FMoT, meaning it becomes available, reliable and more useful through widespread (local) adoption, as well as regular use by each individual user.

FMoT allows users to establish and maintain a self-asserting, anonymous, digital Trust profile or identity, and it requires this Trust identity in order to use it fully. What that means is the ability to demonstrate that I am a real person, with a real life, in the real world, that I am only one person and that I exist and am sovereign, and only I (its creator) can be me and use this identity. My Trust identity cannot be sold, stolen, forged, or hacked, and it cannot be used against me because it has to be earned through the unique complexity of having a complex life in the real world.

The more I use FMoT the greater the trust I earn. My FMoT Trust identity is self-asserting, it needs no permission from anything or anyone to exist and it cannot be turned on or off by anyone but me. Even better, this trusted identity does not require biometric data, or any biographical data from the real world to function, in that sense it is a completely anonymous, secure digital representation of me.

Trust is perpetually earned and maintained by everyday use of FMoT in many different ways, by the processes of my life in my location. If I cease to use and perpetuate it, it diminishes

and eventually ceases to exist. If I move location, or step outside of the established routines and connections of my life, it decreases until I re-establish or earn it through the regular processes of my new life circumstances and through the unbroken connections to my previous life.

This Trust identity is like my digital shadow, it gains its existence only from me, follows me everywhere and simply cannot convincingly be used by anyone but me, because anyone else will inevitably change its ‘shape’ and thus instantly reduce its trust value. If anyone else tries to use my identity it will quickly morph and become the identity of their real-world life and will lose its conformity to me, and thus its Trust value. It derives its form, its unique signature and its resolution from the uniqueness and idiosyncrasy of its creator and user’s everyday process(es). Just as I cannot cast two shadows simultaneously, nor can I convincingly maintain two robust Trust identities.

I cannot sell it or give it away because the recipient is not me and cannot artificially pretend to be me whilst maintaining their own trusted identity, because they would have to inhabit my life in order to represent me. If someone does ‘inhabit’ my life and try to mis-represent me, they must do so in my locality and they must forgo their own original and unique trust identity; users cannot inhabit two identities. Nor can my precious, three dimensional, real world identity be identified by any party interacting with my digital shadow unless i choose to reveal it. If my Trust identity is compromised in any way I can simply turn it off and re-establish a new one.

C. FMoT and the trust blockchain

FMoT will launch an entirely public blockchain dedicated to trust with its own native token to establish and maintain trust for users. The trust blockchain will also act as a gatekeeper, only allowing trusted users to interact with an unlimited number of other, entangled blockchains behind the trust ‘gate’ (cross chain applications). Everybody will be able to become a node and/or use the Trust blockchain not just users and developers of FMoT.

The trust blockchain means that even public blockchains associated or entangled with it can in effect be compartmentalised, or localised, only ever being seen or interacted with by people that have established and maintain trust in a specific locality. Trust on the trust blockchain is established and maintained by users in a particular location, and so the public trust blockchain can be used to direct local users to local blockchains. If a user has not established trust in a given locality they cannot even know what blockchains exist in that location and they cannot be viewed or interfered with remotely. Applications and chains associated with or regulated by the public trust blockchain are in effect public-but-compartmentalised, since anyone can theoretically start or run a node or subscribe to its services but only so long as they are a

trusted user in a specific locality. This will enable the creation of local transactional cryptocurrency without the security overheads of Bitcoin and the other (fully) public blockchain cryptocurrencies, which will leave them to do what they do best, being a secure store of value and means of large and international transactions.

This also negates the risk of quantum attacks as they would have to be very targeted and would only provide a very small segment of information and FMoTential gain.

The candidate pc or device used to create or download and run a local (compartmentalised) transactional blockchain node is an object which has to be assigned trust by a trusted user in order to become a node. To run such a node, the device must be continually trusted by at least one trusted user. If the node device has only one trusted user assigning their trust to it, if that user loses trust then so does the device, and both are excluded until trust is restored. Or, if the node behaves badly and loses trust, the single assigned user will also lose trust; either way both user and node device lose trust and are excluded from the network. However, any object (such as a node device) can be assigned trust by more than one trusted user. In this case, if any single user loses trust the node device will still be trusted by the other trusted users and can continue to run the node. But if the node behaves badly all the associated users of the node will lose trust.

III. SUMMARY

FMoT is a wide-ranging, multi-purpose digital application able to connect people and the processes of the everyday ‘real-world’ with the virtual world, the internet of things and the web. It does not only cover the aspect of integrating data analysis as a formal method, but also utilizing the results and combining this with links to many innovative technologies, giving the formal method its proof of concept as well as outlook for future development.

REFERENCES

- [1] Gregory Piatetsky-Shapiro, *From Data Mining to Big Data and Beyond*; April 18, 2012 [Inside Analysis]; <https://insideanalysis.com/data-mining-and-beyond/>
- [2] IBM – Analytics, Jennifer Clemente: *Big Data in production*, Data & AI, Cloud, IBM 2016 [from use-cases to production], <https://www.ibmbigdatahub.com/blog/bigdata-challenge-transformation-manufacturing-industry> <https://www.ibm.com/analytics/hadoop/big-data-analytics>
- [3] International Data Corporation (IDC) - *Big Data and Business Analytics Forum 2013*, February 2013 [Topics of interest covered on forum: Data quality standardization, Complexity and compatibility of data, Hadoop, MapReduce, and the technologies behind Big Data analytics <https://idccema.com/eng/events/54106-idc-big-data-and-businessanalytics-forum-2013>]

Authors Index

Badra, Hashem, 31

Chotrov, Dimo, 21

Fiedler, Malte, 49

Fritz, Wolfgang, 49

Gankova-Ivanova, Zwetelina, 1

Garabitov, Angel, 55

Hensel, Stefan, 39

Hesapchieva, Simona, 35

Hristov, Daniel, 59

Iltchev, Velko, 13, 17

Ivanova-Vassileva, Svetla, 5, 9

Katsarski, Boyan, 17

Kehret, Christoh, 39

Kirilov, Blagovest, 63

Maleshkov, Stoyan, 21

Marinov, Marin B., 39

Ovtcharova, Jivka, 31

Staiger, Anna-Maria, 49

Topalov, Ivan P., 43

Topalova, Irina, 25

Tsenov, Aleksandar, 13, 55

Uzunova, Zlatka, 21

von der Oelsnitz, Dietrich, 49

Yanachkov, Georgi, 35

Yordanov, Yordan, 21