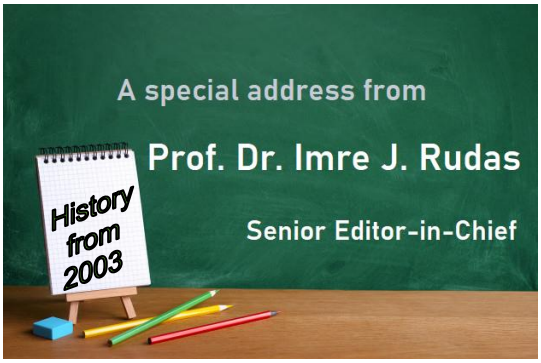
A Milestone of Acta Polytechnica Hungarica



Acta Polytechica Hungarica is a part of my heart, I am proud of it like a father being proud of his child. My first thought, when I am addressing the readers at last with this special issue is that 17 years passed like a single day.

The journal history started in 2003. I was elected as the Rector of Budapest Tech, the successor of three Polytechnic types Colleges namely, Bánki Donát Polytechnic, Kandó Kálmán Technical College and College of Light Industry. I prepared a new institutional strategic plan, where one of the main goals was to establish an international scientific Journal in order to provide a publication platform for our scholars and citizens.

I invited Prof. Dr. András Bakó to be the first Editor-in-Chief, and established the International Editorial Board constituting of highly recognized scientist. Since the beginning till 2014, I was the Chair of the Editorial Board and Honorary Editor-in-Chief. After some discussions with the Board members, I decided to give the name, mission and scope to the new Journal

Acta Polytechnica Hungarica

Journal of Applied Sciences

Mission Statement

Acta Polytechnica Hungarica should be an internationally highly recognized, professionally operated Scientific Journal with high Impact Factor.

Aims and Scope

Acta Polytechnica Hungarica is a multidisciplinary international peer reviewed journal of applied sciences publishing high-quality articles in over 6 issues per year. The journal aims to provide an international forum for the publication of original, fundamental research and applications-based results in engineering sciences as follows:

- Mechanical Engineering
- Electrical Engineering
- Informatics and a
- Special Track for Selected topics in Engineering only for invited papers.

Manuscripts with high standards of scientific quality are only published in the Journal. This is ensured by subjecting each paper to a strict assessment procedure by members of the Editorial Board and the International Advisory Board. The objective is to establish that papers submitted do meet our established requirements. The review is based on the justification of these technical and quality criteria (see Editorial Policy).

After a few early successful years, IEEE Hungary Section joined as a cosponsor of Acta Polytechnica Hungarica, and since that time, has had a great contribution to success.

It took five years to the new Journal to reach an international recognition; in 2008 it was selected for coverage in Thomson Reuters products and custom information services. Beginning with V. 5 (1) 2008, this publication was indexed and abstracted in the following:

- Science Citation Index Expanded (also known as SciSearch®)
- Journal Citation Reports/Science Edition.

Let me recall here a part of our Editorial Policy:

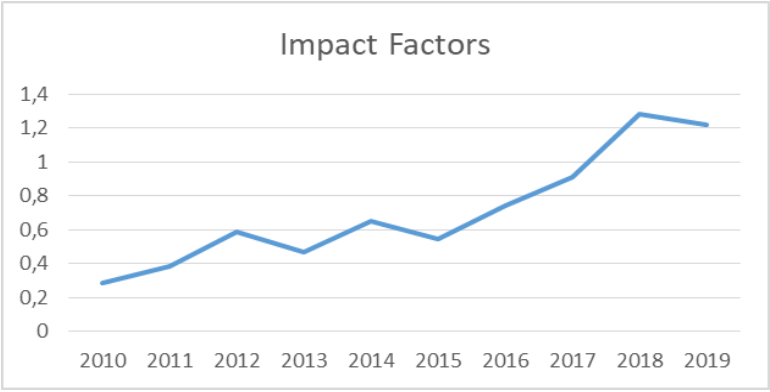
Acta Polytechnica Hungarica has been accepted for coverage in selected Elsevier product(s) starting with 2009 material. Elsevier intends to index and extract data from the full text article and integrate such data in these product(s). Coverage in Elsevier products increases dissemination of authors' work via linking technologies, which drive additional traffic to individual articles, promoting journal brand awareness. The benefits are high visibility to a global audience, fast online dissemination of content and increased exposure and profile for authors and editors.
In 2010 Acta Polytechnica Hungarica was added to Directory of Open Access Journals (DOAJ) which provides many benefits:

- Journals with the content in PDF format will be included in long term preservation project.
- Statistics show more than 6 millions successful requests a month for DOAJ site from all over the world.
- Many web crawlers are fetching the content of DOAJ to be a part of their search engines.
- DOAJ is OAI compliant.

- Once the article is available in DOAJ, it becomes automatically OAI harvestable.
- Many aggregators and database providers are regularly harvesting DOAJ content in order to include it to their commercial databases.

It is cited here to give an example of the long journey and continuous progress Acta Polytechnica Hungarica has made. All that leads to increased visibility and usage of OA journals. The Journal received the Impact Factor first in 2010. The 2019 Impact Factor is 1.219.

The trend of Impact Factor can be seen in figure below.

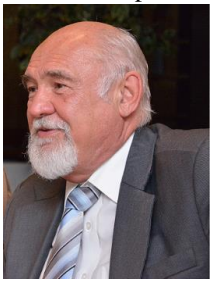


After Prof. Bakó's term, the Editor-in-Chiefs were Profs. János Fodor and Péter Tóth. I thank them all for their efforts devoted to our Journal. I took over the position in 2014.

Since the beginning, the Executive Editor has been Anikó Szakál. Without her contribution of great value, these results would not have been achieved.

After 17 years, I am now ready to step down as the senior Editor-in-Chief. I wish the incoming Editor-in-Chief, Prof. Levente Kovács a good momentum and inspiration to further improve the standard of the Journal and strengthen its international recognition.

My "child" has grown up, it needs fresh blood and should be open to new challenges. As the founding father, I wish further successes and I hope I can always remain proud of Acta Polytechnica Hungarica.



Prof. Dr. Imre J. Rudas

Senior Editor-in-Chief

Budapest, April 6, 2021

Preface

Special Issue recognizing the contributions of Prof. Imre J. Rudas, accompanying Acta Polytechnica Hungarica from inception to its current maturity milestone event

It is hard to assess and show appreciation for such an immense work that Prof. Imre J. Rudas has invested into Acta Polytechnica Hungarica over the years. Over 17 years of tireless work, over 110 volumes published. It is had to count the number of PhDs, dissertations and defenses that were enabled and supported by the highly esteemed scientific articles in Acta Polytechnica Hungarica.

Over the past decades, Prof. Rudas has shown to be an irreplaceable, rock solid and enthusiastic member of the international scientific community, building institutions, ecosystems, programs, organizing events and managing on the top of all a world-class scientific journal. Acta Polytechnica Hungarica has been beneficial for the young professionals and the established researchers alike, documenting and disseminating their key scientific and technological findings.

Acta Polytechnica Hungarica remains committed to its foundations, and the new EiC and co-EiC are dedicated to further improve the reputation and quality of the Journal. More recently, a series of events triggered global changes on the scientific journals' landscape, leading to the endorsement of the Plan S initiative by the European Commission, channeling all publicly-funded research results to Open Access mediums. Thus we have witnessed an unprecedented rise in the number of pay-to-publish journals, initiating a race for quality and recognition. Acta Polytechnica Hungarica needs to preserve its national and regional leadership position in Applied Engineering Sciences, and continue to support its community.

As the incoming leadership, we are fully dedicated to the above goals, while aiming to preserve the rich heritage of the journal. Building on Prof. Rudas' invaluable network, his lasting influence to the field and the ongoing support of IEEE Hungary Section. Understanding the challenges, we will initiate the facelift and slight restructuring of Acta Polytechnica Hungarica to best represent the key strategic research domains of our times, and also to embrace the cloud-based editorial support services currently available. We plan to pull tighter the relation among the IEEE endorsed conference series managed by IEEE Hungary Section, as well as creating lasting cooperations with other scientific communities.

This Special Issue is therefore dedicated to the hard work of Prof. Imre J. Rudas as an Editor-in-Chief, featuring articles from across his research interest domains from robotics to soft computing and education science.

We pledge to be faithful to his principles of quality and scientific eminence and continue to serve our community.

Budapest, April 6, 2021

Prof. Levente A. Kovács

Dr. Tamás Haidegger

Editor-in-Chiefs

Robotics and Intelligent Systems Against a Pandemic

Alaa Khamis[1], Jun Meng[2], Jin Wang[3], Ahmad Taher Azar[4], Edson Prestes[5], Howard Li[6], Ibrahim A. Hameed[7], Árpád Takács[8], Imre J. Rudas[8], and Tamás Haidegger[8,9]

[1]General Motors Canada, 500 Wentworth St W, Oshawa, ON L1J 6J2, alaa.khamis@gm.com
[2] College of Electrical Engineering, Zhejiang University, YuQuan Campus, No.38 ZheDa Road, HangzhouZhejiang, 310027, China, junmeng@zju.edu.cn
[3]Robotics Institute, Zhejiang University, YuQuan Campus, No.38 ZheDa Road,Hangzhou 310027,China, dwjcom@zju.edu.cn
[4]Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt and College of Computer and Information Sciences, Prince Sultan University Riyadh, Saudi Arabia, ahmad.azar@fci.bu.edu.eg, aazar@psu.edu.sa
[5]Informatics Institute, Federal University of Rio Grande do Sul, edson.prestes@ieee.org
[6]Dept. of Electrical and Computer Engineering, University of New Brunswick, Canada, howard@unb.ca
[7] Dept. of ICT and Natural Sciences, Norwegian University of Science and Technology, Larsgårdsvegen 2, 6009 ° Alesund, Norway, ibib@ntnu.no
[8]Antal Bejczy Center for Intelligent Robotics, EKIK, Óbuda University, Budapest, HU, arpad.takacs@irob.uni-obuda.hu
[9]Austrian Center for Medical Innovation and Technology (ACMIT), Viktor-Kalpan-str.2., Wiener Neustadt, A-2700, AT, tamas.haidegger@acmit.at

Abstract: The outbreak of the novel coronavirus and its disease COVID-19 presents an unprecedented challenge for humanity. Intelligent systems and robotics particularly are helping the fight against COVID-19 several ways. Potential technology-driven solutions in this accelerating pandemic include, but are not limited to, early detection and diagnosis, assistive robots, indoor and outdoor disinfection robots, public awareness and patrolling, contactless last-mile delivery services, micro- and nano-robotics and laboratory automation. This article sheds light on the roles robotics and automation can play in fighting this disastrous pandemic and highlights a number of potential applications to transform this challenge into opportunities. The article also highlights the ethical implications of robotics and intelligent systems during the emergency side and in the post-pandemic world.

Keywords: Service robotics at COVID-19; coronavirus response; assistive robots; last-mile delivery; laboratory automation

1 Introduction

Since the outbreak of the novel coronavirus (SARS-CoV-2), the global Robotics and Automation (R&A) community has quickly mobilized and gathered to offer solutions to help preventing, treating and monitoring the pandemic and its effects. Contributions have materialized in several novel applications and services. Identifiying key challenges faced by health care responders and the general population remains a priority, unraveling the role of robotics in infectious disease crises [1]. According to the World Health Organization (WHO) the top priorities in the fight against COVID-19:

1. Early diagnosis;
2. Improved hygiene & disinfection;
3. Social distancing.

The role that robotics and automation (R&A) can play accordingly is getting well recognized, e.g., the WHO–China Joint Mission COVID-19 report cited Artificial Intelligence (AI) and robotics as key elements of the efficient response [2]. The European Commission launched an initiative (managed by the European AI Alliance) to collect ideas about deployable AI and robotics solutions, as well as information on other initiatives that could help solve the ongoing coronavirus crisis: AI-Robotics vs COVID-19 (tinyurl.com/COVID19-AI). The IEEE and other professional organizations are also taking part in the global response; e.g., working on applying the lessons learned to COVID-19 in conjunction with the IEEE Robotics and Automation Society (RAS) Special Interest Group on Humanitarian Technology (SIGHT).

This article describes a number of current and potential future applications of R&A in the battle against the novel coronavirus and its COVID-19 disease. AI solutions are covered in other publications [3, 4]. Applications include—but are not limited to—specific solutions aligned with the WHO priorities (Fig. 1):

1. Telehealth and physical human–robot interaction systems enabling healthcare workers to remotely diagnose and treat patients;
2. Autonomous or teleoperated robots for hospital disinfection and disinfection of public spaces;
3. Social robots for families interacting with patients or with relatives in nursing homes and robots used by public safety and public health departments for quarantine enforcement and public service announcements;
4. General workflow improvements, such as hospital and laboratory supply chain robots for handling and transportation of samples and contaminated materials and robots enabling or assisting humans to return to work or companies to continue to function.

Ethical implications of R&A during the pandemic and in the post-pandemic world are also discussed in the light of the IEEE Global Initiative. Beyond the

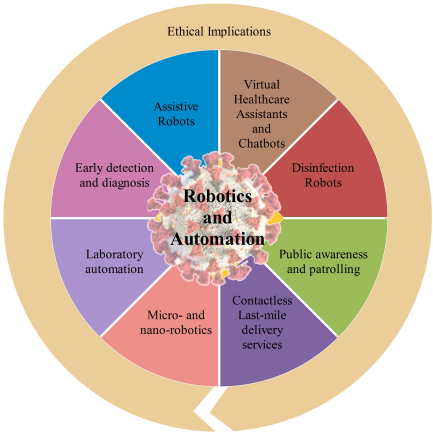


Figure 1
Robotics and Automation in Pandemic Fight.

enlisted new and rising technologies, COVID-19 made its mark on the existing medical–surgical robotic market as well, leaving companies delayed, defunct or emerged due to the new sterility, distancing and patient safety/staff requirements and standards [5].

The article is organized as follows. Section 2 describes the role of R&A in early detection and diagnosis. Potential applications of micro-, and nano-robots in early diagnosis, monitoring and targeted drug-delivery are discussed in Section 3. Hospital and laboratory supply chain robots for handling and transportation of samples and contaminated materials are presented in Section 4. Section 5 covers teleoperated and autonomous robots for hospital disinfection and disinfection of public spaces. The role of assistive robots in supporting humans during the quarantine and in helping them to return to work or companies to continue to function is presented in Section 6, followed by the introduction virtual healthcare assistants and chatbots as another form of automating healthcare. Robots used by public safety and public health departments for quarantine enforcement and public service announcements are described in Section 8 followed by presenting drug, medical supplier, food/grocery delivery robots and last-mile delivery services in Section 9. The ethical implication of R&A related to COVID-19 are discussed in Section 10. Finally, conclusions are summarized in Section 10.

2 Early Detection and Diagnosis

Early detection of the disease and early isolation of the affected are essential. Detecting and tracking transmission on a large scale in a timely manner, and publishing the data of infected and suspected infections are crucial in prevention and control measures, to minimize further transmission. There are numerous effi-

cient approaches to tackle this complex task, from monitoring through AI-driven smart wearable medical devices, to smartphone apps, all the way to telemedicine and telehealth.

2.1 Physical Sampling

A large number of anti-epidemhic methods focus on early detection and isolation, requiring R&A solutions for scalability. Such robots are considered to be medical devices, therefore their regulatory and clearance procedure takes time [6]. Current solutions mainly represent two distinct technical solutions, expressed in their Level of Autonomy (LoA), a scale 0–5 getting applied to the robotics domain more recently [7]:

Brain Navi's NaoTrac neurosurgical navigation robot, which has already been used in over a dozen clinical cases was retrofitted to be able to take nasal swabs for testing autonomously (LoA 4). Thus process normally requires a healthcare professional in full personal protective equipment (PPE). Lifeline Robotics, a spin-off of University of Southern Denmark has developed a similar technology, already available for pre-booking. Similar functionalities are offered by systems developed at the Shenyang Institute of Automation, China and the South Korean Institute of Machinery and Materials, Korea (in a teleoperated robotic setup—at LoA 2) (Fig. 2).

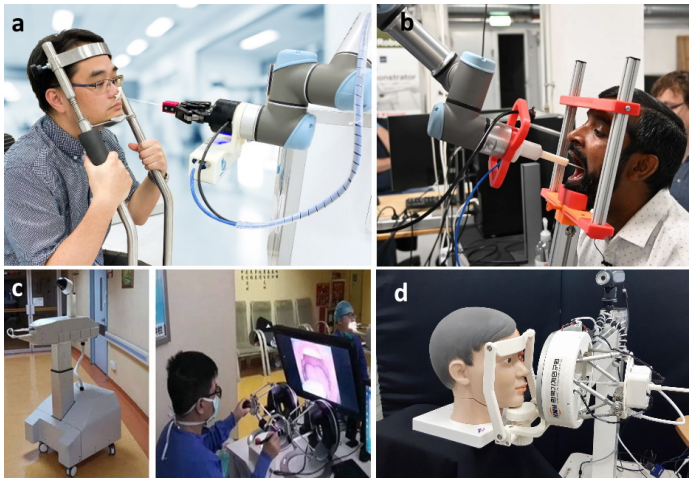


Figure 2
Robotic solutions for automated on teleoperated swabbing: a) Brain Navi; b) Lifeline Robotics; c) Shenyang Institute of Automation; d) South Korean Institute of Machinery and Materials.

Jerusalem's Hadassah Medical Center reportedly use laboratory robots to aid the management of test samples [8]. The number of samples tested has risen by two orders of magnitude, thanks to the custom designed robots by Roche, Tecan and Qiagen. Robots work jointly to extract the RNA of the virus, processing than

95% of the required tests.

A self-administered robot that automates the taking of COVID-19 swab tests has been developed in Singapore, helping to reduce healthcare workers' risk of exposure to the coronavirus. The swabbing process is completed in about 20 seconds, while it remains self-administered: patients can activate and terminate the process at will [9].

2.2 Remote Monitoring

Scaling up in non-invasive monitoring is critical as well. In Japan, SoftBank's Pepper robot has been employed for health checks and temperature monitoring at offices and public places. In China, pandemic drones petrol equipped with temperature and computer vision tool, which are used to detect symptoms of an infectious respiratory disease. Recently, drones have been used to monitor temperature sensors, heart and respiratory levels from a distance up to 10 m, and can detect coughing and sneezing. This may provide researchers with a clear view of infections in public areas and other crowded places such as airports and health care facilities. A drone equipped with infrared thermal imaging lens takes the temperature measurement of each resident in a building without contact, and sends the data back to the disease control center. Once an abnormal body temperature is detected, medical staff will pinpoint the location of the suspected patient [10]. In addition to human body temperature monitoring, health codes have been widely used since late February in China. Each citizen receives a unique online electronic QR code, used for building a database and monitoring health history and status (Figure 3). These measures have serious ethical implications, as discussed in Section 10.

Monitoring data platforms collect information uploaded by medical robots in various communities and store data (time and location of examination for suspected cases). Data mining (mines potential cases based on recorded data such as temperature, reagents, pulse, blood routine, location, etc.), data analysis and calculation (statistical analysis of the epidemic of the entire region) is also commonly performed [12]. Forecasts of the region in the next few days are given, and timely feedback is given to the epidemic units and personnel to facilitate corresponding control decisions. Some diagnostic platforms obtain real-time information through monitoring data platforms.

3 Micro- and Nanorobotics

Considering the current disastrous situation of COVID-19, micro- and nanorobots currently under development will be most useful for drug delivery to several targets inside human body by loading the drug dose to the pharmacyte. Then, it can transport safely to the target and release the dose by nano-injection or progressive cytopenetration. Second major use of nanorobots is body surveillance or

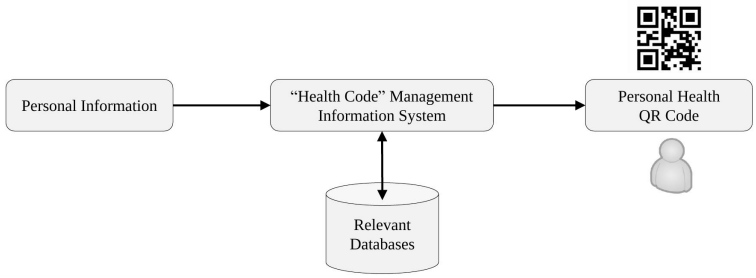


Figure 3
Personal Health QR-based System. The "health code" management information system is based on true data, and connects to relevant databases such as key personnel's dynamic control lists, and issues "health codes" based on the verification and comparison results of the mass declaration information and background data [11].

continuous monitoring of vitals, this provide a frequent response for a detected abnormal behavior. During local drug delivery, the Atomic Force Microscope (AFM) nanorobot sample tip can be used by an antibody or ligands attached to it. Such application can enhance in vitro local drug delivery research into the effects of these chemicals. Two key methods for supplying an AMF tip with certain chemical substances are available. The direct method is dipped into the specified chemical solution by the cantilever or the tip. The commonly recognized approach is the application of a connector molecule that improves the exact targeting in 3D between the AMF tip and the targeted molecule. The idea of this approach is that the chemical and the linker molecule are covalently related [13]. Major challenge of the field is to ensure safety, given the size of the robots, only LoA 4–5 systems could be deployed.

Monitoring the cellular behavior of infected persons by COVID-19 can be also done employing AFM technology. Since the AFM has a tip of about 20 nm or less that can be used for live cell samples; a nanorobot based on AFM may use this tip for accurate operations. Given the many advantages provided by AFM technology, such as high-resolution imagery and vacuum free operating environment, creation of nanobots based on AFM would allow imaging on a nanoscale, precise measurements of mechanical properties, nanoscale tests and the precision of movement control on living cells [13].

4 Laboratory and factory automation for contagious materials

The clinical management of coronavirus related materials and data includes the processes of anticipation, early detection, containment and mitigationeradication. Predictive AI models are urgently needed for different epidemic stages [14].

- In the early stages of a new disease, the R&A community needs to propose methodologies for rapid deployment of technology, to make data collection, model training, testing and wide deployment efficient.
- During the containment and mitigation of COVID-19, data have become increasingly available. It is critical to develop training data sets and accurate AI models that can respond to urgent clinical needs. The AI-assisted diagnostic tools must achieve high diagnostic accuracy in order to provide added value above and beyond existing clinical tools for containment and mitigation.
- During the latter, eradication stage, there is still a need for real-time, convenient and sensitive screening tools and models to control transmission and recognize potential outbreak centers. AI models will play a key role in the critical clinical research during the post-acute phase of COVID-19.

IBM Research Europe researchers built a platform—Deep Search—that could help speeding up the process [15]. The cloud-based platform processes COVID-19 literature and labels the data, table, image, and paragraph, before translating scientific content into a uniform, searchable structure.

4.1 Wearables

AI helmets and AI-powered glasses are used by the Chinese police to identify faces of vehicle occupants and license plates. Alerts are triggered if the vehicle occupant's information can be found in the database of confirmed cases [16]. A smart wearable medical device can be used to obtain a person's body temperature, breathing and heartbeat status, whether or not wearing a mask, and update the position information and infection/suspected infection information in real time. At the same time, the precise position of the device and the path of the terminal can be used to calculate a safe and low-risk travel path for people traveling in real time. Smart wearable medical devices can be useful in the early stage of major infectious disease outbreaks, such as screening population, quickly performing nucleic acid detection, leading to early detection and isolation. The composition of such a wearable medical device is shown in Fig. 4, which is an early concept and prototype developed under the National Nature Science Foundation of China (NSFC) program No. 7204102194.

The three major functions of this system are listed below.

1. Outbreak Visualization: based on individual body temperature charts, spatiotemporal dynamics of personnel mobility, people gathering heatmap,

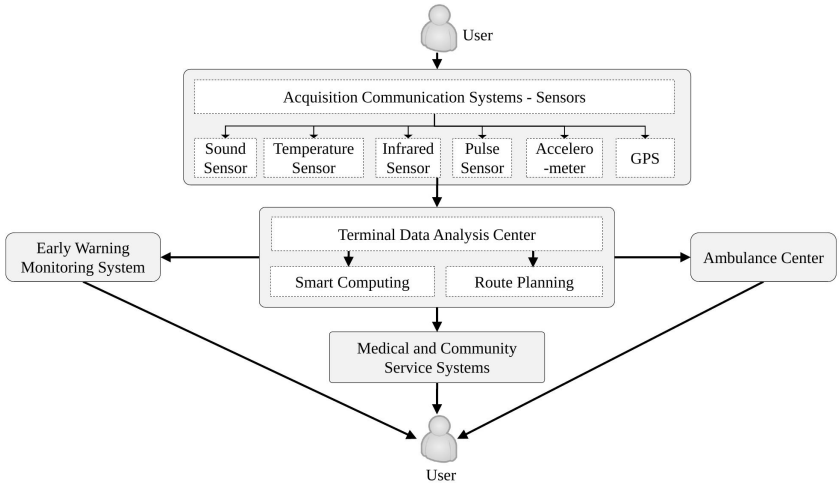


Figure 4
The composition of smart medical wearable devices, which can real-time grasp each person's current body temperature, heartbeat, breathing, confirmed infectious disease status, whether they are in contact with a confirmed/suspected infectious case, whether a mask is worn during going out, the epidemic situation during the period and so on. This information is uploaded to the terminal in real time to generate a spatiotemporal map of the epidemic situation, which quickly, multi-dimensionally and comprehensively displays the current body temperature, breathing, and heartbeat information of the personnel.

spatiotemporal map of confirmed cases;

2. Early Detection and Early Warning: screen out suspicious cases in due time;

3. Pinpoint and Route Planning: propose safe travel paths for travelers.

There are some data-driven approaches to diagnose patients through CT chest scans. Li and Qin, et al [17] proposed a 3D deep learning framework to detect COVID-19 using chest CT, named COVID-19 detection neutral network (COV-Net). University of Waterloo's VIP Lab develops a COVID-Net: a convolutional neural network for COVID-19 detection via chest radiography (13,800 chest X-Ray of 13,725 cases from around the world. In China, there is a supercomputer that provides doctors around the world with free access to an artificial intelligence diagnostic tool for early identification of COVID-19 patients based on a chest scan. The AI system on the Tianhe-1 computer can go through hundreds of images generated by computed tomography (CT) and give the diagnosis in about 10 seconds. Wang et al [18] proposed an innovative method that did not require specialized medical imaging equipment and used footage from Kinect depth cameras, in order to identify respiratory patterns of patients. They first apply a GRU neural network with bidirectional and attentional mechanisms (BI-AT-GRU) to classify six clinically significant respiratory patterns.

4.2 Telehealth

Even beyond the pandemic, emergency response can be greatly enhanced by exploiting the strengths of telehealth, particularly in the case of biological or environmental hazards. Telehealth enables remote care and attending practice, as well as it provides rapid access to information through telecommunication. During the disease outbreak, Singapore has successfully used chatbots to accelerate information flow, while video consultations and remote patient monitoring have been available in the USA since the beginning of the pandemic [19].

There were several occasions in recent history, when government agencies used telehealth solutions to increase healthcare efficiency during large-scale emergencies. The Multinational Telemedicine System developed by the North Atlantic Treaty Alliance (NATO) was deployed multiple times during military missions and crises, partially by deploying portable telemedicine devices. Owing to the coordinated work of private telemedicine companies, remote care was provided to relocated families during the devastating hurricanes Irma and Harvey. In 2003, as a result of the Severe Acute Respiratory Syndrome (SARS) pandemic, China has initiated multiple research project on using telehealth in similar situation. As a response to COVID-19, China is actively providing various remote health services. These include counseling, supervision and monitoring, even training, held by government and academic agencies. Sociometric reports showed that during the quarantine, online support and services were popular among the population, favoring the transition to a more efficient, technologically evolved form of patient care [20].

5 Disinfection Robots

Disinfecting surfaces and the environment, performed routinely in hospitals, has always been essential for infection prevention and control. Numerous AI-based hand hygiene assistance and monitoring solutions have appeared for personal and professional use, such as the Semmelweis Hand Hygiene System (HandInScan Zrt.) and SureWash (Goleta Ltd.) [21]. Beyond regular surface wiping with highly efficient detergents, a key strategy has been to allow the active ingredient to spread to all areas of the room, however this also needs complete lockdown and heavy air ventilation. UV-C irradiation (200–280 nm) is also an efficient disinfectant, if the radiation reaches all critical surfaces. Robots are considered as an effective technology for disinfection, preventing any harm to humans. There are two major types of applications: indoor and outdoor.

5.1 Outdoor Disinfection Robots

Disinfection usually means the spraying of an active material (NaClO, sodium hypochlorite in a very low concentrate). While scientists argue that this method has very limited effect on the spread of a virus [22], engineers have developed

both teleoperated and autonomous ground and aerial vehicles (Fig. 5). Several drone manufacturers have modified their agricultural models in order to spray disinfectant over large areas. These disinfecting drones have been used for the first time in Europe. They have been deployed in China, Chile, Indonesia, the Philippines, Colombia and the United Arab Emirates. XAG Robot has deployed disinfectant-spraying robots and drones in Guangzhou, China. Nevertheless, safety of such systems remains a major concern, and some companies have recommended similar fallback options for these systems that are being developed for self-driving cars [23]since the drones can either fly remotely controlled (LoA 2) or fully autonomously (LoA 4–5).

Figure 5
Outdoor disinfecting robot from Siemens, the Chinese company Aucma working in the joint laboratory for robot applications in Qingdao. The robot was designed using CAD and the actual system was built in just a week. Credit: SIPA Asia.

5.2 Indoor Disinfection Robots

Cleaning robots are available in the market for ward disinfection. The Light-Strike Germ-Zapping robots (XENEX, San Antonio, TX) was the first disinfection system of its kind to deliver intense germicidal action from pulsed UV-C at 200–315 nm over 12 years ago. The Danish UVD Robots ApS was founded in 2016 by Blue Ocean Robotics and its solution offers indoor mobility to deliver the UV-C irradiation. It was the recipient of the IEEE/IFR 2019 Innovation and Entrepreneurship Award (IERA) (Fig. 6).

Shenzhen-based YouiBot was already making autonomous robots, and quickly equipped its existing mobile robot base with thermal cameras and UV-C emitting bulbs. Akara Robotics, a spin-off from Trinity College Dublin is testing and producing a new autonomous mobile robot that is designed primarily to sterilize a room with UV light. The Indian Milagrow offers its Indoor Disinfection RoboCop. The Turkish Milvus Robotics calls its UV disinfecting robot SEIT-UV, with 360 degrees disinfecting coverage, software and sensor-based safety features. The Chinese WellWit Robotics offers its Pulsed Light Germ-Killing Robot (WDR01A). The PBA Group announced the rollout of their Sunburst UV Bots, UV-disinfecting autonomous mobile robots together with Singapore based Frasers Property Group. A similar solution is developed by Anscer Robotics in Bangalore, India. TIAGo robot equipped with UV-C sources can be used for cleaning and disinfection at hospitals and other public areas. A recent devel-

Figure 6
Indoor disinfecting robot UVD, consisting of a mobile base equipped with multiple LIDAR sensors and an array of UV lamps mounted on top. Credit: UVD Robots.

opment from Rovenso targets directly offices and commercial spaces with its focused UV-C disinfectant robot. These are typically LoA 3–5 applications. The Grainger College of Engineering at University of Illinois Urbana-Champaign's Health Care Engineering Systems Center built a low-cost, easily available robot (UVBot) that cleans spaces using UV light. UVBot has proven to be efficient in virus inactivation, the disinfection exposure time and distance for inactivation was tested using an RNA virus similar to COVID-19 [24].

6 Assistive Robots

Robots can perform many assistive tasks during the pandemic to mitigate the risk to healthcare professionals. Applications of assistive robots include, but are not limited to: medical care, nursing, patient monitoring, performing lab work, cooking and serving medication, or meal delivery to patients in isolation wards. Several assistive telecare robots, community medical robots and telemedicine services are used during the outbreak, e.g., doctors use the mounted iPad on Spot to remotely interact with coronavirus patients, ask them question in order to assess the patients. Fig. 7 shows the three major classes of community medical robots. Some more simplistic solutions (e.g., RoboAds) include only a mobile robot platform for autonomous positioning of kiosks on the one hand to reach the target audience with interactive (touchless) billboards, while taking telemetry

and collecting data on people's health condition (primarily body temperature).

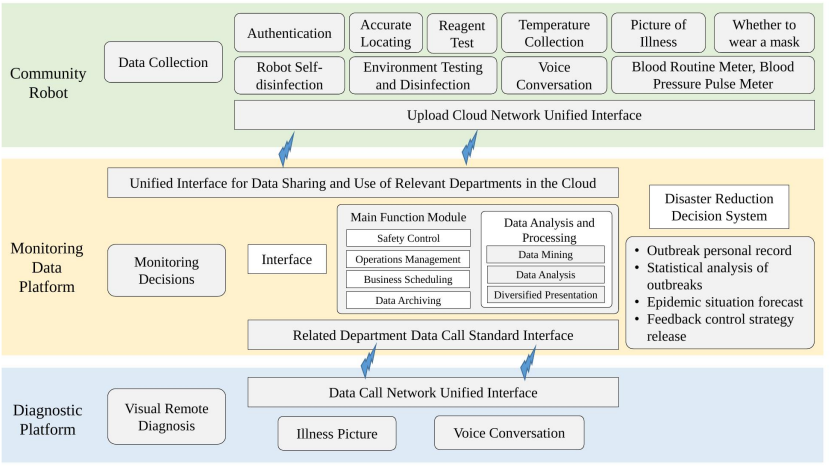


Figure 7
Community Medical Robot types employed in the fight against COVID-19.

Tommy helped Italian medical teams in treating COVID-19 patients (Fig. 8). Fig. 8 shows a remote medicine delivery for the patient using the teleoperated robot in the first affiliated hospital of Zhejiang University School of Medicine (FAHZU). This self-designed teleoperation and telepresence robot system of Zhejiang University is approved and undergoing clinical trials in the Intensive Care Unit (ICU) of FAHZU [25, 26].

The remote daily checkups are conducted via the tablet PC in the front of the robot, which allows the healthcare workers to communicate with the patients with no physical contact (Fig. 9 Left). The teleoperated robot system can also be applied in non-contact interactive diagnosing and medical instrument operating (Fig. 9 Middle and 9 Right), and its functions are undergoing further improvement and validation.

Boston Dynamics has also demonstrated a quadruped robot, Spot (Fig. 10), designed by MIT and Brigham and Women's Hospital, MA. The robot is set to be used as a telepresence-based extension for healthcare workers at remote hospital areas, such as medical tents [27]. Authors argue that compared to a wheeled assistant systems, concrete bumps and curbs are not posing a challenge for a legged robot.

For the past eight months, Boston Dynamics has been trying to find ways in which their friendly yellow quadruped, Spot, can provide some kind of useful response to COVID-19. The company has been working with researchers from MIT and Brigham and Women's Hospital in Massachusetts to use Spot as a telepresence-based extension for healthcare workers in suitable contexts, with the goal of minimizing exposure and preserving supplies of PPE.

Figure 8
Left: Tommy, the robot nurse. Credit: Flavio Lo Scalzo/Reuters. Right: Remote Medicine Delivery Robot of Zhejiang University School of Medicine (FAHZU). Credit: Zhejiang University

Saskatchewan Remote Presence Robots (Fig. 11) are used to deliver health care to rural and remote regions of Saskatchewan, Canada in order to protect physicians on the front lines fighting the COVID-19 pandemic.

The system illustrated in Fig. 12 uses artificial intelligence, big data, robots, the Internet of Things, modern medical technology, cloud computing and other platforms and methods.

According to the users and scenarios, it can be divided into three parts, namely, community robots, the monitoring data platform and the diagnostic platform [12]. Additionally, the RP-Vita platform from iRobot is equipped with the capability of autonomous navigation, which allows a remote clinician or bedside nurse to send the RP-VITA to a target destination with a single click.

Each community entrance can be equipped with several robots, and the position of each robot can be accurately obtained. For each person entering and leaving the community, the identification is verified and physiological data is collected. Blood test can be performed to determine whether a face mask is necessary. For suspected cases, one can further use the pneumonia test kit on the community robot and take pictures of the corresponding disease location. The medical staff is also connected to the remote diagnosis platform through the voice dialogue system. The robot will upload information about the suspected patient to the monitoring data platform, and will record and send data to the monitoring personnel. The robot can self-sterilize at regular intervals and sterilize the air in the surrounding environment and effectively prevent and reduce cross-infection.

At the same time, it can make video and voice calls with the suspects detected by the community robot. It can also check and analyze the photos of the suspects in

Figure 9
Left: Interactive daily consultation of physical and mental conditions using the telepresence system. Credit: Zhejiang University. Middle: Non-contact auscultation using a Doppler Ultrasound Stethoscope based on the teleoperated robot. Credit: Zhejiang University. Right: Remote operation of the medical instruments using the teleoperated robot. Credit: Zhejiang University.

order to provide medical care.

7 Virtual Healthcare Assistants and Chatbots

Virtual healthcare assistants and chatbots are another forms of automating the way of providing healthcare. WHO recognizes innovative solutions to reduce the need for personal contact between medical professionals and patients, and to increase accessibility to healthcare services [28]. Virtual healthcare assistants and chatbots (VHAC) are AI and rule-based systems that interact with humans to perform various tasks, to name a few, banking, answering phone calls, performing data entry, etc. These assistants use cognitive technologies such as machine learning, natural language processing (NLP), and ML to enable interactive communications with the end-users. Virtual assistant technology in the healthcare industry can assist in transforming various health processes and improve healthcare delivery worldwide. VHAC can be considered as the optimum tool to provide healthcare services to vulnerable groups such as the elderly with low immunity and co-morbidity. They can perform regular checkups and answer questions regarding health issues [29]. Recent reports claim that the global healthcare virtual assistant market is expected to grow at a rate of 24.7% from 2018 to 2025 to reach $1.7 billion by 2025 [30].

Virtual assistants have been employed as a communicative agent. Researchers report that the usage of telemedicine increased regularly between 2005 to 2017.

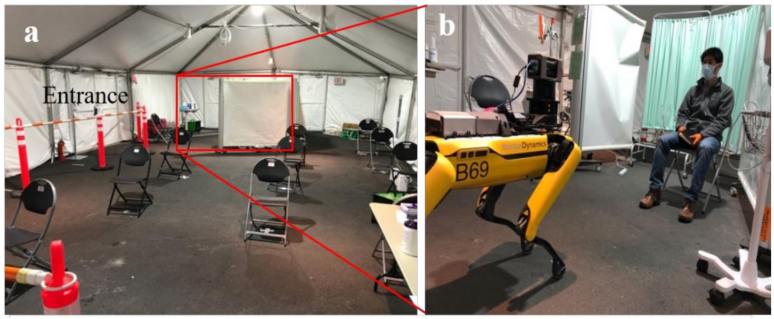


Figure 10
A medical tent outside of the emergency department of the COVID-19 triage area of the Brigham and Women's Hospital (a), and the Spot system using an IR camera for fever screening and respiratory rate detection (b). Credit: Boston Dynamics.

There is a strong demand of more robust and efficient methods and standards for evaluating these technologies [31, 32].

The need for implementing assistive technologies to fight COVID-19 has increased. Centers for Medicare & Medicaid Services (CMS) has permitted people to use interactive applications to visit virtually their physician for a variety of purposes [33]. Canada, the UK, Italy, South Africa and India are some other countries that have deployed virtual health services. Researchers, however, suggest further enhancing the capabilities of technology. "Virtual clinics" enables virtual visits with imaging data (e.g., chest X ray) to be remotely examined [34]. Telemedicine can help establishing equality among all people as well as reducing the number of patients in clinics and healthcare centers [33, 34].

In addition, virtual healthcare assistants can play a major role in preventive measures. "Chatbot" can help screen or identify early symptoms, teach and train users essential hygienic practices (e.g. "Amplify.ai", and "Facebook's Messenger for COVID-19 Program") [35]. Another chatbot designed specifically for COVID-19 is "the Coronavirus Self-Checker" [36], which provides resources to protect people and advises users whether or not they need to visit medical care centers.

8 Public Awareness and Patrolling

Linking to the previous Section, it is obvious that voice assistants can be widely used, being able to complete 200 calls in 5 minutes and offer a statistical result. With the trial of voice AI-agents, the efficiency of epidemic investigation has been greatly improved.

UAVs have been widely used for the fight against COVID-19. Drones can be equipped with high-power speakers and warnings can be broadcasted to the crowd

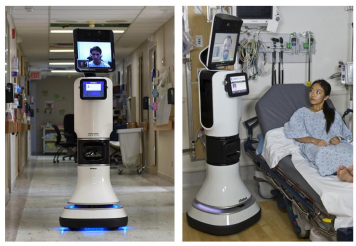


Figure 11
Left: Saskatchewan Remote Presence Robots. Credit: University of Saskatchewan and Saskatchewan Health Authority Remote Presence Robotic Program. Right: Remote Presence Virtual + Independent Telemedicine Assistant (RP-VITA). Credit: iRobot.

or to people not wearing masks. At the same time, drones can detect sick people quickly through thermal imaging and target recognition systems.

Another example is Boston Dynamics' Spot that is used at parks to remind visitors of social distancing measures or patrol at lockdown/curfew areas.

Another example is Knightscope Robotics that can be used for patrolling corporate facilities. Robots are used for remote monitoring of streets and corporate campuses (Figure 13).

9 Contactless Last-mile Delivery Services

Reducing physical contact during the pandemic demands *last-mile delivery* (LMD) robots. Last-mile is a term used in supply chain management and transportation planning to describe the movement of people and goods from a transportation hub to the destination [37]. Various innovative solutions for LMD had been or tested to reduce the delivery cost, increase customer satisfaction and minimize the negative environmental impact. These solutions include, but are not limited to, cargo-bikes, semi and fully autonomous LMD, delivery Droids (Bots), E-Plaette, Postal delivery, driverless deliveries and privately-owned AV. According to Makinsey, semi and fully autonomous LMD will reduce costs by approximately 10–40% [38].

Contactless LMD systems and services can result in avoiding physical contact between caregivers and patients or between delivery workers and the recipients. These benefit from the rapid proliferation of connected technologies and the recent advancements in semi and fully autonomous delivery platforms which revolutionize the urban logistics and provide a safe and efficient delivery methods

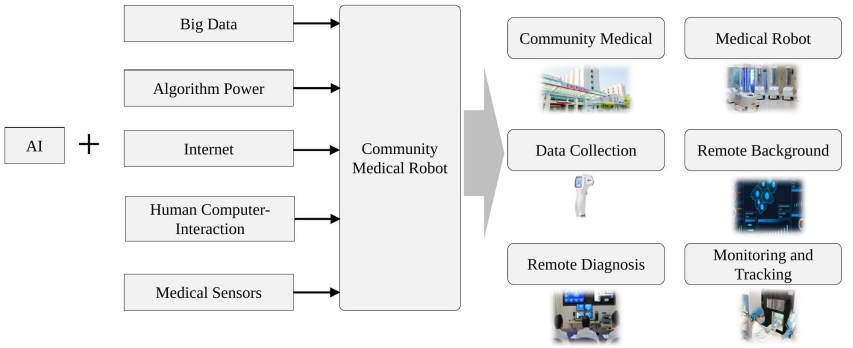


Figure 12
A composition framework diagram of community medical robots.

for medical supplies [39], medications [40], food, grocery and other goods [41, 42]. Fig. 14a shows AIMBOT38, an autonomous mobile robot developed by UBTECH Robotics, performs disinfection tasks at Shenzhen Third Hospital.

The demand for food delivery facilities has never been higher as most restaurants and cafés are closed during lockdown. Fig. **??**b shows Nuro driverless vehicle that has been approved for delivery tests in California in April 2020 [42]. Drone distribution of medication and Personal Protective Environment (PPE) is a valuable option as drones can easily carry samples from hospitals to labs, provide simple security services, and bring medical equipment to patients' homes such that immediate control can be provided. Currently some companies in China and Ghana are actively using this technology.

Shortage of PPE and other medical equipment was a global issue in the first half of 2020, until production volumes could be increased, but some critical issues remain. WHO estimates that one in every six COVID-19 patients needs breathing support through medical ventilators at intensive care units (ICUs). Robots could help in supplies in several ways [43]:

- Increasing medical ventilator production by automating approximately 40% to 50% of assembly tasks;
- Re-purposing existing non-ventilator production—such as airplane, jet engine, automobile, and vacuum production—to ventilator production;
- Reducing the need of human operators;
- Maintaining social distancing at manufacturing facilities;
- Maintaining ability for future production.

Figure 13
The Knightscope lineup of indoor and outdoor telepresence robots. Credit: Knightscope Robotics.

Figure 14
a) AIMBOT, an autonomous mobile robot. Credit: UBTECH Robotics. b) Nuro driverless vehicles. Credit: Nuro.

These functions may apply to other products in need as well.

10 Ethical Implication of Robotics and Automation

Previous Sections highlighted how important Robotics and AI are for addressing COVID-19. At same time, they can pose serious risks for our fundamental rights ranging from the impact on democracy to our right to life [44], [45].

Currently, several Governments are establishing partnerships with private companies to get access and track people's behavior and location. In [46], one can see a live index elaborated by Top10PVN, about the use of cell phone location per country, which ranges from processing anonymized data to the tracking of an

individual person and their contacts. Governments argue that mobile location can provide invaluable information to help understanding the dynamic of COVID-19 and, thus, reducing public threats. The data collected can be used to identify patterns of movements and collect inform on how the virus is being spread in a particular region or, even, at individual level, alert authorities if a particular individual is obeying to home quarantine [47].

The main concern is that these can open the door to further, wider use of such technology, increasing the surveillance of the citizens under different and questionable justifications. This raises serious concerns about potential abuse that would directly impact our privacy, freedom of expression and freedom of association accentuating, e.g., discrimination and marginalization of groups. Recently, a joint statement signed by different organizations urged Governments ensure the use of digital technologies to address the pandemic is in line with human rights [48]. Governments should observe that

- all surveillance measures must be lawful, necessary and proportionate;
- all monitoring and surveillance powers must be time-bound;
- all monitoring may only continue for as long as necessary to address the current pandemic.

Considering the latter, Amnesty International [49], emphasizes ”there is a real danger surveillance measures become permanent fixtures. In the meanwhile, the mounting financial difficulties decrease the companies capabilities to deal with ethical issues and CSR related activities [50].

Another relevant issue is whether an AI-based system should ultimately make critical medical decisions (keeping in mind that trusting ADs with this also puts a huge load on them).There are numerous questions arising on what the data should be used to provide an answers, and whether age, gender, known diseases, race, social status or other should be considered.

In China, researchers developed an AI-based tool to assist doctors in making choices about patients with COVID-19 using blood samples [51]. However, they argue that human agency should never be undermined, i.e., humans should not be obligated to use system decisions blindly without having them scrutinized. The reason is simple: AI-based systems are only math tools. They are neither intelligent nor neutral. They are fairly as good as the quality of the data that is used in their training set. This fact is corroborated with evidences presented in different reports that show gross errors produced by some systems, which a human would never make [52].

A correlated problem is that most people who will use AI-based systems do not know their limitations. This has led to situations where people in key positions anthropomorphized and over-trusted the technology, and, using fallacious arguments, imposed the adoption of these systems in critical and sensitive tasks. Therefore, initiatives like IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems [53] and IEEE/MIT Global Council on Extended Intelligence [54] are very relevant in the resolution of related COVID-19 challenges.

The related IEEE P7000 Ethically Aligned Design standard may have great importance in pandemic-related critical technology development and assessment. P7000 puts emphasis on 5 key domains: accountability;auditability; dependability; predictability and transparency.

While the P7000 aims to educate and empower people to take ethics into consideration during the life cycle of AI-based systems, the latter discusses AI in creating a new narrative for the domain to demystify the hype that surrounds it. The risk for misusing technology is high. Organizations like ProPublica recently pointed out people with disabilities may be denied from lifesaving care during the Pandemic [55]. In Brazil, the Ministry of Health suggested in a press collective [56] that saving an elderly is more financially expensive than saving a young person. Predictions like that can be incorporated in an AI-based system causing irreparable losses, and due to the opaqueness of the system, this discriminatory information would never be uncovered and the responsible would never be held accountable. These domains shall further be regulated and managed, for which we could even take inspiration from SCI-FI and literature [57].

Conclusions

The global battle against the novel coronavirus and COVID-19 has seen robotics and automation playing an extremely important roles in understanding the scope of the pandemic, dealing with the contagious nature of the virus and decelerating its spread. This article structurally reviewed a number of robotics and automation applications in the battle against COVID-19. These applications include early detection and diagnosis, assistive robots, indoor and outdoor disinfection robots, public awareness and patrolling, last-mile delivery services, micro- and nano-robotics and laboratory automation. Ethical implications of robotics and automation were also discussed in the context of the novel principles of ethically aligned design.

Many conclusions can be drawn from this unprecedented crisis that has profoundly impacted humanity on a global scale. There is also a need to fully understand the limitations of automated systems for not over-trusting the technology and creating false arguments especially in critical health-related tasks. Medical regulatory branches have an outstanding responsibility in this. Measures should be taken to protect digital rights, stop the surveillance once the pandemic ends, ban new totalitarianism of surveillance technology and digital totalitarian states. Robotics and AI in general have to be exploited, yet regulated and smartly directed toward the benefit of humanity.

Acknowledgment

T. Haidegger's work is partially supported by the ELKH SZTAKI – OE Cyber-Medical System Development project. T. Haidegger is supported through the

New National Excellence Program of the Ministry of Human Capacities.
T. Haidegger is a Bolyai Fellow of the Hungarian Academy of Sciences, his research is partially supported by the EFOP-3.6.2-16-2017-00016 project.

References

[1] G. Hager, V. Kumar, R. Murphy, D. Rus, and R. Taylor. The role of robotics in infectious disease crises. *arXiv preprint arXiv:2010.09909*, 2020.

[2] World Health Organization. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). Technical report, 2020.

[3] C. Shorten, T. M. Khoshgoftaar, and B. Furht. Deep learning applications for covid-19. *Journal of Big Data*, 8(1):1–54, 2021.

[4] A. A. Hussain, O. Bouachir, F. Al-Turjman, and M. Aloqaily. Ai techniques for covid-19. *IEEE Access*, 8:128776–128795, 2020.

[5] A. Zemmar, A. M. Lozano, and B. J. Nelson. The rise of robots in surgical environments during covid-19. *Nature Machine Intelligence*, 2(10):566–572, 2020.

[6] P. Barattini, F. Vicentini, G. S. Virk, and T. Haidegger. *Human-robot interaction: safety, standardization, and benchmarking*. CRC Press, 2019.

[7] T. Haidegger. Autonomy for surgical robots: Concepts and paradigms. *IEEE Transactions on Medical Robotics and Bionics*, 1(2):65–76, 2019.

[8] L. Luxner. Deluged by pandemic needs, israeli doctors get help from unlikely source: robots. *Jewish Telegraphio Agency*, Jul 2020.

[9] Singapore-made COVID-19 swab test robot could reduce healthcare workers' risk of infection, Sep 2020.

[10] W. Fan. Epidemic promotes surge in demand for infrared temperature measuring instruments: we "reunite families" and they "overtime". *Science and Technology Board Daily.*, 2020.

[11] Shenzhen Municipal Affairs Service Data Administration. Epidemic pass reference architecture and technical guide, 2020.

[12] S. Han and Y. Fan. Discussion on the application of logistics robot distribution in community epidemic isolation. *Digital Communication World*, 2020.

[13] B. Song, R. Yang, N. Xi, K. Patterson, C. Qu, and K. Lai. Cellular-level surgery using nano robots. *J Lab Autom.*, 17(6):425–434, 2012.

[14] Y. Hu, J. Jacob, G. Parker, and et al. The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. *Nature Machine Intelligence*, (2):298–300, 2020.

[15] L. Peskoe-Yang. Ibm's new ai tool parses a tidal wave of coronavirus research. *IEEE Spectrum*, 1 Jul 2020.

[16] Q. Sun. AR glasses check the car, induction disinfection. Technical report, 2020.

[17] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest CT. *Radiology*, 2020.

[18] Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, and N. Yao. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with covid-19 in an accurate and unobtrusive manner. *arXiv preprint arXiv:2002.05534*, 2020.

[19] A. C. Smith, E. Thomas, C. L. Snoswell, H. Haydon, A. Mehrotra, J. Clemensen, and L. J. Caffery. Telehealth for global emergencies: Implications for coronavirus

disease 2019 (COVID-19). *Journal of telemedicine and telecare*, pages 309–313, 2020.

[20] X. Zhou, C. L. Snoswell, L. E. Harding, M. Bambling, S. Edirippulige, X. Bai, and A. C. Smith. The role of telehealth in reducing the mental health burden from COVID-19. *Telemedicine and e-Health*, 26(4):377–379, 2020.

[21] T. Haidegger, V. Varga, A. Lehotsky, P. Róna, R. Pethes, P. Szerémy, L. Szilagyi, T. Ferenci, and L. Kovács. Information technology tools employed in infection control. In *16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 339–344, 2015.

[22] R. F. Service. Does disinfecting surfaces really prevent the spread of coronavirus? *Science Magazine*, Mar 2020.

[23] D. A. Drexler, A. Takács, T. D. Nagy, and T. Haidegger. Handover process of autonomous vehicles-technology and application challenges. *Acta Polytechnica Hungarica*, 16(9):235–255, 2019.

[24] University of illinois urbana-champaign develops autonomous robot to kill covid-19. *Healthcare Hygiene Magazine*, OCT 2020.

[25] G. Yang, Z. Pang, M. J. Deen, M. Dong, Y. Zhang, N. H. Lovell, and A. M. Rahmani. Homecare robotic systems for healthcare 4.0: Visions and enabling technologies. *IEEE Journal of Biomedical and Health Informatics*, 2020.

[26] Z. Pang, G. Yang, R. Khedri, and Y.-T. Zhang. Introduction to the special section: convergence of automation technology, biomedical engineering, and health informatics toward the healthcare 4.0. *IEEE Reviews in Biomedical Engineering*, 11:249–259, 2018.

[27] H.-W. Huang, C. Ehmke, G. Merewether, F. Dadabhoy, A. Feng, A. J. Thomas, C. Li, M. d. Silva, M. H. Raibert, E. W. Boyer, et al. Agile mobile robotic platform for contactless vital signs monitoring. 2020.

[28] World Health Organization. Telemedicine: opportunities and developments in member states. *report on the second global survey on eHealth.*, 17(6):425–434, 2009.

[29] C. Kuziemsky, A. J. Maeder, O. John, S. B. Gogia, A. Basu, S. Meher, and M. Ito. Role of artificial intelligence within the telehealth domain: Official 2019 yearbook contribution by the members of imia telehealth working group. *Yearbook of medical informatics*, 28(1):35, 2019.

[30] Meticulous Research. Healthcare virtual assistant market - global opportunity analysis and industry forecast (2018-2025). page 103, 2019.

[31] S. Hoermann, K. L. McCabe, D. N. Milne, and R. A. Calvo. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *Journal of medical Internet research*, 19(8):e267, 2017.

[32] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau, et al. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258, 2018.

[33] P. Webster. Virtual health care in the era of COVID-19. *The Lancet*, 395(10231):1180–1181, 2020.

[34] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong. Digital technology and COVID-19. *Nature medicine*, 26(4):459–461, 2020.

[35] E. El Habre. A case for switching from self-fulfilling prophecies to rights in ai journalism. 2020.

[36] L. Town and K. Hoffman. *Coronavirus and the Caregiver*. Omega Press, 2020.

[37] R. W. Goodman. Whatever you call it, just don't think of last-mile logistics, last. *Global Logistics & Supply Chain Strategies*, 9(12), 2005.

[38] M. Joerss, J. Schröder, F. Neuhaus, C. Klink, and F. Mann. Parcel delivery: The future of last mile. *McKinsey & Company*, 2016.

[39] T. Niels, M. T. Hof, and K. Bogenberger. Design and operation of an urban electric courier cargo bike system. In *2018 21st international conference on intelligent transportation systems (itsc)*, pages 2531–2537. IEEE, 2018.

[40] L. Nianzhen. How a chinese drone delivery startup is capitalizing on COVID-19. *Nikkei Asian Review*, Mar 2020.

[41] Robot to deliver meals, medication to Covid-19 patients in S'pore. *Bangkok Post*, Mar 2020.

[42] S. Crowe. Nuro driverless vehicles approved for delivery tests in california. *The Robot Report*, Apr 2020.

[43] A. A. Malik, T. Masood, and R. Kousar. Repurposing factories with robotics in the face of COVID-19. *Science Robotics*, 5(43), 17 Jun 2020.

[44] A. Khamis, H. Li, E. Prestes, and T. Haidegger. Ai: A key enabler of sustainable development goals, part 1 [industry activities]. *IEEE Robotics & Automation Magazine*, 26(3):95–102, 2019.

[45] A. Khamis, H. Li, E. Prestes, and T. Haidegger. Ai: A key enabler for sustainable development goals: Part 2 [industry activities]. *IEEE Robotics & Automation Magazine*, 26(4):122–127, 2019.

[46] J. Morley, J. Cowls, M. Taddeo, and L. Floridi. Ethical guidelines for sars-cov-2 digital tracking and tracing systems. *Available at SSRN 3582550*, 2020.

[47] K. Uibu. Poland is making its citizens use a 'selfie' app during the coronavirus crisis. *ABC news Australia*, Apr 2020.

[48] Joint civil society statement: States use of digital surveillance technologies to fight pandemic must respect human rights. *Human Rights Watch*, Apr 2020.

[49] COVID-19, surveillance and the threat to your rights. *Amnesty International*, Apr 2020.

[50] P. Karácsony. Analyzing the relationship between leadership style and corporate social responsibility in hungarian small and medium-sized enterprises. *Acta Polytechnica Hungarica*, 17(7):183–198, 2020.

[51] S. Chen. Should AI help make life-or-death decisions in the coronavirus fight? *South China Morning Post*, Mar 2020.

[52] D. Heaven. Why deep-learning ais are so easy to fool. *Nature*, 574(7777):163–166, 2019.

[53] P. Winch. *Ethics and action*. Routledge, 2020.

[54] G. Adamson, J. C. Havens, and R. Chatila. Designing a value-driven future for ethical autonomous and intelligent systems. *Proceedings of the IEEE*, 107(3):518–525, 2019.

[55] A. Silverman. People with intellectual disabilities may be denied lifesaving care under these plans as coronavirus spreads. *ProPublica (https://www. propublica. org/article/people-with-intellectual-disabilities-may-be-denied-lifesaving-care-under-these-plans-as-coronavirus-spreads). Accessed*, 19, 2020.

[56] Teich já sugeriu que salvar idoso em vez de jovem é mais caro à saúde. *(In Portugese) Carta*, Apr 2020.

[57] R. R. Murphy. Robots and pandemics in science fiction. *Science Robotics*, 5(42), 2020.

Direct Drive Hand Exoskeleton for Robot-assisted Post Stroke Rehabilitation

Márk Ottó Bauer[1], Máté Benjámin Vizi[2], Péter Galambos[3], and Tibor Szalay[4]

[1,4]Department of Manufacturing Science and Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3. H-1111 Budapest, Hungary, {bauer.mark, szalay.tibor}@manuf.bme.hu
[2]Department of Applied Mechanics, Budapest University of Technology and Economics, Műegyetem rkp. 5. H-1111 Budapest, Hungary , vizi@mm.bme.hu
[3]Antal Bejczy Center for Intelligent Robotics, Óbuda University, Bécsi út 96/B. H-1034 Budapest, Hungary, peter.galambos@irob.uni-obuda.hu

Abstract: This article introduces novel rehabilitation hand module development for the physiotherapy of the hand of patients suffering from spastic hemiplegia. Spasm is basically a muscle cramp, it practically involves the sudden, unintended and painful contraction of a muscle or muscle group, which is caused by nerve damage resulting from a stroke. Stroke is the main reason for permanent disability in adulthood, and so the social- and medical care systems require a huge amount of healthcare resources due to the inactivity of the patients concerned. The robotically facilitated rehabilitation assists the physicians in providing repeated therapies of great intensity, and so the patients may enjoy the benefits of rehabilitation, while the therapists may reduce their own workload at the same time. Furthermore, the robotic devices offer an objective and reliable opportunity for tracing and accurately assessing the improvement of the patients' motor skills. This article introduces the electrical- and mechanical design of a therapeutic device and the inverse kinematic and dynamic modules which control this device. The rehabilitation device is capable of moving the thumb, the index-, the middle- and the ring fingers, and allows the rehabilitation of the left- and right hands as well. The device is a completely new design with direct drive approach and several benefits. It has two components: a planar module with serial kinematics of rotational joints with three degrees of freedom (3DoF RRR), and another module with two degrees of freedom (2DoF). The modules integrated load cells, which are built in between each joint to measure the reaction forces. The 3DoF finger moves the index, the middle and the ring fingers, using a load distributor placed above the fingers. The finger orthoses are connected to the load distributor via magnets. The 2DoF finger moves the thumb performing the opening/closing along the plane tilted in two angles.

Keywords: Rehabilitation robotics, Hand rehabilitation, Hand exoskeleton, Wearable structure, Force control

1 Introduction

Stroke is one of the critical causes of disability in cognition or movement. It often results in a combination of cognitive, motor and sensory impairments. Now, it has become one of the main diseases threatening human health and the healthy survival[1]. One of the most common impairment of stroke is the hemiparesis, which results in dyskinesia of some parts of the body. It reflects not only in the upper and lower limbs, but also in the loss of motor function of the hand. Hand rehabilitation requires repetitive task exercises, where a task is divided into several movements and patients need to practice those movements to improve their hand strength, motion accuracy, and range of motion, but high cost of traditional treatments are often prevent patients from spending enough time on the necessary rehabilitation. Few studies show that repetitive movements with robotic assistance can significantly improve the hand motor functions of patients [1]. Hand is one of the most important limb of humans, and the rehabilitation of hand motor function can be assisted by exoskeleton robots, because the application of robotics and related technologies has greatly promoted the development of clinical techniques. At the last 10-20 years many dexterous and advanced mechanisms of hand exoskeletons have been developed and tested. But not all of them are developed for hand rehabilitation. For example some of them are designed as force-feedback devices and some others are designed for master–slave gripping systems. Hand rehabilitation exoskeleton devices are still seeking to achieve key features such as bi-directional actuation, low cost, low complexity, compactness, portability, safe interaction between human and machine, and simple and reliable control.

The robot-assisted hand/finger rehabilitation is recently in the focus of researchers, so it has a broad literature. There are several approaches to solve the difficulties appearing in the case of hand and finger movement, the main constraints are the available space for the robot and the relatively large forces required to move the spastic fingers. The most common approaches considering the actuator types are the pneumatic gloves and the electric motors with indirect drive through a wire/cable system. Also worth to mention a new direction in the this field, rapid development of material sciences results in new possibilities, shape memory alloys can be used as the active actuators [2] in hand exoskeletons. Pneumatic gloves use air-pressurized chambers on the palmar side of the hand for finger extensions; if these are the only actuators, then only the finger extension is controlled by the robot [1, 3] (see Figure 1). The finger flexion can be achieved by a supplemented motor-cable system [4] or by pre-loaded passive springs [5] on the dorsal side of the hand. The other common approach uses electric motors, the produced torques are applied on the finger joints through a cable/wire system [6, 7] (see Figure 1). Separate motors are necessary for flexion and extension because of the soft cables transmitting the force, so two motors must be used for each controlled joints. These solutions result in a relatively thin exoskeletons [8, 9] or gloves [10, 11], but the overall required space is large [12, 13] due the high number of motors. All of mentioned studies threat each of the five

[1] https://www.healthline.com/health/top-10-deadliest-diseases

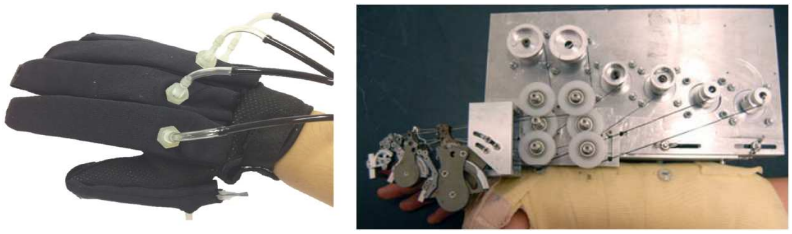


Figure 1
Pneumatic glove [3] on the left and cable driven robot finger [7] on the right

fingers totally individually, the thumb itself may require separate design due to its special function of opposition [14].

Other important property is number of separately controlled joints or the number of degrees of freedom. Pneumatic gloves control the extension of the whole fingers with a single actuator. The motor-cable systems may control each joint of each finger separately, this provides a great flexibility for the realized motion [9, 15]. However, it is also possible control the last two joints [6] or all the three joints (common in the case of simple gloves) [4] together requiring fewer motors, but also resulting in less flexibility for the possible motions.

Our design utilizes the direct-drive approach, meaning that the motor are placed near the fingers and the force is applied directly on the middle phalanx (distal phalanx in the case of the thumb) through the robot body without any cable system. The movement of index, middle, ring and pinkie fingers are grouped, they are moved together using a single robotic arm with only three motors; so the number of necessary motors is highly reduced, still, the free motion of separate joints along the fingers is preserved. The thumb has its own robot finger with two motors, overall, this hand module has five degrees of freedom.

The module introduced in the article will be applied during the augmentation and further development of the REHAROB robotic physiotherapeutic equipment [16, 17] built in the Hungarian Medical Institute for Rehabilitation (OORI). The further development involves supplementing the REHAROB physiotherapeutic equipment with a hand module and making it capable of supporting active kinesiotherapy. The basic system assists the stressful tasks of physiotherapists engaged in the kinesiotherapy of patients who suffer from spastic hemiplegia - called spasm in medical terminology - by applying industrial robots. The system consists of two industrial robots manufactured by ABB, a PC-based high-level controller, a motorised treatment seat, an object table, one hand- and several finger-orthoses and a custom-designed hand module. The system has two modes of operation, the first is the passive mode: in this mode, the therapist trains the robots to perform a given exercise. The robot is force-controlled during the training (in impedance mode), it records the trajectory of the movements, and then it is capable of replaying the saved trajectory, which means repeating the tasks (taught by the therapist) on the patient. The second is the active mode, where the exercises consist of five predetermined activities of daily living movements

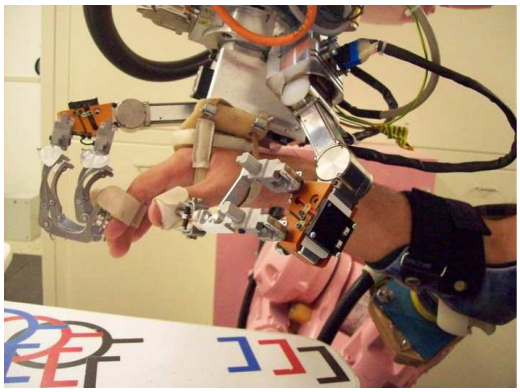


Figure 2
Structure of the REHAROB system (hand module, ABB robot arms)

(drinking, using zipper, wiping mouth with a sponge, using telephone, etc.).

The robot has a pre-written program for the sequence of movements, and performs the exercises on the basis thereof. The modified SDH (SCHUNK Gmbh. Dextrous Hand 2.0) hand module consists load cells, two robot fingers, and the mediator elements and finger orthoses mounted on the former. One finger is responsible for moving the thumb, the other is responsible for moving the index-, the middle- and the ring fingers. The little finger is not involved. Figure 2 shows the structure of the REHAROB system and the old hand module mounted onto the end of the IRB1600 robot mounted on the "ceiling", which implements the moving/rehabilitation of the hand.

2 Design considerations

The input data of the design (specifications) is based on the experiences obtained by clinical examinations [18] performed with the REHAROB robotic physiotherapeutic system shown in the Figure 2. The so-obtained design aspects are also based on the opinions of the dedicated medical team.

The proven features of the previous system are the following:

- Replaceable hand orthosis (standard sizes customizable for the patient).
- replaceable finger orthoses (orthoses with customizable sizes).
- Changeable left/right side (it is possible to perform physiotherapy both for the left and right side, ie. side change is possible).
- Adjustable orthoses.
- Absolute angle sensors installed in the joints.
- Applied admittance control.

- Drive integrated in the central unit.
- Drive of each joint in stress-free condition (no self-locking transmissions).

The functions missing from the previous device and the its defects are summarized as follows:

- Overly complicated assembly.
- The removable connections are inadequate, the thumb finger module needs to be replaced for left-right side change. The unit moving the thumb needs to be replaced when changing the side, which entailed not only the mechanical reassembly of the modules, but also replacing the electrical wires. The connections were damaged often and the cables were broken when connected. The damages of the connections and cables resulted in short-circuits.
- Few implemented degrees of freedom (2 controlled joints), under-actuated movement.
- Low performance (the folding-out torque of the joints is low: 1.4[Nm] and 2.1[Nm]).
- The robot finger moving the three fingers (index-, middle- and ring finger) is asymmetric, its allocation is eccentric, so the module needs to be placed to the other side when side is changed.
- The modules - due to their geometrical sizes - are uncapable of encompassing the standard anatomical dimensions (60 percentiles for the female and 95 percentiles for the male hand).
- The modules - due to the physical dimensions of the load distributors mounted on them - are unable to implement gentle motions (grabbing and releasing objects).

The defects/deficiencies listed above made the development of a new construction necessary:

- Electrical wiring needs to be established within the finger.
- The degrees of freedom of the three-finger module has been increased: 3 driven joints, full closing has been solved.
- The three-finger module has been redesigned to be symmetrical: the module has been aligned to the centreline of the lower arm and of the hand. The module does not need to be replaced when changing the side, only the thumb-moving unit's position needs to be changed.
- The hand orthosis is replaceable: it can be removed and fixed with an locating pin.
- The unit moving the thumb is provided with a built-in electrical connection. No wiring is necessary when changing the side.

- The load distributor is now part of the end of the three-finger robotic finger, so the encompassing size is minimized.
- A two-directional force measurement has been implemented at the end of both robot fingers.
- Built-in force measurement: the inner joint torques are now measured directly.
- The performance of the motors has been increased: application of 20[W] and 60[W] brushless direct current motors with electronic commutation.
- The overload protection of the measuring cells is implemented via mechanical collision: the built-in load cells are now protected against overload.
- Power electronics integrated in the joints (force measurement, position measurement, speed- and position control organized in the lower level).

3 System Overview

The structure of the modules consisting rotational joints with serial kinematics is as described below. Each rotational joint is actuated by harmonic drives, and directly-driving brushless DC motors (BLDC). The torques on the joints are sensed by small load sensors with a measuring limit of 100[N], which have been placed between the joints. Each force cells sense radial - ie. normal direction - forces, the torque on the joints can be calculated from the radial forces, in knowledge of the fixed geometrical distances between the joints. The occasional overload protection of the load cells is implemented by the mechanical collision of the housing. The force component perpendicular to the normal direction is sensed by micro force sensor placed at the end of the finger, which is capable of sensing compression and tension forces as well. The angle sensor of the joints is sensed by Hall type 14-Bit Programmable Magnetic Rotary Position Sensor. The joints are provided with integrated power electronics placed at the end of the BLDC motors, and receive the signals from the ADC SPI which is processing the absolute transducers and the force measurement cells. The AS5047 absolute encoder and the ADC122S101 ADC communicate with the dsPIC33EP512GM304 micro controller via two independent SPI lines. The controlling of the brushless motors is performed by the DRV8308 Brushless DC Motor Controller. The integrated power electronics panels use a full duplex RS-422 communication line to communicate with the communication distribution unit which is connected to the higher-level controller. The distributor unit receives the measured data of the joints (torque/forces, absolute positions, current rpms), and sends the set-point calculated by the controller (speed reference) and the list of the related controlling parameters (values of the PID speed regulator's parameters). RS-422 hardware setup was applied due to the symmetrical and differential structure, in order to reduce sensitivity to communication disturbances. The slave units are scheduled by the synchronizing signal coming from the master, so the scheduling of the controller can be perpetuated. The communication jitter resulting from

the communication of the system can be minimized. The power electronics panels integrated in the joints are powered by a 24[VDC] power supply, which also powers the final drive of the brushless motor and the DC-DC switching converter which generates the inner power supply required for the operation. The power electronics panel is located at the end of the motor, and an FFC connector links it to the inner communication PCB located in the joints of the robot finger. The signals between the rotation joints - including the RS-422 level-shifted separated RX, TX communication signals of the rotation joints, the communication lines of the absolute angle sensor of the drive and the communication lines of the ADC SPI which is processing the load cell - are forwarded by the flexible printed circuit board. The flexible PCB is rolled around the joint to ensure free movement, so its length can change during rotation. The absolute encoder IC is placed on this circuit board as well. The aluminium console of the HAL IC is placed above the IC, fixed to the moving part of the joint with two bolts, so the bipolar magnet placed in the console is rotated together with the joint.

The power supply of the power electronics modules is solved within the joint by using flexible silicone cables (Stäubli FlexiVolt-E), since the guide strip's cross-section required due to the total power load of the electrical motors and the space available inside the joint does not allow forwarding the required power via a flexible circuit board. The flexible silicone cables are located above the flexible circuit board, their position and guiding are secured by the aluminium consoles of the HAL IC, and the cables are connected to the distributor panels in the joints. The inverse kinematic (mathematical) module implementing the robot finger's movement, as well as the admittance controller implementing the trajectory recording and power control are implemented in the higher-level controller. Real-time calculation capacity can be ensured easier and even the algorithm implementation is also faster using a PC-based architecture and a real-time operating system, than using an embedded device. The parameters of the admittance controller (required strength, attenuation) can be adjusted during the physiotherapy, so that the robot finger can achieve the desired results. Cascade controlling has been implemented in the joints, while speed regulation has been applied in each motor - in addition to subordinated voltage regulation and super-ordinated position regulation. The admittance controller provides position setpoint to the joints of the robot finger, the position trajectory determines the reaction force of the module, so the prescribed torque can be interpreted as a position setpoint within the regulatory chain. The prescribed position is generated by the motion of the circular spline of the harmonic drive, so the feedback of the position regulation is implemented on the folding-out part of the transmission, and so a dead time element is created in the regulatory chain, which has a relatively large time constant. This is the amount of time that needs to pass in the system to have the effect of the input appear on the output.

Figure 3 shows the flexible PCB rolled up in the joint (its 3D-printed test version) and the flexible silicone cables. The image shows the external final position of the joint. The length of the flexible cables is changing during the rotation of the joint, because the cables are not rotating along the same curve due to their diameter (they are located abreast). Their length change inside the housing is allowed by

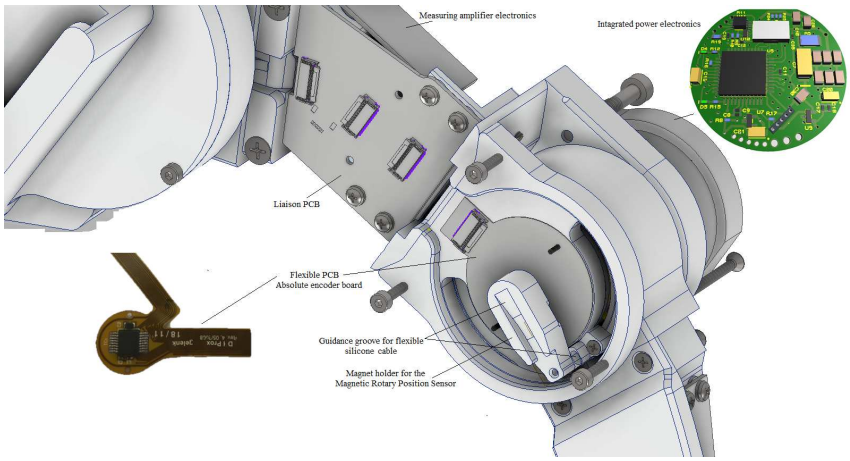


Figure 3
The second joint of the robotic hand with the integrated electronics and flexible PCB

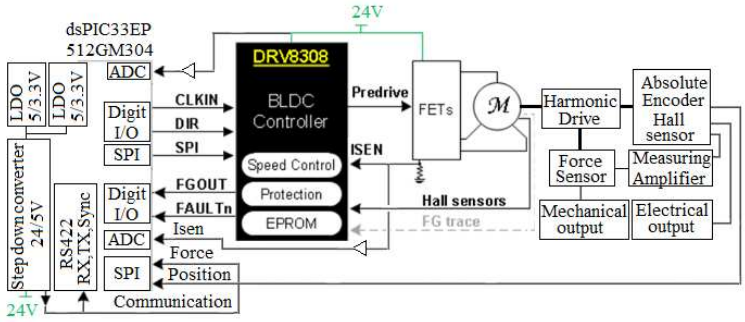


Figure 4
Schematics of the actuation hardware

the free movement of the cables. The inner size and shape of the grooves on the housing was established according to the cables' bend radius. The flexible cables are connected to the inner circuit boards that connect the joints. Figure 4 shows the integrated power electronics schematic.

4 Mechanical Design

The rehabilitation device serving therapeutic purposes has two components: a planar module 3DoF RRR, and another module with 2DoF. The robot finger with three degrees of freedom moves the index-, the middle and the ring fingers, using a load distributor placed above the fingers. The finger orthoses are connected to the load distributor via magnets. The robot finger with two degrees of freedom moves the thumb according to closing, along the plane (tilted in two angles). We have built a measuring system to determine the required plane; this system consists of two perpendicular cameras. We've obtained the spatial location of the motion's trajectory by using the markers fixed to the thumb, from the cross-section of the two cameras' perspectives and from the images recorded frame-by-frame during the motion. The coordinate values of the points recorded by the cameras could be imported into a CAD system. The motion's trajectory results in a conical surface; this surface can be approximated well by moving along a plane tilted in the proper degree, by implementing two degrees of freedom. Grabbing with the thumb is ensured by the orthosis located on the distal phalanx bone. The motion occurring upon closing the finger is allowed by the joint provided with a bearing.

The robot finger with three degrees of freedom allows the full opening and closing of the hand, and the module encompasses the hand's anthropometric size range between 60 and 95 percentile size. Hence, the device is fitting for the rehabilitation of male and female patients as well. The rehabilitation device is fixed to

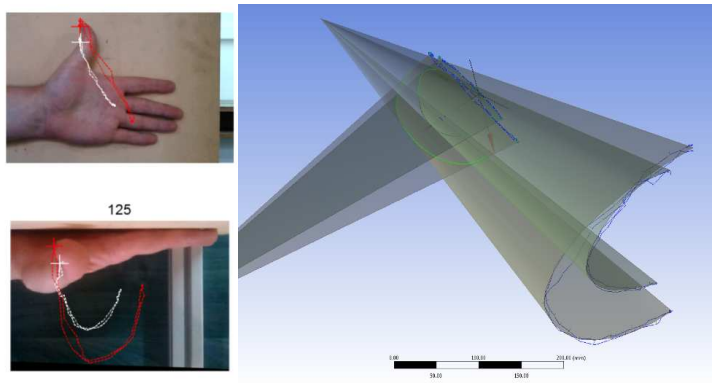
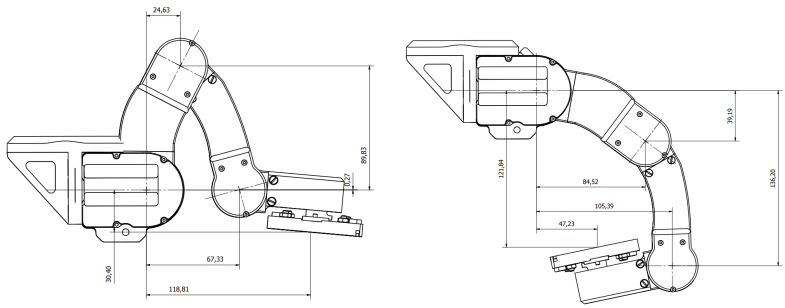


Figure 5
Thumb finger trajectories (measuring layout), and CAD model

the hand by a custom-designed hand orthosis, which can be fitted to the patient's hand in accordance with the size standards (Small, Medium, Large), while the orthosis ensures the hand's stable grasp by a spring prestressing. The design of the inner surface's curve follows the curve of the back of the hand. The rehabilitation device allows changing between the left- and the right hand, which allows the rehabilitation of both sides.

Two sizes of drives and two kinds of motors have been used in the robot finger. The smaller joints producing a torque of 4.8[Nm], while the torque of the larger joints is 11[Nm]. The position is measured on the side of folding-out, directly in relation to the fixed pillow block (see Figure 3); this provides the feedback signal of the position control. The load cells integrated between each joint measure the reaction forces, the torques of the joints can be calculated in knowledge of the geometry. The implemented mechanical construction allows a rigid and well-controlled structure, which has a smaller loss and is much more accurate than the indirect cable-controlled tendon actuation. A two-component force meter (two load cell) has been installed at the end of the fingers where the orthoses are fixed, so in addition to the torque of the joints, we can even measure the force component perpendicular to the torque.

Figure 7 shows the 3D design of the rehabilitation device serving therapeutic purposes. The device and the UR-10 e-Series robot which moves the upper limb are linked by an adapter which is fitted to the flange-end of the robot. The image shows the T-nut-shaped fixing insert which allows interchangeability and implements the tilt angles of the 2DoF thumb module. It is clearly visible in the image that the thumb module is fixed in a different angle on the left side of the device than it is fixed in horizontal position on the right side, because of the side change. The module fixing T-nut insert is provided with a built-in electrical connection, which solves the module's electrical connection after a mechanical stabilisation when sides are changed. Side change is easily achieved by the telescopic indexing plungers. The finger orthoses on the patient's fingers are fixed to the rehabilitation device by using magnets.



(a) Fully opened (the hand is stretched)

(b) Fully closed (fists clenched)

Figure 6
The design of the direct drive robot hand

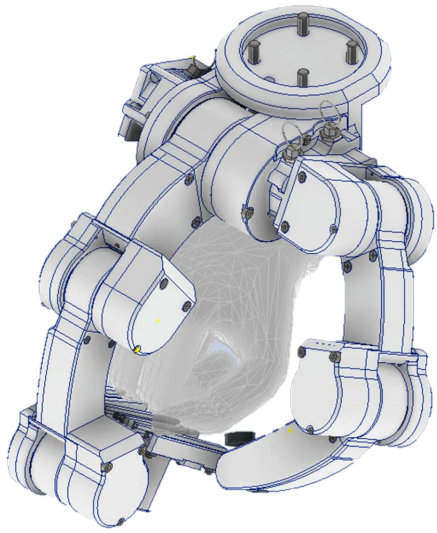


Figure 7
The proposed arrangement of the hand exoskeleton

5 System modelling

In this Section, the kinematics and dynamics of the proposed equipment is discussed.

5.1 Kinematics

The relative joint angles are denoted as θ_1, θ_2 and θ_3, the end point coordinates are denoted by x and y, while φ is the absolute angle of the last segment. The mechanical model is shown in Figure 8.

The forward kinematics can be derived using basic geometry, the (l_1, l_2, l_3) parameters are the lengths of the corresponding segments:

$$\begin{bmatrix} x \\ y \\ \varphi \end{bmatrix} = f(\theta_1, \theta_2, \theta_3) = \begin{bmatrix} l_1 \cos(\theta_1) + l_2 \cos(\theta_1 + \theta_2) + l_3 \cos(\theta_1 + \theta_2 + \theta_3) \\ l_1 \sin(\theta_1) + l_2 \sin(\theta_1 + \theta_2) + l_3 \sin(\theta_1 + \theta_2 + \theta_3) \\ \theta_1 + \theta_2 + \theta_3 \end{bmatrix}. \tag{1}$$

The Jacobian matrix of the forward kinematics f is:

$$\mathbf{J} = \begin{bmatrix} -l_1 \sin(\theta_1) - l_2 \sin(\theta_1+\theta_2) - l_3 \sin(\theta_1+\theta_2+\theta_3) & -l_2 \sin(\theta_1+\theta_2) - l_3 \sin(\theta_1+\theta_2+\theta_3) & -l_3 \sin(\theta_1+\theta_2+\theta_3) \\ l_1 \cos(\theta_1) + l_2 \cos(\theta_1+\theta_2) + l_3 \cos(\theta_1+\theta_2+\theta_3) & l_2 \cos(\theta_1+\theta_2) + l_3 \cos(\theta_1+\theta_2+\theta_3) & l_3 \cos(\theta_1+\theta_2+\theta_3) \\ 1 & 1 & 1 \end{bmatrix} \tag{2}$$

The determinant of the Jacobian matrix $\mathbf{J}$ is:

$$\det \mathbf{J} = l_1 l_2 \sin\theta_2, \tag{3}$$

so the singular positions are $\theta_2 = 0, \pm\pi$ rad or $0°, \pm 180°$, which means that the middle segment is straight with respect to the first segment. This position should be avoided during the control, because the inverse kinematics calculation is non-unique here, this requirement is naturally fulfilled by the planned motions (human-hand physiotherapy), so no further treatment is necessary. The inverse

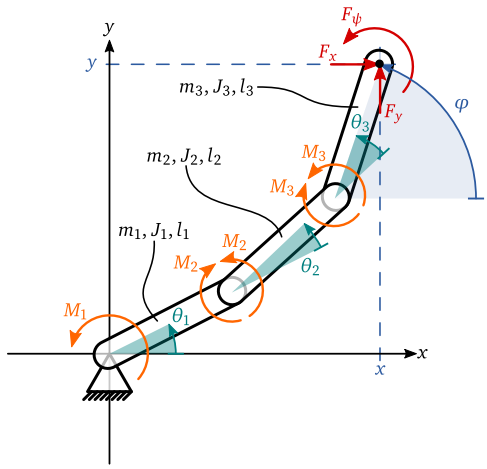


Figure 8
Mechanical model of the 3DoF exoskeleton finger

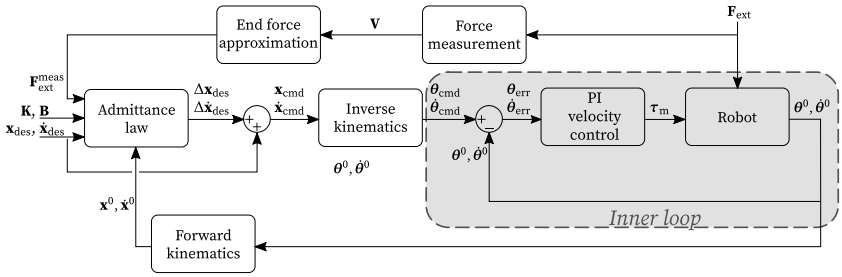


Figure 9
Block diagram of the admittance control

kinematics can be described locally by the inverse Jacobian matrix:

$$\mathbf{J}^{-1}(\theta_1,\theta_2,\theta_3) = \begin{bmatrix} \frac{\cos(\theta_1+\theta_2)}{l_1\sin(\theta_2)} & \frac{\sin(\theta_1+\theta_2)}{l_1\sin(\theta_2)} & \frac{l_3\sin(\theta_3)}{l_1\sin(\theta_2)} \\ -\frac{l_1\cos(\theta_1)+l_2\cos(\theta_1+\theta_2)}{l_1l_2\sin(\theta_2)} & -\frac{l_1\sin(\theta_1)+l_2\sin(\theta_1+\theta_2)}{l_1l_2\sin(\theta_2)} & -\frac{l_3(l_1\sin(\theta_2+\theta_3)+l_2\sin(\theta_3))}{l_1l_2\sin(\theta_2)} \\ \frac{\cos(\theta_1)}{l_2\sin(\theta_2)} & \frac{\sin(\theta_1)}{l_2\sin(\theta_2)} & \frac{l_2+\frac{l_3\sin(\theta_3)}{\tan(\theta_2)}+l_3\cos(\theta_3)}{l_2} \end{bmatrix} \tag{4}$$

5.2 Control

A cascade controller is designed for the robot fingers, the inner (fast) loop is velocity controller separately for each motor. The outer (slow) loop is a PC based admittance controller which provides the reference signal for the inner loop, the high level controller is responsible for the compliant behavior.

The deviation $\Delta\mathbf{x}_{des}$ from the reference trajectory $\mathbf{x}_{des}$ can be obtained by the admittance law:

$$\Delta\mathbf{x}_{des} = \mathbf{K}^{-1}(\mathbf{F}_{ext} - \mathbf{B}\dot{\mathbf{x}}), \tag{5}$$

this term causes the compliant behavior of the end effector. The $\mathbf{K}$ and $\mathbf{B}$ matrices are the corresponding stiffness and damping matrices, respectively; the compliance can be tuned through these parameters.

The end effector velocity is calculated with the Jacobian $\mathbf{J}(\theta_0)$ as $\dot{\mathbf{x}} = \mathbf{J}(\theta_0)\dot{\theta}$. The external force is obtained from the measured force signals $\mathbf{V}$ by the transformation $\mathbf{F}_{ext}^{meas} = \mathbf{T}(\theta_0)\mathbf{V}$.

The inverse kinematics is approximated by the inverse Jacobian $\mathbf{J}^{-1}(\theta_0)$ as

$$\Delta\theta = \mathbf{J}^{-1}(\theta_0)\Delta\mathbf{x}, \tag{6}$$

where $\Delta\mathbf{x} = \mathbf{x}_{\text{cmd}} - \mathbf{x}_0$ is the difference of command position $\mathbf{x}_{\text{cmd}}$ and current position of the end effector $\mathbf{x}_0$.

Finally, the deviation $\Delta\theta$ and the reference joint angles θ_{cmd} can be expressed as

$$\Delta\theta = \mathbf{J}^{-1}(\theta_0)\mathbf{K}^{-1}\left(\mathbf{T}(\theta_0)\mathbf{V} - \mathbf{B}\mathbf{J}(\theta_0)\dot{\theta}\right), \tag{7}$$

$$\theta_{\text{cmd}} = \theta_0 + \Delta\theta\,, \tag{8}$$

the reference joint angles are used in the inner control loop.

The equivalent torsional stiffness matrix with respect to the measured forces $\mathbf{V}$ can be defined as:

$$\mathbf{K}_{\text{eqv}}^{-1} = \mathbf{J}^{-1}(\theta_0)\mathbf{K}^{-1}\mathbf{T}(\theta_0)\,, \tag{9}$$

This may be used in the inner control loop to improve the performance of the overall system with doing more calculations in the fast inner control loop.

5.3 Simulation

Beside the kinematic model, the dynamical model was also developed in order to test the performance of the system via numerical simulations. The simulations were carried out by using the Python programming language with the scientific libraries NumPy, SciPy, SymPy and Matplotlib. The setup was the following, at $t = 0s$ start the simulation at a realistic position of the robot, at $t = 1$s the end effector is pushed by an $F = 10$N force either in the x or in the y direction until the end of the simulation at $t = 5$s.

Figure 10 shows the initial configuration of the robot at the beginning of the simulation and at the final configuration at the end of the simulation. Figures 11 and 12 contain the joint angles and end point coordinates during the simulation, respectively. Minor error can be seen in the end point coordinates, when the external force is in the y direction there is a minimal change in the x direction as well. This is related to two reason: one is the approximated inverse kinematics and the relatively large position change, the other is the change in the static force/moment equilibrium caused by the external force. This small error is fully acceptable in the case of human hand physiotherapy; overall, the system behaves as expected.

6 Conclusion

We developed a hand exoskeleton for robotic post-stroke rehabilitation. In contrast to the different structures that exist, we proposed a unique direct drive solution with integrated force measurement.

Advantages of the proposed exoskeleton:

- Compact, lightweight design

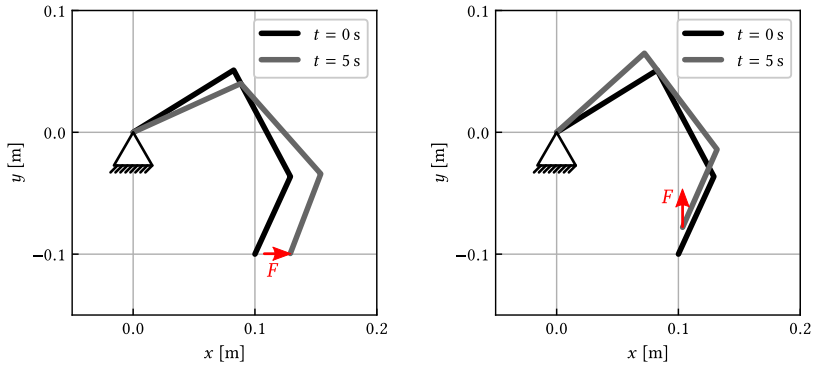


Figure 10
The robot arm configuration at the beginning of the simulation $t = 0$s and at the end of the simulation $t = 5$s; the external force F is in the x direction (on the left) and in the y direction (on the right)

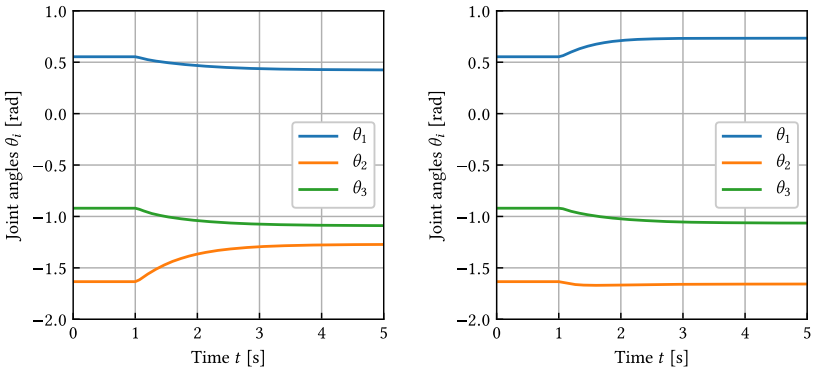


Figure 11
Time evolution of the relative joint angles, the external force F is in the x direction (on the left) and in the y direction (on the right)

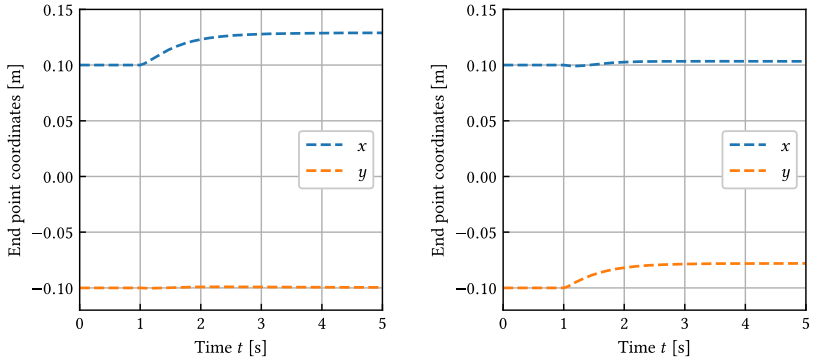


Figure 12
Time evolution of end effector coordinates, the external force F is in the x direction (on the left) and in the y direction (on the right)

- Changeable left/right side (it is possible to perform physiotherapy both for the left and right side, i.e. side change is possible) tegration to lower limb rehabilitation systems
- Wide anatomical size coverage
- High effort
- Built-in force measurement: the inner joint torques are measured directly
- High performance motors

The clinical trials will be realized in the next phase of the research, when the exoskeleton are completed and after the assembly are analyzed experimentally.

Acknowledgment

This work has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the GINOP-2.3.3-15-2016-00032 funding scheme. Péter Galambos thankfully acknowledge the financial support of this work by the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 project.

Also, we would like to express our appreciation to assistant professor Dr. Bálint Magyar for the implementation of the measurements (BME, Faculty of Mechanical Engineering Department of Applied Mechanics).

References

[1] L. Connelly, Y. Jia, M. L. Toro, M. E. Stoykov, R. V. Kenyon, and D. G. Kamper. A pneumatic glove and immersive virtual reality environment for hand rehabilitative training after stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(5):551–559, 2010.

[2] Y. Kim, S. S. Cheng, A. Ecins, C. Fermüller, K. P. Westlake, and J. P. Desai. Towards a robotic hand rehabilitation exoskeleton for stroke therapy. In *ASME 2014 Dynamic Systems and Control Conference*. American Society of Mechanical Engineers Digital Collection, 2014.

[3] H. K. Yap, J. H. Lim, J. C. H. Goh, and C.-H. Yeow. Design of a soft robotic glove for hand rehabilitation of stroke patients with clenched fist deformity using inflatable plastic actuators. *Journal of Medical Devices*, 10(4), 2016.

[4] A. Stilli, A. Cremoni, M. Bianchi, A. Ridolfi, F. Gerii, F. Vannetti, H. A. Wurdemann, B. Allotta, and K. Althoefer. Airexglove—a novel pneumatic exoskeleton glove for adaptive hand rehabilitation in post-stroke patients. In *2018 IEEE International Conference on Soft Robotics (RoboSoft)*, pages 579–584. IEEE, 2018.

[5] B. Wang, A. McDaid, M. Biglari-Abhari, and K. C. Aw. Design and development of a glove for post-stroke hand rehabilitation. In *2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 1047–1051. IEEE, 2017.

[6] M. Cempini, S. M. M. De Rossi, T. Lenzi, M. Cortese, F. Giovacchini, N. Vitiello, and M. C. Carrozza. Kinematics and design of a portable and

wearable exoskeleton for hand rehabilitation. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE, 2013.

[7] C. L. Jones, F. Wang, R. Morrison, N. Sarkar, and D. G. Kamper. Design and development of the cable actuated finger exoskeleton for hand rehabilitation following stroke. *IEEE/ASME Transactions on Mechatronics*, 19(1):131–140, 2012.

[8] T. Worsnopp, M. Peshkin, J. Colgate, and D. Kamper. An actuated finger exoskeleton for hand rehabilitation following stroke. In *2007 IEEE 10th international conference on rehabilitation robotics*, pages 896–901. IEEE, 2007.

[9] A. Chiri, N. Vitiello, F. Giovacchini, S. Roccella, F. Vecchi, and M. C. Carrozza. Mechatronic design and characterization of the index finger module of a hand exoskeleton for post-stroke rehabilitation. *IEEE/ASmE Transactions on mechatronics*, 17(5):884–894, 2011.

[10] M. A. Delph, S. A. Fischer, P. W. Gauthier, C. H. M. Luna, E. A. Clancy, and G. S. Fischer. A soft robotic exomusculature glove with integrated semg sensing for hand rehabilitation. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pages 1–7. IEEE, 2013.

[11] S. Mohamaddan and T. Komeda. Wire-driven mechanism for finger rehabilitation device. In *2010 IEEE International Conference on Mechatronics and Automation*, pages 1015–1018. IEEE, 2010.

[12] C. L. Jones, F. Wang, C. Osswald, X. Kang, N. Sarkar, and D. G. Kamper. Control and kinematic performance analysis of an actuated finger exoskeleton for hand rehabilitation following stroke. In *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 282–287. IEEE, 2010.

[13] Y. Fu, Q. Zhang, F. Zhang, and Z. Gan. Design and development of a hand rehabilitation robot for patient-cooperative therapy following stroke. In *2011 IEEE International Conference on Mechatronics and Automation*, pages 112–117. IEEE, 2011.

[14] F. Wang, M. Shastri, C. L. Jones, V. Gupta, C. Osswald, X. Kang, D. G. Kamper, and N. Sarkar. Design and control of an actuated thumb exoskeleton for hand rehabilitation following stroke. In *2011 IEEE International Conference on Robotics and Automation*, pages 3688–3693. IEEE, 2011.

[15] C. L. Jones, F. Wang, R. Morrison, N. Sarkar, and D. G. Kamper. Design and Development of the Cable Actuated Finger Exoskeleton for Hand Rehabilitation Following Stroke. *IEEE/ASME Transactions on Mechatronics*, 19(1):131–140, February 2014. Conference Name: IEEE/ASME Transactions on Mechatronics.

[16] G. Fazekas, M. Horvath, and A. Toth. A novel robot training system designed to supplement upper limb physiotherapy of patients with spastic hemiparesis. *International Journal of Rehabilitation Research*, 29(3):251–254, 2006.

[17] G. Fazekas, M. Horvath, T. Troznai, and A. Toth. Robot-mediated upper limb physiotherapy for patients with spastic hemiparesis: a preliminary

study. *Journal of rehabilitation medicine*, 39(7):580–582, 2007.

[18] O. Peter, I. Tavaszi, A. Toth, and G. Fazekas. Exercising daily living activities in robot-mediated therapy. *Journal of physical therapy science*, 29(5):854–858, 2017.

Developing Strategies in System Level Model of Smart Cyber Physical System

László Horváth

Óbuda University, Doctoral School of Applied Informatics and Applied Mathematics and John von Neumann Faculty of Informatics, Bécsi u. 96/b, H-1034 Budapest, Hungary, horvath.laszlo@nik.uni-obuda.hu

Abstract: It was long way to current highly integrated and smart modeling and simulation in engineering. This way started with separated problem solving for representation, manufacturing control, and visualization of mathematical surfaces. By now, autonomous engineering model system (EMS) serves lifecycle innovation and engineering of systems operated autonomous intelligent industrial and commercial products. Modeling platforms constitute one of the most important and extensive application areas in information technology. Advanced EMS is highly integrated, multidisciplinary, and multipurposed model structure. Any contribution to EMS must use contextual object modeling (CM) and continuous engineering (CE) which are within the most important achievements in engineering during the past two decades. Considering the above situation, this paper introduces the novel organized content model (OCM) approach, concept, and methodology as contribution to integrated and lifecycle serving model system of highly automated products. OCM organizes driving contexts for lifecycle of modeled product. It is an extension to the currently prevailing EMS structure to contribute fulfilling the new requirement of active context based integration of virtual (EMS) with physically operating cyber physical system (CPS) products. Other essential requirement is that relevant objects in EMS and CPS must be contextually driven by higher level decisions, intellectual property (IP) items, and human intervention attempts while EMS and CPS are increasingly autonomous. Therefore, main purpose of OCM is to support active driving contextual connections for EMS and CPS. In this way, this paper discusses integrated model of highly automated complex product system, extended OCM for situation controlled product system, and Support of Situation Awareness by OCM. Research for OCM is done at the Laboratory of Intelligent Engineering Systems (IESL). In this context, preliminary research at IESL and concept of configuring professional cloud organized modeling platform for the purpose of virtual research laboratory (VRL) are also included in this paper.

Keywords: smart cyber physical system; engineering model system; system level modeling; situation based modeling; organized content model

1 Introduction

This paper introduces latest methodological and communication contributions to system level engineering modeling. Work for these contributions was motivated by more than two decades development of integrated model systems to represent aircraft, car, and other very complex, strongly multidisciplinary, systems operated, and situation controlled industrial product structures in autonomous engineering model system (EMS). The internationally recognized Laboratory of Intelligent Engineering Systems (IESL) joined to research in very extended and comprehensive virtual technology for EMS. The main objective was to achieve integrable principles, methods, and structures.

Due to its fantastic multidisciplinary development including recently emerged new essential modeling issues, the latest engineering technology is a challenge to understand for experts with classical modeling knowledge and experience. Autonomous operation of EMS requires consistent structure of knowledge based driving contextual connections, which includes all object parameters. Modeling related meaning of the word “driving” may be unusual for readers of this paper. Meaning here is setting value of an object parameter by value of related parameters in EMS objects using contextual connection definitions established between the relevant parameters. It must be also noted that word “definition” is used for a human or systems activity which inserts objects and their parameters with actual values in EMS.

Advanced EMS includes generic, self-adaptive, theory-experience integrated, active knowledge driven, systems representation enabled and, autonomous decision capable component models. Accommodating huge amount of new and changed contexts is unavoidable during the integrated innovation and life cycle of a represented product. Variants, developments, and improvements of product must be handled using the same EMS during lifecycle. By now, product structure is often composed by several millions of systems organized components. Shifting industrial engineering to much higher level enforced integrating fundamental, problem solving purposed, and industrial and commercial product related research activities into EMS. Considering so many new terms and new interpretation of old terms, author of this paper apologizes if the reader cannot find explanation for all of them. References in this paper and references in the cited papers include all necessary explanations.

Product model was a main concept in the ISO supported integrated product information model (IPIM) during the 90s. Product model system is called as EMS in this paper. It is consistent representation of generic product system which recently includes cyber units controlled physical units and sensor networks. In this context, term “product system” refers to product, which is modeled as system or system of systems. EMS must be capable of behaving as planned or analyzed product. EMS level model structure was first applied in model based product

lifecycle management (PLM). Structure of a system level EMS requires levels for the extension of classical physical level product model to representation of systems. Driving an EMS is the capability to operate consistent structure of active outside and inside contexts. Situation-controlled product system is developed to achieve its autonomous features. Cyber.units are capable to recognize situation, decide physical actions in accordance with recognized situation then real time control those physical actions. Situation awareness is an essential new capability at systems and intervening humans.

IESL has been active in the organization of Óbuda University for fifteen years. It was established in year 2005 [1] to support university research and graduate higher education in theory, methodology and systemics of engineering modeling. Research in EMS was intended as main area considering integrated application of mathematics, physics, systems engineering, intelligent informatics, and area specific industrial object modeling. Recently, IESL is undergone a laboratory development which aims availability of research enabled systems level engineering modeling platform (EMP) capabilities to introduce experimental model system based research in wide context. In this level, EMP is very complex and provides modeling and other needed organizational capabilities for lifecycle engineering of system operated product using EMS. Availability of suitable engineering modeling platform makes establishing the Virtual Research Laboratory (VRL) at the Doctoral School of Applied Informatics and Applied Mathematics (DSAIAM) possible. Main concept of VRL is experimental model centered research [3] mainly for doctoral research serving Ph.D. projects. VRL is main development of IESL.

This paper introduces results from research in organized content model (OCM) which fills the gap between EMS and driving contexts which connect EMS with world outside of it. OCM includes, among others, representations for decision making, human intervention, knowledge, and former experience. Several former OCM structure variants were proposed among others for driving of systems enabled [4] and cyber physical system (CPS, [36]) connected [5] EMS during the past decade. Main issues in this paper are integrated model of highly automated complex product system, analysis of contexts in OCM, extended OCM for situation controlled product system, and support of situation awareness by OCM. Research for OCM is done at the IESL. Preliminary research at IESL and concept of configuring professional cloud organized EMP for the purpose of virtual research laboratory (VRL) are also included in this paper.

Although the author of this paper tried writing with professional comprehensibility, the results presented here inherently require relatively depth knowledge about latest engineering modeling achievements as well as leading engineering platforms applied at implementation of these achievements in leading practice.

2 Preliminary Research

Research in knowledge driven definition, generation, and application of engineering models to enhance development, production, and application of industrial and commercial products is long-term program at the IESL and its predecessor laboratories from the 90s. This program systematically renewed several times to follow new milestones during the above period. To achieve this, the modeling software system was regularly upgraded to represent the latest theories, methodologies, and systemics. This sector of paper is a selection of former contributions which have effect on current and future research at the IESL [11-32].

In the second half of the 90s, one of the main issues was modeling of process objects in connection with product component features. Generic process features were defined and represented by Petri net in [11] together with built-in knowledge for the evaluation of Petri net. Evaluation procedure defined actual status of Petri net to achieve a process instance. This method was applied at manufacturing process. As contribution to solutions in model integration, a methodology was proposed to attach designer intent knowledge to form feature representation, among others to improve communication between product designer and production engineer [12]. A method was introduced in [13] to enhance integration of process modeling with simultaneous and collaborative engineering methodology. Actual issues in intelligent computing were studied regarding computer assisted engineering activities in [14]. Research results in associative connections for models of mechanical parts were published in [15] and in integration of expert knowledge in behavioral feature enabled product model in [16].

Importance of product behavior and its human intent origin were recognized, and relevant modeling method was proposed in [17]. Book [18] introduced and organized model based problem-solving methods in representative areas of engineering. Paper [19] discussed a new model representation of human intent to achieve capability of product model objects for self-modification in accordance with environmental changes.

When intelligent computing supported decision-assistance had become a main concern in model-based engineering, related issues were analyzed in [20]. A new approach to handle product changes in product model and a method for the related processing in product model assisted engineering were introduced in [21]. Content which represents knowledge for driving the generation of model information is proposed in [22] as one of the main initiatives during the long-term research. Developing strategies in engineering placed new emphasis on model representation of human intent at the beginning of the new century. Joining to this trend, human intent representation was proposed for product model with the emphasis on the intensive knowledge processing capabilities [23]. In this context,

knowledge technology methods were proposed for application in industrial product modeling [24], for integration the aspect of engineering process in virtual product definition [25], and for adaptive knowledge driving at definition of product objects [26].

Recently, research for OCM at the IESL has changed for issues in driving of object parameters in case of model systems serving engineering of highly automated and increasingly autonomous industrial products. This changed situation needed concept for system level OCM. Considering current industrial modeling practice, driving object parameters in requirements, functional, logical, and physical (RFLP) structured EMS was decided [8]. The RFLP structure is IEEE standardized and it is one of the essentials in systems engineering (SE). It integrates SE methodology with the former physical level product modeling. Flexibility and reliability of an OCM installation is highly depends on consistent structure of contextual connections [6]. Anyway, contextual modeling has outstanding importance in recent achievements in engineering modeling [38].

By the second decade of our new century, modeling of situation controlled, and systems operated product became key issue. Hundreds of systems can cooperate among others in cars, aircrafts, robot based assembly systems, and hardware-software systems. Analysis of possibilities to apply active knowledge for situation based driving at definition of product motivated the development of new situation driven definition processes and entities in OCM [27]. Paper [28] introduced the initiative, behavior, context and action structured driving content representation and its application at control of component generation in RFLP structure.

Multilevel abstraction was applied at definition method for self adaptive model and relevant model structure [29] as contribution to higher abstraction based calculation of component parameters in RFLP structure. Organic integration of the proposed model structure into PLM model structure was considered.

Paper [30] discusses the need to improve the formerly proposed initiative, behavior, context, and action content structure, proposes extended and restructured version of this structure, and introduces new approach to driving parameters in contextual product model components. A methodology for intelligent driving of parameters in functional and logical components of RFLP structured EMS in system level engineering is introduced in [31]. Formerly published concept of virtual engineering space (VES) is applied at development of new method for the generation of self-adaptive generic model objects using contextual initiatives for knowledge content background in [32].

3 Integrated Model of Highly Automated Complex Product System

Current industrial practice requires highly integrated autonomous EMS to support development, production, and application activities at engineering of very complex and highly automated products. This EMS is also very complex, and it is developed and applied in product lifecycle organized integrated engineering process. Huge number of product objects and object parameter contexts are defined and represented in highly integrated, generic, and autonomous EMS. Representations, algorithms, decision support procedures, and other advanced computing related entities are applied at definition of consistent context structure in EMS. Beyond contexts between inside object parameters, EMS as product model system (PMS) is driven by outside contexts for human intervention, higher level decisions, former results, and intellectual property. Therefore, relevant EMS object parameters are in contextual connection with outside driving object parameters. Finally, relevant PMS object parameters are in contextual connection with field operating product and production system.

Situation-controlled and systems-organized autonomous product includes extended cyber units which are authorized for real time control of physical units. PMS is continuously changed during lifecycle of product. Conventional sense of product does not apply any more in advanced industry. Contextual connection between PMS and cyber units of CPS has increasing role. For analysis of the above complex system, model of extended product system (EPS) was conceptualized (Fig. 1). In EPS, PMS is the virtual form of product and interacts with physical product for mutual update, upgrade, intelligent cooperation, and intelligent control. Consistency of context structure must be assured in the full EPS during its lifecycle. As it can be seen in Fig 1, EPS receives outside driving contexts in five groups. These groups are participant interventions, higher level decisions, laws and standards, intellectual property, and authorized operator actions. Development of PMS needs application of powerful contextual modeling methodology, which is considered as one of the most important recent achievents in virtual engineering. In the EPS scenario, role of OCM is coordinated diving of PMS and cyber units of CPS with the possibility of bypassing. As it can be seen in Fig. 1, the driving connection is with the situation recognition unit of CPS cyber sector referring to the unavoidable situation recognition and its OCM support in the future. Recent methodology of continuous engineering (CE, [6]) has critical role at lifecycle development of EPS.

Interactions for context accommodation in OCM or/and PMS include procedures to receive, check, and accommodate contexts for EPS. This is one of the two main contributions in this paper (see the next section). The other main contribution is the organized content model (OCM) itself. OCM fills the gap between context accommodation interactions and RFLP structured PMS as well as situation

recognition. As it can be seen in Fig. 1, authorized engineer can decide bypass of OCM at product which does not require representation of complex context structure. The remaining of Fig. 1 shows relevant functional units in cyber units of a CPS product. Situation recognition unit is responsible for situation awareness of highly automated product with autonomous functions. Extended OCM is proposed in section 5 of this paper to assist situation awareness both in approved systems and at authorized humans on duty. Physical activities are decided by decision unit and executed by relevant control in accordance with commands deployed by execution unit.

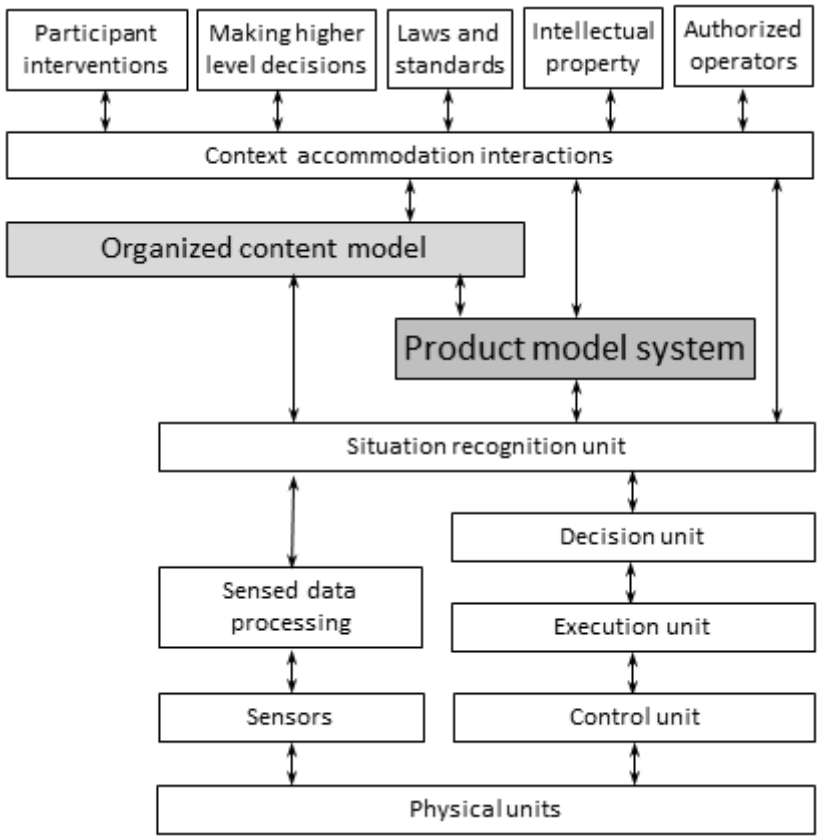


Figure 1
Model of the extended product system (EPS)

In the past, classical product was operated by chains of mechanical connections with power drives at given points. By the development of mechatronics, electronic units were included to control electrical so that indirectly mechanical units. The next milestone was the computer control which step-by-step integrated the formerly individual controls. This trend has led to systems organized product.

In development of a systems-organized product, first the systems are defined on the levels of functions and logical connections. It is obvious that systems operated product is multidisciplinary [37]. In a highly automated product, cyber units control existing physical units and accommodate control of newly announced physical units during the life cycle of product. Sensors carry actual physical operation parameter values in accordance with inquiries from cyber units. Cyber units flexibly reconfigured and modified during the life cycle of product.

RFLP structured PMS is defined for systems operated generic industrial or commercial product or product family and it includes functional (F) and logical (L) structures of product components to establish and virtually execute behavior driven concept model of product [7]. Each component includes appropriate behavior representation to allow virtual execution of the above conceptual product model. This model includes contexts for the generation of physical component models as feature structures. In RFLP structured model, requirements (R) level elements drive functional (F) level components, F level components drive logical (L) level components, finally L level components drive physical (L) level components in a contextual structure.

Beyond improved capability of OCM to assist cyber units of CPS product, these cyber units are capable to carry actual information and even processed experience about physical object and process parameters to OCM and/or RFLP structured PMS. Continuous engineering (CE) processes are capable of real time consequence evaluation at any model or software modification. The above benefits motivated integration of OCM, RFLP structured PMS, and cyber units of CPS products in EPS (Fig. 1). This integration improves optimal as well as secure physical activities by the assistance of CE procedures.

4 Contexts in OCM

One of the main purposes of OCM is to assure consistency of actual active context structure for EPS. This is essential for the capability of continuous development, which is realized in physically operating product. When development of a product requires any change of cyber or physical units, OCM must communicate considering the changed configuration. Recognizing this new style of product development motivated the EPS concept and the related analyses which are reported in this paper below. It must be noted that OCM related key words such as information, content, and contexts are applied with diverse meanings in other technical areas.

Before further characterizing the role of OCM in EPS, some words must be devoted to the content, which is organized and represented in OCM. Representation of systems constituting PMS (EMS) requires higher abstraction in

content for OCM than conventional physical level representation. This recognition was published thirteen years ago in [9] where content was defined as information content acting as background of model information for information level product model. As it was explained above in this paper, content item represents or generates driving context.

Although context item may come from various sources, its origin is always human brain as in case of any information. Connection chain between original source and the context item, which is communicated with EPS generally includes transformations. Russell Ackoff classified human brain content into data, information, knowledge, understanding, and wisdom categories [10]. In this categorization, data is content which does not have any meaning. Information is content which has meaning by relational connections of data. Information answers questions like *who, what, where*, and *when*. In this way, data with contextual connection is considered as information. Information content in OCM is required and suitable to drive relevant target parameters in relevant target object.

However, information does not answer two main OCM item related questions that *how* and *why*. Ackoff proposed knowledge to answer *how* while understanding to answer *why*. In this concept, understanding is in interposition between knowledge and wisdom. Although wisdom is not an issue in this paper, its representation probably will be increasingly important for future autonomous products in the industry so that for future OCM. Considering the current product development trend, high-level automation related concepts will be essential for everyday engineering in the future.

It can be concluded from the above that one of the main purposes of OCM development must be to provide answer the questions *how* and *why*. Driving of PMS and cyber unit object parameters by actual OCM content must be analyzed in this level. It is obvious that OCM item represents active driving knowledge. The next question is how can understanding be represented as OCM content. This question will be answered by future research at the IESL.

Main steps of context item processing in OCM for driving relevant parameters of RFLP structure objects can be followed in Fig. 2. As it can be seen in Fig. 1, OCM receives contexts from the context accommodation interactions. This context is passed to RFLP structured product model object parameter(s) after suitable processing. At conventional human or formal object parameter connection based definition of objects in RFLP structure, this connection is direct. This is denoted in Fig. 2 by dashed line. Definition of RFLP structure objects using many contexts and tracking the changes of these contexts during life cycle of product are practically would become impossible in direct way considering high product complexity as in case of CPS system. Fig. 2 shows a proposed process which can replace manual or other direct definition of contextual objects of RFLP structure.

The context accommodation interactions function passes any eligible context to the space of received driving contexts. Eligibility of a context item to pass into the

OCM is checked using relevant protocol, which is effective in the actual engineering environment. OCM related protocols are not issues in this paper and are subjects of future research at IESL. Definition of contexts for OCM is transformation to make contexts suitable for the definition or extension of relevant OCM substructure element(s). Structure of OCM is explained in the next section of this paper. New context in an OCM element become active and new or modified contextual connection with other related elements is activated in coordination with existing contexts along contextual chains of object parameters. In this way, any new or modified context is coordinated with all related active contexts and actual coordinated contexts drive related object parameters in RFLP structure as driving context.

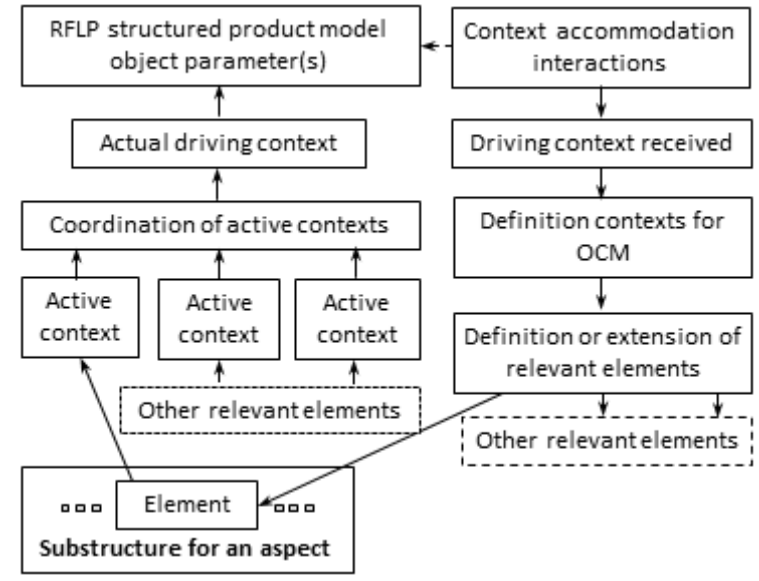


Figure 2
Processing contexts in OCM for RFLP structure

Eligibility check must be included to assure suitability and security of context item received for the accommodation in OCM. Main steps of the proposed eligibility check are explained in Fig. 3. After receiving, driving context is undergone a procedure to recognize the relevant aspects then elements to be originally defined or completed within each aspect in OCM. Aspects are represented in substructures of OCM levels. This step includes checking the new context in accordance with protocol in effective. As a result of the above procedure, any obviously erroneous context is refused. Because new contexts enrich important content during continuous development of EMS, it is essential not to trash any valuable context. In argued case, it is advised to allow online context correction for trusted sources. Status of context is recorded, and

appropriate notice is generated for the owner of the context. It must be recognized that communication with autonomous sources must be increasingly considered in the future.

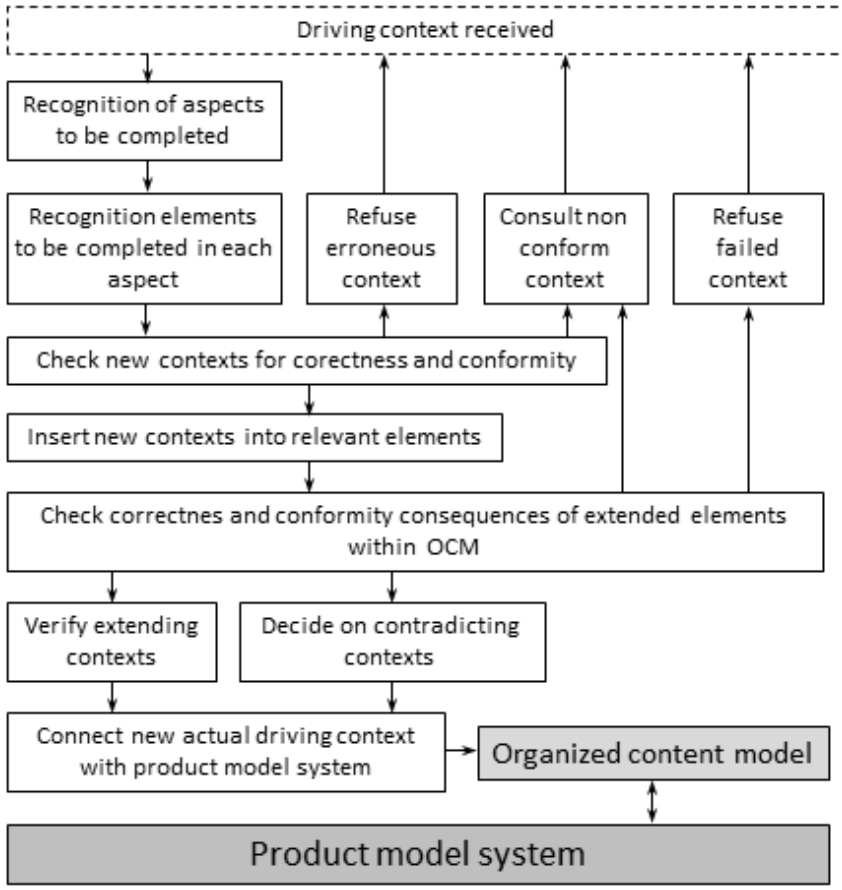


Figure 3
Essential accommodation of contexts in OCM

When it is placed in the appropriate elements within the relevant aspect(s), any new context is checked for correctness and conformity. Correctness is checked using a checklist provided by the effective protocol. Protocol includes a set of firm contexts to be fulfilled. Conformity is more complicated than correctness because it can depend on any existing context with which the checked context is connected or interdependent. Any obviously not correct context is refused with notice sent to its owner. Non conform context is discussed by OCM responsible and owner humans or system. In case of disagreement, checked context is passed to higher level decision making or refused in accordance with the effective protocol. Anyway, discussion is required except for obviously correct and conform cases.

After successful check of correctness and conformity, insert of context in relevant elements of OCM is finalized. The next step is checking correctness and conformity consequences on extended elements within OCM along relevant active context chains. Appropriate extending contexts are verified while additional decisions are required on contradicting contexts. In final decision, previously accepted contexts in context chain are often need modification or even refuse. A finally approved new actual context in OCM is actively connected in EMS with permission for its driving effect.

The above procedure may seem as very complicated together with strict development of hard to survey OCM structure. Really, the recent practice in EPS has become very complex and different from previous practice. OCM which is proposed in the next section of this paper is an attempt to organize so many different contributions from so many different sources into consistent driving structure to facilitate decisions which aware of all essential as well as valuable contributions. This is a contribution to new level of engineering owing to the capabilities of contextual autonomous engineering modeling. OCM is intended as a possible contribution to solve recently typical problems caused by locally correct results which are failed at its integration in the demanded much wider context.

5 Extended OCM for Situation Controlled Product System

A new OCM model structure was conceptualized to be eligible to provide context support for RFLP structured PMS and cyber units of physically operating product. Formally, OCM consists of levels and includes sequence of substructures on each level. Each substructure serves an aspect of content and consists of dynamic structure of port connected blocks. A block represents an element of substructure. Levels and substructures within a level are placed in mandatory contextual chains. Each level serves closely connected group of aspects. Any contextual element can be connected by additional context if the new context does not break any effective context. Context item is represented in OCM and it is placed in one or more element in one or more substructures. Structure of OCM is planned to realize using functionality, which is available for RFLP structure. Life cycle development of OCM is a great challenge for engineers and it is one of the highest-level engineering activities in product engineering.

Publications about former relevant OCM variants which are specially designed for driving of systems enabled [4] and CPS connected [5] PMSs introduce basic theory, methodology, and systemics of OCM. This sector of paper is restricted to the new OCM structure variant which extends the above formerly developed versions to drive PMS which represents situation-controlled product.

A specific feature of the proposed OCM structure is placing new emphasis on situation awareness. Situation awareness is equally important at smart system which applies it at its autonomous activities and at human who is on duty at operating this system. Human on duty should be available to correct a failed autonomous decision at any time. Self-awareness, self-adaptation, and self-optimization were emphasized as main interacting capabilities for situation awareness in [33]. In the concept of [33], self-awareness recognizes situation using behavior, self-adaptation modifies objects acting on situation, and self-optimization improves quality of decisions on adaptive modifications.

The proposed OCM consists of five levels in a contextual chain (Fig. 4). This contextual chain is enforced by procedures at development and application of OCM for saving integrity and consistency of OCM. This also applies to the sequence of substructures within each level. In the proposed OCM, levels are defined to contextually organize aspects for *model construction, product behavior, systems which operate the product, control of automatic equipment and devices, and human-system connections*. Elements in sublevels are in driving contextual connection with sources from which contexts are accommodated for driving object parameters in RFLP structured EMS, for human inspectors who are on duty at operating of product, and for connected cyber units of CPS system. Aspects and related content on levels are characterized below.

Construction aspects organize content for the construction of PMS. While required function and component contexts of product and the related process and procedure contexts are in a contextual sequence, outside contexts also drive elements in all these substructures.

Behavioral aspects have critical position in smart systems because virtual execution on system level of PMS relies upon behavior representation in components within RFLP structured PMS. On the other side, behavior is defined in the context of circumstances for situation. This level is an attempt to coordinated driving by these three groups of objects. It is critical at real time operation of systems within smart product structures. Behaviors are among main areas of future research at IESL.

Level *Systems aspects* is contribution to represent driving contexts for model of systems in RFLP structure. Current RFLP structures inherently represent systems implicitly on their functional and logical levels. Level *Systems aspects* organizes system related driving contexts in substructures for cooperating component systems. Units of a component system are defined for a given system, but they are organized in common substructure to represent their contexts mainly for their interaction. Finally, aspect of system operation is defined for system functions and related operational parameters.

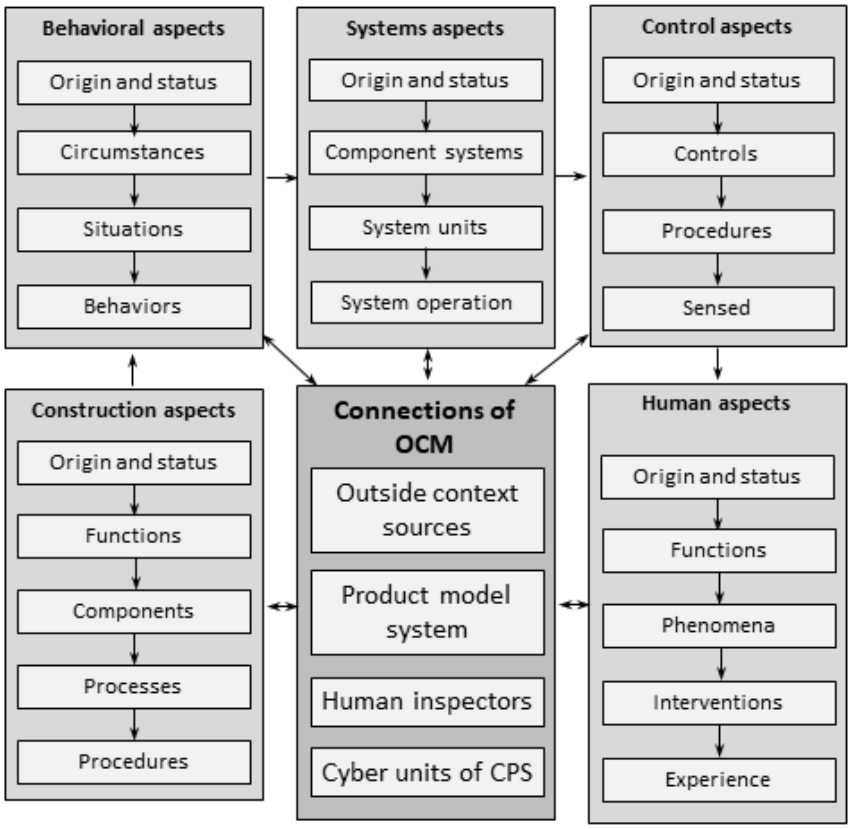


Figure 4
Main structure of the extended OCM

Level *Control aspects* is attempted to organize contexts for control equipment and devices. Activities of these objects are coordinated by cyber units in smart CPS product. Control equipment and associated devices are represented in PMS. While substructure *Controls* organizes driving contexts for control equipment and devices, driving contexts are also organized for procedures in PMS and CPS cyber units, and for sensed parameters which are real time processed in cyber units.

Finally, level *Human aspects* is devoted as contribution to solutions for the problem of human situation awareness with special attention to system fails at situation recognition when human on duty must correct the recognized situation to improve systems operation and prevent malfunction or even danger. Functions of product which are executed according to valid decisions in smart products are structured then phenomena contexts are organized in the context of functions. The next sub-level organizes possible interventions at system related functions. Finally experience organizes content items from past cases in the context of phenomena through relevant interventions.

OCM is model level environment to achieve organized driving content for the other units in EPS. In this way, OCM is integrated to EMS so that in EPS. For that reason, implementation of OCM must be done in appropriately conditioned professional engineering modeling platform as it will be discussed in sector 7 of this paper. As in case of smart engineering [34] and smart CPS products, engineering with autonomous OCM requires participants who are in possession of all knowledge and experience needed by system level, multidisciplinary, large structure related engineering modeling activities.

6 Support of Situation Awareness by OCM

By now, situation awareness is one of main concerns in highly automated situation-controlled product systems especially when autonomous functions are supported. OCM related discussion of situation awareness needs answer for three essential questions. The first one is that how connection between autonomous PMS and CPS cyber sector can be enhanced by appropriately configured autonomous OCM. The second one is that how situation extension of OCM assists development and operation of autonomous product characteristics. The third one is that how to automate the process in Fig. 3 to cope with demands posed by autonomous features of product. Answers to these questions need research which must be extended to whole EPS (Fig. 1). Considering the above questions, OCM support of situation awareness needs suitably configured sets of contextual connections within OCM, between OCM and PMS, and between OCM and cyber units of product.

Three aspect chains with three aspects (substructures) in each, play important role at support of situation awareness by OCM (Fig. 5). Elements in *Circumstances* aspect represent content for correct recognition of situation and can be improved by situation-based experiences in CPS. Actual behavior can be correctly represented in this way. Level *Control aspects* provides equipment and device control drives together with the related procedures to enhance control by cyber units. In the other direction, sensed parameter values contribute to field improved driving content for models and procedures. Level *Human aspects* supports situation recognition by phenomena content, which is the context for required human intervention and, in this chain, enriches content with past proven and failed experience.

Process of situation recognition support by OCM can be followed in Fig. 5. Relevant cyber units provide decision chain for controllable physical units in CPS. This chain includes cyber functions for execution of physical actions and for deploy decided and prepared control on physical units (Fig. 5). In the meantime, sensor information about physical unit parameters is collected and processed.

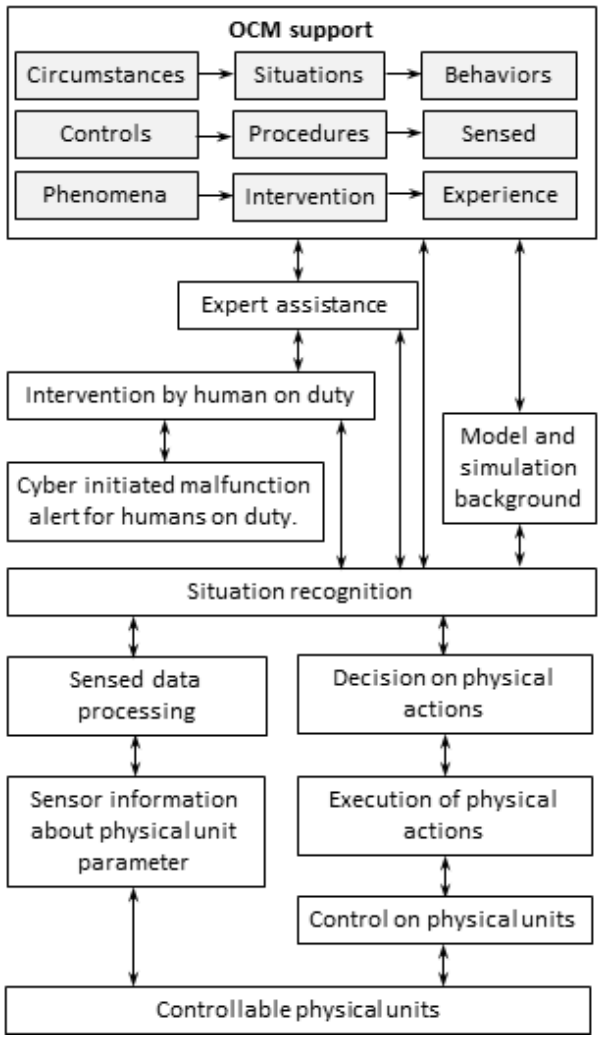


Figure 5
OCM support of situation awareness

In the schema in Fig. 5, responsibility of human on duty is continuous for intervention in case of CPS cyber initiated malfunction alert. This alert can be communicated with OCM directly, or through expert assistance. OCM is capable of expert assistance using content on the level *Human aspects*. Engineer on duty can be assisted by information about former similar cases with of without expert assistance. Any autonomous and human intervention is done through situation recognition. The challenge here is not only recognition and correction of situation but also the enforced immediate intervention. Therefore, the autonomous response of OCM is very important.

7 Implementation Issues

Current high-level model support of engineering activities is the result of dynamic continuous development of the relevant virtual technology during the past two decades. More and more industrial products apply cyber decisions with real time physical execution. Any contribution such as OCM requires firm and reliable knowledge about actual modeling technology and platform at any participant. Availability of professionally proven engineering modeling platform is essential for all participants. Platform must be configured for all modeling and other capabilities which represent the required level. Any restriction of platform capabilities may result conservative and not actual content in OCM and can lead to malfunction even dangerous physical activity in smart CPS system.

DSAIAM decided establishing its virtual research laboratory (VRL) to serve research for OCM and other model centered studies by suitable virtual environment. DSAIAM joined to the 3DEXPERIENCE platform for Academia on the Cloud program of Dassault Systémes. This leading engineering modeling platform provides software capabilities in APPs for modeling (SaaS), means for customizing VRL for Ph.D. and other university research projects in platform (PaaS), and storage, virtual machine, language and other infrastructure (IaaS) services.

The main structure of VRL is introduced in Fig. 6. Platform services include platform management among others to grant roles for participants. VRL specific roles can be surveyed in the first column. Experimental model developer defines model system for research project. Researcher contextually places results and virtually executes them in experimental model system using an experiment plan. Research supervisor can follow research at any time. Industrial adviser brings valuable experience into the research. Modeling adviser gives assistance for researchers in model system related affairs. Finally, research process supervisor manages research project or a group of closely related research projects. Owing to the cloud environment, participants can access the VRL from anywhere and anytime in accordance with the authorization by the granted roles. Any participant has their own workspace.

Thematic sets of modeling capabilities (APPS) are available in modeling, simulation, information intelligence, and social and collaborative groups in the 3DEXPERIENCE. University and research specific functions and research-oriented modeling capabilities make it possible to configure the platform for the demands posed by specific tasks in VRL. Serving university research, VRL highly relies on fundamental, problem related, and experimental configuration related research capabilities in platform. More essential research purposed capabilities are demanded to provide integration with Mathlab, Simulink and Dymola systems and surfaces for Modelica and logic control modeler (LCM) languages in the platform.

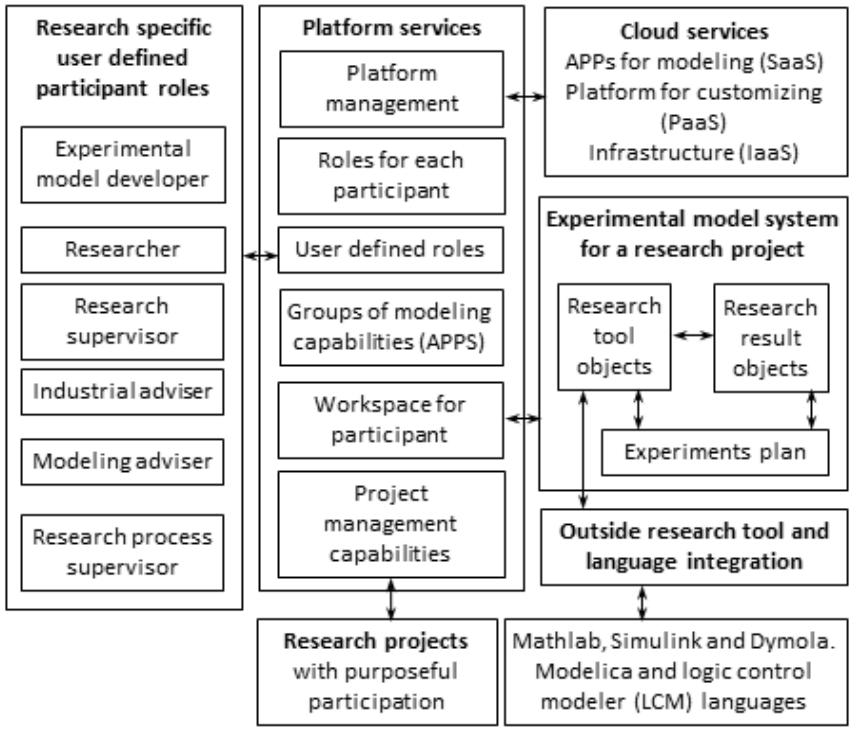


Figure 6
Virtual Research Laboratory (VRL)

As it was stated above, implementation of OCM related and other relevant university and Ph.D. project related research are planned for the VRL. These research programs demand wide area context representations in experimental model. Modeling environment, which is required to do this will be available in the VRL.

The next challenge is to configure experimental model in laboratory circumstances to include details from EPS enough for the experiments when the real world EPS may consist of hundreds of systems realized by a structure consisting of millions of mechanical, electrical, electronic, and software components. VRL is intended to be a flexible virtual research environment which allows wide variety of experimental configurations to develop and simulate. The selected platform provides research tools to integrated handling of theories, methodologies, practices, and experiences. This is a worldwide trend in most advanced model based engineering.

Conclusions

High level automation together with system structure, high complexity, and relying upon several disciplinaries as mathematics, physics, intelligent computing

and systemics brought new era and new challenges in research for engineering. Latest integrated autonomous generic modeling and simulation have become essential at development and operation of advanced industrial and commercial products. In this scenario, product lifecycle developed consistent structure of driving contexts is essential glue in extended product system (EPS). Always actual structure of wide area contexts is firm requirement to assure renewable virtual and field operating smart CPS product.

Recognizing the above change in engineering, this paper introduces results of research in extended organized content model (OCM). OCM is suitable for representation of content behind driving contexts for an EPS in which driving contexts of object parameters establish active connection of OCM with PMS and cyber units of field operating smart CPS. Beyond role of filling the gap between outside contexts and EPS, the proposed OCM includes substructures to organize driving content for situation recognition related aspects. OCM is developed for lifecycle making continuous engineering (CE) possible with actual content and availability of content preliminaries. It must be emphasized that OCM needs continuous exceptional level work for life cycle of EPS. OCM is a specific organized storage of context positioned knowledge, which is used at EPS related engineering.

Experimental model centered VRL is under installation at the DSAIAM as virtual laboratory for realization research among other in OCM using Dassult Systémes's 3DEXPERIENCE engineering modeling platform on the cloud. DSAIAM intends to prepare itself to the future in this way.

Acknowledgement

The author gratefully acknowledges the financial support by the Óbuda University and the Doctoral School of Applied Informatics and Applied Mathematics.

References

[1] L. Horváth: In the Main Stream of Emerging Engineering, in proc of the 16th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 2015, pp. 11-20

[2] V. Gomez-Jauregui, F. Cue-Palencia, C. Manchado, C. Otero: Education for the Industry of the Future (IoF) with the 3D Experience Platform, in book Advances on Mechanics, Design Engineering and Manufacturing II, Springer, 2019, pp. 761-769

[3] L. Horváth: Virtual Research Laboratory for Smart Engineering in the Cloud, in proc. of the IEEE 13th International Symposium on Applied Computational Intelligence and Informatics, Timişoara, Romania, 2019, pp. 179-184

[4] L. Horváth: New Method for Definition of Organized Driving Chains in Industrial Product Model, in Proc. of the 2017 IEEE International

Conference on Industrial Technology, Toronto, Canada, 2017, pp. 1183-1188

[5] L. Horváth: Intellectual Content Driving for Model of Industrial Cyber Physical System, in Proc. of the 2019 IEEE International Conference on Systems, Man and Cybernetics, Bari, Italy, 2019, pp. 1394-1399

[6] B. Fitzgerald, K-J. Stol: Continuous software engineering: A roadmap and agenda, Journal of Systems and Software, Vol. 123, 2017, pp. 176-189

[7] K. Baughey: Functional and Logical Structures: A Systems Engineering Approach," SAE 2011 World Congress, SAE Technical Paper 2011-01-0517, 2011

[8] S. Kleiner, C. Kramer: Model Based Design with Systems Engineering Based on RFLP Using V6, in Smart Product Engineering, Springer, 2013, pp. 93-102

[9] L. Horváth: Supporting Lifecycle Management of Product Data by Organized Descriptions and Behavior Definitions of Engineering Objects, In Journal of Advanced Computational and Intelligent Informatics, Tokyo, Vol. 11, No. 9, 2007, pp. 1-7

[10] R. L. Ackoff: From Data to Wisdom, in Journal of Applied Systems Analysis, Vol. 16, No. 1, 1989, pp. 3-9

[11] I. Rudas, L. Horváth: Modeling of Manufacturing Processes Using Petri Net Representation, Engineering Applications of Artificial Intelligence, Vol. 10, No. 3, 1997, pp. 243-255

[12] L. Horváth, I. Rudas: Intelligent Computer Methods for Modeling of Manufacturing Processes and Human Intent, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 2, No. 3, 1998, pp. 111-119

[13] L. Horváth, I. Rudas: Modeling of Manufacturing Processes in Simultaneous Engineering Using Collaborative Methods and Tools, In book Simultaneous Engineering Methodologies and Applications, Gordon and Breach Publisher, 1999, pp. 321-357

[14] L. Horváth, I. Rudas: Emerging Intelligent Technologies in Computer Aided Engineering, Journal of Advanced Computational Intelligence and Intelligent Informatics Vol. 4, No. 4, 2000, pp. 268-278

[15] L. Horváth, I. Rudas: Virtual technology based associative integration of modeling of mechanical parts, Journal of Advanced Computational Intelligence and Intelligent Informatics Vol. 5, No. 5, 2001, pp. 269-278

[16] L. Horváth, I. Rudas: Integrating expert knowledge in behavioral feature based product models, in Proc. of the 28th Annual Conference of the IEEE Industrial Electronics Society, Sevilla, Spain, 2002, pp. 2479-2484

[17] L. Horváth, I. Rudas: Modeling Behavior of Engineering Objects Using Design Intent Model, in Proc. of the 28th Annual Conference of the IEEE Industrial Electronics Society, Roanoke, VA, USA, 2003, pp. 872-876

[18] L. Horváth, I. Rudas: Modeling and Problem Solving Techniques for Engineers, Elsevier Academic Press, 2004

[19] L. Horváth, I. Rudas: Human Intent Description in Environment Adaptive Product Model Objects, Journal of Advanced Computational Intelligence and Intelligent Informatics Vol. 9, No. 4, 2005, pp. 415-422

[20] L. Horváth, I. Rudas: Intelligence for Assistance of Engineering Decisions in Modeling Systems, Studies in Informatics and Control, Vol. 15, No. 3, 2006, pp. 297-306

[21] L. Horváth, I. Rudas: An approach to processing product changes during product model based engineering, IEEE International Conference on System of Systems Engineering, San Antonio, TX, USA, 2007, Paper No.: 4304267

[22] L. Horváth, I. Rudas: Towards the information content driven product model, IEEE International Conference on System of Systems Engineering, Monterey, CA, USA, 2008, Paper No.: 4724183

[23] L. Horváth, I. Rudas: Human Intent Representation in Knowledge Intensive Product Model, Journal of Computers, Vol. 4, No. 10, 2009, pp. 954-961

[24] L. Horváth, I. J. Rudas: Knowledge Technology for Product Modeling, in book Knowledge in Context-Few Faces of the Knowledge Society, Wolters Kluwer, 2010, pp. 113-137

[25] L. Horváth, I. J. Rudas: Integrated Engineering Processes in Virtual Product Definition, in proc of the IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, USA, 2011, pp. 2996-3001

[26] L. Horváth, I. J. Rudas: Adaptive Method for Knowledge Driven Definition of Products in Modeling Systems, in proc. of the World Automation Congress, Puerto Vallarta, Mexico, 2012, pp. 1-6

[27] L. Horváth, I. J. Rudas: Active Knowledge for the Situation-driven Control of Product Definition, Acta Polytechnica Hungarica, Vol. 10, No. 2, pp. 217-234 (2013)

[28] L. Horváth, I. J. Rudas: New Method for Generation of RFLP Structure Elements in PLM Model, in book New Trends in Software Methodologies, Tools and Techniques, IOS Press, 2014, pp. 310-324

[29] L. Horváth, I. J. Rudas: Product Behavior Definition for Element Generation in RFLP Structure, in book New Trends on System Sciences and Engineering, IOS Press, 2015, pp. 485-498

[30] L. Horváth, I. J. Rudas: New Approach to Multidisciplinary Content Driving of Engineering Model System Component Generation, in book New Trends in Software Methodologies, Tools and Techniques, IOS Press, 2016, pp. 38-49

[31] L. Horváth, I. J. Rudas: Driving Engineering Model Generation on Functional and Logical Levels, in book New Trends in Software Methodologies, Tools and Techniques, IOS Press, 2017, pp. 678-690

[32] L. Horváth, I. J. Rudas: Information Content Driven Model for Virtual Engineering Space, Acta Polytechnica Hungarica, Vol. 15, No. 2, 2018, pp. 7-32

[33] L. Horváth: Smart Engineering Modeling for Smart Industrial Products, Acta Polytechnica Hungarica, Vol. 16, No. 10, 2019, pp. 11-30

[34] M. Abramovici, J. C. Göbel, M. Neges: Smart Engineering as Enabler for the 4th Industrial Revolution, in book Integrated Systems: Innovations and Applications, Springer, 2015, pp. 163-170

[35] M. R. Endsley: Situation Awareness in Future Autonomous Vehicles: Beware of the Unexpected, in Proc. of the 20th Congress of the International Ergonomics Association, Florence, Italy, 2018, pp 303-309

[36] P. Leitao, A. W. Colombo, S. Karnouskos: Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges, in Computers in Industry, Vol. 81, 2016, pp. 11-25

[37] J Lefèvre, S Charles, M Bosch-Mauchand, B Eynard, É Padiolleau: Multidisciplinary modelling and simulation for mechatronic design, Journal of Design Research, Vol. 12, No. 1-2, 2014, pp. 127-144

[38] A Canedo, E Schwarzenbach, E. M A Al Faruque: Context-sensitive synthesis of executable functional models of cyber-physical systems, In Proc. of the 2013 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS), Philadelphia, PA, USA, 2013, pp. 99-108

Mass-Measurement-based Automatization of the Engler-Viscometer

Katalin Harangus, András Kakucs

Sapientia Hungarian University of Transylvania, Faculty of Technical and Human Sciences, Op. 9, Cp. 4, 540485 Târgu-Mureș, Romania
e-mail: katalin@ms.sapientia.ro, kakucs2@ms.sapientia.ro

Abstract: The Engler-viscometer, as one of the efflux-type viscometers, in general, was introduced and standardized, more than 100 years ago. The principle of the measurements is based on the efflux time, of a certain volume of liquid trough a short capillary tube, placed on the bottom of the cup of the viscometer. The viscosity of the liquid is given in Engler-degrees as the ratio of the efflux time of the measured liquid and that of the distillated water at 20 ºC. There is no direct proportionality between the efflux time and viscosity because the Reynolds-number is not the same. This usually is transformed in "true" viscosity units using some semi-empirical formulas. The technical literature of these types of viscometers, especially of the Engler-viscometer, is very poor, but in these papers and books, certain doubts are often mentioned, concerning the accuracy of these formulas. One of the sources of the error is volume measurement, especially for the less-viscous liquids, for example, water. We have proposed a new, mass-measurement based approach that does not need the modification of the viscometer and it also eliminates the negative effect of the thermal expansion of the liquid. This makes possible, a fully automatized measurement and because it provides the evolution in time, of the collected quantity, it also helps in the deeper investigation of the efflux time - viscosity relation.

Keywords: Engler-viscometer; modernization; mass-measurement

1 Efflux-Type Viscometers

The Engler viscometer is a "capillary" type, categorized as "efflux type", mainly used in determination of the viscosity of petroleum products. Based on some earlier viscometers, it was developed by the German chemist, Carl Oswald Victor Engler (1842-1925) around 1890, following his studies in petrochemistry, in 1884. He published the description of this apparatus in "Normen für das Viscosimeter" (Standards of the viscometer) in "Zeitschrift für Angewandte Chemie" (Journal of Applied Chemistry) in 1892. This viscometer became the standardized one, widely used in the world, especially in the continental European Countries.

A few years earlier, in 1885-1886, sir Thomas Boverton Redwood (1846-1919), focusing on some technical problems regarding petroleum products, had improved an older efflux type viscometer in use. This improved version became a standard laboratory apparatus in England and in many other countries. For very viscous fluids, this type of viscometer was modified leading to the "Redwood No. 2": The main differences entail the increase of the diameter of the evacuating tube (that of the "jet") and in the modification of the dimensions of the reservoir (those of the "cup") in order to reduce the pouring time of the target fluid.

In the United States of America, a third efflux-type apparatus was standardized, the Saybolt Universal viscometer that also has two versions with enlarged diameter of the jet for more viscous fluids, the Saybolt Furol and Saybolt Asphalt viscometers.

All of these viscometers work, based on the same principle: The efflux time of a certain standardized volume of liquid, is directly related to the viscosity of that liquid, so the viscosity is computed or appreciated, based on the measured efflux time.

They work in a similar manner, only some of the construction details are different [4].

They consist of a cup (vertical cylindrical reservoir, denoted "*A*" in case of the Engler viscometer in figure 1) with standardized sizes, opened on the top and closed by a standardized shape bottom, covered by a thermo-insulated lid (*B*). The cup of the early Engler and Redwood viscometers was plated inside with gold or silver, nowadays they are made of cheaper corrosion-resistant stainless steel. The cup of the Saybolt viscometers is made of brass or other corrosion-resistant metal. This cup is placed in a thermostatic chamber filled with water or oil (*C*) and provided with a thermometer (*D*) and a stirrer (*E*) to obtain a uniform temperature in the thermostatic bath. The fluid usually is heated electrically, in the case of some versions, it can also be cooled.

The internal diameter of the Engler viscometer cup is $a = 106\,\text{mm}$ and the height of its cylindrical part is $b = 25\,\text{mm}$. The volume of this cylindrical part is $220.5\,\text{cm}^3$ that is 10% larger than the volume to be collected during the measurements (this standardized volume is $200\,\text{cm}^3$).

On the bottom of the cup there is a central hole where a short vertical capillary (the jet, *F*) is fitted. This also has some standardized dimensions and it is made of non-corrosive material. The jet of the early Engler viscometers was made of platinum or brass covered inside with thick platinum coating. Nominally, its length is 20 mm and its internal diameter is 2.8 mm at the lower end. On the upper end its diameter is 0.1 mm larger (2.9 mm). Its external diameter is 4.5 mm. The lower end is $c = 52\,\text{mm}$ under the maximum level in the cup and $d = 3\,\text{mm}$ under the bottom of the thermostatic bath [3].

The traditional material of the jet of the Redwood viscometer is agate. The Saybolt viscometers are made entirely of corrosion-resistant metal.

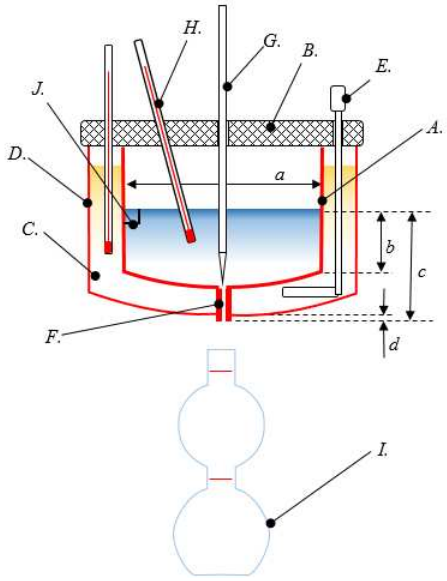


Figure 1
Schematics of the Engler viscometer

The jet in the case of the Engler viscometer is closed in its upper end, inside the cup, by a valve rod (*G*) that can be vertically lifted through the lid.

The jet of the Redwood viscometer is controlled by a small metallic ball resting in a hemispherical cavity of the agate, on its superior end. This can be lifted using the attached wire.

So given, at the beginning of the measurements, the capillary tube in case of these viscometers is empty.

The Saybolt viscometer has a cork stopper placed under the jet that can be pulled out downward using a small chain or a rope. This stopper has to be fitted as well as it prevents escaping the air; this is evidenced by the absence of the measured liquid on the stopper when it is removed at the beginning of the measurement but it does not guarantee that the jet is empty.

All viscometers are provided with a thermometer placed inside the cup (*H*), a three-legged stand with leveling screws (not shown in Figure 1) and a measuring flask (*I*).

The cup has to be filled with a certain volume of liquid. In the case of the Engler viscometer the corresponding level is indicated by three pointers (*J*) used also in leveling, while the Redwood viscometer has only a single gauge. This pointer is a

piece of sharpened nail (wire) bent upwards, fixed on the inner surface of the cup. The volume in the Saybolt viscometer is controlled by an overflow-rim.

The measuring flask of the Engler viscometer is a two-bulb glassware with a short narrow neck between the bulbs and a longer one at its mouth. Between the two bulbs there is a mark corresponding to 100 cm^3 of volume. Above the upper bulb on the neck there is another mark indicating the 200 cm^3 of the collected liquid poured out from the cup.

The flask of the Redwood viscometer is a 50 cm^3 volumetric one. The Saybolt viscometer has a flask of 60 cm^3 capacity that has a funnel-like mouth.

The standard testing method begins with preparing the sample (e.g. by filtering if this is necessary), cleaning and drying the cup, the jet and the flask of the apparatus. The thermostatic bath must be set up by filling with oil or water. The volumetric flask has to be placed beneath the jet. The cup has to be filled to the level indicated by the pointer(s) or set by the overflow rim, while the jet is closed by the valve rod, the metallic ball or the cork, respectively. The thermostatic bath has to be heated or cooled applying some thermal control and stirring to reach and keep the desired temperature in an allowable range. When the temperature is set, the jet is opened and a chronometer is started in the same time. This chronometer is stopped when the level of the effluent liquid reaches the mark on the measuring flask.

In the case of the Engler viscometer, the viscosity is given as the ratio of the t efflux time of the measured liquid and that t_0 of the distillated water at 20 ºC. This is expressed in Engler degrees (ºE) and usually it is transformed in "true" viscosity units (e.g. poise or m^2/s for cinematic viscosity) using some experimental tables, charts or relations like:

$$\nu_k = A \cdot E - \frac{B}{E}, \; E = t / t_0 \tag{1}$$

Are obtained by calibration using liquids with known viscosity (Ubbelohde's formula).

In the case of Redwood viscometer, the efflux time is given in "Redwood seconds". The efflux time of the Saybolt viscometer is corrected with an experimental calibration factor obtained using some standard oils, obtaining the measure of the cinematic viscosity in "Saybolt X seconds" or "Saybolt X viscosity" units, where "X" stands for "Universal", "Furol" or "Asphalt" depending on the apparatus used. The "Redwood seconds" and "Saybolt seconds" are transformed in kinematic viscosity by the formula of above as:

$$\nu_k = A \cdot t - \frac{B}{t} \tag{2}$$

2 Disadvantages of Efflux-Type Viscometers

There are often enumerated, some technical issues, as disadvantages of these viscometers, since the difficulty of cleaning the capillary tube and so on, but studying how these are working, some other issues can be enumerated.

A century ago (in 1921), Winslow H. Herschel has written a short monography entitled "The Redwood viscometer", published by the U.S. Government Printing Office in 1922 (also republished by the Forgotten Books in 2019) [1], concerning the standardization of this type of viscometers. He paid attention to the "cooling error" that is the negative effect of the changes in temperature of the effluent liquid. The principle of these viscometers is based on the measuring the efflux time of a certain volume of fluid on certain set temperature, but the volume is measured using a collecting flask that is non thermostatic as the cup is. At the end of the measurement, the difference between the nominal temperature of the test and that of the liquid in the flask depends upon a series of factors. As the difference between the testing temperature increases and the viscosity is higher, the error will be higher due to the shrinking of the liquid and to the dilation of the collecting flask. This error in the mentioned monography is up to 5% in case of oils tested at 100 °C.

To reduce this error, some solutions were proposed, such as, using a double walled glassware as a collecting flask or providing it with a thick thermo-insulator cover.

Another disadvantage of these viscometers, is the reading error of the level in the collecting flask. The chronometer must be stopped when the effluent liquid reaches the mark that is appreciated by human eye reading. The impediments are the meniscus, the foam, the presence of air bubbles and those of the liquid drops on the wall in case of rapidly pouring liquids with reduced viscosity (this is the case of the water used in calibration of the Engler viscometer, Figure 2).

Figure 2
Bubbles and foam in the collecting flask

The human reaction time in starting-stopping the chronometer is also a source of the reading errors: it is said that normally this is 0.7 s that is more than 1% of the cca. 50 s efflux time of the water in the calibration process of Engler viscometer.

The main disadvantage of efflux viscometers, is found in the difficulty of calibration. The concept of the hydraulic similitude cannot be applied in their calibration because the outflowing of two different liquids happens with different Reynold's numbers, so the viscosity cannot be computed using scale ratios.

3 A New Method in Measuring the Efflux Time

Because of the difficulties generated in the detection of the exact volume of the liquid collected in the flask, we have proposed a new method based on mass (weight) measurement, without any modification of the viscometer. This is done by using a scale placed under the collecting flask. This scale is built with a load-cell, having four strain-gauges, wired as a full Wheatstone-bridge. This is connected to a specialized load-cell amplifier, that has an analog-to-digital converter circuit (e.g. HX711). The digital output of this circuit is read and saved in a file by a data acquisition and controlling system built around a microcontroller (e.g. ATmega2560 or ESP32). To avoid the acquisition of useless data a micro-switch ought to be placed on the viscometer to detect the position of the valve rod: as the rod is lifted the data acquisition starts and this is stopped when the valve is pushed back into its closed position. Using the same microcontroller and some circuitry, there is also the possibility to resolve the controlling of the thermostatic bath.

In this way the out flown volume of the liquid can be computed as:

$$V(t) = \frac{m}{\rho(t)} \tag{3}$$

Where, m is the acquired mass and $\rho(t)$ is the density of the liquid in the cup (at the set temperature t). When t is not equal to the room temperature the volume collected in the flask is not equal to this computed value. In this way the problems regarding the reading of the level and the "cooling error" mentioned before are resolved but the $\rho(t)$ density must be known.

The viscosity is computed from the time t corresponding to $V(t) = 200\ \mathrm{cm}^3$. Because the mass is acquired with a certain rate, $V(\tau)$ is obtained by interpolation based on the registered data.

4 Pros and Cons of the New Method

The main advantages of the proposed mass-measurement based volume detection are:

- It can be applied without any modification of the apparatus
- It can be automatized in a very simple way
- It eliminates the errors related to the thermal expansion of the liquid (the cooling error)
- It eliminates the level-reading errors, including the human reaction time, those produced by bubbles and by the foam
- There is no need of well-cleaned and dried calibrated collecting flask
- It makes possible to track the whole process as the liquid flows out from the cup (this may be useful in calibrating and theoretical investigations)

The main disadvantages are:

- The density of the liquid at the testing temperature must be known (measured)
- The acquired data is influenced by noise (because the collected and measured mass is continuously growing there is not possible to get a mean value from repeated samples as is usually done in case of digital balances)
- The efflux time of the standardized $200\,\mathrm{cm}^3$ is obtained by processing the acquired data
- The mass indicated by load cell is slightly biased by the momentum p of the falling jet or that of the drops
- In case of less viscous liquids (as water at $20\,^{\circ}\mathrm{C}$), the lower part of the jet may be turbulent and this turbulence may cause some noise
- In case of higher viscosity, instead of a continuous jet the liquid falls in individual drops that can also generate some noise as the dynamic excitation of the low-damped scale mechanism

The negative influence of the phenomena mentioned in the last three disadvantages, may be reduced, by using a flat dish-like collecting vessel, as a flat beaker or crystallizer, placed close under the bottom of the viscometer.

5 Efflux Time vs. Viscosity

The process of outflowing is not a stationary one because of the decreasing head as the liquid leaves the cup during the measurements, so the exact analytical computation is not possible because we need to know the streamlines at each moment of the computation. So the first approximation of this computation can be done by using the Bernoulli's principle in its simplest form for stationary flows:

$$\frac{\alpha_1 \cdot v_1^2}{2 \cdot g} + \frac{p_1}{\gamma} + z_1 = \frac{\alpha_2 \cdot v_2^2}{2 \cdot g} + \frac{p_2}{\gamma} + z_2 + h_{r1\text{-}2} \tag{4}$$

In this equation:

– α_i are the Coriolis coefficients adjusting the kinetic energy computed from the mean velocity to that computed by taking the true velocity distribution on the cross section of the flow. In laminar flow in circular tubes the value of this coefficient is computed based on the Hagen-Poiseuille velocity distribution and it is $\alpha = 2$

– v_i are the mean velocities on the two cross sections. If we take the first one on the free surface of the liquid in the cup and the second one on the lower end of the evacuating capillary tube, the ratio of cross sections with $d_1 = 106\,\text{mm}$ of the cup and $d_2 = 2.8\,\text{mm}$ as the diameter of the lower end of the capillary is:

$$\frac{A_1}{A_2} = \left(\frac{106}{2.8}\right)^2 \approx 1433 \tag{5}$$

and the v_1 velocity is less than 0.1% of v_2. So given, the specific kinetic energy on the first cross section is less than one millionth part of that at outflowing and it can be neglected by taking $v_1 = 0$. Let denote the v_2 mean velocity as "v", computed with the volumetric flow rate and the cross section of the capillary as:

$$v = 4 \cdot Q_V / (\pi \cdot d^2) \tag{6}$$

– p_i are the hydrostatic pressure that is the atmospheric one on both sections;

– $h = z_1 - z_2$ is the hydraulic head as the vertical distance from the free surface in the cup to the lower end of the tube. This varies from 52 mm at the beginning of the measurements to 29.3 mm at the end when the 200 cm^3 of the liquid is collected in the flask;

– $h_{r\,1-2}$ is the head loss. Because inside the cup the velocity is very low and the cross section is relatively large, this loss is produced at the evacuating capillary tube. Using the Darcy-Weissbach equation this is computed related to the kinetic energy as:

$$h_r = \zeta \cdot \frac{v^2}{2 \cdot g} \tag{7}$$

where:

$$\zeta = \zeta_t + \zeta_l \tag{8}$$

is the sum of the flow coefficient of the tubes and those locally.

In our case there is only one local source of the energy losses that is the contraction at the entrance in the capillary, which has a flow coefficient of: $\zeta_l = 0.5$

The flow coefficient for circular tube with the length l and diameter d is expressed using the λ friction factor as:

$$\zeta_t = \lambda \cdot \frac{l}{d} \tag{9}$$

Based on the Hagen-Poiseuille velocity distribution the friction factor in laminar flow for circular tubes is:

$$\lambda = \frac{64}{\mathrm{Re}} \tag{10}$$

where the expression of the Reynold's number is:

$$\mathrm{Re} = \frac{v \cdot d}{\nu_k} \tag{11}$$

By taking an average diameter $d = 2.85$ mm of the capillary and $\mathrm{Re} = 2000$ (this is the approximate value of the Reynold's number at the beginning of testing with water), with its length of $l = 20$ mm the flow coefficient of the evacuating tube is

$$\zeta_t = \frac{64}{\mathrm{Re}} \cdot \frac{l}{d} = 0.228 \tag{12}$$

that is less than that for the contraction at the entrance. For higher viscosity oils the Reynold's number is greater and the head loss of the capillary became predominant over the local loss of above.

With all these relations we obtain the formula of the head loss with:

$$h_r = \frac{32 \cdot l}{g \cdot d^2} \cdot \nu_k \cdot v + \frac{\zeta_l}{2 \cdot g} \cdot v^2 \tag{13}$$

and the Bernoulli's equation takes the form:

$$h = \frac{(\alpha + \zeta_l) \cdot v^2}{2 \cdot g} + \frac{32 \cdot l}{g \cdot d^2} \cdot v \cdot \nu_k \tag{14}$$

or, with the flow rate Q_V

$$h = \frac{8 \cdot (\alpha + \zeta_l) \cdot Q_V^2}{\pi^2 \cdot d^4 \cdot g} + \frac{128 \cdot l}{\pi \cdot g \cdot d^4} \cdot Q_V \cdot \nu_k \tag{15}$$

So the kinematic viscosity can be expressed as:

$$\nu_k = \frac{\pi \cdot g \cdot d^4 \cdot h}{128 \cdot l} \cdot \frac{1}{Q_V} - \frac{\alpha + \zeta_l}{16 \cdot \pi \cdot l} \cdot Q_V = A \cdot \frac{1}{Q_V} - B \cdot Q_V \tag{16}$$

that can be regarded as a detailed form of the Ubbelohde's formula (1).

The main problem with this formula is that nothing guarantees that the flow in the capillary tube follows the Hagen-Poiseuille velocity distribution. It is deduced using the hypothesis of the steady-state flow by neglecting the

$$\int_1^2 \frac{\partial \boldsymbol{v}}{\partial t}\, d\boldsymbol{s} \tag{17}$$

quantity (because of the variable cross section and of the decreasing level in the cup, this integral cannot be computed analytically). This formula gives the value of the kinematic viscosity at a certain moment τ for the actual values of $h = h(\tau)$ and $Q_V = Q_V(\tau)$. Following the standardized process of the measurements these values are unknown, only the efflux time t of the volume V is determined. With this time an average value of the volumetric rate flow can be computed as:

$$\bar{Q}_V = \frac{V}{t} \tag{18}$$

and some attempts were made to establish a mean value $\bar{h}$ of the head. For the Redwood viscometer by using these mean values and a corrected value $\bar{l}$ of the length of the tube it was proposed the

$$\nu_k = \frac{\pi \cdot g \cdot d^4 \cdot \bar{h}}{128 \cdot \bar{l}} \cdot \frac{1}{\bar{Q}_V} - \frac{m}{8 \cdot \pi \cdot \bar{l}} \cdot \bar{Q}_V = A \cdot \frac{1}{Q_V} - B \cdot Q_V \tag{19}$$

formula [1], where B is corresponding to $\alpha + \zeta_l = 2 \cdot m$ by referencing the (17^{th}) formula. Here m is a correction coefficient of the kinetic energy. The corrected value $\bar{l}$ is an increased one of the length of the tube that includes the effect of the contraction at the entrance in the capillary. Some details are given in [2] too. In the case of the Engler viscometer the variable diameter of the tube also has to be taken with an average $\bar{d}$ value.

With the (18) average rate flow this can be transformed into a relation that gives the kinematic viscosity as the function of the efflux time:

$$\nu_k = \frac{\pi \cdot g \cdot d^4 \cdot \bar{h}}{128 \cdot \bar{l} \cdot V} \cdot t - \frac{m \cdot V}{8 \cdot \pi \cdot \bar{l}} \cdot \frac{1}{t} = A \cdot t - \frac{B}{t} \tag{20}$$

Accepting the (14) formula the mean velocity on the cross section of the capillary tube can be computed using a step-by step approximation:

$$v(t_{i+1}) = \frac{-c_1 + \sqrt{c_1^2 + 2 \cdot c_2 \cdot h(t_{i+1})}}{c_2} \tag{21}$$

where:

$$c_1 = \frac{32 \cdot l \cdot \nu_k}{g \cdot d^2}$$

$$c_2 = \frac{m \cdot (\alpha + \zeta_l)}{g} \tag{22}$$

$$h(t_{i+1}) = h(t_i) - \frac{V(t_i)}{A}$$

When we use $l = 20\,\text{mm}$ (the real length of the capillary), $d = 2.85\,\text{mm}$ (averaged diameter), $A = 8.824734 \cdot 10^{-3}\ \text{m}^2$ (the cross section of the cup calculated using its diameter $D = 0.106\,\text{m}$), $h(t = 0) = 0.052\,\text{m}$ (the head of the filled viscometer), $\nu_k = 1.0034 \cdot 10^{-6}\ \text{m}^2/\text{s}$ (water at 20 ºC), $\alpha = 2$ (Hagen-Poiseuille), $\zeta_l = 0.5$ (contraction) and $m = 1$ (as correction factor) the efflux time results $t = 59.3\,\text{s}$, much more than the expected $52\,\text{s}$. With the value of $m = 0.75$ of the correction factor the computed efflux time is $51.8\,\text{s}$, closer to the desired one, leading to the

$$A = 1.072 \cdot 10^{-5}\ \text{m}^2/\text{s}, \quad B = 7.187 \cdot 10^{-6}\ \text{m}^2/\text{s}$$

coefficients of the (1) formula.

The coefficients published by Ubbelohde are quite different [2] [3]:

$$\nu_k = 0.073 \cdot E - \frac{0.0631}{E}\,[\text{centistokes}] = 0.73 \cdot 10^{-5} \cdot E - \frac{6.31 \cdot 10^{-6}}{E}\,[\text{m}^2/\text{s}] \tag{23}$$

referred in [2] as "theoretically derived" ones. If we introduce a correction factor of the coefficient *A* (originally it was applied to get a "corrected length" of the capillary tube) there is a hope to get some results closer to the Ubbelohde's formula but these won't be purely theoretical ones. To investigate the role of the coefficients we rewrite eq. (16) as:

$$\nu_k = A' \cdot h \cdot \frac{1}{Q_V} - B \cdot Q_V \tag{24}$$

with,

$$A' = \frac{\pi \cdot g \cdot d^4}{128 \cdot l} = \frac{A}{h} \tag{25}$$

Accepting that A' and B are time-independent constants, for any $Q_{V,1} = Q_V(t_1)$ and $Q_{V,2} = Q_V(t_2)$ corresponding to the $t_1 \neq t_2$ instants we must have,

$$\nu_k = A' \cdot h_1 \cdot \frac{1}{Q_{V,1}} - B \cdot Q_{V,1} = A' \cdot h_2 \cdot \frac{1}{Q_{V,2}} - B \cdot Q_{V,2} \tag{26}$$

leading to:

$$\frac{A'}{B} = \frac{Q_{V,1} \cdot Q_{V,2} \cdot (Q_{V,1} - Q_{V,2})}{Q_{V,2} \cdot h_1 - Q_{V,1} \cdot h_2} = \text{const.} \tag{27}$$

Using some acquired data we can observe that this ratio is not constant (Figure 3, drawn using smoothed experimental data for water at 20 ºC, where $t_2 = t_1 + 0.1\,\text{s}$), demonstrating the fact that the steady-state version of the Bernoulli's equation cannot be used in exact modelling of the relation between the kinematic viscosity and the efflux time, so Ubbelohde's relation may not fit the actual data. This fact was experimentally observed [5] and a new formula was proposed:

$$\nu_k = \frac{E}{100} \cdot 7.6^{1-1/E^3} \quad \text{[centistokes]} \tag{28}$$

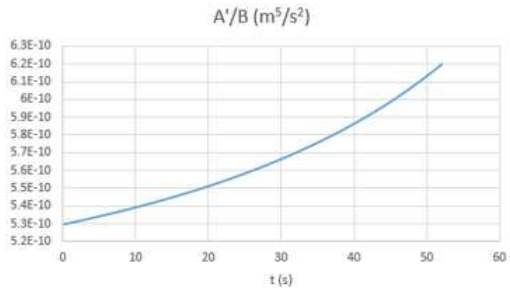


Figure 3
Evolution of the A'/B ratio

Conclusions

Experimental data was obtained for water, 10W40 motor oil and T90 transmission (gear) oil, all tested at room temperature (20 °C). The efflux time was measured using both the "classic" volumetric method and that based on mass measurement.

As primary conclusions we can state the following:

- In the case of water, which has a higher outflow rate, even if there are some errors in the measured mass, the obtained results are closer to each other than those obtained by visual observation of the collected volume.
- In the case of higher viscosity oils the outflow rate is so small that changes can only be seen over a long period of time during the level reading.
- The noise of the electronic apparatus can be reduced by smoothing the acquired data.

The mean efflux time obtained for water was $50.6\,\mathrm{s}$, $2034\,\mathrm{s}$ for motor oil, and $5051\,\mathrm{s}$ in the case of transmission oil. Those lead to a value of $40.2\,°\mathrm{E}$ for the viscosity of the motor oil and that of $99.8\,°\mathrm{E}$ for the transmission oil. For comparison, the standardized minimum of the kinematic viscosity of the tested motor oil at $20\,°\mathrm{C}$ is $2.42\cdot10^{-4}\ \mathrm{m^2/s}$ (its maximum value is not specified by the standards, but the minimum for the 15W40 motor oil is $2.87\cdot10^{-4}\ \mathrm{m^2/s}$). In the case of the transmission oil at this temperature, the range of the kinematic viscosity is given as $(5...14)\cdot10^{-4}\ \mathrm{m^2/s}$ (standardized values are specified only for higher temperatures).

Using Ubbelohde's (23) formula the kinematic viscosity is obtained as $2.93\cdot10^{-4}\ \mathrm{m^2/s}$ for the motor oil and $7.28\cdot10^{-4}\ \mathrm{m^2/s}$ for the transmission oil.

With (28), the obtained viscosities are slightly higher: $3.06\cdot10^{-4}\ \mathrm{m^2/s}$, respectively $7.58\cdot10^{-4}\ \mathrm{m^2/s}$.

The "nominal" density of the water on this temperature is $998.23\,\mathrm{kg/m^3}$. The density of the oils was obtained by measurements: $875\,\mathrm{kg/m^3}$ for the motor oil and $896\,\mathrm{kg/m^3}$ for the transmission oil.

Figures 4 and 5 show the volumetric flow-rate obtained with the $m(t)$ collected mass as:

$$\bar{Q}_V(t) = \frac{m(t)/\rho}{t} \tag{29}$$

The curve obtained for the transmission oil is similar to that of the motor oil; (the outflow time of this oil was much longer).

Eq. (29) is a mean value at the time t introduced based on the observation that for higher values t the relative error in $m(t)$ is smaller. These graphs show an unsteady behavior in the initial phase that can be explained by the neglected component (17) of Bernoulli's equation and by the dynamic excitation of the scale mechanism.

The observed noise of the used apparatus built around the HX711 load cell amplifier is approx. $0.03\,\mathrm{g}$ (as r.m.s. value), with peaks around $0.1\,\mathrm{g}$ (data obtained in steady conditions). The apparatus has the precision and the sensitivity of the technical scales, so it does not need special enclosure as an analytical balance does. Even in the case of a low-density liquid as the gasoline is ($\rho \approx 700\,\mathrm{g/cm^3}$), the relative error in the measured mass of the $200\,\mathrm{cm^3}$ liquid is under 1 ‰ (in steady conditions).

The theoretical value of the impact force of the jet can be computed as $F = \rho \cdot Q_V \cdot v$. One of the worst cases is that of the water at the beginning of the measurements because of the relatively high density and of the highest outflow velocity, when the jet hits the solid bottom of the collecting flask. In this case the estimated magnitude of the impact force is $F = 1000\,\mathrm{kg/m^3} \cdot 4.8 \cdot 10^{-6}\ \mathrm{m^3/s} \cdot 0.75\,\mathrm{m/s} = 0.0036\,\mathrm{N}$. This leads to a computed bias of approx. $0.37\,\mathrm{g}$ in mass measurement and also induces some vibrations of the load cell assembly. As the collected volume in the flask increases the outflow velocity becomes smaller and we can observe a continuously falling amplitude of the vibrations due to the increasing mass and of the viscous damping, and the effect of the falling jet or drops is buried into the general noise of the acquired data.

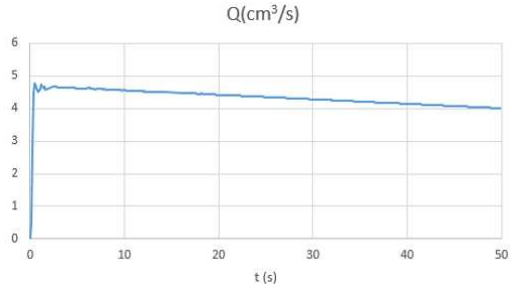


Figure 4
Mean flow-rate in case of 20 ºC water

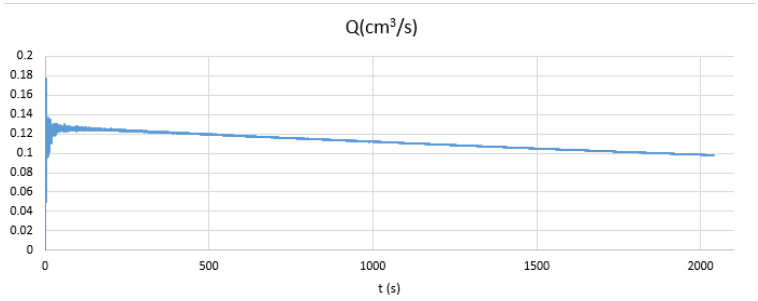


Figure 5
Mean flow-rate in case of a 10W40 motor oil at 20 ºC

To reduce these negative influences we focused on the outflow of the last $100\ \text{cm}^3$ of liquid. It seems that the observed $V(t) = m(t) \cdot \rho$ volume is well approximated by a second order polynomial. The coefficients of this polynomial are smaller for higher values of the efflux time (Figure 6). Using some non-dimensional coordinates

$$t_{rel} = \frac{t - t_0}{T}, \quad v_{rel} = \frac{V(t) - V(t_0)}{V(T)}, \tag{30}$$

the comparison of the observed curves can be done in a much easier way (Fig. 7). In (30) t_0 is the efflux time of the first $100\ \text{cm}^3$ (that is $V(t_0)$) and T is the efflux time of the standardized $200\ \text{cm}^3$ (that is $V(T)$) of the studied liquid.

The slope of the non-dimensional curves is a kind of non-dimensional volumetric flow-rate, as the slope of the curves represented on Figure 6 is the true $Q_V(t)$ function.

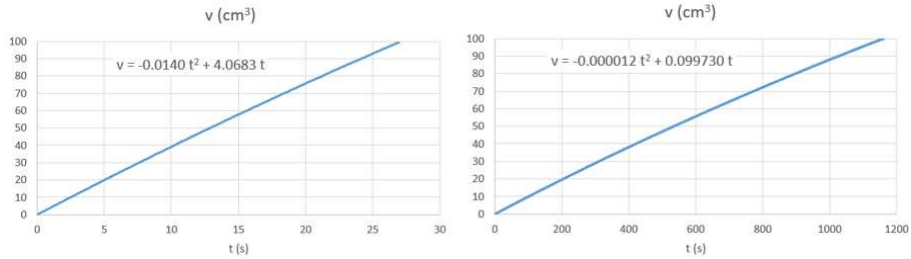


Figure 6
Outflow vs. efflux time of the last 100 cm3 for water and 10W40 motor oil at 20 ºC

Both the true and the non-dimensional volume-time curves (their shape), indicate that the flow rate and the efflux time of a certain volume, are given by some-non-linear functions and there is no proportionality between these quantities and the viscosity of the liquid.

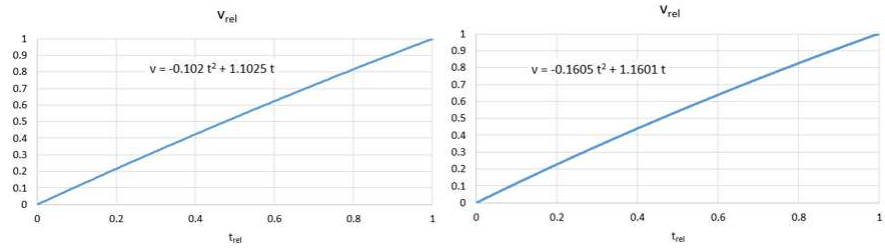


Figure 7
Outflow vs. efflux time of the last 100 cm3 for water and 10W40 motor oil at 20 ºC in non-dimensional coordinates

As a final conclusion, the following can be stated:

- By implementing the mass measurement the efflux-type viscometers can be fully automatized without their structural modification. In this way the measuring errors can be reduced and the viscosity can be obtained with higher accuracy because the human factor (reading errors, reaction time) is completely eliminated. The "cooling error" and those related to the precision of the measured volume using the collecting flask are also eliminated.
- Computing the Engler-viscosity from the efflux-time of the last $100\ \mathrm{cm}^3$ of the liquids is a more consistent method because it reduces the effects of the unsteady flow, especially for less viscous liquids, such as, water.
- A properly accurate relation between the "true" kinematic viscosity and efflux time can be obtained only with a multipoint calibration using liquids with known viscosity. This can be obtained only with a multipoint calibration using liquids with known viscosity.
- Acquired data is the evolution in time of the outflow mass and volume. This makes possible a detailed analysis of the outflow. The physical meaning of the coefficients of the polynomial approximation of the observed $V(t)$ curves, will be investigated in future research work.

References

[1] Winslow H. Herschel: The Redwood viscometer, U.S. Government Printing Office, 1922 (republished by the Forgotten Books, 2019, ISBN 978-1528511605)

[2] Winslow H. Herschel: Determination of Absolute Viscosity by Short-tube Viscosimeters in "Technologic Papers of the Bureau of Standards", U.S. Government Printing Office, 1917 (republished by the Forgotten Books, 2018, ISBN 978-1390929072)

[3] S. V. Gupta: Viscometry for Liquids: Calibration of Viscometers, Springer Series in Materials Science, Vol. 194, 2014, ISBN 978-3-319-04858-1, https://doi.org/10.1007/978-3-319-04858-1

[4] Viswanath, D. S., Ghosh, T., Prasad, D. H. L., Dutt, N. V. K., Rani, K. Y: Viscosity of Liquids. Theory, Estimation, Experiment, and Data, Springer, Dordrecht, 2007, ISBN 978-1-4020-5481-5, https://doi.org/10.1007/978-1-4020-5482-2

[5] L. Schiller: The Engler Viscometer and the Theory of Laminar Flow at the Entrance of a Tube, Journal of Rheology, 3(2), 1932, pp. 212-216, https://doi.org/10.1122/1.2116453

[6] Avia Semiconductor: Datasheet of the HX711 load cell amplifier https://cdn.sparkfun.com/datasheets/Sensors/ForceFlex/hx711_english.pdf

A convergence analysis of the Nelder-Mead simplex method

A. Galántai

Óbuda University
John von Neumann Faculty of Informatics
1034 Budapest, Bécsi u. 96/b, Hungary
galantai.aurel@nik.uni-obuda.hu

Abstract: We give a sufficient condition for a certain type of convergence behavior of the Nelder-Mead simplex method and apply this result to several examples. We also give two related examples for the case of repeated shrinking which indicates a kind of local character of the method.

Keywords: Nelder-Mead simplex method, convergence, counterexamples

1 Introduction

The Nelder-Mead simplex method [12] is a direct search algorithm for the unconstrained minimization problem

$$f(x) \to \min \quad (f : \mathbb{R}^n \to \mathbb{R})$$

using a sequence of simplices and function evaluations of their vertices and some related points. Since its publication, the Nelder-Mead simplex method gained high popularity in various application areas and derivative-free optimization [16], [8], [5], [2], [1]. The Nelder-Mead paper [12] has a citation number over 30000 (google scholar, 30-03-2020). In spite of the great number of related papers and variants of the original method quite a few theoretical results are known on the convergence of the Nelder-Mead method (see [4], [7], [6]). A famous two dimensional example by McKinnon [10] shows that Nelder-Mead simplex algorithm may fail to converge to a stationary point of f, even if f is strictly convex and has continuous derivatives. In this paper we also give a few new examples which provide further insights into the convergence properties of the method. In the next section we define the Nelder-Mead simplex method and summarize the main convergence results. In Section 3 we analyze in detail the decision structure of the method in two dimension. In Section 4 we give a sufficient condition for a certain type of convergence behavior of the Nelder-Mead simplex method. We apply this result to several examples in Section 5. In the last section we give two related examples for the case of repeated shrinking

which indicate a kind of local character of the method. It has been suspected that the choice of the initial simplex may influence the performance of the simplex method (for experimental results, see e.g. [13], [17]). The results presented in this paper clearly support this assumption as well.

2 The Nelder-Mead simplex method

We use the following generally accepted form of the original method [7]. The vertices of the initial simplex are denoted by $x_1, x_2, \ldots, x_{n+1} \in \mathbb{R}^n$. It is assumed that the vertices $x_1, \ldots, x_{n+1}$ are ordered such that

$$f(x_1) \le f(x_2) \le \cdots \le f(x_{n+1}).$$

Define $x_c = \frac{1}{n}\sum_{i=1}^{n} x_i$. The related evaluation points are

$$x_r = (1+\alpha)x_c - \alpha x_{n+1}, \qquad x_e = (1+\alpha\gamma)x_c - \alpha\gamma x_{n+1},$$

$$x_{oc} = (1+\alpha\beta)x_c - \alpha\beta x_{n+1}, \qquad x_{ic} = (1-\beta)x_c + \beta x_{n+1},$$

where $\alpha = 1$, $\beta = 1/2$, $\gamma = 2$. Then one iteration step of the method is the following.

Operation	Nelder-Mead simplex method
1: Ordering	$f(x_1) \le \cdots \le f(x_{n+1})$
Reflect	if $f(x_1) \le f(x_r) < f(x_n)$, then $x_{n+1} \leftarrow x_r$ and goto 1
Expand	if $f(x_r) < f(x_1)$ and $f(x_e) < f(x_r)$, then $x_{n+1} \leftarrow x_e$ and goto 1. If $f(x_e) \ge f(x_r)$, then $x_{n+1} \leftarrow x_r$ and goto 1.
Contract outside	If $f(x_n) \le f(x_r) < f(x_{n+1})$ and $f(x_{oc}) \le f(x_r)$, then $x_{n+1} \leftarrow x_{oc}$ and goto 1.
Contract inside	If $f(x_r) \ge f(x_{n+1})$ and $f(x_{ic}) < f(x_{n+1})$ then $x_{n+1} \leftarrow x_{ic}$ and goto 1.
Shrink	$x_i \leftarrow (x_i + x_1)/2$, $f(x_i)$ (for all i) and goto 1

There are two rules that apply to reindexing after each iteration. If a nonshrink step occurs, then x_{n+1} is discarded and a new point $v \in \{x_r, x_e, x_{oc}, x_{ic}\}$ is accepted. The following cases are possible:

$$f(v) < f(x_1), \quad f(x_1) \le f(v) \le f(x_n), \quad f(v) < f(x_{n+1}).$$

Let

$$j = \begin{cases} 1, & \text{if } f(v) < f(x_1) \\ \max_{2\le \ell \le n+1} \{ f(x_{\ell-1}) \le f(v) \le f(x_\ell) \}, & \text{otherwise} \end{cases}.$$

Hence

$$x_i^{new} = x_i \ (1 \le i \le j-1), \ x_j^{new} = v, \ x_i^{new} = x_{i-1} \ (i = j+1, \ldots, n+1).$$

This type of selection inserts v into the ordering of Step 1 with the highest possible index. If shrinking occurs, then

$$x_1' = x_1, \quad x_i' = (x_i + x_1)/2 \quad (i = 2, \ldots, n+1)$$

plus a reordering takes place. By convention, if $f(x_1') \leq f(x_i')$ $(i = 2, \ldots, n)$, then $x_1^{new} = x_1$.

Here we can also write $x(\lambda) = (1+\lambda)x_c - \lambda x_{n+1}$ and so

$$x_r = x(1) = 2x_c - x_{n+1}, \qquad x_e = x(2) = 3x_c - 2x_{n+1},$$

$$x_{oc} = x\left(\tfrac{1}{2}\right) = \tfrac{3}{2}x_c - \tfrac{1}{2}x_{n+1}, \qquad x_{ic} = x\left(-\tfrac{1}{2}\right) = \tfrac{1}{2}x_c + \tfrac{1}{2}x_{n+1}.$$

Lagarias, Poonen and Wright [6] defined a restricted version of the above method, where expansion steps are not allowed.

Kelly [4], [3] developed a sufficient decrease condition for the average of the object function values (evaluated at the vertices) and proves that if this condition is satisfied during the process, then any accumulation point of the simplices is a critical point of f. For other variants of the Nelder-Mead algorithm, see Tseng [15], Nazareth-Tseng [11], Pryce-Coope-Byatt [14].

Lagarias, Reeds, Wright and Wright [7] proved that if the function f is strictly convex on $\mathbb{R}^2$ with bounded level sets and the initial simplex is non-degenerate, the function values at all simplex vertices converge to the same value. They also proved that the simplex diameters are converging and the simplices in the standard Nelder-Mead algorithm have diameters converging to zero ([7] Theorems 5.1, 5.2).

For the restricted version of the Nelder-Mead method, Lagarias, Poonen, Wright [6] showed also in $\mathbb{R}^2$ that for any non-degenerate starting simplex and any twice-continuously differentiable function with everywhere positive definite Hessian and bounded level sets, the algorithm always converges to the minimizer.

In the light of the above convergence results McKinnon's nonconvergence example is particularly interesting (see, e.g. [6], [18]). McKinnon [10] constructed the function

$$f(x,y) = \begin{cases} \theta\phi\,|x|^\tau + y + y^2, & \text{if } x \leq 0 \\ \theta x^\tau + y + y^2, & \text{if } x \geq 0 \end{cases} \tag{1}$$

where ϕ, θ, τ are positive constants. This f is strictly convex and has continuous first derivatives if $\tau > 1$. For this function, the Nelder-Mead simplex algorithm may fail to converge to a stationary point. In particular, with $\phi = 6$ and $\theta = 60$, the counterexample works for $0 \leq \tau \leq \widehat{\tau}$, and it does not work for $\tau > \widehat{\tau}$, where $\widehat{\tau} \approx 3.0605$.

Wright [18] raises several open questions concerning the Nelder-Mead method such as

- Why is it sometimes so effective (compared to other direct search methods) in obtaining a rapid improvement in f?

- One failure mode is known (McKinnon [10]) – but are there other failure modes?
- Why, despite its apparent simplicity, should the Nelder-Mead method be difficult to analyze mathematically?

Our purpose here is to show other failure modes indicating the complicated convergence structure or behavior of the method.

3 The Nelder-Mead method in two dimensions

Assume that $f:\mathbb{R}^2\to\mathbb{R}$ is continuous and the i^{th} simplex $S^{(i)}=\left[x_1^{(i)},x_2^{(i)},x_3^{(i)}\right]$ is such that $f\left(x_1^{(i)}\right)\le f\left(x_2^{(i)}\right)\le f\left(x_3^{(i)}\right)$ for all $i\ge 0$. The related refection, expansion and contraction points are denoted by $x_r^{(i)}$, $x_e^{(i)}$, $x_{oc}^{(i)}$ and $x_{ic}^{(i)}$, respectively. For the given parameters, $x_c^{(0)}=\frac{1}{2}\left(x_1^{(i)}+x_2^{(i)}\right)$,

$$x_r^{(i)}=2x_c^{(0)}-x_3^{(i)}, \qquad x_e^{(i)}=3x_c^{(0)}-2x_3^{(i)},$$

$$x_{oc}^{(i)}=\tfrac{3}{2}x_c^{(0)}-\tfrac{1}{2}x_3^{(i)}, \qquad x_{ic}^{(i)}=\tfrac{1}{2}x_c^{(0)}+\tfrac{1}{2}x_3^{(i)}.$$

Taking all possible cases into account, the $(i+1)^{\text{th}}$ iteration of the Nelder-Mead method can be written as follows:

1. If $f\left(x_1^{(i)}\right)\le f\left(x_r^{(i)}\right)<f\left(x_2^{(i)}\right)$, then $S^{(i+1)}=\left[x_1^{(i)},x_r^{(i)},x_2^{(i)}\right]$.

2a) If $f\left(x_e^{(i)}\right)<f\left(x_r^{(i)}\right)<f\left(x_1^{(i)}\right)$, then $S^{(i+1)}=\left[x_e^{(i)},x_1^{(i)},x_2^{(i)}\right]$.

2b) If $f\left(x_r^{(i)}\right)<f\left(x_1^{(i)}\right)$ and $f\left(x_e^{(i)}\right)\ge f\left(x_r^{(i)}\right)$, then $S^{(i+1)}=\left[x_r^{(i)},x_1^{(i)},x_2^{(i)}\right]$.

3) If $f\left(x_2^{(i)}\right)\le f\left(x_r^{(i)}\right)<f\left(x_3^{(i)}\right)$ and $f\left(x_{oc}^{(i)}\right)\le f\left(x_r^{(i)}\right)$, then three cases are possible:

3a) If $f\left(x_{oc}\right)<f\left(x_1^{(i)}\right)$, then $S^{(i+1)}=\left[x_{oc}^{(i)},x_1^{(i)},x_2^{(i)}\right]$.

3b) If $f\left(x_1^{(i)}\right)\le f\left(x_{oc}^{(i)}\right)<f\left(x_2^{(i)}\right)$, then $S^{(i+1)}=\left[x_1^{(i)},x_{oc}^{(i)},x_2^{(i)}\right]$.

3c) If $f\left(x_2^{(i)}\right)\le f\left(x_{oc}^{(i)}\right)\le f\left(x_r^{(i)}\right)$, then $S^{(i+1)}=\left[x_1^{(i)},x_2^{(i)},x_{oc}^{(i)}\right]$.

4) If $f\left(x_r^{(i)}\right)\ge f\left(x_3^{(i)}\right)>f\left(x_{ic}^{(i)}\right)$, then three cases are possible:

4a) If $f\left(x_{ic}^{(i)}\right)<f\left(x_1^{(i)}\right)$, then $S^{(i+1)}=\left[x_{ic}^{(i)},x_1^{(i)},x_2^{(i)}\right]$.

4b) If $f\left(x_1^{(i)}\right)\le f\left(x_{ic}^{(i)}\right)<f\left(x_2^{(i)}\right)$, then $S^{(i+1)}=\left[x_1^{(i)},x_{ic}^{(i)},x_2^{(i)}\right]$.

4c) If $f\left(x_2^{(i)}\right)\le f\left(x_{ic}^{(i)}\right)<f\left(x_3^{(i)}\right)$, then $S^{(i+1)}=\left[x_1^{(i)},x_2^{(i)},x_{ic}^{(i)}\right]$.

5) Shrinking occurs if and only if

$$f\left(x_2^{(i)}\right) \leq f\left(x_r^{(i)}\right) < f\left(x_3^{(i)}\right) \text{ and } f\left(x_{oc}^{(i)}\right) > f\left(x_r^{(i)}\right)$$

or

$$f\left(x_r^{(i)}\right) \geq f\left(x_3^{(i)}\right) \text{ and } f\left(x_{ic}^{(i)}\right) \geq f\left(x_3^{(i)}\right)$$

holds. Then $S^{(i+1)} = \left[x_1^{(i+1)}, x_2^{(i+1)}, x^{(i+1)}\right]$, where

$$\left\{x_1^{(i+1)}, x_2^{(i+1)}, x^{(i+1)}\right\} = \left\{x_1^{(i)}, \frac{1}{2}\left(x_2^{(i)} + x_1^{(i)}\right), \frac{1}{2}\left(x_3^{(i)} + x_1^{(i)}\right)\right\}$$

whose order is determined by the requirement

$$f\left(x_1^{(i+1)}\right) \leq f\left(x_2^{(i+1)}\right) \leq f\left(x^{(i+1)}\right).$$

McKinnon investigated the Nelder-Mead method concerning a repeated case 4b), that is the behavior

$$f\left(x_1^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_2^{(i)}\right) < f\left(x_3^{(i)}\right) \leq f\left(x_r^{(i)}\right) \quad (i \geq 0). \tag{2}$$

His construction keeps one vertex ($x_1^{(0)} = (0,0)$) fixed, while vertices $x_2^{(i)}$ and $x_3^{(i)}$ converge to $x_1^{(0)}$.

In the next two sections we give a sufficient condition under which the Nelder-Mead method repeats case 4c) and also show its application to several functions in various situations. Our construction keeps two vertices ($x_1^{(0)}$ and $x_2^{(0)}$) fixed, while the third one converges to the midpoint of the line segment $\overline{x_1^{(0)} x_2^{(0)}}$.

4 A condition for repeated inside contraction

Here we give a sufficient condition under which the inside contraction (case 4c)

$$f\left(x_1^{(i)}\right) \leq f\left(x_2^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_3^{(i)}\right) \leq f\left(x_r^{(i)}\right) \tag{3}$$

is repeated.

Assume that $f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f\left(x_3^{(0)}\right)$. The points $x_r^{(0)}, x_{ic}^{(0)}$ are located on the straight line defined by the points $x_c^{(0)}$ and $x_3^{(0)}$, where $x_c^{(0)} = \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$. The equation for this line is given by

$$\varphi(t) = (1+t)x_c^{(0)} - t x_3^{(0)}. \tag{4}$$

For $t \in [-1,1]$, we have $x_3^{(0)} = \varphi(-1)$, $x_{ic}^{(0)} = \varphi\left(-\frac{1}{2}\right)$, $x_c^{(0)} = \varphi(0)$, $x_r^{(0)} = \varphi(1)$.

If

$$f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) \le f\left(x_{ic}^{(0)}\right) < f\left(x_3^{(0)}\right) \le f\left(x_r^{(0)}\right), \tag{5}$$

that is

$$f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) \le f\left(\varphi\left(-\frac{1}{2}\right)\right) < f(\varphi(-1)) \le f(\varphi(1)), \tag{6}$$

then $x_{ic}^{(0)} = \varphi\left(-\frac{1}{2}\right)$ is selected in the first iteration of the Nelder-Mead simplex method and

$$x_1^{(1)} = x_1^{(0)}, \quad x_2^{(1)} = x_2^{(0)}, \quad x_3^{(1)} = x_{ic} = \varphi\left(-\frac{1}{2}\right). \tag{7}$$

Since $x_1^{(0)}$ and $x_2^{(0)}$ do not change, $x_c^{(0)}$ also remains and the next $x_r^{(1)}$ and $x_{ic}^{(1)}$ will be on the line segment $x(t)$ ($t \in \left[-\frac{1}{2}, \frac{1}{2}\right]$).

Let $t_k = \frac{1}{2^k}$ and assume that we have performed k consecutive steps so that

$$x_1^{(k)} = x_1^{(0)}, \quad x_2^{(k)} = x_2^{(0)}, \quad x_3^{(k)} = \varphi(t_k) \tag{8}$$

and

$$f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) < f\left(x_3^{(k)}\right). \tag{9}$$

Since $x_c^{(k)} = x_c^{(0)}$ and

$$\begin{aligned} x(\lambda) &= (1+\lambda)x_c^{(0)} - \lambda x_3^{(k)} = (1+\lambda)x_c^{(0)} - \lambda\left((1+t_k)x_c^{(0)} - t_k x_3^{(0)}\right) \\ &= (1-\lambda t_k)x_c^{(0)} + \lambda t_k x_3^{(0)}, \end{aligned}$$

we have

$$x_r^{(k)} = x^{(k)}(1) = (1-t_k)x_c^{(0)} + t_k x_3^{(0)} = \varphi(-t_k)$$

and

$$x_{ic}^{(k)} = x^{(k)}\left(-\frac{1}{2}\right) = \left(1+\frac{t_k}{2}\right)x_c^{(0)} + \frac{t_k}{2}x_3^{(0)} = \varphi\left(\frac{t_k}{2}\right).$$

If

$$f\left(x_2^{(0)}\right) \le f\left(\varphi\left(\frac{t_k}{2}\right)\right) < f(\varphi(t_k)) \le f(\varphi(-t_k)), \tag{10}$$

then

$$x_1^{(k+1)} = x_1^{(0)}, \quad x_2^{(k+1)} = x_2^{(0)}, \quad x_3^{(k+1)} = \varphi(t_{k+1}) \quad \left(t_{k+1} = -\frac{1}{2^{k+1}}\right). \tag{11}$$

If the above conditions hold for all k values (and f is a continuous function), that is

(i) $f\left(x_1^{(0)}\right) \le f\left(x_2^{(0)}\right) < f\left(x_3^{(0)}\right)$;

(ii) $f(\varphi(-t_k)) \ge f(\varphi(t_k)) > f\left(\varphi\left(\frac{t_k}{2}\right)\right) \ge f\left(x_2^{(0)}\right)$ ($t_k = -\frac{1}{2^k}$, $k = 0, 1, \ldots$),

then $x_3^{(k)} \to x_c^{(0)}$, $f\left(x_3^{(k)}\right) \to f\left(x_c^{(0)}\right)$, while $x_1^{(k)}$ and $x_2^{(k)}$ ($f\left(x_j^{(k)}\right) = f\left(x_j^{(0)}\right)$, $j = 1,2$) remain fixed. Hence the simplices converge the line segment $\overline{x_1^{(0)} x_2^{(0)}}$, while $\lim_{k\to\infty} x_3^{(k)} = x_c^{(0)}$ is the midpoint of this line segment. Also, the diameters of the simplices do not converge to 0.

A sufficient condition for the requested behavior can be formulated as follows.

Theorem 1. Assume that $S^{(0)} = \left[x_1^{(0)}, x_2^{(0)}, x_3^{(0)}\right]$ is such that

$$f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f\left(x_3^{(0)}\right).$$

If, in addition, f is such that
(a) $f(\varphi(t))$ is continuous on $[-1,1]$;
(b) $f(\varphi(t)) \geq f(\varphi(-t))$ for $t \in [0,1]$;
(c) $f(x(t))$ is strictly monotone decreasing on $[-1,0]$;
(d) $f(\varphi(t)) > f(\varphi(0)) = f\left(x_c^{(0)}\right) \geq f\left(x_2^{(0)}\right)$ ($t \in [-1,1]$, $t \neq 0$),
then

$$f\left(x_1^{(i)}\right) \leq f\left(x_2^{(i)}\right) \leq f\left(x_{ic}^{(i)}\right) < f\left(x_3^{(i)}\right) \leq f\left(x_r^{(i)}\right) \tag{12}$$

holds for all $i = 0,1,2,\ldots$, $x_3^{(i)} \to x_c^{(0)}$, and $f\left(x_3^{(i)}\right) \to f\left(x_c^{(0)}\right)$.

Proof. Assume that for some $-1 \leq t < 0$, $x_3 = \varphi(t)$ and

$$f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f(\varphi(t)).$$

Then $x_r = \varphi(-t)$, $x_{ic} = \varphi\left(\frac{t}{2}\right)$, and (b) and (c) imply that

$$f(x_r) = f(\varphi(-t)) \geq f(\varphi(t)) = f(x_3) > f\left(\varphi\left(\frac{t}{2}\right)\right) = f(x_{ic}).$$

Condition (d) implies that $f\left(x_1^{(0)}\right) \leq f\left(x_2^{(0)}\right) < f\left(\varphi\left(\frac{t}{2}\right)\right)$. □

In the next section we apply this sufficient condition to a few functions and show some different types of convergence behavior.

5 Examples of repeated inside contractions

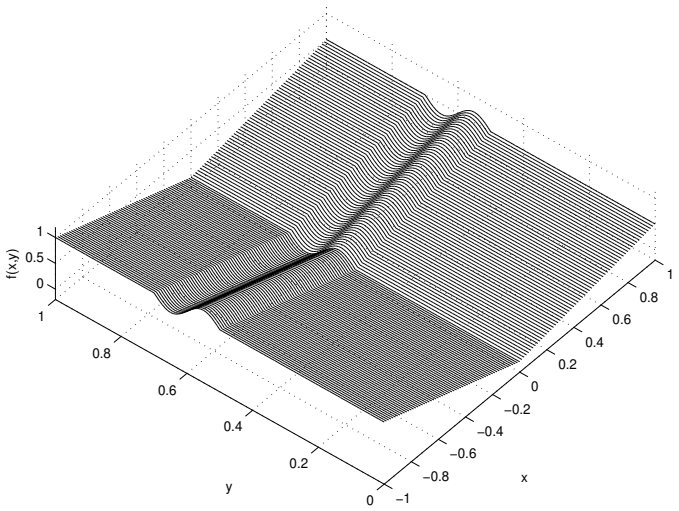
Define the first function as

$$f(x,y) = \frac{1}{4}(x+|x|) + \frac{1}{2}|x-|x|| + g(y), \tag{13}$$

where

$$g(y) = \begin{cases} 0.2\sin(10\pi y - 5\pi), & \text{if } 0.5 \leq y \leq 0.7 \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

This function is shown on the next figure.



Define the initial simplex vertices as $x_1^{(0)} = (0, 0.5), x_2^{(0)} = (0, 0.7), x_3^{(0)} = (0.5, 0.6)$. Then $f\left(x_1^{(0)}\right) = 0, f\left(x_2^{(0)}\right) = 0, f\left(x_3^{(0)}\right) = 0.25, x_c^{(0)} = [0, 0.6], \varphi(t) = [-0.5t, 0.6]$ and

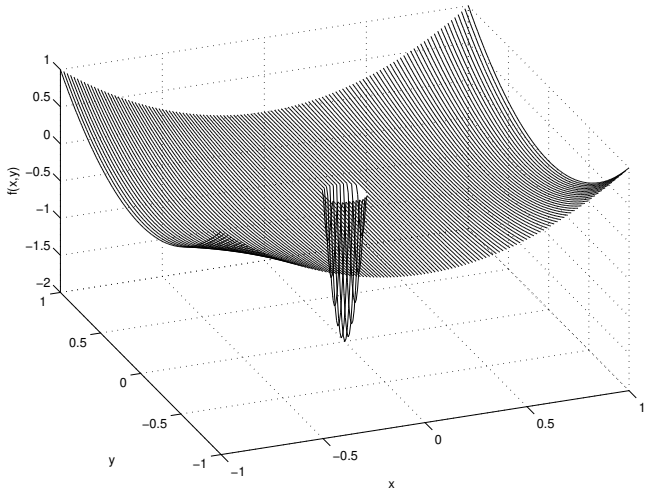
$$f(\varphi(t)) = \begin{cases} \frac{|t|}{4}, & \text{if } t < 0 \\ \\ \frac{t}{2} & \text{if } t > 0 \end{cases}.$$

This function clearly satisfies conditions (a)-(d). Hence we have the repeated x_{ic} behavior and the convergence $x_3^{(i)} \to x_{ic}^{(0)}$. Note that the limit point $x_c^{(0)}$ is not a local minimum point of $f(x, y)$!

Our next example is the function

$$f(x,y) = x^2y^2 + \frac{1}{\varepsilon^2}\left(x^2+y^2\right) - 1 - \left|\frac{1}{\varepsilon^2}\left(x^2+y^2\right) - 1\right| \quad (\varepsilon > 0), \qquad (15)$$

where $\varepsilon < 1$ is small enough. For $\varepsilon = 0.1$, this function is shown on the next figure.



This function has a unique global minimum point at the origin and a continuum number of non-isolated local minimum points along the x and y axes. Select the initial vertices as $x_1^{(0)} = (0,0)$, $x_2^{(0)} = (0,2\varepsilon)$, $x_3^{(0)} = (\varepsilon,\varepsilon)$. Then $f\left(x_1^{(0)}\right) = -2$, $f\left(x_2^{(0)}\right) = 0$, $f\left(x_3^{(0)}\right) = \varepsilon^4$, $x_c^{(0)} = (0,\varepsilon)$, $\varphi(t) = (-t\varepsilon,\varepsilon)$ and $f(x(t)) = \varepsilon^4 t^2$. Hence Theorem 1 applies. Note that $x_1^{(0)}$ is the global minimum point and $x_2^{(0)}$ is a non-isolated local minimum point, while $x_3^{(i)}$ converges to $x_c^{(0)}$ which is not a local minimum point. As ε can be chosen arbitrarily small, this problem can occur arbitrarily close to the global minimum point.

The third example is related to saddle points. Assume that $f(x,y)$ is separable in the form

$$f(x,y) = g(x) - h(y), \tag{16}$$

where g and h are continuous real functions, $g(x) > 0$ for $x \neq 0$, $g(0) = 0$, $g(x)$ is strictly monotone increasing for $x \geq 0$, $g(x)$ is strictly monotone decreasing for $x < 0$, $g(-x) \geq g(x)$ $(x \geq 0)$, $h(y) \geq 0$ for $y \neq 0$, $h(0) = 0$ and $h(-y) \geq h(y)$ for $y \geq 0$. Select the initial vertices as $x_1^{(0)} = (0,-a)$, $x_2^{(0)} = (0,a)$ and $x_3^{(0)} = (b,0)$ with $a,b > 0$. Then $x_c^{(0)} = (0,0)$,

$$f\left(x_1^{(0)}\right) = -h(-a) \leq -h(a) = f\left(x_2^{(0)}\right) \leq 0 = f\left(x_c^{(0)}\right) < g(b) = f\left(x_3^{(0)}\right).$$

Since $\varphi(t) = (-bt,0)$ and $f(\varphi(t)) = g(-bt) \geq g(bt)$ for $t \in [0,1]$, Theorem 1 implies that $x_3^{(i)}$ converges to the saddle point $x_c^{(0)} = (0,0)$. The same result holds, if the above conditions are restricted to an open neighborhood of the origin.

This saddle point phenomenon also appears if $f(x,y)$ is not separable. Using Theo-

rem 1 it is easy to check this for the function

$$f(x,y) = x^2 - y^2 + \frac{1}{\varepsilon^2}\left(x^2 + (y-2\varepsilon)^2\right) - 1 - \left|\frac{1}{\varepsilon^2}\left(x^2 + (y-2\varepsilon)^2\right) - 1\right| + \frac{1}{\varepsilon^2}\left(x^2 + (y+2\varepsilon)^2\right) - 1 - \left|\frac{1}{\varepsilon^2}\left(x^2 + (y+2\varepsilon)^2\right) - 1\right|$$

with $\varepsilon > 0$ small enough and initial vertices $x_1^{(0)} = (0, 2\varepsilon)$, $x_2^{(0)} = (0, -2\varepsilon)$ and $x_3^{(0)} = (1,0)$. Note that $x_1^{(0)}$ and $x_2^{(0)}$ are are the two local minimum points of f.

6 A note on repeated shrinking

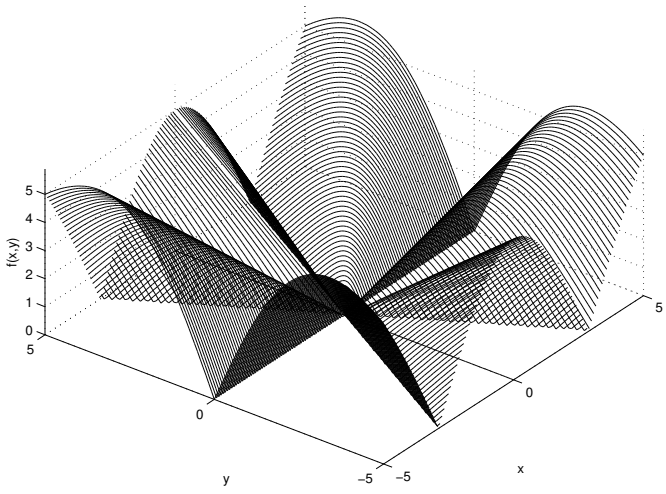
In the case of repeated shrinking the algorithm behaves as a contraction procedure and the vertices of the simplex sequence converge to a common limit point. The following simple examples indicate that this fail safe convergence may have unwanted consequences if we seek for the global minimum point.

Assume that $f\left(x_1^{(i)}\right) \le f\left(x_2^{(i)}\right) \le f\left(x_3^{(i)}\right)$. We use the second condition for shrinking, that is

$$f\left(x_r^{(i)}\right) \ge f\left(x_3^{(i)}\right) \ \wedge f\left(x_{ic}^{(i)}\right) \ge f\left(x_3^{(i)}\right).$$

Consider the function

$$f(x,y) = \begin{cases} \left|\frac{y^3 - 3x^2 y}{x^2 + y^2}\right|, & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0). \end{cases} \quad . \tag{17}$$



Observe that $0 \le f(x,y) \le 3|y|$, $f(x,0) = 0$, $f\left(x, \sqrt{3}x\right) = f\left(x, -\sqrt{3}x\right) = 0$, and f has a continuum of non-isolated global minimum points.

Let

$$x_1^{(0)} = (0,0), \quad x_2^{(0)} = \left(-1, \sqrt{3}\right), \quad x_3^{(0)} = \left(1, \sqrt{3}\right).$$

Then $f\left(x_1^{(0)}\right) = f\left(x_2^{(0)}\right) = f\left(x_3^{(0)}\right) = 0$, $x_c^{(0)} = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, $x_r^{(0)} = (-2,0)$, $x_{ic}^{(0)} = \left(\frac{1}{4}, \frac{3\sqrt{3}}{4}\right)$. Since $f\left(x_r^{(0)}\right) = f\left(x_3^{(0)}\right) = 0$, and $f\left(x_{ic}^{(0)}\right) = \frac{9\sqrt{3}}{14} > f\left(x_3^{(0)}\right) = 0$, we perform a shrink operation:

$$x_1^{(1)} = (0,0), \quad x_2^{(1)} = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \quad x_3^{(1)} = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

where $f\left(x_j^{(1)}\right) = 0$ $(j = 1,2,3)$ and the previous argument can be applied again. After the i-th iteration we have

$$x_1^{(i)} = (0,0), \quad x_2^{(i)} = \frac{1}{2^i}\left(-1, \sqrt{3}\right), \quad x_3^{(i)} = \frac{1}{2^i}\left(1, \sqrt{3}\right),$$

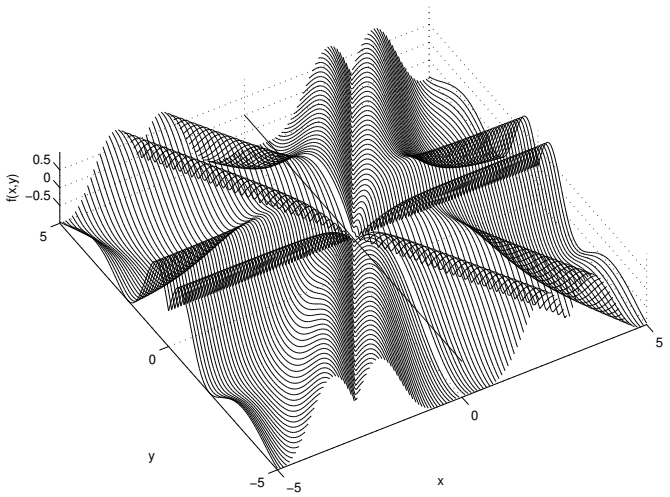
where $f\left(x_j^{(i)}\right) = 0$ $(j = 1,2,3)$. Then

$$x_c^{(i)} = \frac{1}{2^{i+1}}\left(-1, \sqrt{3}\right), \quad x_r^{(i)} = \left(-\frac{1}{2^{i-1}}, 0\right), \quad x_{ic}^{(i)} = \left(\frac{1}{2^{i+2}}, \frac{3\sqrt{3}}{2^{i+2}}\right),$$

$f\left(x_r^{(i)}\right) = 0$ and $f\left(x_{ic}^{(i)}\right) = \frac{9\sqrt{3}}{14} 2^{-i} > f\left(x_3^{(i)}\right) = 0$. Hence we have a repeated shrinking and the convergence $x_2^{(i)}, x_3^{(0)} \to x_1^{(0)}$.

Consider the following modification of function (17)

$$f(x,y) = \begin{cases} \sin\left(\left|\frac{y^3 - 3x^2 y}{x^2 + y^2}\right|\right), & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0). \end{cases}$$



The modified function has non-isolated local minimum points and several global minimum points. For this function, $f(x,0)=0$, $f\left(x,\sqrt{3}x\right)=f\left(x,-\sqrt{3}x\right)=0$ also hold. Take again the initial simplex

$$x_1^{(0)}=(0,0)\,,\quad x_2^{(0)}=\left(-1,\sqrt{3}\right),\quad x_3^{(0)}=\left(1,\sqrt{3}\right).$$

Then $f\left(x_1^{(0)}\right)=f\left(x_2^{(0)}\right)=f\left(x_3^{(0)}\right)=0, x_r^{(0)}=(-2,0), x_{ic}^{(0)}=\left(\frac{1}{4},\frac{3\sqrt{3}}{4}\right), f\left(x_r^{(0)}\right)=0$, and $f\left(x_{ic}^{(0)}\right)=\sin\frac{9\sqrt{3}}{14}>f\left(x_3^{(0)}\right)=0$. So we must perform a shrinking and we obtain a repeated shrinking with the same sequences $x_2^{(i)}$, $x_3^{(i)}$, $x_c^{(i)}$, $x_r^{(i)}$, $x_{ic}^{(i)}$ with one difference: $f\left(x_{ic}^{(i)}\right)=\sin\left(\frac{9\sqrt{3}}{14}2^{-i}\right)$.

The obtained result is essentially the result of the previous example. However $x_2^{(i)}, x_3^{(i)}$ converge to the local non-isolated minimizer x_1, while the global minimum of f is not achieved.

We note that the Nelder-Mead simplex algorithm is also used in the context of global optimization (see, e.g. [9]). The last two examples indicate a kind of local character of the method.

References

[1] Audet, C., Hare, W: Derivative-free and Blackbox Optimization, Springer, 2017

[2] Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimizations, SIAM, 2009

[3] Kelley, C.T.: Iterative Methods for Optimization, SIAM, 1999

[4] Kelley, CT.: Detection and remediation of stagnation in the Nelder-Mead algorithm using an sufficient decrease condition, SIAM Journal of Optimization, 10 (1999) 33–45

[5] Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods, SIAM Review 45, 3 (2003) 385–482

[6] Lagarias, J.C., Poonen, B., Wright, M.H.: Convergence of the restricted Nelder-Mead algorithm in two dimensions, SIAM J. Optim., Vol. 22, No. 2, pp. 501-532, 2012

[7] Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the Nelder-Mead simplex method in low dimensions, SIAM J. Optim., Vol. 9, No. 1, pp. 112-147, 1998

[8] Lewis, R.M., Torczon, V., Trosset, M.W.: Direct search methods: then and now, Journal of Computational and Applied Mathematics 124 (2000) 191–207

[9] Matlab Global Optimization Toolbox User's Guide, Version R2016a, MathWorks, 2016

[10] McKinnon, K.I.M.: Convergence of the Nelder-Mead simplex method to a nonstationary point, SIAM j. Optim. 9, 1 (1998) 148-158

[11] Nazareth, L., Tseng, P.: Gilding the Lily: A variant for the Nelder-Mead algorithm based on golden-section search, Computational Optimization and Applications, 22 (2002) 133–144

[12] Nelder, J. A., Mead, R.: A simplex method for function minimization, Computer Journal, 7 (1965) 308–313, doi:10.1093/comjnl/7.4.308

[13] Parkinson, J.M., Hutchinson, D.: An investiagtion into the efficiency of variants on the simplex method, in: F.A. Lootsma (ed): Numerical Method for Non-linear Optimization, Academic Press, 1972, pp. 115–135

[14] Price, C.J., Coope, I.D., Byatt, D.: A convergent variant of the Nelder-Mead algorithm, Journal of Optimization Theory and Applications, 113, 1 (2002), 5–19

[15] Tseng, P.: Fortified-descent simplicial search method: A general approach, SIAM J. Optim. 10, 1 (1999) 269–288

[16] Walters, F.H., Morgan, S.L., Parker, L.R., Deming, S.N.: Sequential Simplex Optimization, CRC Press LLC, 1991

[17] Wessing, S.: Proper initialization is crucial for the Nelder-Mead simplex search, Optimization Letters volume 13 (2019) 847–856

[18] Wright, M.H.: Nelder, Mead, and the other simplex method, Documenta Mathematica, Exra Volume: Optimization Stories (2012) 271-276

Development Level of Engineering Students' Inductive Thinking

Péter Tóth, Kinga Horváth, Katalin Kéri

J. Selye University
Bratislavská cesta 3322, SK-94501 Komárno, Slovakia
e-mails: tothp@ujs.sk, horvathki@ujs.sk, kerik@ujs.sk

Abstract: In Hungary, the rate of drop-out in technical higher education, exceeds the average of the European Union. The affective, sociological and socio-economic factors hiding behind the reasons have already been examined by several researches, however, less attention has been paid to the students' cognitive deficiencies although inductive reasoning is of decisive importance in gaining knowledge e.g. in mathematics and natural sciences. Based on these findings, the object of this research, was on one hand to describe the development level of the inductive reasoning and thinking of students starting their studies in technical higher education, and on the other hand, to prove or disprove the correlation between the task solution time and the result achieved in the inductive test. The research implemented by means of an online test involved 253 students; of the components of inductive thinking, we examined the development level of abstract, analogue and diagrammatic thinking.

Keywords: drop-out; measuring competences; inductive thinking; engineering education

1 Introduction

One of the vital problems for Hungarian higher education, and within that technical higher education, is a high drop-out rate. The question arises as to how much advanced logical thinking is required, for the successful acquisition of real subjects, which can cause serious difficulties in technical higher education. It is hypothesized that underdeveloped logical thinking leads to drop-out due to the ineffective fulfillment of these subjects. Advanced inductive thinking is essential for learning mathematics and basic technical subjects. Knowing this, at the beginning of studies, it would be possible to differentiate the development of the students, to fill their gaps, and through these to reduce the drop-out rate.

The Hungarian data exceeds the EU members' 25-30% average drop-out rate by almost 10%. According to the data of 2016, the rate of those gaining a degree in the technical field was 41.6% (completion rate), the level of drop-out was 39.6%

while the proportion of those not yet having completed their studies (e.g. because of the lack of a language exam) and having changed their training was 18.8%. The numbers in the non-technical field show this picture: 53.5%, 33.6% and 13.0% [1] [2]. As regards to the completion rate, the picture seen varies by types of training. In technical higher education, 35.7% of full-time students dropped out, while in correspondence training this rate was 54.6%, and the rate of those changing their trainings or having not yet graduated was 19.2 and 17.6%. Regarding financing, the drop-out rate was 35.3% with state-funded students and 53.9% with fee-paying students. [2]

In students' registries, the following reasons of drop-out are enlisted in most cases:

(1) Attributable to deficiencies in the studies:

  (a) the number of retake or repeat exams has reached the maximum

  (b) the number of passivated semesters has reached the maximum

  (c) non-completion of training requirements

(2) Financial problems: e.g. non-payment of the training fee

(3) Financial problems originating in study insufficiencies: reclassification from state-financed to fee-paying status

(4) Unknown reasons: e.g. voluntary interruption of studies. [2]

Searching for the real reasons of drop-out, several empiric researches have been implemented during the latest 40 years. [1] [3] [4] Analyzing these, the following categories can be distinguished:

(1) Cognitive deficits (existing knowledge, competences, skills and capabilities)

(2) Affective deficits (commitment for the chosen course, interest, motivation, scale of values, career prospects, prestige and stress tolerance)

(3) Social reasons (financial problems, living, student work, residence, lack of morally supporting environment or disadvantageous socio-economic status)

(4) Reasons originating in the insufficient selective functions of the educational system (e.g. entrance exam)

(5) Reasons originating in the insufficient operation of the institution (deficiencies in learning support, insufficient level of student services, qualitative deficiencies of teaching or deficits in the equipment)

By the early 2010s, drop-out became an acute problem in Hungarian higher education, as well. Several of the papers published during the latest 10 years [2]

[5] [6] [7] have revealed the affective, sociological and socio-economic reasons of drop-out. When categorizing the reasons, they described:

(1) Institutional ones (reputation of the institution, infrastructure supporting learning, committed lecturers, student-friendly nature of study classes, etc.)

(2) Social and cultural ones (social status, socio-economic status, decreasing impact of degree in the student's micro environment, etc.)

(3) Economic ones ("it is more worth working than staying at the university" working in parallel, training fee, etc.)

(4) Personal, pedagogical, psychological and learning-related ones

In terms of drop-out, the lack of ICT competences [25] and the inadequate organization of the learning process [26], inadequate level of comprehension [27], and lack of self-confidence [28] should be highlighted.

In the research implemented in Hungary, less attention has been paid to the students' cognitive deficiencies albeit this is the one lecturers meet in most cases during teaching their subjects. This is exactly why the Ministry for Innovation and Technology, which is responsible for the development of higher education, set a target to introduce competence measurement in higher education. However, the question of what competences connectible to learning difficulties should be measured arises.

As in natural scientific, and so in engineering training, mathematical and natural scientific cognition is of decisive importance, we selected certain components of inductive thinking and reasoning as the focus of our competence measurement.

There is a further aspect that enhances the relevance of the research. A strong demand is being formed on behalf of the labor market that can be best expressed by the term soft skills (labor market key competencies). Several studies point to the fact that career starter employees arrive from higher education to the labor market unprepared especially in terms of these competences [6], and it is not rare that companies consider these skills more important than professional preparedness [7], or call the trainers to develop these competences, as well [8]. According to the Manpower Group, 16% of the jobs in Europe remained unfilled for reason of soft skill deficiencies.

Opinions strongly differ concerning the question of what competences belong to soft skills. Mainly intra- and inter-personal cognitive competences and attitudes are mentioned. One of these is the higher-grade cognition, the components of which are problem solving and critical thinking as well as judgement and decision making ability. In the world of labor, they appear as a demand in their organic relation system [9] [10] [11] [12].

Therefore, in some sense, our research can be considered a "situation survey" since our interest was focused at the development level of the students entering higher education *in terms of logical thinking, which is one of the labor market key competences, i.e. what is the initial level of their skills in this field.*

2 Inductive and Abstract Reasoning and Thinking

When identifying the various aspects of reasoning and thinking, we use plenty of attributives. Possible forms include but are not limited to formal, informal, everyday, logical, deductive, inductive, abstract, analytical, verbal and quantitative while with regards to cognition, the list includes, inter alia, higher rank, critical, creative, problem solving, rational, lateral, inductive and deductive. On the following pages we will discuss the important field of Inductive reasoning and cognition.

Inductive cognition can be explained as the recognition of regularities and irregularities in a process during which we identify similarities, differences as well as similarities and differences by comparing attributes and relations in the tasks containing verbal, pictorial, geometrical and numerical figures. [13] Thus, induction can be related to characteristics and relations. Similarities in characteristics result in generalization, while differences result in discrimination and the common result of the two will be classification. By way of similarity of the relations, relationships will be recognized, while their differences lead to the differentiation of relationships, and the common result of the two will be system construction. Analogue can be considered one of the most important problem groups of induction and it has components like cause and reason or association-based whole and parts relations. [14]

Abstract thinking is a significant form of human knowledge. It allows us extract the essence of complicated and abstract things and to recognize interrelations. And this is of basic importance in gaining knowledge. Human differs from other living beings also in having the capacity to explain and get to know the surrounding world in a way transcending sensation and perception. For example, an engineer is able to describe the rules determining certain technical phenomena by abstraction, is able to create an abstract system by applying them, and is also able to think within this system. Technical modelling is a good example of this.

Understanding, reconstructing and creating technical models requires developed abstract cognition of the students, which in most cases means a problem for the students in learning mathematics and natural scientific (including technical scientific), i.e. STEM subjects. Thus it is not by chance that drop-out shows the highest level in relation to these subjects.

3 Objective and Means of the Research

As presented above, inductive cognition plays an important role in understanding the learning content and in performing well on the labor market. Therefore, the double demand of reducing drop-out and offering better preparation for the labor market makes our research really timely.

The speed of cognition reflects the time the experimental person needs to solve a certain test item. Carroll presumes that there is some relationship between the competence level and the speed of cognition and declares that in general, the person with developed cognitive skills solves a mathematical problem faster than the one with weaker capacities. [16] Some query this statement. [17] According to Sternberg and Pellegrino, the time used for finding a good answer can predict performance more precisely than the skill level itself. [15]

Based on the above, the research aimed to:

(1) Describe the development level of the inductive thinking of the first-grade students starting their studies in technical higher education

(2) Prove or disprove the relationship between the time used for task solution and the performance achieved in the inductive test.

A question arises here: how can the development level of the students' inductive reasoning and reasoning be reliably measured without specific subject knowledge and skills (e.g. in mathematics or physics)? Several methods can be used ranging from certain intelligence tests (e.g. Raven) and inductive reasoning tests to special measurement tools focusing at the given skill component.

In this research we used the measurement tools developed by Psychometric Success WikiJob Ltd. (UK, London) who pay special attention to labor market demands. [18] When developing the test, the experts based their measurement tools on one- and multifactor intelligence tests. [15] The sample tasks of the test are available here: https://psychometric-success.com/

To examine cognition based on inductive thinking the developer elaborated the abilities presented in Figure 1.

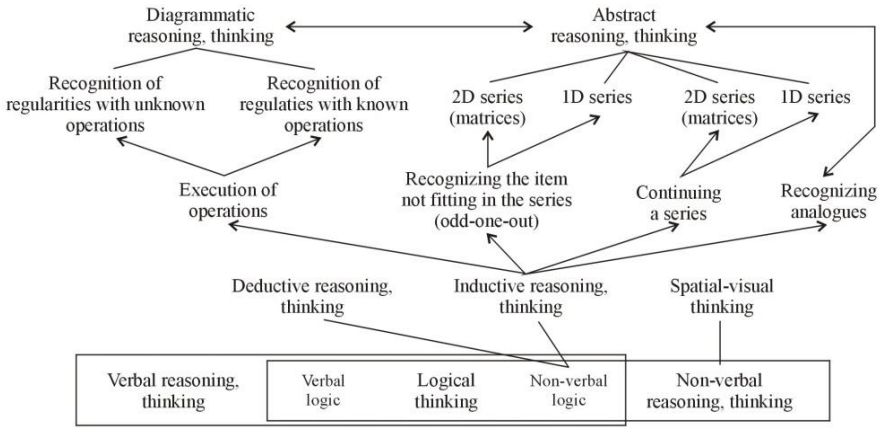


Figure 1

Task system examining inductive cognition

The problem for the respondent lies in the difficulty to recognize the logical relationships hiding behind the patterns included in the tasks (https://psychometric-success.com/). Problems root in the hardness of recognizing the change or repetition of these characteristics: (1) shape, (2) size, (3) color and (4) pattern. The tasks consist of visual patterns and geometrical figures and three types of problems must be solved by recognizing the logical relationships hiding behind them.

The following are the types of tests applied in this research.

- *One-dimension series* (Task1, Task2) require the ability to recognize relationships that are in many cases not obvious at the first glance. The recognition of the connections between the geometrical figures can be separated from the identification of the certain figures. The latter one must be clear for each of the experimenters. While solving a task, each characteristic of the geometrical figures must be identified and their correlations understood simultaneously. [19] Solving this type of problems requires such high level logical skills that promote the recognition of complex situations and thinking about them. In most cases, these processes are non-verbal and highly depend on sudden recognition and understanding. These are so fast that they do not have the time to get assimilated into the language. [20]

  Several researches examined whether it was possible to conclude difficulty and performance from the time used for solution by items. According to Jacobs and Vanderventer, the difficulty level of the items is in connection with the number of the characteristics that must be compared and kept in mind simultaneously [19], so those achieving better results are able to keep several items in mind in parallel, however it can also mean that they laid

more effort in the given item. [21] Vodegel Matzen et al. refer to Home and Habon's work according to which it is not possible to draw conclusions in terms of the difficulty level of the items from time consumption. [22]

Regarding one-dimension series, two types of exercises were included in the test: Task1: continuing a series, Task2: recognizing the item not fitting in the series (odd-one-out).

- With regards to *recognizing analogues* (Task3), already Sternberg pointed it out that the difficulty of the problems hides in recognizing the regularities rooting in the changes of the characteristics of the certain objects (A, B, C, X); to recognize these, the relation(s) (R) between the objects must be realized in terms of A and B and be applied in terms of C and X [23]:

  A – R – B :: C – R – X

  The experimental person may realize two strategies when selecting object X: (1) sequential search, or (2) alternative search.

- The *diagrammatic reasoning* test measures the skill of how logically the person is able to follow the arranged sign series. Although the test consists of simple flow charts, its solution requires the experimenter to be able to track the changes taking place in the shape, color and size of the objects. This ability is outstandingly important, i.e. when analyzing certain system processes, in error correction and in system planning [24].

  Diagrammatic reasoning was examined by means of two types of exercises in the test: Task4: recognition of regularities with unknown operations, Task5: recognition of regularities with known operations.

The online test, which was applied during the research implemented in 2019 at the Alba Regia Technical Faculty and the Rejtő Sándor Faculty of Light Industry and Environmental Engineering of Óbuda University, consisted of 6 items by task types, and allowed the measurement of the time used for the solution of the certain items. The latter one was of great important because of the second objective of this research.

4 Research Participants

The research involved 253 first-grade students, 171 men and 82 women; this rate conforms to the rate of the students at the two examined faculties. 79.05% of the students (200 persons) were less than 21 years old and 19.76% of them (50 persons) were between 21 and 25, which means that they had their secondary school leaving exam earlier, while 3 people (1.19%) were older than 25.

54.5% of the students (138 persons) continued their studies immediately after they had completed their secondary school studies, 26.1% (66 people) took their school leaving exams one year earlier, 10.7% (27 persons) did them 2, 5.5% (14 people) completed them 3 years earlier and 3.2% (8 persons) had finished their secondary studies more than 3 years before entering university. 51.7% of the students (131 persons) took their secondary school leaving exams at secondary grammar schools, 23.7% (60 people) took them at secondary technical schools and 23.7% (60 people) attended vocational education after having completed their secondary studies. Regarding those having studied at secondary grammar schools, 100 persons (39.5%) studied at 4-5-grade schools, while 31 people (12.2%) studied at 6-8th grade talent development schools.

In engineering training, mathematics and natural scientific subjects, and because of text understanding, Hungarian language and literature are of outstanding importance, thus we asked the students about their school leaving exams results, as well. Almost one third of the students achieved the excellent grade in Hungarian and/or mathematics. By way of cross table analysis, we explored that only 12.3% of the students (31 persons) had achieved an excellent result in both subjects and 59.29% (150 people) achieved a good or excellent result.

Regarding the students' higher education studies, 15.4% of them (39 persons) are involved in land surveying and land management engineering training, 11.5% (29 persons) study mechanical engineering, 23.7% (60 persons) attend industrial design engineering, 4.7% (12 students) are at light industrial engineering, 9.9% (25 persons) at environmental engineering, 7.5% (19 people) at electrical engineering and 27.3% (69 persons) at information technology engineering BSc.

We also examined the secondary school results by programs. In general, most of the students achieved 60-79% results (mark 4) both in mathematics and Hungarian at most programs. However, the rate of those having achieved the best results (80-100%) is high amongst the students of the industrial design engineering (36.7%) and the mechanical engineering (41.4%) programs, and as for the land surveying and land management programs, the rate of those with medium results (40-59%) was also high (46.2%). Regarding the exam results in Hungarian language and literature, the picture is even more splendid. It is only the land surveying and land management program where the results are a bit weaker, 30.8% of the students gained grade 3 (40-59%), at the same time, at the industrial design engineering program, 50.0% of the students and 30.8% at the environmental engineering specialization gained grade 5 (80-100%) in this subject.

5 Results

5.1 Development Level of Inductive Reasoning

The statistical indicators of the whole sample are summarized in Table 1. The best results were achieved in the analogue recognition task (Task3), while the weakest ones in the recognition of regularities including known operations (Task5). Probably, as for the latter one, the difficulty hid in the considerable memory burden as the results of the certain operations had to be stored in the short-term memory to modify them again by new operations and select the proper final result from the given list. These sensual – perceptional shifts may have caused difficulties to the students.

We compared the students' results by the various sub-samples, as well.

As for the sexes, men achieved significantly better results in continuing series (Task1: Mann-Whitney U= 5622.500; $p<0.05$) and in recognizing regularities containing known operations (Tas5: Mann-Whitney U= 4987.000; $p<0.05$), i.e. in the tasks where the average results were the lowest for the whole sample.

Table 1
Descriptive statistical indicators of the whole sample by tasks

Descriptive statistical indicators			Task1	Task2	Task3	Task4	Task5
Whole sample	N		253	253	251	244	228
	M		3.64	4.14	4.75	4.09	2.54
	SD		1.270	1.363	1.329	1.887	1.664
	95% Conf. int.	Low.	3.48	3.97	4.58	3.86	2.32
		Up.	3.79	4.31	4.91	4.33	2.75
	%	25%	3	3	4	2	1
		50%	4	4	5	5	2
		75%	5	5	6	6	4
Sex	male (171 p.)	M	3,78	4.06	4.78	4.04	2.70
		SD	1.240	1.388	1.360	1.895	1.704
	female (82 p.)	M	3.34	4.30	4.67	4.21	2.23
		SD	1.288	1.302	1.267	1.876	1.551
Place of residence	parents (136 p.)	M	3.89	4.21	4.94	4.47	2.80
		SD	1.120	1.367	1.187	1.697	1.656
	sublet (23 p.)	M	3.61	4.43	4.87	4.35	2.39
		SD	1.340	1.080	1.180	1.555	1.469
	hostel (77 p.)	M	3.32	3.97	4.51	3.46	2.15
		SD	1.352	1.460	1.466	2.088	1.623

We also found significant differences in terms of the series continuation task (Task 1: χ^2= 9.820; p<0.05) and the two tasks requiring the recognition of regularities (Task4: χ^2= 11.856; p<0.05 and Task5: χ^2= 6.993; p<0.05) according to the students' place of residence (with parents, in sublet or hostel). The two examined faculties are situated in big cities (Budapest and Székesfehérvár) where the rate of those living in hostels is near to 30-30%; in Budapest, almost 40% of the students live with their parents, while this number exceeds 60% in Székesfehérvár. The proportion of those living in sublets approaches 20% in Budapest, which can be reasoned by the low number of hostel rooms. 75.4% of the boarders but only 55.9% of those staying with their parents live in villages or small towns, and it is probably the impact of the countryside secondary schools with differing capabilities that is reflected in the students' results: the students staying with their parents were more successful in solving the more difficult tasks.

In terms of the secondary school leaving exams in mathematics, we found significant differences in the average results in the certain tasks (p<0.05). The better result the person achieved during his/her mathematics exam, the better result could be expected in the tasks of the inductive test (Figure 2).

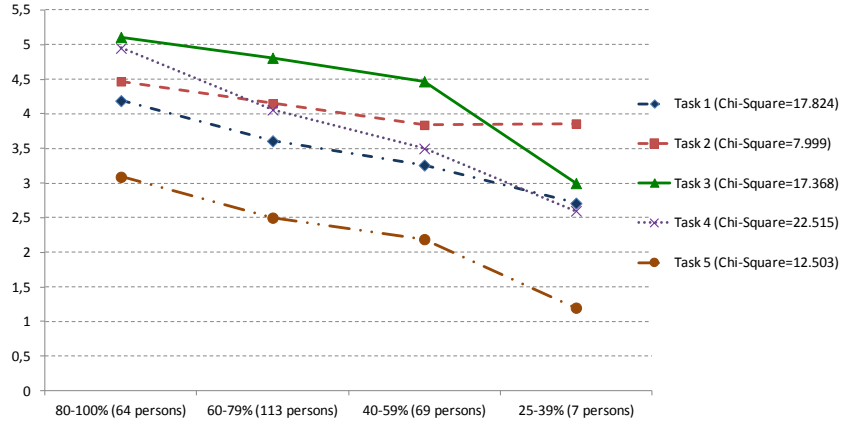


Figure 2

Average results by tasks in terms of the results of the school leaving exam in mathematics

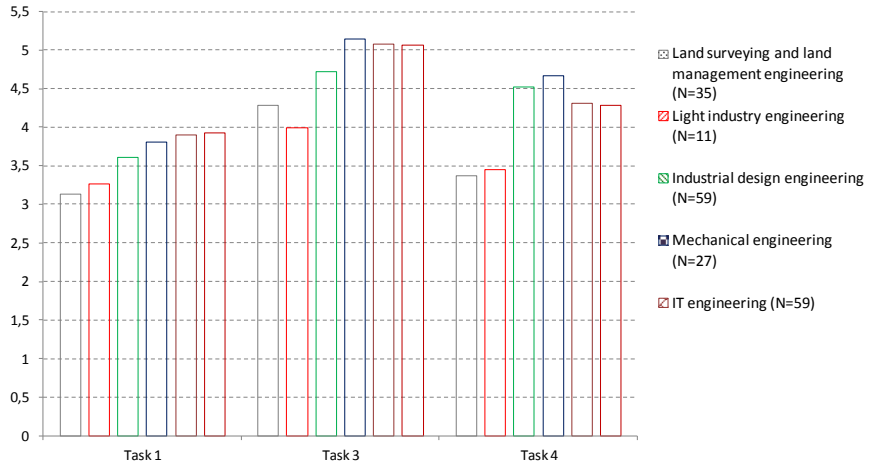


Figure 3

Average task results in terms of programs

Regarding the parents' educational level and the type of secondary school (secondary grammar or technical), we did not find significant differences in the averages of the certain types of tasks.

Figure 3 presents the average results of the certain tasks by programs. Concerning the tasks of continuing a series (Task1: χ^2= 15.044; p=0.010), the analogue (Task3: χ^2= 15.706; p=0.008) and the recognition of regularities containing unknown operations (Task4: χ^2= 14.583; p=0.012), we found significant differences by programs. The best results were achieved by the students involved in mechanical engineering, information technology engineering and electrical engineering.

5.2 Time Consumption on Task Solution

The students were given 1800 seconds to solve the tasks. Online measuring allowed us to register the time spent on task solution by items, thus it became possible to analyze and compare them with the achieved results.

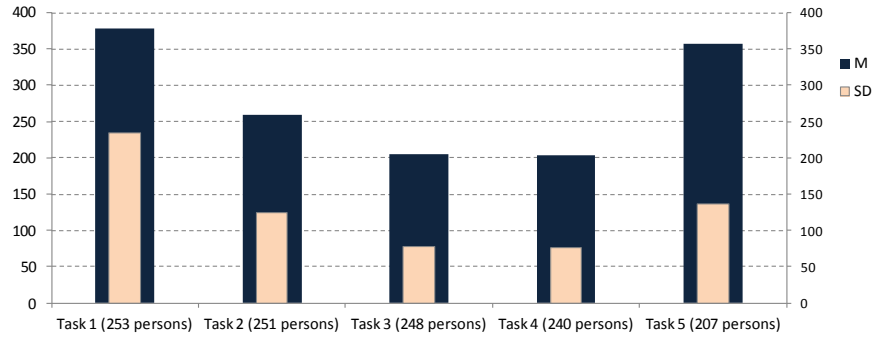


Figure 4

Mean and standard deviation of the time used to solve the given task types

The students consumed the most time for solving the series continuation (Task1) and the diagrammatic task containing known operations (Task5), and used the least time for the recognition of the analogue (Task3) and the diagrammatic items containing unknown operations (Task4) (Figure 4). If we compare these to the data presented in Table 1, reciprocal proportion-like relations will appear between the average results of certain task types and time consumption. However, high standard deviation rates imply considerable personal differences. We considered it important to analyze these and compare them to the achieved results. In case we concentrate at the first and the last type of tasks and analyze time consumption by items, as well, we will find a rather diverse phenomenon. The items of the series continuation task (Task1) differed in difficulty. It was easy for the students to understand what the task was, and time consumption clearly reflects the difficulty level of the item (Figure 5). It can also be clearly seen that the more difficult an item appeared to be, the bigger personal differences can be observed, which is proved by the high standard deviation levels.

Another phenomenon noticed, regarding the diagrammatic items using known operations (Task5) (Figure 6). According to Figure 4, is that the average time consumption of this task type almost agrees with that of the series continuation task (Task1). However, it can be clearly seen that the first item meant a real challenge for the students, i.e. at first they had faced problems of understanding. Later this decreased gradually. Personal differences were less typical here, which is reflected by much more regular standard deviation values, thus in this case, it was difficult to solve the items for everyone.

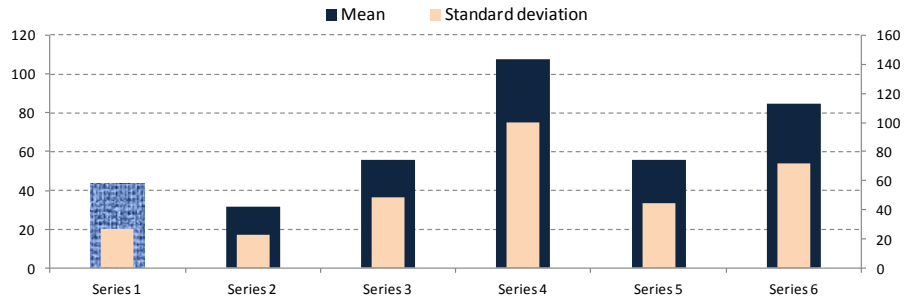


Figure 5
Mean and standard deviation of the time used to solve series continuation items

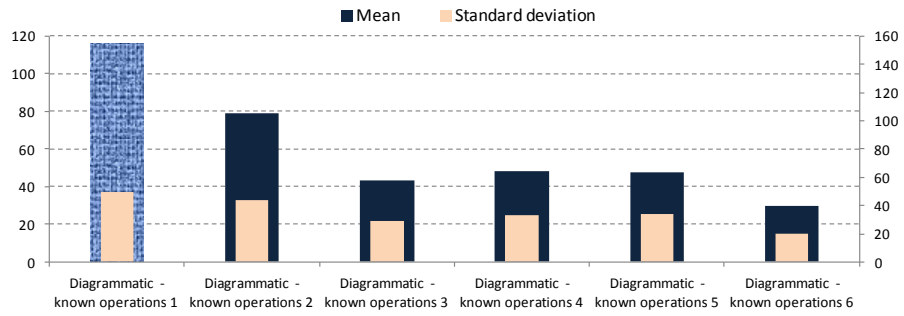


Figure 6
Mean and standard deviation of the time spent on solving the diagrammatic items containing known operations

Time consumption of Task2 and Task3 by items showed a picture similar to Figure 5 while that of Task4 was like Figure 6.

Regarding background variables, we found significant differences only in terms of the programs (Task1: χ^2=30.168; p<0.05; Task2: χ^2=21.828; p<0.05; Task3: χ^2=13.172; p<0.05; Task4: χ^2=14.284; p<0.05; Task5: χ^2=11.975; p<0.05). In line with the average results, it was the mechanical engineer and IT engineering students who spent the most time on solving the tasks (Figure 7). Behind the better average results by programs we find higher average time consumption.

According to Table 1 and Figure 4 it is clear that the diagrammatic task containing known operations (Task5) caused serious difficulties for the students, therefore, we have analyzed its items more deeply. Online measuring allowed the simultaneous storage of the students' solutions and time consumption as well as their analysis. Concerning item 1 (proper answer: A), whichever answer the student selected (the number of proper answers was very low), the time spent on solution was equally high, and was the highest for one of the bad solutions (E) (Figure 8, upper graph).

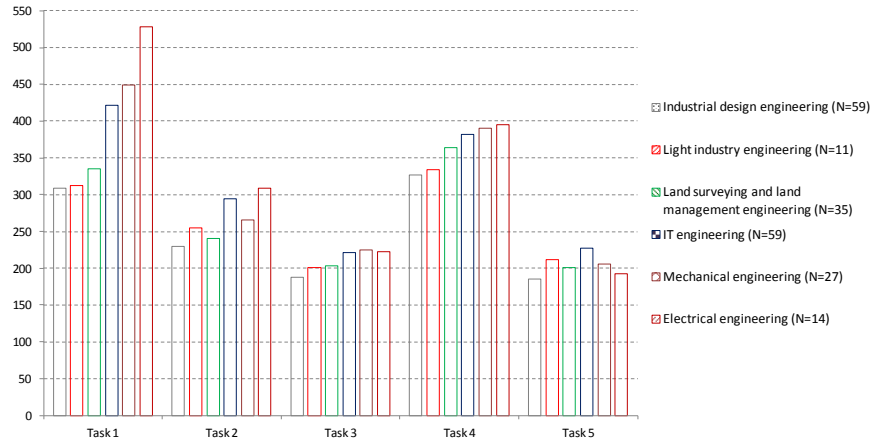


Figure 7
Average time used for task types by programs

However, as for items 5 and 6, the students chose the good answers at a higher rate although time consumption was much lower than it was at item 1, however, time used for the good solutions was the highest (Figure 8, middle graph). Based on the standard deviation analysis (Figure 8, lower graph) we can state that concerning the three items, it is the time consumption of the students having selected the good solutions where the difference is the lowest, so they came to the good solution within almost the same time. Higher time consumption and standard deviation values of item 1 and the higher number of bad solutions can be reasoned, in addition to its difficulty, by its highly novel nature.

Next, we will analyze the results of the students having achieved the highest total scores as compared to the whole sample (10% top), i.e. who are expected to have the most developed level of inductive cognition. In this case, students having collected 25 points or more are included in this category, altogether 32 persons. These students' scores (filled rectangles) and time consumption (empty rectangles) are presented in Figure 9, in terms of time consumption, in a growing order from left to right. Regarding the first five students, it is well visible that the students achieved high scores with relatively low time consumption. These respondents achieved 25-27 points within less than 18 minutes (quick-witted ones). The next category is made of 18 students who achieved similar results within less than 23 minutes (considered clever ones). The rest of the students used the available time almost fully (28-30 minutes) and reached good results (slow clever ones).

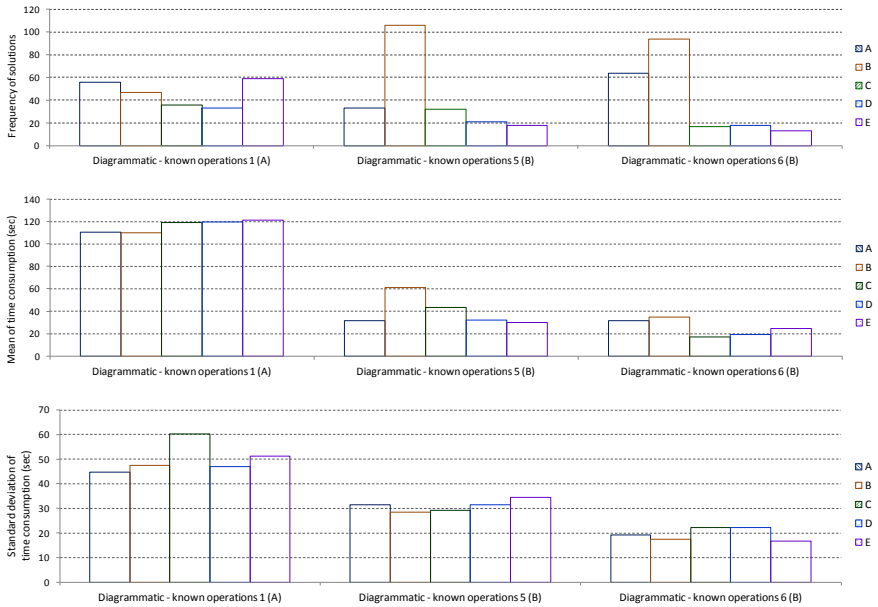


Figure 8

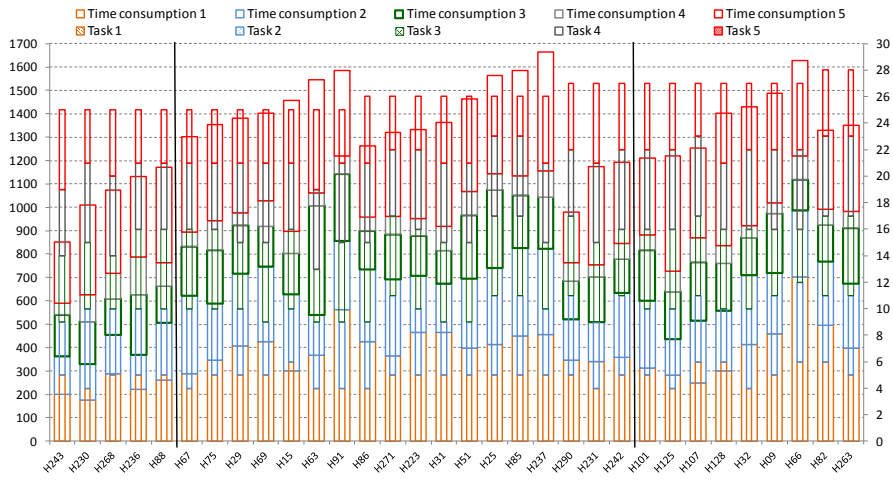
Solutions and time consumption of diagrammatic items (known operations)

To examine the relation between time consumption and inductive reasoning, the concept of *specific performance* was introduced by means of which the performance of the students with the best results was compared. In the inductive test, a specific performance was understood as the time necessary to achieve a unit of score, which was defined as the ratio of consumed time and the score achieved by items:

$$Time\ consumption_x / Score_x$$

where time consumption$_x$ indicated the time spent on solving task x (6 items) by seconds, while score$_x$ represented the score reached during this time.

The students with the highest scores were ranked based on specific performance (Figure 9). 250 sec/point as high specific performance was considered, which means that the students achieved a high number of points using little time. The values between 250 and 300 sec/point were assessed as specific performance of medium level, while the values above were considered low specific performance, i.e. a high amount of time was used to reach one unit of points. Specific performance can be used for further differentiation of the students and also expresses the pace of cognition. Students H243 and H91 achieved the same score, but the specific performance of the former one is better. The two students having gained the highest score (H82 and H263) could just be ranked into the group of those with good specific performance as they used more time for good solutions.



Remark: The left hand vertical axle indicates consumption time while the right hand one shows scores.

Figure 9

Time consumption of the students with highest total scores

We ranked the students according to the achieved scores, as well. The time consumption of the students with high scores appeared within a wide scale, and time consumption increased in parallel with the growth of the score achieved (e.g. the average time of those with 25 points was 1272, while that of the ones with 26 points was 1444 seconds), i.e. better performance required more time. At the same time, quick-witted students appeared in each score category.

Based on the background variables of the 32 students reaching the best results in the test, the following statements can be made: In the group of the students with the best results, the average of those that:

- Have parents have a degree
- Live in villages
- Live with their parents during their studies
- Took their secondary school leaving exams in four-grade schools
- Study in the IT engineering program
- Gained a mark 5 in Mathematics and Hungarian language and Literature at their school leaving exams was higher.

We also examined the relation between time consumption and achieved points for the whole sample (Figure 10). The relationship can be described in a reasonable way by a power function:

$$Score = 0.557 * Time\ consumption^{0.490}$$

This model accounts for 34.9% of all of the variances. The ANOVA test indicates a significant regressive relation (F=109.348; p<0.05). The result of the t-probe of the coefficients of the model shows that time consumption plays an irregularly significant part in the model (t= 10.480; p<0.05).

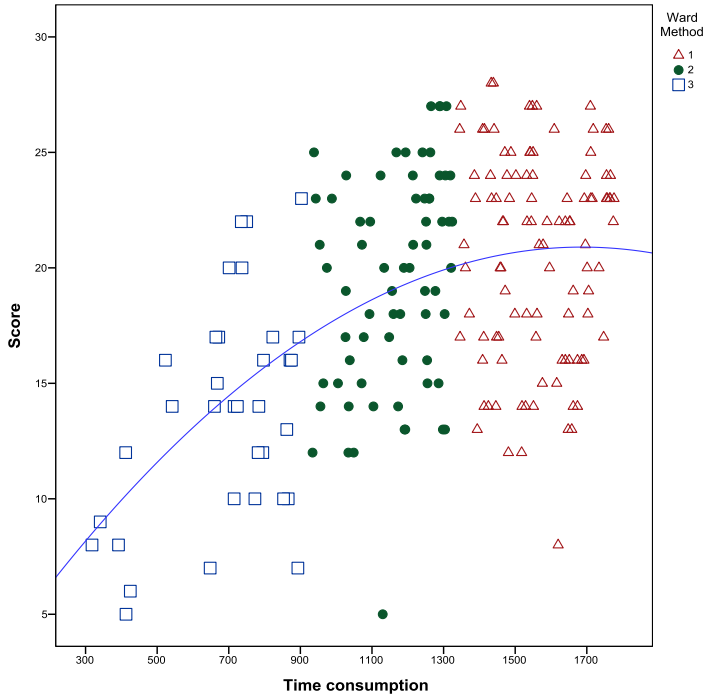


Figure 10

Clusters created in terms of the whole sample and relationship between the time consumption and the total score

In case school leaving exams in Hungarian language and literature and in mathematics are involved as independent variables, the new model will account for 39.1% of all of the variances (F=43.463; p<0.05), i.e. they have a relatively low impact. Regarding the coefficients of the linear model and according to the t-probe, both time consumption and the school leaving exam results in the two subjects have a significant weight in the relationship (t_{time}= 8.414; $t_{mathematics}$= -5.163; $t_{Hungarian}$= -3.101; for each variable p<0.05):

*Score= 16.367 + 7*10^{-3}*Time consumption – 1.954*Maths exam result – 1.211*Hungarian exam result*

The method of least squares assures that the distribution of the standardized random errors is normal (p>0.05), and standard deviation does not depend on the value of the target variable.

Based on the score achieved in the test and the time needed, we created student groups by cluster analysis (Figure 10), and as it can be seen, the basis of classification was the time needed to solve the tasks. The first group was made of the thorough ones using the whole of the time available (△), the second one involved the considered ones (●), while the third group included the impatient and shallow ones striving to be quick (□). The clusters were examined according to the cluster centroids (Table 2). The means were analyzed by ANOVA. We found significant differences between the center of certain clusters in terms of both the achieved scores ($F= 34.165$; $p<0.05$) and time consumption ($F= 577.226$; $p<0.05$). Time spent accounts for 85.0% of the standard deviation. The reliability of the results was checked by K-means algorithm but no major difference was found.

Table 2

Cluster centroids and standard deviation

Cluster		Score achieved in the test	Time used for task solution
1	M	20.57	1567.50
	N	103	103
	SD	4.491	124.255
2	M	19.41	1165.04
	N	68	68
	SD	4.662	118.523
3	M	13.25	660.75
	N	36	36
	SD	4.972	212.904
Total	M	18.92	1277.60
	N	207	207
	SD	5.306	364.164

We examined the composition of the clusters for the whole sample in terms of the background variables, as well, and proved it with the χ^2 probe that there was significant correlation between the place of living during the studies ($\chi^2= 32.538$; $p<0.05$), the program ($\chi^2= 34.442$; $p<0.05$), the gender ($\chi^2= 9.366$; $p<0.05$), the level of language exam ($\chi^2= 14.090$; $p<0.05$) and belonging to a certain cluster. In terms of these correlations, the contingency coefficient indicating the strength of the relation took these values: 0.369; 0.378; 0.208; 0.252. These are weak-medium values. The description of the clusters by main and background variables is summarized in Table 3.

The results the students of the five programs achieved in the inductive test taken as a function of the time used for task solution. The "focal points" defined according to the average results of the certain programs are also indicated. C1 mainly consists of students in mechanical engineering and IT engineering, C2 of industrial and product designers, while environmental engineering and land survey and land management engineering students mostly belong to cluster C3.

Table 3
Introduction of clusters

Cluster	C1	C2	C3
Description of cluster	thorough and clever	considered	impatient and shallow
Time used for task solution	1400 – 1800 sec	1000 – 1400 sec	600 – 1000 sec
Achieved result	16 – 26 points	15 – 23 points	10 – 19 points
Sex	male and female	female and male	male
Program	technological programs	designer programs	environment related programs
Residence during studies	with parents	in sublet or hostel	in hostel
Language exam	medium and/or advanced level	advanced level	none

6 Conclusions

Since in mathematical and natural scientific knowledge, inductive cognition plays a decisive role, this research aimed to:

(1) Describe the development level of inductive cognition of first grade students leaving secondary education and entering technical higher education.

(2) Prove or disprove the correlation between the time used for task solution and the result achieved in the inductive test.

The research implemented via an online test involved 253 students. By using the measuring tools [18] developed by the Psychometric Success WikiJob Ltd. (UK, London) we came to the following conclusions, in terms of the research participants.

6.1 Development Level of Students' Inductive Thinking

The students participating in the survey reached very much differing results in the five tasks. They performed the best in Task3 requiring analogue thinking and the weakest in Task5 containing known operation and requiring diagrammatic reasoning with high standard deviation values in terms of the latter. Difficulties here were probably caused by sensual – perceptual shifts.

We found significant differences in the average results of the certain tasks in line with the results of the school leaving exam in mathematics, i.e. the better result someone had achieved in mathematics, the better result could be expected in the

test. This concludes that the result of the secondary school leaving exam, in mathematics, closely reflects inductive cognition.

In the two task series requiring diagrammatic reasoning (Task4 and Task5), as well as the series requiring rule induction it was male students living with their parents during their studies and mechanical engineer, electric engineer and engineering IT students who achieved significantly better results.

6.2 Relationship between the Results Achieved in the Inductive Test and Time Consumption

Online measurement allowed the registration of time consumption by items, so it was possible to analyze and compare them to the achieved results.

We found close relationship in terms of the time spent on solving the certain tasks – each task consisted of six items – and the average results. Students spent the most time and, with high standard deviation (considerable personal differences), reached the weakest results in the tasks requiring rule induction (Task1) and developed diagrammatic cognition (Task4 and Task5). These phenomena may have several reasons and can be observed well by the applied time analysis by items. In the first case, the standard deviation of time consumption indicates the difficulty level of the given item, while it reflects the problems of understanding the task, as well as, the novel and unusual nature in the latter.

By taking into account the score achieved in the test and the time used for solution, three clusters were created:

(1) The thorough ones utilizing the available time

(2) The considered ones

(3) The impatient and shallow ones striving to be quick

The significant differences of these groups were proved by hierarchic and the K-means algorithm, as well. Regarding the composition of the certain groups and in terms of the whole sample, we found that there was significant relationship between the place of residence during the studies, the program, the gender and the level of language exam and belonging to the given cluster.

To describe success, we introduced the notion of specific performance, which we explained as the time needed to achieve a unit of score. We came to the conclusion that it is suitable, for further differentiation of the students' results and expresses the speed of thinking.

The relationship between the two variables is described by a power function and the model accounts for almost 35% of all of the variances:

$$Score = 0.557 * Time\ consumption^{0.490}$$

We made an independent analysis of the results of the students with developed inductive cognition (upper 10%). In this case, the three clusters gave this picture:

(1) Slow clever ones

(2) Considered to be clever ones

(3) Quick-witted ones

The characteristics of the latter group: their parents have a degree, they are living at their parents' during their studies, they took their secondary school leaving exams at four-grade schools, most of them are studying IT engineering and had marks 5 in mathematics and Hungarian language and literature at their school leaving exams.

The biggest challenge for engineering students is meeting the implications of math and science subjects. Secondary school mathematics graduation results only partially, unless it is an advanced level exam, reliably predict the development of logical thinking. Based on all this, it can be seen that at the beginning of higher education there is a need for an input competence measurement that establishes the development of logical thinking. These measurements can be linked to specific subjects or examine general abilities. In the first case, it could be a special math test, and in the second case, it would be a test measuring inductive thinking. The present research has chosen the latter path. Input competence measurement can provide a good basis for differentiated student development, exploration of learning difficulties, and talent development. Students who perform poorly in input measurement may be offered an optional course specifically designed to develop thinking, while students who achieve good results may be involved in Scientific Students' Associations (Tudományos Diákkör, TDK in Hungarian).

It is important to determine how effectively such an input competency measurement predicts drop-out. To prove this, the results measured in the inductive test must be compared with the students' recorded study results and study progress. This will be the goal of future research.

It is also important to talk about how closely the results obtained in the inductive test, agree with the time spent on the solution. Spending time indicates, on the one hand, the speed of thinking and, on the other hand, perseverance, diligence, patience, concentration, persistence of attention, that is, many attitudes that also play an important role in technical higher education studies.

Acknowledgements

This research has been supported by the project titled "VEGA-1/0663/19 Analysis of science and mathematics education in secondary schools and innovation of teaching methodology".

References

[1] European Commission/EACEA/Eurydice: The European Higher Education Area in 2018: Bologna Process Implementation Report. Luxembourg, Publications Office of the European Union, 2018

[2] Felsőoktatási elemzési jelentések (Analytical reports on higher education): Vol. 2, No. 3, September 2018

[3] Spady, W. G.: Dropouts from higher education: An interdisciplinary review and synthesis. Interchange, 1(1), 1970, pp. 109-121

[4] Bean, J. P.: Interaction effects based on class level in an explanatory model of college student dropout syndrome. American Educational Research Journal, Vol. 22, No. 2, 1985, pp. 35-64

[5] Kocsics, Zs., Pusztai, G.: Student Employment as a Possible Factor of Dropout. Acta Polytechnica Hungarica, Vol. 17, No. 4, 2020, pp. 183-199

[6] Engler, Á.: Non-traditional students in higher education. Hungarian Educational Research Journal, Vol. 9, No. 3, 2019, pp. 560-564

[7] Veroszta, Zs., Nyüsti, Sz.: Institutional effects on Bachelor-Master-level transition. International Journal of Social Sciences, Vol. 4, No. 1, 2015, pp. 39-61

[8] Kautz, T. D., Heckman, J., Diris, R., der Weel, B., Borghans, L.: Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. National Bureau of Economic Research, Cambridge, 2014

[9] Balcar, J.: Soft skills and their wage returns: Overview of empirical literature. Review of Economic Perspectives, Vol. 14, No. 1, 2014, pp. 3-15

[10] Carnevale, A. P.: 21st Century competencies for college and career readiness. National Career Development Association, Broken Arrow, 2013

[11] Eger, H., Grossmann, V.: Noncognitive abilities and within-group wage inequality. Institute for the Study of Labour, Bonn, 2004

[12] Manpower Group: Talent Shortage Survey Research Results. Manpower Group, Milwaukee, 2015

[13] Klauer, K. J., Phye, G. D.: Inductive Reasoning: A Training Approach. Review of Educational Research, Vol. 78, No. 1, 2008, pp. 85-123

[14] Phye, G. D.: Inductive problem solving: Schema inducement and memory-based transfer. Journal of Educational psychology, Vol. 82, No. 4, 1990, pp. 826-831

[15] Mackintosh, N. J.: IQ and human intelligence. Oxford University Press, Oxford, 1998

[16] Sternberg, R. J., Ben-Zeev, T. (Eds.): The nature of mathematical thinking. Lawrence Erlbaum, Mahwah, 1996

[17] Winch, C.: Language, ability and educational achievement. Routledge, New York, 1990

[18] Newton, P., Bristoll, H.: Numerical reasoning, verbal reasoning, abstract reasoning, personality tests. Psychometric Success. https://www.psychometric-success.com/ Accessed on March 2019

[19] Georgiev, N.: Item Analysis of C, D and E Series from Raven's Standard Progressive Matrices with Item Response Theory Two-Parameter Logistic Model. Europe's Journal of Psychology, Vol. 4, No. 3, 2008, https://doi.org/10.5964/ejop.v4i3.431

[20] Raven, J.: The Raven's Progressive Matrices: Change and Stability over Culture and Time. Cognitive Psychology, Vol. 41, No. 1, 2000, pp. 1-48

[21] Jacobs, P. I., Vandeventer, M.: Progressive Matrices: An experimental, developmental, nonfactorial analysis. Perceptual and Motor Skills, Vol. 27, No. 3, 1968, pp. 759-766

[22] Vodegel Matzen, L. B. L., Van der Molen, M. W., Dudink, A. C. M.: Error analysis of Raven test performance. Personality and Individual Differences, Vol. 16, No. 3, 1994, pp. 433-445

[23] Sternberg, R. J.: Component processes in analogical reasoning. Psychological Review, Vol. 84, No. 4, 1977, pp. 353-378

[24] Stieff, M., Hegarty, M., Dixon, B.: Alternative strategies for spatial reasoning with diagrams. In: Goel, A., Jamnik, M., Narayanan, N. H. (Eds.): Diagrammatic Representation and Inference. Springer, Berlin, Heidelberg, 2010, pp. 115-127

[25] Eger, L., Klement, M., Tomczyk, Ł, Pisoňová, M., Petrová, G.: Different User Groups of University Students and their ICT Competence: Evidence from three Countries in Central Europe. Journal of Baltic Science Education. Vol. 17, No. 5, 2018, pp. 851-866

[26] Szőköl, I., Nagy, M.: The Qualitative Transformation of Understanding the Essence of the Education Process and Learning. In: Chova, L. G., Martínez, A. L. (eds.): Proceedings of 12th Annual International Conference on Education and New Learning Technologies, Palma de Mallorca, 2020, http://doi.org/10.21125/edulearn.2020

[27] Harangus, K.: Examining the relationships between problem-solving and reading comprehension skills. New Trends and Issues Proceedings on Humanities and Social Sciences. Vol. 6, No. 5, pp. 66-74, https://doi.org/10.18844/prosoc.v6i5.4375

[28] Csehi, A., Kanczné Nagy, K., Tóth-Bakos, A.: Experience education to alleviate fears of educator candidates in their university studies. AD ALTA: Journal of Interdisciplinary Research, Vol. 10, No. 2, 2020, pp. 170-175, www.doi.org/10.33543/1002

Factors Affecting the Decision of Adoption Cloud Computing Technology: The Case of Jordanian Business Organizations

Thabit Atobishi[1], Miriam Bahna[1], Katalin Takács-György[2*], Csaba Fogarassy[3]

[1]Szent István University, PhD School of Business Management and Administration, Páter Károly u. 1, 2100 Gödöllő, Hungary
Thabet.Tubashi@moj.gov.jo; Miriam.Bahna@phd.uni-szie.hu

[2]Óbuda University, Keleti Károly Faculty of Business and Management, Institute of Management and Organization, H-1084 Budapest, Hungary
Takacsnegyorgy.Katalin@kgk.uni-obuda.hu

[3]Szent István University, Climate Change Economics Research Centre, Páter Károly u. 1, 2100 Gödöllő Hungary; Fogarassy.Csaba@gtk.szie.hu

*Corresponding author

Abstract: This research, is a clear case for developing countries, that realize the profits of cloud computing, to enhance their business and develop cutting-edge technology, but first, they need to address the concerns of decision-makers toward cloud computing technology. This research is based on a Technology Acceptance Model, which is modified to fit the context. The factors applied in this research are: Perceived Usefulness, Perceived Ease of Use, Perceived Risk/Security, Cloud Computing Awareness, compatibility, cost and an aim to use where it is a dependent variable in this study. Within this study, the Likert Scale (five-point scale) survey was produced, to get data from IT employees of the Jordanian business organization. Data was collected from 175 IT employees. Linear regression was used to examine the data using SPSS software. The outcomes revealed that perceived ease of use, perceived usefulness, cost, perceived risk were statistically significant, while compatibility and cloud computing awareness were found not to be statistically important. These outcomes explain that cloud computing systems can be quickly deployed in developing countries (avoiding unnecessary innovation); if users accurately understand their utility, it is clear that operating costs might be reduced. The perceived risk of system usage is low. We consider, of particular importance, the research line that technological compatibility is not significant, with the successful introduction of cloud applications. Fast deployment of cloud applications are good opportunities for developing countries, because of the lower cost of new software applications and use of big data.

Keywords: cloud technology; cloud computing; cloud security; Technology Acceptance Model (TAM); Technology-Organization-Environment framework (TOE); Jordan

1 Introduction

The new big difference in the information technology sector, currently takes place everywhere, using information technology services called "cloud computing." This new paradigm shift expects to reform all information technology industries and the way of delivering information technology services to the business organizations; it's expected that the leaders in the cloud computing industry will reach 160 billion revenue [1]. Many advantages organizations can gain by this new technology such as on-demand, on-service, more mobility access, shared resource pooling, rapid elasticity, and pay measured service, which ultimately drives to cost reduction and more efficiency [2]. Hence in the coming future, information technology services will deliver to a business organization as a fifth utility to be like water, telephone, gas, and electricity [3].

Despite its great benefits, the choice to utilize cloud computing holds several concerns about this latest technology. The perceived ease of use and perceived usefulness are essential factors in adopting new innovative technology [4, 5]. But also in cloud computing adoption, different parts can be classified as technology, human, and organizational elements [6]. Jordan, as a developing country still suffering a lack of cutting-edge technology, and the digital divide is high compared with developed countries. So that cloud computing considers as promised with its vast possibilities to support business organizations to overcome this digital divide, but like any new technology, many matters regarding the adoption of this technology. Trends in different years show that cloud computing will play an active role in business development in both developed and developing countries. One of the common critical questions is whether there is a variety in the dynamics of the introduction of new systems, and there are vital differences between the advantages and disadvantages of cloud computing in developed and emerging countries. What effects are expected on cloud computing innovation in Jordan or countries with similar levels of growth to Jordan?

Information and communication technology (ICT) sector is a promising sector in Jordan that can help in developing the economy and reduce the impact of its crisis specially the unemployment issue [7]. IT expenditures are growing exponentially for enterprises, which are primarily the purchase of hardware and software licenses, but their education and training costs are similarly rising [8]. Also 82% of the employees in business organizations use the internet in their work and have an access and about 70% of the business formations have an access to the internet as well as more than 3000 business organizations use the internet from outside of their workplaces [8]. The above information means Jordanian business organizations are capable to get the cloud computing services in the future [9]. Cloud computing can be an effective solution to allow the companies and entrepreneurs to utilize the cutting-edge IT solutions at a reasonable price through the cloud computing solutions: infrastructure as a service, data platform as a service and software as a service.

Since the internet advent, in the late 20th Century to the ubiquitous computing facilities of the present day, the internet has drastically changed the whole world of computing. It has moved from the parallel computing concept to distributed, grid, and now to cloud computing [10]. Conforming to [1], cloud computing is relevant to as the software applications and several other resources available online via the internet to multiple users, rather than being installed in local computers of certain users only. [11] also defined cloud computing as an act of performing business activities on off-premises and shared computing systems. The traditional models of software in which organizations were tied to applications and which were supplier-specific, restricted by the license of users, required software upgrades or patches, and incurred annual license fees are being displaced by the cloud computing solutions [12]. Also, the cloud computing system provides monetary benefits, which cannot be neglected by business organizations. However, many scholars identified privacy and security as the two biggest concerns in cloud computing [10, 13, 14]. It is because some people fear to yield their important confidential data to another company [15]. The authors inferred that customers pay cloud service vendors, when they find them reliable and trustworthy in security measures. Another factor that affects cloud computing is privacy. Data can be derived from any place, so the privacy of clients cannot be compromised [16]. In addition, reliability is also seen as a problem in cloud computing [17]. According to a report by the National Institute of Standards and Technology (NIST), cloud computing is enabled when needed and provides convenient network access to IT resources and applications. This includes storage, various servers, data networks, services, and other applications that need to build up to provide to develop and maintain services. A general question related to the use of cloud systems is what level of trust the systems can operate in, what are the security levels of data storage and data sharing in the business and private spheres, and whether their mixing is typical when using the system?

2 Issues that Impact the Decision to Utilize Cloud Computing

Multinational companies are focusing more on adopting information technology to compete in an ever-changing market. The customers are increasingly demanding more product choices and outstanding product quality at lower prices. That one way to get such needs of customers, organizations are utilizing more sophisticated IT systems. It is because reliable and timely information is the key to achieving efficiency in performance, and information technology are significantly considered as a tool to obtain the needed capability to endure competitively [18] [19] for instance the cloud computing technique supports an individual to access their work at any time and from anywhere with any portable device by reducing administrative cost to perform the business activities [20]. Such approach supports

suppliers to offer several virtual or real resources in a dynamic manner in the cloud [21]. However, Davis, Bagozzi & Warshaw [22] inferred that organizational performance could not be improved by computer systems if they are not used properly. In this connection, [23] identified that an essential factor that contributes to the contradiction of productivity is the low usage of installed information systems, which also defines less financial returns on investments in IT. Also, it seems that people are unwilling to adapt to the changes [24]. Moreover, risk presents in all projects of IT and the resistance of users can increase such risk. Therefore, information systems' successful implementation that ranges from simple applications, like spreadsheets and word places into more complicated apps, requires acceptance of users.

In the above context, a model has been proposed by Davis, Bagozzi & Warshaw [22] to estimate the manner in which users accept new technologies, which is known as TAM (Technology Acceptance Model). The particular model was produced to forecast further IT use and individuals' level of adaptation with it. TAM suggests two beliefs, including 'perceived ease of use' and 'perceived usefulness,' determining the behavioral aim of individuals to use information technology. Perceived usefulness is indicating to as the degree to which a person assumes that utilizing IT will develop his/her performance at the job while perceived ease of use is indicated to a degree to which a person considers that adopting a new IT system in his/her routine operations will require no additional efforts [25]. The model also provides that external variables (such as, design characteristics) affect behavioral intention that is mediated by two variables, 'perceived ease of use' and 'perceived usefulness.

Professional opinion written by Agarwal et al. [26] [27], in their diffusion of innovation theory, studied individuals' perception and their concerns regarding new innovative technical characteristics, which is not restricted to variables of TAM but comprise and evaluated two outcomes in such context in the manner of users' perception. First is new technology utilization and eagerness to keep on utilizing the same. Second is the perceived volunteer variable, which means whether the new technology adopters feel that supervisors are mandatory to deploy and use the latest technology or not. Tornatzky and Fleischer developed a multi-perspective TOE (technology-organization-environment) framework in 1990. It is an organizational-level theory, which characterizes only one part of the change process, such as how the company's context influences its capability to adapt and perform new technology [28]. In the subject of this framework, there are three factors of an organization's context, which influence its process to utilize technological innovation [29]. First is a technical context that represents the external and internal technologies of the companies, i.e., both the technology available in the market but not currently in use as well as the technology already in use by the firm. Such techniques might include either practice or equipment. Second is organizational context, which is relevant to the firm's characteristics and sources, like its managerial structure and size. For instance, correlated to small

firms, big companies have more capital to fund new technology adoption and implementation [30] [31]. The third is the environmental context, which is about the arena wherein the company carries out its business activities.

Based on the analyzes in the literature, it can be well defined that the technological conditions for the introduction of cloud systems do not pose a major financial challenge for businesses, but operation can be a major competitive advantage, especially for small businesses. The present research provides an answer to how user support can be used to use cloud systems.

The main question is; will the perceived usefulness of users increase the use of cloud computing? An important research question is the relationship between perceived simplicity and perceived usefulness of cloud computing. The subject of our study is whether the perceived ease of use, perceived utility, cost, and perceived risk are statistically significant during the operation of cloud computing systems.

2.1 Computing Awareness

Technological awareness “keeps abreast of available technology, understands applicability and limitations of technology to the work of the office, actively seeks to apply technology to appropriate tasks, shows a willingness to learn new technology” (glosbe.com, "technological awareness in English," 2019). Tarmidi, Rasid, Alrazi, & Roni, (2014) [32] assured that the level of awareness is a crucial factor in the adoption of cloud computing technology. So, when the level of awareness increases, then adoption and using the level of cloud computing will increase as well [33]. A study by [34] conducted on students showed a low level of awareness about cloud computing because of not enough necessary resources.

H1. Cloud computing awareness to use has a straight impact on Behavioral intention to use

2.2 Computing Perceived Ease of Use

First, Rogers [35] argued that perceived ease of use is a term that indicates the difficulty innovation is encountered, learned, or operated. He also declared that perceived ease of use is the extent to which users perceive a new product or service as a substitute [35]. The researchers said that the perceived ease of use means that a person considers it is realistic that applying a rigorous method does not cost that person. Perceived ease of use is when a user thinks that the use of a particular technology or system is open to mental and physical support. Also, it is equal to self-efficacy, which is related to 'judgments about how effectively one can implement a required course of actions, that one may carry out potential situations [22] [36].

H2. Perceived ease of use has a straight impact on behavioral intention to apply.

H8. Perceived ease of use has a straight impact on perceived usefulness

2.3 Perceived Usefulness

Belonging to the TAM, a perceived utility is the point to which one accepts that utilizing a precise system would improve work performance. According to [22], perceived usefulness is considered as a level to which user-perceived that a particular technology will advance his/her performance at the job. According to [22], perceived usefulness is considered as a level to which user-perceived that a particular technology will advance his/her performance at the job [1].

H3. Perceived usefulness has a straight impact on behavioral intention vital intention to use

2.4 Perceived Risk in Security and Privacy

Risks in cloud computing are mainly about security and privacy concerns [37]. According to [36], data security is the most critical barrier to choose cloud computing, as it goes along with several other significant challenges, including privacy, trust, legal matters, and compliance. As in the cloud computing mechanism, data are divided into various locations of machines like PCs, mobile phones, servers, and storage devices, the problem of data security converts more serious. The study has identified several aspects of security and privacy concerns in cloud computing environment such as accountability, confidentiality, preservability of privacy due to failure of physical control on data, integrity due to manipulation of data or dishonest computation on remote servers, and availability issues of bandwidth and pricings [36].

H6. Perceived risk has a straight negative impact on behavioral intention to apply.

2.5 Compatibility and Incompatibility

This variable came from the dissemination of innovation theory, the concept of the compatibility tool: "the degree to which an innovation is perceived as being consistent with the existing values, needs, and past experiences of potential adopters" [26]. Incompatibility in between needs of potential adopters and business process and features of innovation is perceived as the key barrier that affects the adoption of new technology [38]. Two factors determine the compatibility: the internal organizational infrastructure (e.g., current values organization's strategy and needs) and internal information systems environment (e.g., IT infrastructure)). Historically, in new technology adoptions, it showed that first adopters tend to point upon the expected benefits of the latest technology and

on the compatibility between the new technology and the current organizational environment and infrastructure [39].

H4. Compatibility has a straight impact on behavioral intention to use

H7. Compatibility has a straight effect on perceived usefulness

2.6 Cost as a Vital Factor

Previous studies about the adoption cloud computing indicated that cost factor is a key determinate of cloud adoption [40]. Compared to end-users and business executives who have been generally researched in previous research, IT employees may make different decisions about utilizing the technology. For example, many IT employees do not have a high level of technological knowledge (no info about the price of technology) despite their general competence and learning ability [41] [42]. Cost is a vital factor when the organization decides to adopt new technology. Many types of research have indicated that TAM needs to be integrated with other variables that may give a better adoption model [43]. Behavioral decision theory emphasized that the cost-benefit pattern is significant to both perceived usefulness and ease of use in the choice to utilize new technology [43].

H5. The cost has a straight negative result on behavioral intention to use

3 Method and Data

TAM model was mainly developed to assess how technology affects the user's behavior, and it uses two factors: perceived usefulness and perceived ease of use, but it has been examined because it gives only general information about the adoption level by the individuals [44]. In our proposed model, we investigate the individual adoption level within the organizational context. So, we need to extend the TAM model and use additional variables from other theories that adjust on the organizational scale, such as TOE theory. The suggested research model was derived based on other studies that extended the TAM model: a study by [45] (Gangwar, Date & Ramaswamy, 2015) and a study by [37].

In the first step, we have defined the essential factors that influence cloud computing decisions (Figure 1). By analyzing the previous models and reviewing existing literature, the following variables (these variables can be further expanded based on sectoral differences) are recognized as the most critical to be investigated throughout this study (based on TAM):

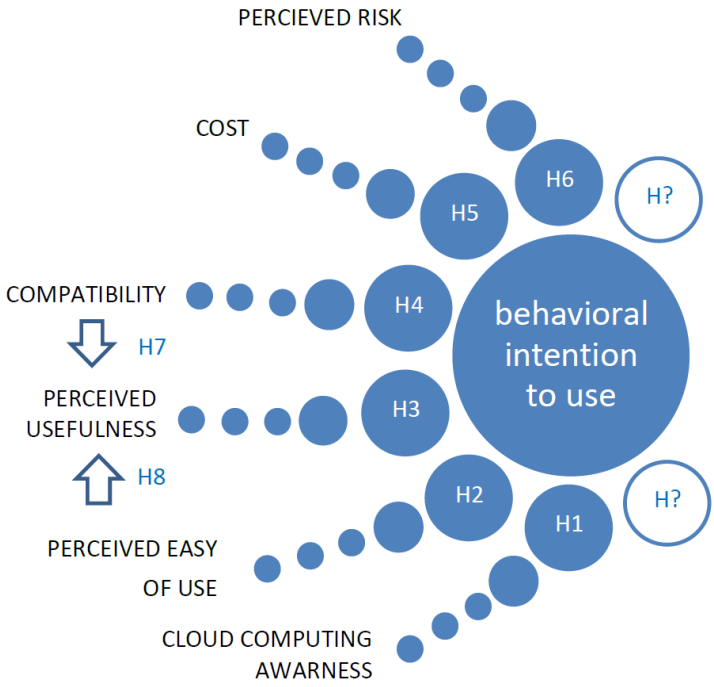


Figure 1
Essential factors of cloud computing decision

3.1 Quantitative Research

This is quantitative research depends on a regression model and descriptive analysis. The cause and impact relationship of variables is investigated. The study involved a sample of N = 175IT employees from different business organizations in Jordan. The data was collected using a 23-item quantitative scale, which assessed the factors that influence the decision making for using cloud computing services. The questionnaire was based on seven contracts; Cloud Computing Awareness, Perceived Risk, Cost, Compatibility, Perceived Usefulness, Perceived Ease of Use, and Behavioural Intentions to Use. The scale reliability was 0.74, and among the constructs, Perceived Risk and Perceived Ease of Use have High reliability, Cost, Perceived Usefulness and Behavioural Intensions to Use, have Above Average reliability while the subscales Cloud Computing Awareness and Compatibility have average reliability. Data collection and analysis was done in line with the research model presented herein. Though it would be a systematic procedure to ensure all techniques address the research problem along with hypotheses outlined for research. Also, a criterion to answer questions in research undertaking focused at, in a pure, comprehensive atmosphere to easily be understood by decision-makers in Jordan investments sector at the national level.

We, therefore, present the following intended methods to be used in primary data collection. They are the methods identified as the best methodology in the course of data collection by the point that data collection should be of high accuracy, as well as address the significant scope of the study. Interview sessions with IT specialists, computer systems engineers, and CIOs. Nearly 220 questionnaires were supplied to IT employees, in the main commercial and industrial cities of jordan: Amman, Zarqa and Irbid. We got back 175 valid questionnaires. Mainly the survey was submitted in traditional form in hand, but in some cases it was send by email. The questionnaire was built on previous studies and reviewed by the supervisors and by an external consultants (appendix A).

By way of the goal of our research, we chose quantitative research methods to meet the objectives of this study. The questionnaire, based on the study model and by using previously tested and validated instruments, was utilized to get the IT employee s 'responses regarding the overall intention to use cloud computing technology, and the same data is also used for validation of the model using statistical analysis.

After all these methods, mainstreaming of data from all primary sources was on course. Badges shall be provided to guarantee the right people are on the ground conducting research. Since the data collected for this research is quantitative; therefore, quantitative data analysis techniques would be applied. Moreover, different statistical measures would be calculated based on the nature of the analysis. If the analysis addresses the system evaluation, one of the research questions of this study, then descriptive statistics would be applied. For model validation, statistical measures for hypothesis testing calculated. The test of the structural model includes the estimation of the route coefficients in conjunction with coefficients of determination. Path coefficients show the strengths of the correlations between the dependent and independent variables [41] [46].

3 Results and Discussion

3.2 Demographic Characteristics

The questionnaire included questions associated with demographic characteristics such as age, gender, and position in an organization. Following Tables 1 and 2 show a summary of those characteristics.

Table 1 shows the age of the members and records that the most number of members were from age 30-39. And the least number of members was from the age range of more than 39. There were 175 members, and no data values were missing. It shows the gender of the members, and the values show that there were

more male participants with a percentage of 46.4% and fewer female participants with a percentage of 42.9%.

Table 1
Demographic Characteristics

		Frequency	**Percent**
Age	Less than 30	57	32.9
	30-39	66	37.7
	More than 39	52	29.7
	Total	**175**	**100**
Experience	Less than five years	41	23.4
	5-10 years	69	39.4
	More than ten years	65	37.1
	Total	**175**	**100**
Gender	male	91	52
	female	84	48
	Total	**175**	**100**
Position	top-level	46	52.2
	middle level	69	35.2
	non-management	60	30.6
	Total	**175**	**100**
Size of Organization	Small	46	26.2
	Medium	66	37.7
	Large	63	36.1
	Total	**175**	**100**

The table also displays the position of employees in their respective organizations. The table presents that the most number of members from the middle level of management and the least of members were from the top level of management, which means that the study results can be concluded on the population of Middle-Level Management positions. The size of organizations is also presented in the table in which the participants have been working. The table records that the most number of members were from Medium size organizations. A medium-sized organization is the one that has 50-250 employees.

3.3 Reliability of Scale

The reliability of the scale reflects the size to which the measure of a scale is dependable or consistent. Reliability analysis was used as a scale to assess the reliability in this study. The following are the results of the reliability analysis.

Table 2
Reliability of Scale

No	Domain	Alpha	Item No
1	Cloud Computing Awareness	6.73	3
2	Perceived Risk	8.46	4
3	Cost	7.54	3
4	Compatibility	6.98	3
5	Perceived usefulness	7.87	3
6	Perceived Ease of Use	8.23	3
7	Behavioral Intentions to Use	7.16	4
	Value of Cronbach's Alpha	**0.74**	**24**

The above Table 2 shows that the questionnaire had a reliability value of Cronbach's Alpha=0.74, which reflects the Good reliability of the scale. This shows that the model is an excellent fit for data. It shows that among the subscales, Perceived Risk and Perceived Ease of Use have High reliability, Cost, Perceived Usefulness, and Behavioral Intensions to Use, have Above Average reliability. In contrast, the subscales Cloud Computing Awareness and Compatibility have average reliability.

3.4 The Regression Analysis

To assess the impact of factors that are likely to affect decision making in using cloud computing, Regression analysis was used. The study hypothesized impacts of different factors which are stated below:

H1. Cloud computing awareness to use has a straight impact on behavioral intention to use

Table 3
Regression analysis for Cloud computing awareness variable

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Cloud Computing	1.655	0.100	0.600	0.737	0.642	110.736	0.325

Table 3 shows that there is no statistically significant effect at a considerable level ($\alpha \leq 0.05$) of Cloud computing awareness on behavioral intention to use, where "f" value reached (110.736) by statistically insignificant (0.325). (R) Value reached (0.737), and (R^2) value reached (0.642).

H2. Perceived ease of use has a straight impact on behavioral intention to use

Table 4

Regression analysis for Perceived ease of use variable

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Perceived ease of use	0.761	0.001	0.021	0.684	0.743	110.823	0.000

The above Table 4 shows that there is a vital impact of Perceived ease of use on behavioral intention to apply at a considerable level ($\alpha \leq 0.00$), where "f" value reached (110.823) by statistically insignificant (0.00). (R) Value reached (0.684) and (R^2) value reached (0.743).

Table 5

Regression analysis for Perceived usefulness variable

H3. Perceived usefulness has a direct effect on behavioral intention to use

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Perceived usefulness	0.987	0.000	0.029	0.579	0.637	112.464	0.001

The Table 5 shows that there is a statistically vital effect at a considerable level ($\alpha \leq 0.05$) of Perceived usefulness of use on behavioral intention to use, where "f" value reached (112.8464) by statistically insignificant (0.01). (R) Value reached (0.579), and (R^2) value reached (0.637).

Table 6

Regression analysis for Compatibility variable

H4. Compatibility has a direct effect on behavioral intention to use

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Compatibility	0.001	0.975	0.034	0.639	0.742	110.346	0.253

The above Table 6 shows that there is no statistically significant effect at a considerable level ($\alpha \leq 0.05$) of compatibility on behavioral intention to use, where "f" value reached (110.346) by statistically insignificant (0.253). (R) Value reached (0.639) and (R^2) value reached (0.742).

Table 7

Regression analysis for Cost variable

H5. The cost has a straight negative result on behavioral intention to apply

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Cost	-15.131	0.00	-0.540	0.755	0.570	228.953	0.00

The above table shows that there is a statistically vital effect of cost on behavioral intention to use at a considerable level (α≤0.05), where "f" value reached (228.953) by statistically substantial (0.00). (R) Value reached (0.755) and (R^2) value reached (0.570).

Table 8

Regression analysis for Perceived risk variable

H6. Perceived risk has a straight negative result on behavioral intention to use

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Perceived risk	-14.895	0.00	-0.492	0.749	0.561	220.799	0.00

The above Table 8 shows that there is a statistically significant effect at a substantial level (α≤0.05) of Perceived risk on behavioral intention to use, where the "f" value reached (220.799) by statistically significant f sig (0.00). (R) Value reached (0.749) and (R^2) value reached (0.561).

Table 9

Regression analysis for Compatibility variable

H7. Compatibility has a straight impact on perceived usefulness

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Compatibility	3.014	0.768	0.025	0.752	0.643	112.248	0.560

The above Table 9 shows that there is no statistically significant effect at a substantial level (α≤0.05) of compatibility on perceived usefulness, where "f" value reached (112.248) by statistically insignificant (0.560). (R) Value reached (0.752), and (R^2) value reached (0.643).

Table 10

Regression analysis for Perceived ease of use variable

H8. Perceived ease of use has a straight impact on perceived usefulness

Independent variable	"t" value	"t" sig	B	R	R2	"f" value	"f" sig
Perceived ease of use	36.357	0.000	0.761	0.793	0.647	116.375	0.000

The above Table 10 shows that there is a statistically significant effect at a substantial level (α≤0.05) of Perceived ease of use on perceived usefulness, where "f" value reached (116.375) by statistically insignificant (0.00). (R) Value reached (0.793), and (R^2) value reached (0.647).

Based current literature, the context of an organization has three factors that affect the method of introducing technological innovation. The first is the technical

environment, which represents the external and internal technologies of the companies. Based on the answers to questions about technology compatibility, we believe that there is no vital relationship between the successful introduction of cloud applications and some degree of familiarity with technology. The second is the organizational context, which is relevant to the characteristics and resources of the company, like its management structure and size.

According to the studies, there is a clear correlation, meaning that more significant and more powerful companies are implementing new technology solutions faster and more successfully. The third is the business environment, which belongs to the business arena, where the company conducts business. The answers to the question perceptual usefulness are significant, so we can say that business partners in the studied environment do influence the successful introduction of cloud applications.

Conclusions

By discussing all of the above factors, it can be concluded that cloud computing is an emerging and advanced technology, within the industrial set-up. It has many benefits related to cost reduction, time consumption, compatibility of staff, equal platform provision, virtualization, flexibility, security, etc. All these factors greatly contribute to the path of success, for not only the organizations, but also for staff members. It reduces their burden, as well as, the organizational load. Hence, to reach the goal formulated by the companies and to compete within the market, cloud computing adoption becomes a certain factor. Moreover, in the current study, hypothesis are approved that perceived usefulness increases the use of cloud computing. And there is also link with perceived ease and perceived usefulness of cloud computing. The results of our research shows that perceived ease of use, perceived usefulness, cost, perceived risk are statistically significant, while compatibility and awareness of cloud computing are not statistically vital. The result is a important achievement because, as we have described, in the literature, compatibility and cloud computing awareness are the conditions that have been recognized, to enhance the spread of cloud computing systems. Therefore, there may be a vital distinction between developed and developing countries in the success rate of cloud-based systems. Studies have shown that ease of use and perceived usefulness in Jordan, and in countries with a similar economic development to Jordan, should be the focus of decision-making mechanisms. It also clearly shows that the attitude of companies, is market-oriented and it is the demand, rather than technological momentum and innovation, that defines the speed of adaptation of new systems. Future cross company comparisons should be executed for the evaluation of the adoption of cloud computing. The results show that, according to IT professionals, cloud computing systems can be deployed quickly in developing countries and perhaps the avoidance of costly and risky technological innovations. If users have a clear understanding of their usefulness, it is clear that operating costs can be reduced. The perceived risk of using the system is low. We consider it a particularly

important research outcome that technology compatibility is not that significant with the successful deployment of cloud applications. The rapid deployment and deployment of cloud applications offers good growth opportunities for businesses in developing countries, due to lower costs for new software applications and big data usage. The present study is realized from just one culture. Future research may include comparisons of cross-cultures, so that this phenomenon can be explored around the globe. The sample size herein was small, future sample sizes should be increased. Equal opportunity should be given in the sampling means and male/female participation ratios should be similar. The present research quantitatively assesses the phenomenon only. It is recommended in the future, to apply further qualitative means of exploration and a comparison should be made between professional and beginner levels, for the use of cloud computing and their attitude. A comparison should also be made between private and public sectors.

Acknowledgment

The authors would like to show the acknowledgment of Szent Istvan University Climate Change Economics Resreach Centre and the National Research, Development and Innovation Office – NKFIH, OTKA 131925.

References

[1] Behrend, T. S., et al., Cloud computing adoption and usage in community colleges. 2011. 30(2): pp. 231-240

[2] Mell, P. and T. Grance, The NIST definition of cloud computing. 2011

[3] Rodríguez Monroy, C., et al., The new cloud computing paradigm: the way to IT seen as a utility. 2012. 6(2): pp. 24-31

[4] Davis, F. D., A technology acceptance model for empirically testing new end-user information systems: Theory and results. 1985, Massachusetts Institute of Technology, pp. 5-12

[5] Rudnák, I. and J. Garamvölgyi, Hungary's young, technically educated workforce engaged in working abroad. Hungarian Agricultural Engineering, 2016(29): pp. 40-43

[6] Lian, J.-W., D. C. Yen, and Y.-T. J. I. J.o. I. M. Wang, An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. 2014. 34(1): pp. 28-36

[7] Al-Soud, A. R. Towards the adoption of cloud computing for boosting the Jordanian e-government. in European, Mediterranean & Middle Eastern Conference on Information Systems 2013 (EMCIS2013), Windsor. 2013

[8] DOS, Jordan statistical year book 2016 available : http://dosweb.dos.gov.jo/ar/publications/. Accessed 16 septemper 2019, Department of statistics, 2016

[9] DOS, Department of statistics ,Amman - Jordan . available : http://dosweb.dos.gov.jo/ar/. Accessed 15 March 2020, Department of statistics, 2020

[10] Jadeja, Y. and K. Modi. Cloud computing-concepts, architecture and challenges. in 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET) 2012, IEEE

[11] Rader, D. J. S. and Leadership, How cloud computing maximizes growth opportunities for a firm challenging established rivals. 2012, 40(3): pp. 36-43

[12] Hinde, C., J.-P. J. I. J. o. C. S. Van Belle, and E. Engineering, Cloud computing in South African SMMEs: Risks and rewards for playing at altitude. 2012, 1(1): pp. 1-10

[13] Kshetri, N. J. T. P., Privacy and security issues in cloud computing: The role of institutions and institutional evolution. 2013, 37(4-5): pp. 372-386

[14] Liu, F., et al., Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications. 2013, 20(3): pp. 14-22

[15] Boros, A. and C. Fogarassy, Relationship between Corporate Sustainability and Compliance with State-Owned Enterprises in Central-Europe: A Case Study from Hungary. Sustainability, 2019, 11(20): p. 5653

[16] Pearson, S., Privacy, security and trust in cloud computing, in Privacy and security for cloud computing. 2013, Springer, pp. 3-42

[17] Sanaei, Z., et al., Heterogeneity in mobile cloud computing: taxonomy and open challenges. 2013, 16(1): pp. 369-392

[18] Laudon, K. C., &Laudon, J. P, Management information system. 2016: Pearson Education India

[19] Huang, P.-Y., et al., The role of IT in achieving operational agility: A case study of Haier, China. 2012, 32(3): pp. 294-298

[20] Ma, X., Y. Cui, and I. J. P. C. S. Stojmenovic, Energy efficiency on location based applications in mobile cloud computing: a survey, 2012, 10: pp. 577-584

[21] Hamdaqa, M. and L. Tahvildari, Cloud computing uncovered: a research landscape, in Advances in Computers. 2012, Elsevier, pp. 41-85

[22] Davis, F. D., R. P. Bagozzi, and P. R. J. M. s. Warshaw, User acceptance of computer technology: a comparison of two theoretical models. 1989, 35(8): pp. 982-1003

[23] Sichel, D. E., The computer revolution: An economic perspective. 2001: Brookings Institution Press

[24] Gonçalves, J. M. and R. P. J. P. T. da Silva Gonçalves, Overcoming resistance to changes in information technology organizations. 2012, 5: p. 293-301

[25] Sentosa, I. and N. K. N. J. R. W. Mat, Examining a theory of planned behavior (TPB) and technology acceptance model (TAM) in internetpurchasing using structural equation modeling. 2012, 3(2 Part 2): p. 62

[26] Agarwal, R. and J. J. D. s. Prasad, The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. 1997, 28(3): pp. 557-582

[27] Sen, S. K., and Agarwal, R. P. Perusing the minds behind scientific discoveries, Studies in Systems, Decision and Control, Volume 179, 2019, pp. 3-33

[28] Baker, J., The technology–organization–environment framework, in Information systems theory. 2012, Springer, pp. 231-245

[29] Dobos, P. and K. Takács-György, The Factors Influencing The Emergence of Unethical Business Behavior. International Journal of Contemporary Management, 2018, 17(3): pp. 51-75

[30] Saedi, A. and N. A. Iahad. An Integrated Theoretical Framework for Cloud Computing Adoption by Small and Medium-Sized Enterprises. in PACIS. 2013

[31] Z. Bártfai, Z. B., L. Kátai and I. Szabó Synergy of robotics and ict in the smart farming. Hungarian Agricultural Engineering, 2019, N° 36/2019 pp. 22-28

[32] Tarmidi, M., et al., Cloud computing awareness and adoption among accounting practitioners in Malaysia. 2014, 164: pp. 569-574

[33] Pikkarainen, T., et al., Consumer acceptance of online banking: an extension of the technology acceptance model. 2004, 14(3): pp. 224-235

[34] Bagish, S. J. I. J. o. E. D. and Research, Student's Awareness of Cloud Computing: Case Study Faculty of Engineering at Aden University, Yemen. 2014: pp. 1122-1129

[35] Rogers, E. M., Diffusion of Innovations. Free Press, New York, 1983: p. pp. 15-45

[36] Xiao, Z., Y. J. I. c. s. Xiao, and tutorials, Security and privacy in cloud computing. 2012, 15(2): pp. 843-859

[37] Zeqiri, A., et al., An empirical investigation of cloud computing usage in education. 2017(3): pp. 77-85

[38] Nedev, S., Exploring the factors influencing the adoption of Cloud computing and the challenges faced by the business. Enquiry, 2014

[39] Lin, A. and N.-C. J. I. J. o. I. M. Chen, Cloud computing as an innovation: Percepetion, attitude, and adoption. 2012, 32(6): pp. 533-540

[40] Dwivedi, Y. K., et al., Small firm e-business adoption: a critical analysis of theory. Journal of enterprise information management, 2009

[41] Chau, P. Y., P. J.-H. J. I. Hu, and management, Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. 2002, 39(4): pp. 297-311

[42] Bálint, K., Cvetković, D., Takács, M., Holik, I. and Tóth, A. Connecting Bitcoin Blockchain with Digital Learning Chain Structure in Education. Acta Polytechnica Hungarica, Vol. 16, No. 1, 2019, p. 79

[43] Wu, J.-H., S.-C. Wang, and L.-M. Lin. What drives mobile health care? An empirical evaluation of technology acceptance. in Proceedings of the 38th Annual Hawaii International Conference on System Sciences. 2005, IEEE

[44] Liu, I.-F., et al., Extending the TAM model to explore the factors that affect Intention to Use an Online Learning Community. 2010, 54(2): pp. 600-610

[45] Gangwar, H., H. Date, and R. Ramaswamy, Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. Journal of Enterprise Information Management, 2015

[46] Garai, A., Péntek, I. and Adamkó, A. Revolutionizing Healthcare with IoT and Cognitive, Cloud-based Telemedicine. Acta Polytechnica Hungarica, Vol. 16, No. 2, 2019, p. 165

Appendix (A) Questionnaire

Introduction

In the frame of a PhD research program we are conducting a study entitled with measuring factors that affecting the decision of adoption of cloud computing services in Jordanian business organizations. Please answer all the enclosed items of the questioner by adding (x) under the suitable alternatives given in each item, knowing that your answers will be used only for scientific research purposes.

Thanks for your cooperation.

Note:

Cloud computing is the new generation of IT services paradigm. Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Essential Characteristics of this technology: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service.

First part: personal information.

Gender: Male Female

Age: less than 30 30-39

more than 39

Experience: 5-10 years less than 5 years

More than 10 years

The occupying position : ☐ lower level IT management

☐ Middle level IT management

☐ Top level IT management

Second part: Organization information

Organization size: ☐ small (Number of employees >50)

☐ Medium (50 <=number of employees<=249)

☐ Large (Number of employees>=250)

What is the primary type business conducted by your organization?

☐ Agriculture, forestry and fishing ☐ Mining

☐ Manufacturing ☐ Electricity

☐ Construction ☐ wholesale trade

☐ Retail trade ☐ accommodation, hospitality, food/beverage services

☐ Banking and insurance service ☐ real state service

☐ Health care ☐ public administration

☐ Education and training ☐ transportation

Where is the market scope of your organization?

☐ Local ☐ national ☐ international

Item	strongly agree	agree	neutral	disagree	strongly disagree
Cloud computing awareness					
I have received enough information about cloud computing					
I have received enough information about benefits of using cloud computing					
I have enough information about services that are offered through cloud computing					
Perceived risk					
I think using cloud computing in monetary transactions has potential risk					
I think using cloud computing in merchandise services has potential risk					
I think using cloud computing in product purchases has potential risk					
I think using cloud computing puts my privacy at risk					

Cost					
I think the equipment cost is expensive of using cloud computing to offer integrated services					
I think the access cost is expensive of using cloud computing architecture of my organization					
I think the transaction fee is expensive of using cloud computing would be easy					
Compatibility					
Using cloud computing is compatible with most aspects of my tasks					
Using cloud computing fits my work					
Using cloud computing fits well with the way I like to engage in doing my work					

Perceived usefulness					
I think Using cloud computing allow me to manage business operation efficiently					
I think Using cloud computing allow me to increase business productivity					
I think Using cloud computing enables me to do my organizational task more quickly					
Perceived easy of use					
I think Learning to run and use the cloud services would be easy for me					
I think my interaction with the cloud computing services would be clear and understandable					
I think that cloud computing services is easy to use.					

Behavioral intension to use					
Generally, I think that using cloud computing services is advantageous					
Generally, I am in favor of using the cloud computing services					
Assuming I had access to cloud computing, I intend to use it					
Given that I had access to cloud computing , I predict that I would use it					

Any additional notes

……
……
……
……
……..

Lower Leg Characteristics Influance on Hopping Height

Bálint Kovács, Örs Sebestyén, József Tihanyi

Deparment of Kinesiology, University of Physical Education
Alkotás u. 44, H-1123 Budapest, Hungary
e-mail: kovacs.balint@tf.hu, aytw4p@neptun.tf.hu, tihanyi@tf.hu

Abstract: It has been suggested that certain lower leg variables affects running and jumping performance. However, it is unclear how these variables interact with each other, and how they jointly affect hopping performance. We assume that AT-MA influences the physical performance differently depending upon biomechanical variables of the plantar flexor muscles. We hypothesize that no single variable can explain hopping height, rather a combination of lower leg variables can explain the variation in hopping height. Healthy young adults (n = 28, age 21.8 ±4.0 yrs) performed serial hops on a force plate during which we recorded right leg joint kinematics, lateral gastrocnemius fascicle behavior, and plantar flexor electromyography activity. We found no correlation between hopping height and AT-MA (r=0.28, p=0.14). Multiple regression analyses revealed that variations in AT-MA, ankle dorsiflexion amplitude, and peak ground reaction force explained 53% of the variation in hopping height. We concluded that even a combination of selected biomechanical variables can only moderately account for hopping performance.

Keywords: Human ankle; moment arm; EMG; vertical jump

1 Introduction

Biomechanical properties of the plantar flexors (i.e. ankle joint range of motion, muscle electrical activity, fascicle, and tendon behavior) and the structure of the ankle joint (i.e. moment arm length), are key determinants of vertical hopping performance. Because the product of muscle force and moment arm length is muscle moment, it is reasonable to expect that individuals with longer AT-MA would generate higher plantar flexors moments with lower and faster force generation hence greater efficiency. Against such expectations, experimental data seem to suggest that sprint performance (Kumagai et al., 2000; Lee and Piazza, 2009; Baxter et al., 2012) and running economy are associated with shorter instead of longer AT-MA [1, 36, 41].

Achilles tendon moment arm length (AT-MA) can increase or decrease the magnitude of ankle joint power in various types of joint movement [3, 37, 45, 46].

Because of these contradictory results it remains uncertain how the AT-MA length affecting joint output. We can assume that different types of joint movements might benefit from a long AT-MA while other types of movements benefit from a short one. Also, it is reasonable to assume that other important variables which are also influence joint output (i.e. muscle strength, neuromuscular activation pattern (force generation capacity), joint range of motion (ROM), fascicle, and tendon behavior) affected by the AT-MA length.

Two-legged hopping mainly uses ankle thrust which makes it an ideal task to experimentally examine the effects of the ankle joint properties (AT-MA, fascicle length, plantar flexor length changes, muscle activation) on motor performance under dynamic condition. Repetitive vertical hopping has been extensively studied [4, 6, 9, 10, 15, 16, 25, 42], but only a few studies have focused on the effects of AT-MA length on hopping performance [44, 45]. Volleyball players and runners have shorter moment arm than controls and while volleyball players jump the highest, such differences in jumping performance were independent of AT-MA length but other ankle properties were not examined [44]. In contrast, there was a strong correlation of $r = 0.74$ between heel length, a proxy measure of AT-MA length, and maximal jump height, a measure of mechanical power [45]. However, it should be noted that in this study the vertical jump was carried out with plantar flexion only (static jump), i.e., the plantar flexors contracted concentrically, which can explain the benefit of a longer moment arm based on the force-velocity-power relationship [18].

The magnitude of plantar flexor force and ankle ROM can affect hopping height [29, 30]. While longer AT-MA length can produce greater muscle moments [46], individuals with a short AT-MA length can rotate the ankle through a greater ROM and jump higher [41, 44]. The larger ROM is also associated with greater negative and positive accelerations of the center of mass during hopping and thus increase jump height [29, 45]. Thus, greater ankle ROM lengthens the muscle-tendon unit more, producing greater shortening velocities, positive work, and ultimately jump height. Indeed, athletes with greater jump height operated muscle fascicle close to resting length, which increased tendon lengthening during joint flexion [40]. The longer tendon stretch increases the amount of the stored elastic energy, which can be reused in the subsequent joint extension [33]. When fascicles operate in a quasi-isometric state, the greater tendon lengthening requires a greater muscle tensile force, resulting in an increase in muscle activation. In particular, during a powerful and fast stretch of the MTU, such as during hopping, the short-latency stretch reflex can become active [28]. Such a cascade of events, i.e., increased muscle activation and muscle force generation underlies higher jumps.

The purpose of the study was to investigate what/which set of variables could predict the most accurately hopping height. In addition to previous studies, we aimed to investigate also the relation between AT-MA and 1) ankle joint ROM, 2) plantar flexor electromyography activity, 3) LG fascicle behavior and 4) Force-

related variables derived from a force platform. We hypothesized that longer AT-MA would associate with lower plantar flexor electromyography activity, greater maximal vertical ground reaction force, and force development ratio. In addition, we do not expect that AT-MA correlates with hopping height, rather than correlate with the aforementioned variables which contribute to achieving superior hopping height.

2 Methods

2.1 Participants

Healthy young adult volunteers participated in the study (n=28, age 21.8 ±4.0 yrs, height 1.80 ±0.8 m, body mass 80.2 ±6.1 kg). They were free of pain and had no injuries to the lower extremities over the past two years. Participants gave written informed consent prior to the start of the study, which was performed in accordance with the Declaration of Helsinki and was approved by the local ethics (TE-KEB/No5/04/2017).

2.2 Experimental Protocol

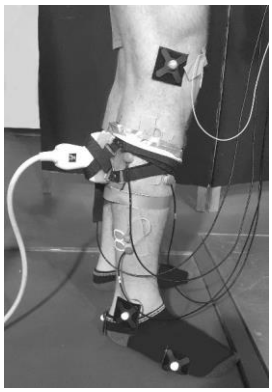


Figure 1
Installation and fixation of US probe, EMG electrodes on the right shank, and location of the markers on the foot, ankle and knee joint

After participants received a detailed explanation of the protocol, they warmed up by cycling on an ergometer for 10 minutes. Surface EMG electrodes and an ultrasound probe were attached to the lower leg. After several minutes of practice of repetitive two-legged hopping, participants stood barefoot on a force platform

with arms akimbo and were asked to hop 10-12 times with gradually increasing effort to reach maximal hopping height. Participants were instructed to keep their knee and hip joints as straight as possible while hopping. Knee and ankle joint angular displacement, ground reaction force (GRF), electrical activity (EMG) of ankle plantar flexor muscles, and fascicle behavior of the lateral gastrocnemius were recorded during the hops (Fig. 1). AT-MA was estimated. All measurements were carried out on the dominant (right) lower extremity.

2.3 Kinetic and Kinematic Data Collection and Analyses

Participants performed hopping on a force plate (Kistler Force Platform System 92-81B, Switzerland, sampling rate 1 kHz) while being videotaped at 120 Hz by a digital camera set at 1m height on a tripod, four meters from the force platform perpendicular to the sagittal plane of the participant. To measure ankle and knee ROM, four retroreflective markers (1.5 cm diameter) were placed on the skin on the right leg on the: greater trochanter of the hip; lateral condyle of the tibia, and lateral malleolus of the ankle, and on the 5th metatarsal head of the foot. Video images were digitized in 2D (Skillspector software, v. 1.2.4, Denmark). From the force-time curve, contact and flight times, and the peak GRF were determined. Rebound height was calculated from the flight time [40]. The reactive strength index was computed as the ratio of the flight and contact times [13] to select the best three hops. Data of these three hops were time normalized and averaged. The braking and take-off phase and their duration was determined from the ankle angular displacement-time curve. The anatomical angle of the knee joint is zero when the longitudinal axis of the thigh and shank is aligned. However, we used the other angle determination, where this angular position represents 180 degrees. In terms of the ankle joint, the angle of the joint was considered zero when the shank was perpendicular to the foot base (sole). Ankle ROM was calculated from the angular displacement-time curve. The peak GRF was normalized to body weight (rGRF).

2.4 EMG Data Collection and Analyses

EMG signals were recorded during hopping at a sampling frequency of 1 kHz (TeleMyo, Noraxon U.S. Inc., Scottsdale, Az, USA). After skin preparation, silver-silver chloride bipolar surface electrodes (Blue Sensor M-00-S/25, Ambu, Denmark) with a 10 mm diameter and an inter-electrode distance of 20 mm (center-to-center) were placed on the medial gastrocnemius (MG), lateral gastrocnemius (LG), and soleus (SOL). For LG and SOL, the electrode placements were defined using ultrasonography. The electrodes were placed on the lateral side of the SOL to minimize cross-talk between SOL and adjacent muscles [38]. For LG, EMG electrodes were placed slightly lateral to the muscle midbelly so that the ultrasound probe could be placed correctly (ultrasound

preparation detailed below). For MG, the electrodes were aligned parallel with the fascicle orientation and were placed according to SENIAM guidelines [17]. The reference electrode was placed on the ipsilateral patella. To minimize movement artefacts, EMG cables were taped over the skin.

The EMG signals were band-pass filtered (20-450 Hz) with a 4th order zero-lag Butterworth filter to remove movement artefacts and signal noise. Root-mean-square values (RMS) were calculated the pre-activation, eccentric, and concentric phases of each hop. The RMS envelopes were integrated for each phase and divided by the integration time to calculate the average EMG activity [12].

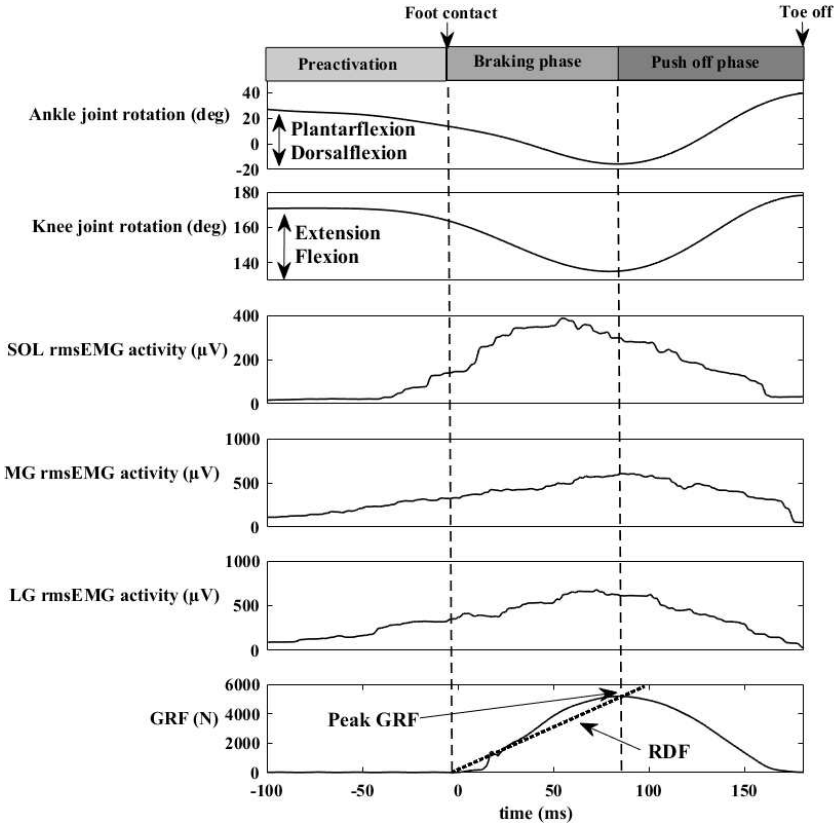


Figure 2

Ankle angular displacement, electric activity (EMG) for gastrocnemius medialis (MG) and lateralis (LG), soleus (SOL) muscles, and the vertical ground reaction force (GRF) curves in the function of time in the case of one participant. The first area refers to the preactivation time (100 ms). The second and third area represents the braking and take-off phase duration during the ground contact pGRF represents the peak ground reaction force and RFD is the rate of force development, i.e., pGRF divided by the time from zero to pGRF. Neutral knee joint angle was considered 180° between shank and thigh and neutral ankle joint angles was considered 90° between foot and shank.

Pre-activation was the 100-ms period before foot contact [26]. EMG ratios were calculated by dividing average EMG in the stretching phase by average EMG in the pre-activation phase, and average EMG in the shortening and stretching phases [21]. The EMG records were synchronized with the kinematic records and phases were determined from the joint angle data. (Fig. 2.)

2.5 Ultrasonography Data Collection and Analyses

Ultrasonography was used to estimate architecture and architectural changes in LG during hopping (50 frames·per sec^{-1}, 6 cm field-of-view, B-mode linear array probe, 7.5 MHz scanning frequency, Hitachi-Aloka EUB 405 plus, Japan). The ultrasound probe was placed over the right LG in the plane of the fascicles and was secured using a custom-made cast and tightly fixed around the shank to minimize probe movement relative to the skin (see in supplementary material). A custom-made synchronization module was used to synchronize the kinematic, kinetic, EMG, and ultrasonography recordings. LG fascicles were manually outlined and fascicle length measured (Fig. 3). Multiple contour lines were fitted to follow fascicle curvature. If part of the fascicles was outside the field-of-view, fascicle length was estimated by linear extrapolation. Images were processed in ImageJ (v.1.44b National Institutes of Health, USA).

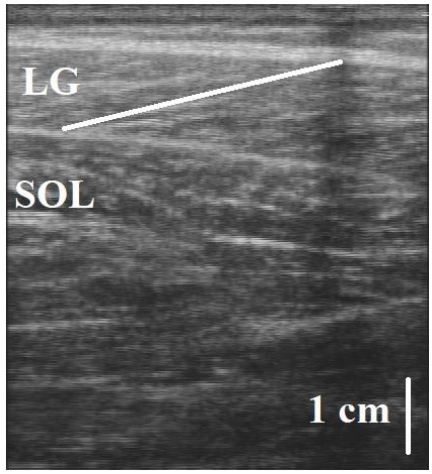


Figure 3

Representative ultrasound image of soleus in sagittal plane for estimate fascicle length. The image was taken at 50% of the muscle length because that region possibly contains the longest fascicles of the muscles. Fascicle length (solid white line along fascicles), are drawn in the images.

2.6 Estimation of AT-MA

We used a photograph-based method to estimate AT-MA length [41]. Briefly, the participants were seated with the right foot on a reference block and the ankle in a position. The foot was photographed from both the lateral and medial sides in the sagittal plane. The most prominent tip of the lateral and medial malleoli was marked on the skin and the horizontal distance from the markers to the posterior aspect of the Achilles tendon were measured on both sides [41] using ImageJ software (v. 1.8.0_112, USA 2006). The mean of these two distances was defined as the length of the AT-MA (Fig. 4).

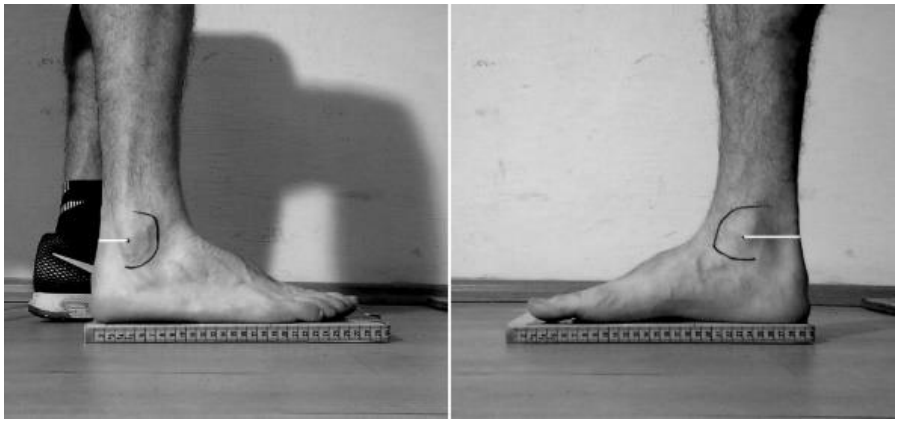


Figure 4
A typical sagittal photograph of the right ankle placed on and aligned with a reference block. The horizontal distance from the medial and lateral malleolus to the Achilles tendon was determined (white lines). Moment arm was calculated as the mean of these two distances

2.7 Statistical Analysis

Data are presented as mean and standard deviation (±SD). The distribution of the data was checked with the Shapiro–Wilk test. Relationship between variables was estimated with Pearson's correlation coefficients (r), which was categorized as small (0.0-0.1), moderate (0.1-0.3), medium (0.3-0.5), large (0.5-0.7), and very large (0.9-1.0) [22]. Additionally, the 95% confidence intervals for each coefficient was calculated. In cases of non-Gaussian data distribution, a Spearman rank correlation was used. Multiple regression was used to determine the variables that predicted most accurately hopping height. The level of significance was set at $p \leq 0.05$.

3 Results

Tables 1-2 show descriptive data for all outcomes. Table 3 shows the correlation between the hopping height and the selected variables. Hopping height only correlated with knee ROM (extension amplitude) but not with ankle and plantar flexor variables, ankle joint kinematics, and EMG variables. There was a strong relationship between hopping height and force variables, i.e., peak ground reaction force (pGRF), normalized peak ground reaction force (pGRFr), and rate of force development (RFD) (Table 3).

Table 1
Hopping kinematics and ground reaction force (GRF), mean SD

Variables	mean ±SD
Hopping height (m)	0.24±0.06
Contact time (s)	0.186±0.02
Flight time (s)	0.435±0.05
Contact to flight time ratio	2.36±0.41
Braking phase duration – [Tcon] (s)	0.082±0.016
Push off phase duration – [Tecc] (s)	0.102±0.015
Ankle flexion amplitude (rad)	0.51±0.18
Ankle extension amplitude (rad)	0.83±0.19
Knee flexion amplitude (rad)	0.34±0.14
Knee extension amplitude (rad)	0.68±0.11
Peak GRF (N)	4498.7±995.2
Normalized peak GRF (N/kg)	65.72±8.26
Rate of force development (N/s)	58220±20302

Table 2
Achilles tendon moment arm length, gastrocnemius lateralis fascicle, and plantar flexor electromyography activity variables. EMG activity measured during ankle flexion (STR) and divided by EMG in preactivation (PRA) (STR to PRA) and EMG during ankle extension (SHO) divided by EMG in ankle joint flexion phase (SHO to STR). Muscles: soleus (SOL), medial gastrocnemius (MG), lateral gastrocnemius (LG)

Variables	mean ±SD
Achilles tendon moment arm (cm)	4.29±0.75
Normalized Achilles tendon moment arm (ratio)	16.16±2.56
Fascicle length	4.24±0.85
Fascicle length change during joint flexion (cm)	0.03±0.54
Fascicle length during joint extension (cm)	-0.38±0.47
STR to PRA SOL	4.67±3.38
STR to PRA MG	2.66±1.62
STR to PRA LG	1.76± 0.5

SHO to STR SOL	0.87±0.44
SHO to STR MG	0.90±0.34
SHO to STR LG	0.83±0.28

Table 3
Correlation between hopping height and variables of plantar flexor, Achilles tendon, joint kinematics, ground reaction, and EMG activity. Asterisks indicate a significant correlation

Hopping height		Independent variable	AT-MA length	
r	p value		r	p value
0.28	0.14	Achilles tendon moment arm	-	-
0.23	0.23	Resting fascicle length	0.06	0.74
-0.09	0.65	Fascicle length change (flexion)	0.18	0.35
0.14	0.47	Ankle flexion amplitude	-0.48	0.009*
0.25	0.19	Ankle extension amplitude	-0.39	0.037*
0.18	0.25	Knee flexion amplitude	-0.41	0.03*
0.53	0.001*	Knee extension amplitude	-0.20	0.31
0.08	0.66	EMG STR to PRA ratio SOL	-0.17	0.32
-0.14	0.47	EMG STR to PRA ratio MG	-0.2	0.31
0.23	0.23	EMG STR to PRA ratio LG	-0.08	0.88
-0.16	0.41	EMG SHO to STR ratio SOL	-0.02	0.99
-0.15	0.42	EMG SHO to STR ratio MG	-0.03	0.98
-0.19	0.32	EMG SHO to STR ratio LG	-0.06	0.92
0.69	0.001*	Peak ground reaction force	0.64	0.001*
0.48	0.008*	Normalized ground reaction force	0.63	0.001*
0.57	0.001*	Rate of force development	0.48	0.001*

*significant correlation

Additional correlations

Ankle dorsiflexion and plantar flexion amplitude correlated with joint flexion (Tecc) and extension (Tcc) time, respectively ($r=0.64$ $p<0.001$ and $r=0.67$ $p<0.001$). Knee flexion and extension time and amplitude were also interrelated ($r=0.61$ $p< 0.001$and $r= 0.76$ $p<0.001$). There was an inverse relationship between rGRF and ankle dorsiflexion amplitude ($r=-0.43$ $p<0.01$) and between RFD and ankle dorsiflexion ($r=-0.56$, $p=0.01$).

Summarizing the correlations AT-MA did not correlate with hopping height ($r=-0.28$, $p<0.14$), but it was related to peak GRF and RFD so that both force-related variables correlated with hopping height. AT-MA, negatively correlated with joint kinematic variables (ankle and knee flexion amplitude) which correlated also negatively with peak GRF and rate of force development. Concentric contraction duration did not correlate with GRF and RFD. Only knee extension amplitude correlated with hopping height (Fig. 5).

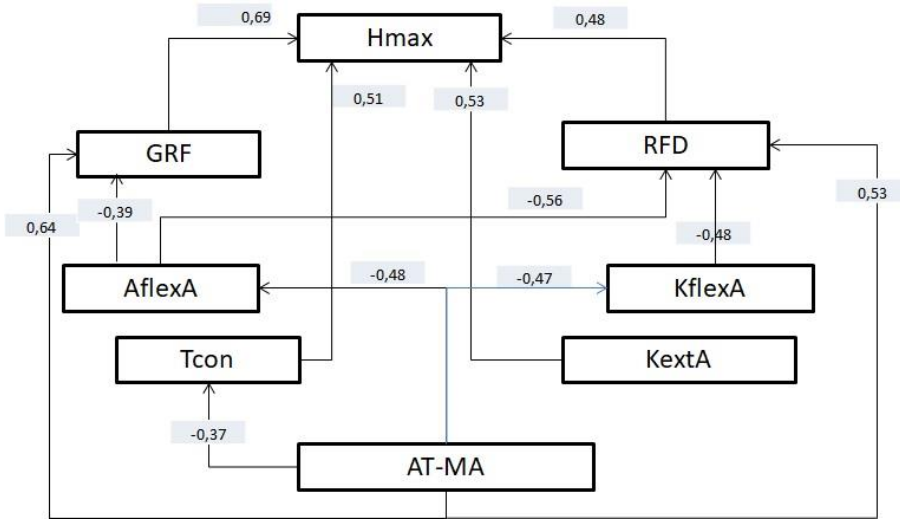


Figure 5

Correlation between dependent and independent variables. Numbers shows the correlation coefficients between the variables. Abbreviations: Hmax – maximal hopping height, GRF – ground reaction force, RFD – rate of force development, AflexA – ankle flexion amplitude, KflexA – knee flexion amplitude, Tcon - concentric contraction time, KextA – knee extension amplitude, AT-MA – Achilles tendon moment arm length

In a multiple linear regression analysis, AT-MA length, ankle dorsiflexion amplitude, and peak GRF explained 53% of the variation in hopping height ($F(3,24) = 5.85$, $p = 0.001$. $R^2 = 0.53$). The model met the assumptions of the multiple regression model described by Field [11]. The Durbin-Watson statistic (2.23) revealed no autocorrelation in the residuals. The tolerance and VIF values also showed that there is no multicollinearity between the variables (see in supplementary materials).

4 Discussion

The purpose of the study was to determine the relationship between AT-MA, joint kinematics, ground reaction forces, plantar flexor electromyography activation, lateral gastrocnemius fascicle behavior, and hopping height.

The hypothesis was that no single variable is related sufficiently strongly to hopping height rather a certain combination of variables predicts hopping height.

Over the past 20 years, a number of studies has sought to determine the relationship between AT-MA or calcaneus length and sprint and distance running

performance. We thought that repetitive hopping is ideal to study the effect of the selected ankle characteristics on hopping height because it results in a real stretch-shorten muscle contraction of the plantar flexor muscles in which mostly Achilles tendon carry out both negative and positive mechanical work. On the other hand, continuous rebound jump provides the greatest muscle stretch versus other jumps (i.e. countermovement jump, squat jump, static jump).

Interestingly, there have been relatively few studies, in which hopping was used to test the connection among these variables [44, 45]. The hopping protocol in our experiment was similar to that of Watanabe et al. [44]. Namely, we used repetitive vertical bilateral hopping.

Watanabe et al. [44] did not find a correlation between the AT-MA and hopping height or with other muscle-tendon complex variables. Werkhoven and Piazza [45] found that subjects having longer heel length (indicate a longer AT-MA) were able to jump higher using a single static jump restricting knee and hip joint rotation. In this case, the relationship can be explained, because only plantar flexor muscles force is used to accelerate the body mass. To produce the same moment less muscle force is needed when the lever arm is longer and the same muscle should exert less force it can shorten faster. However, repetitive vertical hopping differs very much from single static hop. Namely, during static hop, the plantar flexor muscles produce force by muscle shortening, while in repetitive hopping the muscle is stretched during the ankle dorsiflexion resulting in enhanced motor unit activation and elastic energy storage in the Achilles tendon [4, 6]. The question arises whether the long AT-MA is also beneficial during muscle stretch? Theoretically, short-moment arm can be advantageous because the muscle should exert greater force, and therefore the tendon stretch, ergo elastic energy storage can be greater. It is well documented in the literature that during SSC movement the fascicles operate under isometric contraction and therefore the tendon is stretched mostly during joint flexion [23, 24, 27, 32]. In our study, we could not measure the Achilles length change, but the short, not significant length change of the fascicle length in lateral gastrocnemius may indicate that the Achilles tendon was stretched predominantly. However, it should be noted the inter-individual variation in fascicle length change was very high that may indicate fascicle behavior adaptation to joint kinematics and AT-MA length.

Although the AT-MA length changes in a function of joint angle and force exerted by the muscle [14, 43], it cannot be assumed that the Achilles tendon length changes could be shorter during dorsiflexion and longer during joint plantar flexion to meet the quality criteria in both contraction type. Nagano and Komura [37] in a model design suggested that AT-MA length and also contraction speed influence the mechanical moment and mechanical power differently during eccentric and concentric contraction. In contrast to the results of Wekhoven and Piazza [45] they found that a shorter moment arm increases ankle joint moment during concentric contraction.

During hopping when joint flexion and extension occurs the muscle undergoes a stretch shorten type of contraction, it is not clear either a long or a short AT-MA would be beneficial. We think that can be because other important factors have also influenced the joint output as well, thus a different combination of these variables can explain more accurately the joint performance. We thought that subjects may use different joint kinematic strategies depending upon the AT-MA length to jump as high as possible. Therefore, we assumed that there might be no correlation between AT-MA and hopping height. Indeed, in this experiment, we did not find an association between AT-MA length and hopping height, but the strong correlation was between AT-MA and ankle and knee flexion amplitude, i.e. the greater the AT-MA length the smaller was the joint flexion. This could lead a greater deceleration at the end of the joint flexion and therefore the peak GRF and RFD became greater. This assumption is supported by the strong correlation between peak joint flexion amplitude and GRF and RFD, i.e. the shorter the joint flexion the greater was the peak GRF and RFD, and as a consequence the greater the peak force and RFD the highest was the hopping height. These relationships were supported by the multiple regression analysis because the three independent variables (MA-TA, ankle dorsiflexion amplitude, and peak GRF) jointly influenced the hopping height. In addition, AT-MA showed a strong correlation with both peak GRF and RFD indicating that longer AT-MA is beneficial for muscles to exert a greater force on the ground and develop force within a shorter time.

It could be assumed that if the ankle and knee joint angular deceleration was increased then plantar flexor muscles should generate greater force and therefore muscle activation is enhanced indicated by EMG activity, i.e. in our case greater stretch to pre-activation EMG ratio. Sano et al. [40] in a similar study reported that Kenyan runners jumped higher than the control group with a greater stretch to pre-activation EMG ratio during hopping.

We could not detect a correlation between EMG ratios and other variables that is hard to be explained. However, we have to stress that the EMG ratio that we used may not be the best indicator of muscle activation level, since we found a significant correlation between pre-activation EMG and stretching EMG. Namely, when the pre-activation EMG increased the stretching EMG also increased. So that, we can speculate that soleus and gastrocnemius muscle had neuromuscular potentiation due to the muscle stretch because there was a correlation between RFD and ankle dorsiflexion amplitude and as a consequence, the greater hopping height was associated with greater RFD and smaller ankle dorsiflexion amplitude.

So that, we can speculate that soleus and gastrocnemius muscle had neuromuscular potentiation due to the muscle stretch because there was a correlation between RFD and ankle dorsiflexion amplitude and as a consequence, the greater hopping height was associated with greater RFD and smaller dorsiflexion amplitude.

From the results of the correlation analysis and the related measured variables (i.e. small change in fascicle length), we can assume that the shorter ankle dorsiflexion amplitude and longer AT-MA made the elastic energy storage and reuse possible in the Achilles tendon, which is very important in enhancing jumping height [4, 6, 10, 15].

Comparing our results with those of others concerning the achieved hopping height in our study is heavily relies on a well-coordinated joint movement [5, 7, 8, 35], and our measured ankle and knee joint ROM were similar to previous reports [19, 40, 44]. The measured hopping heights performed by our participants are also similar to those reported in earlier studies [19, 20, 40] thus we can conclude that participants in this study performed the hopping similarly. However, we have noted that the variables involved in our study may have different associations if the knee and hip joints were fixed. In this respect we have to mention that knee extension amplitude, in the present study, had a direct and independent correlation with hopping height that probably modified the effect of ankle-related variable.

This experiment has some methodological limitations that should be addressed. To estimate the AT-MA length we used a low-tech method similar to previous reports [41, 45, 47]. The measured lengths were similar to those in previous reports and not differ remarkably from MRI-based estimations [39]. However, the AT-MA length is changing constantly with joint rotation [14, 43], therefore, it can influence the results of our report. Furthermore, the investigation of tendon characteristics and length changes during the hops could provide more information about the efficiency of the elastic strain energy store and reuse mechanism. To study the length changes of LG, a two-dimensional ultrasound technique was used despite the three-dimensional structures of this muscle. This method has potential errors that occur during hopping and other dynamic movement and it is not possible currently to quantify the amount of error and may not be consistent throughout the hop cycle. The mean fascicle length changes are in accordance with previous results. However, there is a large inter-individual variance in these data which must be considered. The ankle plantar flexor muscle strength and muscle morphology (volume, mass) were not assessed in this study, and these parameters obviously determine the hopping performance. Thus, it would be ideal to include these variables in future studies.

Conclusions

Our results suggest that AT-MA length does not affect the flying height when hopping is carried out with stretch-shorten cycle contraction, using repetitive double-legged hopping. However, we may conclude that AT-MA has an indirect association with hopping height via maximum force generation, force development, and pushing off phase time. Also, our results indicate that AT-MA length can influence hopping height jointly with the electromyography activity of plantar flexor muscles (gastrocnemius medialis in our case), maximum force generation, and rate of force development.

Acknowledgement

The authors wish to thank all participants for volunteering in this study.

References

[1] K. R. Barnes, M. R. McGuigan, A. E. Kilding: Lower-body determinants of running economy in male and female distance runners. J. Strength Cond. Res., 28 (5), pp. 1289-1297, 2014

[2] J. R. Baxter, T. A. Novack, H. Van Werkhoven, D. R. Pennell, S. J. Piazza: Ankle joint mechanics and foot proportions differ between human sprinters and non-sprinters. Proc. R. Soc. B Biol. Sci., 279 (1735), pp. 2018-2024, 2012

[3] J. R. Baxter, S. J. Piazza: Plantar flexor moment arm and muscle volume predict torque-generating capacity in young men. J. Appl. Physiol., 116 (5), pp. 538-544, 2014

[4] A. Belli, C. Bosco: Influence of stretch-shortening cycle on mechanical behaviour of triceps surae during hopping. Acta Physiol. Scand., 144 (4), pp. 401-408, 1992

[5] M. F. Bobbert, G. J. van Ingen Schenau: Coordination in vertical jumping. J. Biomech., 21 (3), pp. 49-262, 1988

[6] C. Bosco, I. Tarkka, P. V. Komi: Effect of elastic energy and myoelectrical potentiation of triceps surae during stretch-shortening cycle exercise. Int. J. Sports Med., 3 (3), pp. 137-140, 1982

[7] G. A. Cavagna: Storage and utilization of elastic energy in skeletal muscle. Exerc. Sport Sci. Rev., 5 (1), pp. 89-130, 1977

[8] P. Cormie, J. M. Mcbride, G. O. Mccaulley: Power-time, force-time, and velocity-time curve analysis of the countermovement jump: Impact of training. J. Strength Cond. Res., 23 (1), pp. 177-186, 2009

[9] P. Dyhre-Poulsen, E. B. Simonsen, M. Voigt: Dynamic control of muscle stiffness and H reflex modulation during hopping and jumping in man. J. Physiol., 437, pp. 287-304, 1991

[10] C. T. Farley, R. Blickhan, J. Saito, C. R. Taylor: Hopping frequency in humans: A test of how springs set stride frequency in bouncing gaits. J. Appl. Physiol., 71 (6), pp. 2127-2132, 1991

[11] A. Field: *Discovering statistics using IBM SPSS statistics*. Los Angeles: Sage, 2013

[12] T. Finni, P. V. Komi, V. Lepola: In vivo muscle mechanics during locomotion depend on movement amplitude and contraction intensity. Eur. J. Appl. Physiol., 85 ((1-2)), pp. 170-176, 2001

[13] E. P. Flanagan, W. P. Ebben, R. L. Jensen: Reliability of the reactive

strength index and time to stabilization during depth jumps. J. Strength Cond. Res., 22 (5), pp. 1677-1682, 2008

[14] J. R. Fletcher, B. R. MacIntosh: Estimates of achilles tendon moment arm length at different ankle joint angles: Effect of passive moment. J. Appl. Biomech., 34 (3), pp. 220-225, 2018

[15] S. Fukashiro, P. V. Komi, M. Järvinen, M. Miyashita: In vivo achilles tendon loading' during jumping in humans. Eur. J. Appl. Physiol. Occup. Physiol., 71 (5), pp. 453-458, 1995

[16] K. Funase, T. Higashi, A. Sakakibara, K. Imanaka, Y. Nishihira, T.S. Miles: Patterns of muscle activation in human hopping. Eur. J. Appl. Physiol., 84 (6), pp. 503-509, 2001

[17] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, G. Rau: Development of recommendations for SEMG sensors and sensor placement procedures. J. Electromyogr. Kinesiol., 10 (5), pp. 361-374, 2000

[18] A. V. Hill: The heat of shortening and the dynamic constants of muscle. Proc. R. Soc. London. Ser. B - Biol. Sci., 126 (843), pp. 136-195, 1938

[19] H. Hobara, K. Kimura, K. Omuro, K. Gomi, T. Muraoka, M. Sakamoto, K. Kanosue: Differences in lower extremity stiffness between endurance-trained athletes and untrained subjects. J. Sci. Med. Sport, 13 (1), pp. 106-111, 2010

[20] H. Hobara, K. Kanosue, S. Suzuki: Changes in muscle activity with increase in leg stiffness during hopping. Neurosci. Lett., 418 (1), pp. 55-59, 2007

[21] M. Hoffrén-Mikkola, M. Ishikawa, T. Rantalainen, J. Avela, P. V. Komi: Neuromuscular mechanics and hopping training in elderly. Eur. J. Appl. Physiol., 115 (5), pp. 863-877, 2015

[22] W. G. Hopkins, S. W. Marshall, A. M. Batterham, J. Hanin: Progressive statistics for studies in sports medicine and exercise science. Med. Sci. Sports Exerc., 41 (1), pp. 3-13, 2009

[23] M. Ishikawa, J. Pakaslahti, P. V. Komi: Medial gastrocnemius muscle behavior during human running and walking. Gait Posture, 25 (3), pp. 380-384, 2007

[24] M. Ishikawa, P. V. Komi: Muscle fascicle and tendon behavior during human locomotion revisited. Exerc. Sport Sci. Rev., 36 (4), pp. 193-199, 2008

[25] G. M. Jones, D. G. D. Watt: Observations on the control of stepping and hopping movements in man. J. Physiol., 219 (3), pp. 709-727, 1971

[26] P. V. Komi, A. Golihofer, D. Schmidtbleicher, U. Frick: Interaction between man and shoe in running: Considerations for a more

comprehensive measurement approach. Int. J. Sports Med., 8 (3), pp. 196-202, 1987

[27] P. V Komi, C. Nicol: *Stretch-Shortening Cycle of Muscle Function*. International Olympic Committee, 2011

[28] P. V. Komi, A. Gollhofer: Stretch reflexes can have an important role in force enhancement during SSC exercise. J. Appl. Biomech., 13 (4), pp. 451-460, 1997

[29] B. Kopper, Z. Csende, L. Trzaskoma, J. Tihanyi: Stretch-shortening cycle characteristics during vertical jumps carried out with small and large range of motion. J. Electromyogr. Kinesiol., 24 (2), pp. 233-239, 2014

[30] B. Kopper, Z. Csende, S. Sáfár, T. Hortobágyi, J. Tihanyi: Muscle activation history at different vertical jumps and its influence on vertical velocity. J. Electromyogr. Kinesiol., 23 (1), pp. 132-139, 2013

[31] M. M. Kumagai K., Takashi A., Brechue W. F., Ryushi T., Takano S.: Sprint performance is related to muscle fascicle length in male 100-m sprinters muscle performance Sprint performance is related to muscle fascicle length in male 100-m sprinters. J Appl Physiol, 88 (3), pp. 811-816, 2000

[32] A. Lai, G.A. Lichtwark, A. G. Schache, Y.-C. Lin, N. A. T. Brown, M. G. Pandy: In vivo behavior of the human soleus muscle with increasing walking and running speeds. J. Appl. Physiol., 118 (10), pp. 1266-1275, 2015

[33] M. Lamontagne, M. J. Kennedy: The biomechanics of vertical hopping: A review. Res. Sport. Med., 21 (4), pp. 380-394, 2013

[34] S. S. M. Lee, S. J. Piazza: Built for speed: musculoskeletal structure and sprinting ability. J. Exp. Biol., 212 (22), pp. 3700-3707, 2009

[35] J. M. Mcbride, G. O. Mccaulley, P. Cormie: Influence of preactivity and eccentric muscle activity on concentric performance during vertical jumping. J. Strength Cond. Res., 22 (3), pp. 750-757, 2008

[36] M. Mooses, B. Tippi, K. Mooses, J. Durussel, J. Mäestu: Better economy in field running than on the treadmill: Evidence from high-level distance runners. Biol. Sport, 33 (2), pp. 155-159, 2015

[37] A. Nagano, T. Komura: Longer moment arm results in smaller joint moment development, power and work outputs in fast motions. J. Biomech., 36 (11), pp. 1675-1681, 2003

[38] A. Péter, A. Hegyi, L. Stenroth, T. Finni, N. J. Cronin: EMG and force production of the flexor hallucis longus muscle in isometric plantarflexion and the push-off phase of walking. J. Biomech., 48 (12), pp. 3413-3419, 2015

[39] S. G. Rugg, R. J. Gregor, B. R. Mandelbaum, L. Chiu: In vivo moment arm calculations at the ankle using magnetic resonance imaging (MRI). J. Biomech., 23 (5), pp. 495-501, 1990

[40] K. Sano, M. Ishikawa, A. Nobue, Y. Danno, M. Akiyama, T. Oda, A. Ito, M. Hoffrén, C. Nicol, E. Locatelli, P. V. Komi: Muscle-tendon interaction and EMG profiles of world class endurance runners during hopping. Eur. J. Appl. Physiol., 113 (6), pp. 1395-1403, 2013

[41] M. N. Scholz, M. F. Bobbert, A. J. van Soest, J. R. Clark, J. van Heerden: Running biomechanics: shorter heels, better economy. J. Exp. Biol., 211 (20), pp. 3266-3271, 2008

[42] M. Voigt, F. Chelli, C. Frigo: Changes in the excitability of soleus muscle short latency stretch reflexes during human hopping after 4 weeks of hopping training. Eur. J. Appl. Physiol. Occup. Physiol., 78 (6), pp. 522-532, 1998

[43] F. E. Wade, G. S. Lewis, S. J. Piazza: Estimates of Achilles tendon moment arm differ when axis of ankle rotation is derived from ankle motion. J. Biomech., 90, pp. 71-77, 2019

[44] T. Watanabe, K, Koyama, K, Yanagiya: *Functional characteristics of Achilles tendon moment arm length measured using ultrasonography in vivo*. In: Conference Proceedings Archive, 26th International Conference on Biomechanics in Sports. p. 2008, 2008

[45] H. Van Werkhoven, S. J. Piazza: Foot structure is correlated with performance in a single-joint jumping task. J. Biomech., 57, pp. 27-31, 2017

[46] H. Van Werkhoven, S. J. Piazza: Does foot anthropometry predict metabolic cost during running? J. Appl. Biomech., 33 (5), pp. 317-322, 2017

[47] L. S. Wessbecher, A. N. Ahn: Lower leg morphology in runners: forefoot strikers have longer heels but not bigger muscles than rearfoot strikers. J. Exp. Biol., 222 (4), p. 2019

The Relationship of Maximal Strength with the Force-Velocity Profile in Resistance Trained Women

Gergely Pálinkás, Bettina Béres, Zsófia Tróznai, Katinka Utczás, Leonidas Petridis

Research Centre for Sport Physiology, University of Physical Education
Alkotás u. 44, H-1123 Budapest, Hungary
E-mails: palinkas.gergely@tf.hu, beres.bettina@tf.hu, troznai.zsofia@tf.hu, utczas.katinka@tf.hu, petridis.leonidas@tf.hu

Abstract: Force – velocity (F-v) profiling is used as a diagnostic tool to evaluate ballistic performance and to optimize strength and power training. Previous research examining the lower limbs' neuromuscular qualities of female athletes has reported a large prevalence of force-deficit, which raises the question about the level of maximal strength, which may be associated with an optimal F-v profile. In this study, we investigated the relationship between maximal strength abilities and the F-v profile of female athletes. 27 resistance-trained female athletes (age: 29.7 ± 5.0 years) from three different sports (Olympic weightlifting, CrossFit and recreational bodybuilding) participated in the study. The F-v profile of the lower limbs was assessed using vertical jumps on a force platform against various external loads, while maximal strength was measured with the one-repetition maximum in the back squat. Back squat relative to body weight was calculated and used in the statistical analysis. The results showed a significant correlation between maximal back squat/body weight and the jump height during the unloaded jumps ($r=0.73$; $p<0.01$), the theoretical maximal power ($r=0.62$; $p<0.01$), and the ratio of the measured to the optimal F-v slope ($r=0.56$; $p<0.01$). An optimal F-v profile was observed around 1.2x bodyweight squat. Our results confirm the force-oriented profile of highly trained resistance athletes and contribute to the existing literature by providing indicative values associated with an optimal profile. Such data can be used by the coaches and athletes to evaluate the athletes' strength qualities and design training according to the individual needs.

Keywords: Olympic weightlifting; CrossFit; one repetition maximum; ballistic performance

1 Introduction

Force, velocity, and power are the most common measures to evaluate human muscle's mechanical properties during dynamic motor tasks. The relationship of

force and muscle shortening velocity was first described in the early studies of Hill [1]. Based on his fundamental theorem, force and shortening velocity have an inverse relationship meaning that an increase in force causes a decrease in velocity and vica-versa. The Maximum force (F_0) is obtained when the shortening velocity is zero (which corresponds to an isometric contraction), whereas maximum velocity (v_0) is obtained against no external load. Additionally, by multiplying force and velocity we can calculate power. Usually, maximal power occurs at submaximal force or velocity intensities and varies between individuals. Understanding the relationship of force-velocity-power is essential in athletic training since the development of these qualities requires different training methodology.

From force and velocity measures we can define the Force-velocity (*F-v*) slope of an athlete during ballistic movements. For lower limbs, force and velocity are typically measured using vertical jumping, more specifically, with a squat jump (SJ) or countermovement jump (CMJ), with greater power values obtained during CMJ [2; 3]. Recent studies have proposed a novel approach in the evaluation of the lower limb's main mechanical properties (force, velocity, and power) by comparing the measured slope of the force-velocity relationship (S_{fv}) to the theoretical optimal slope (S_{fv}opt) [4; 5]. This allows for establishing an athlete's force-velocity profile, which has been introduced as a diagnostic tool to optimize strength and power training and to monitor training adaptations [5].

The theoretical optimal *F-v* slope represents the optimal relationship of force and velocity and is calculated by using measured indices, thus capabilities that an individual is anywise capable of. By extrapolating the regression line of the measured values we get the theoretical maximal force and the theoretical maximum velocity (v_0), which correspond respectively to the maximum force (F_0) and the maximum shortening velocity of the Hill-equation [1].

The percentage difference between the measured and the optimal *F-v* slope is called force-velocity imbalance (FV_{imb}). Based on the imbalance we can quantify the relative contribution of the main components of power output (force or velocity). In strength and power training, such information may be used to determine the optimal load that is the most effective for the athletes [6], and which of the central components of power should be prioritized: force, velocity [7] or mixed methods [8; 9]. Besides the calculation of the FV_{imb}, we can also calculate the theoretical maximum jump height. This represents the jump height that an athlete could reach with no changes in the strength and power qualities, but just having an optimal force-velocity relationship.

Experimental evidence has shown that *F-v* profile is a determinative factor in ballistic, explosive movements independently of maximal power [10]. It has been proposed that at a given maximal power output, optimal *F-v* profile maximizes performance, whereas an imbalance in force-velocity profile (FV_{imb}) can cause ~30% decrease in performance [10] and negatively affects ballistic performance

[4]. As suggested, *F-v* profile can be a better indicator to describe the power capabilities of an athlete than jump height [11].

Athletes with notable FV_{imb} are suggested to first decrease the imbalance and then target further development in maximal power [12]. The type of training to reduce the imbalance is individual and depends on the magnitude and the direction of the FV_{imb}.

Available *F-v* profile data with female athletes have reported mostly force-deficit. For example, Marcote-Pequeno et al.[13] using squat jump and not CMJ, found FV_{imb} of 64.5±16.3% of force-deficit in female soccer athletes. The *F-v* profile of elite handball and volleyball players [14] revealed FV_{imb} values of 31.1±14.0% and 22.1±9.6% of again mainly force-deficit. Such results raise the question of the level of maximal strength, which may be linked to an optimal *F-v* profile in female athletes. The only study examining the relationship of the *F-v* profile for the lower limbs with the maximal strength describes the positioning of the one-repetition maximum (1RM) on the force-velocity curve [15]. As expected, the 1RM is positioned below the theoretical maximal force (F_0) and above the force exerted against the heaviest load during the loaded jumps, which seems reasonable since F_0 represents the maximal force under isometric conditions (when velocity is zero) and in general is larger than the 1RM. However, this is a theoretical approach with limited transferability to coaching practice.

From a practical point of view, it would be useful to examine the connection of *F-v* profile with the maximal strength as measured with the one-repetition maximum. 1RM is a quite common measure of maximal strength in everyday practice and its assessment does not require expensive equipment. The results can be expressed either in absolute or relative to bodyweight values making comparisons possible between athletes of different body weights. Due to its similar movement pattern with vertical jump and to its close connection with the lower limbs' power production ability, back squat exercise is widely used for strength training and testing purposes [16]. Such results could help to evaluate the maximal strength capabilities providing a better understanding of the athletes' strength qualities. In turn, this may support the strength training of the athletes, not only in traditional strength-based sports but also in team sports [17; 18], where our earlier findings showed considerable force deficit in Hungarian elite female athletes [14]. While force-oriented training uses mostly heavy resistance loads targeting the force-end of the force-velocity curve, velocity oriented training uses light (or no external) resistance loads targeting the velocity-end of the force-velocity curve. In power training, it has been suggested that weaker athletes can benefit from improving strength (use of heavier loads), while stronger athletes can benefit from a more specialized, ballistic type training [19]. Overall, a mixed approach is recommended, which encompasses various loads affecting the entire range of the force-velocity spectrum.

Therefore, the purpose of this study was to examine the relationship of the lower limbs' Force-velocity profile with the relative to bodyweight maximal back squat in resistance-trained women. It was hypothesised that maximal performance in back squat correlates with the *F-v* profile with higher strength levels being associated with a force-oriented profile.

2 Material and Methods

2.1 Participants

27 resistance-trained female athletes (age: 29.7 ± 5.0 years; body height: 166.7 ± 4.6 cm, bodyweight: 66.0 ± 8.2 kg; training load: 5.9 ± 2.4 hours/week; training history: 2.5 ± 1.8 years) from three different sports (Olympic weightlifting, CrossFit and recreational bodybuilding) and different competition levels (intermediate to international level) with a minimum of one year experience in strength training participated in the study. None of the subjects reported illness or injury at the time of the measurements. All participants were informed about the type and the risks of the measurements and gave their written consent to participate in this study. The study was approved by the University's Research Ethics Committee (approval number: TE-KEB/NO34/2019).

2.2 Measurements

Measurements took place at the laboratory of the University of Physical Education Budapest. After arrival, participants went through basic anthropometric measurements. Body height was measured with a stadiometer to the nearest 0.1 cm. Bodyweight (BW) was measured with a digital scale (Seca 888) to the nearest 0.1 kg.

A vertical jump test was performed on a force platform (FP4, HUR Labs Oy, Tampere, Finland) with a sampling rate of 1000 Hz. First, subjects performed a unified warm-up protocol which contained mobilization exercises and running. Then, they performed five submaximal countermovement jumps (CMJ) to ensure proper technique. Participants performed three unloaded jumps, then two loaded jumps against external loads of 25, 50, 75, and 100 percent of their BW. During unloaded jumps hands were kept on the hips, while during loaded jumps on the bar. Loaded jumps were performed with free weights. The rest time between jumps was two minutes to ensure total recovery. In this study, the best jumps were used in the statistical analysis. Participants were asked to refrain from any exercise 24 hours prior to the measurements.

A one-repetition maximum (1RM) full back squat protocol was used to assess maximal strength of the lower limbs. Measurements protocol consisted of two sets of three repetitions at 50, one set of two repetitions at 70 then one set of one repetition at 80, 90, and 95 percent of their previous 1RM back squat. Then, they attempted to hit the maximum weight they can squat with. Rest time was two minutes between sets. Full squats were performed with a high bar technique. To accept a proper technique of full squat, the subjects' hip had to go lower than the horizontal line of the knees and they had to be able to hold the bar without losing tension of the upper body. Testing of the 1RM back squat was performed 72 hours before the jumping protocol.

2.3 Calculations

Mean force values for every loaded condition were extracted from the force platform during the push-off phase (when ground reaction force becomes greater than the bodyweight to the instant of takeoff). Push off distance (hpo) was defined as the vertical displacement of the centre of mass (COM) moved from the lowest point of the countermovement to the instant of takeoff and it was extracted from the vertical displacement-time curve of the force platform. For the calculation of the jump height, we used the impulse-momentum method [20].

The average speed of the jumping movement was calculated from jump height according to [5]:

$$v = \sqrt{\frac{9.81 * Jump\ Height}{2}}$$

where v is the average speed of the jump, jump height is the height of the jump based on the impulse-momentum method.

The theoretical maximal power (P_{max}) was calculated from the theoretical maximal force (F_0) and the theoretical maximal velocity (v_0) according to the following formula [21]:

$$P_{max} = \frac{F_0 * v_0}{4}$$

Based on the P_{max} and h_{po} values, we calculated the optimal slope ($Nskg^{-1}m^{-1}$) for force and velocity. Once we have the optimal and the measured slope, F-v imbalance can be computed based on the following formula [10]:

$$FV_{imb} = 100 * \left|1 - \frac{S_{fv}}{S_{fv}opt}\right|$$

where FV_{imb} is the difference (in %) between optimal ($S_{fv}opt$) and measured (S_{fv}) slope of the force–velocity relationship.

The theoretical maximum height (h_{max}) of the vertical jump was calculated based on S_{fvopt} with no changes in P_{max} and h_{po} values. We used the following recently validated formula to calculate h_{max} [10]:

$$h_{max} = \frac{{h_{PO}}^2}{2g}\left(\sqrt{\frac{S_{fv}opt^2}{4} + \frac{2}{h_{PO}}\left(\sqrt{2 - P_{max} * S_{fv}opt} - g\right)} + \frac{S_{fv}opt}{2}\right)^2$$

where h_{po} is the height of push-off phase, $S_{fv}opt$ is the optimal slope of the Force-velocity curve, P_{max} is the theoretical maximal power, g is the acceleration due to gravity.

2.4 Statistical Analysis

Values are expressed as means ± SD. Intraclass correlation coefficient (ICC) was used to assess reliability between jumps in each loaded condition. Shapiro-Wilks test was used to assess normality. Outliers for vertical jump height were removed under P25-1.5 * (P75-P25) or above P75+1.5 * (P75-P25) value, where P25 is the 25th and P75 is the 75th percentile. Pearson's correlation was used to quantify the correlation between relative back squat and power, jump height, and FV_{imb}. Correlation coefficients were classified as 0.0-0.1 (trivial); 0.1-0.3 (small); 0.3-0.5 (moderate); 0.5-0.7 (large); 0.7-0.9 (very large); 0.9-1.0 (near perfect) [22]. Statistica 13.5 for windows (TIBCO Software Inc, Palo Alto, CA, USA) statistical package was used for the statistical analysis.

3 Results

Relative values of the subjects in back squat 1RM showed normal distribution. ICC was 0.95 (95% CI 0.90 to 0.98) for unloaded jumps, 0.99 (95% CI 0.96 to 0.99) for loaded jumps with 25% of BW, 0.97 (95%CI 0.92 to 0.99) with 50% of BW, 0.99 (95% CI 0.97 to 0.99) with 75% of BW and 0.97 (95% CI 0.92 to 0.99) with 100% of BW, respectively. The descriptive results for the unloaded CMJ and for the *F-v* profile can be found in tables 1 and 2, respectively.

Table 1
Descriptive results for the unloaded countermovement jump

Variable	mean	SD	min	max
Jump Height (cm)	29.0	4.8	19.9	43.1
Push off distance (m)	0.37	0.04	0.28	0.43
Average Force abs. (N)	867.6	108.4	642.3	1034.5
Average Force rel. (N/kg)	13.2	0.5	12.4	14.5
Average Power abs. (W)	777	126	576	1074
Average Power rel. (W/kg)	11.8	1.6	8.7	15.3

abs. = absolute values, rel. = relative values

Table 2
Descriptive results of the Force-velocity profile

Variable	mean	SD	min	max
Theoretical F_0 (N/kg)	35.58	5.67	23.80	46.33
Theoretical v_0 (m/s)	2.47	0.36	1.90	3.70
Theoretical jump height (m)	0.31	0.05	0.21	0.43
Measured *F-v* slope (Ns/m/kg)	-14.84	3.65	-22.06	-6.84
Optimal *F-v* slope (Ns/m/kg)	-14.66	0.74	-16.89	-13.51
F-v imbalance (%)	20.1	14.9	0.3	55.2

F_0: maximal force; v_0: maximal velocity; *F-v slope:* slope of the force-velocity relationship; *F-v* imbalance: force-velocity imbalance

FV_{imb} values showed velocity deficit in 12 cases (44%) and force deficit in 15 (55%) cases. 22% of the subjects (6 athletes) had normal (less than 10%), 59% (16 athletes) had low (between 10% and 40%) and 19% (5 athletes) had high (greater than 40%) imbalance. On average, measured jump height corresponded to 94% of the theoretical maximum. Three subjects' jump height was greater than their theoretical maximum.

A strong correlation was found between the back squat and weekly training load ($r=0.86$; $p<0.05$), while the correlation between back squat and training history of the subjects was moderate to large ($r=0.69$; $p<0.05$).

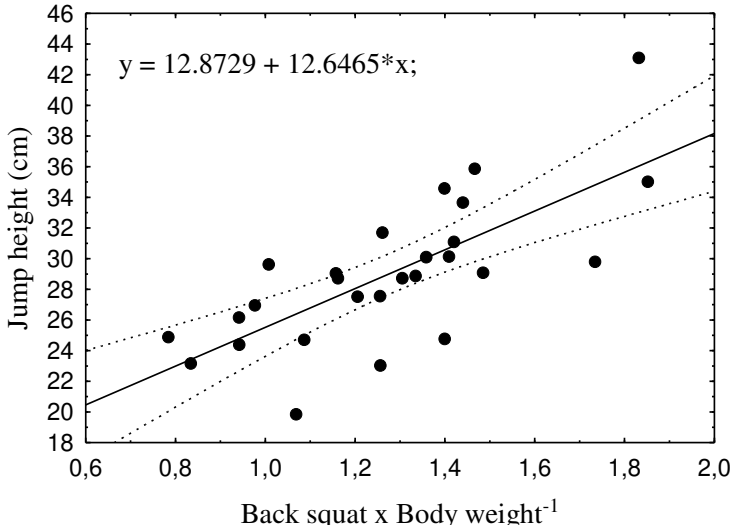


Figure 1
Correlation between the jump height during unloaded jumps and the relative values of one repetition maximum back squat. Dotted lines represent 95% CI.

Figures 1, 2, and 3 show the significant correlations between the relative strength and the height of the unloaded jumps, (r=0.73; p<0.01), the P_{max} (r=0.62; p<0.01), and the ratio of the measured to optimal *F-v* slope (r=0.56; p<0.01). Subjects with better back squat*body weight^{-1} ratio reached higher values during the jumps. Optimal FV_{imb} values can be found around the 1.2 value of back squat * body weight^{-1}.

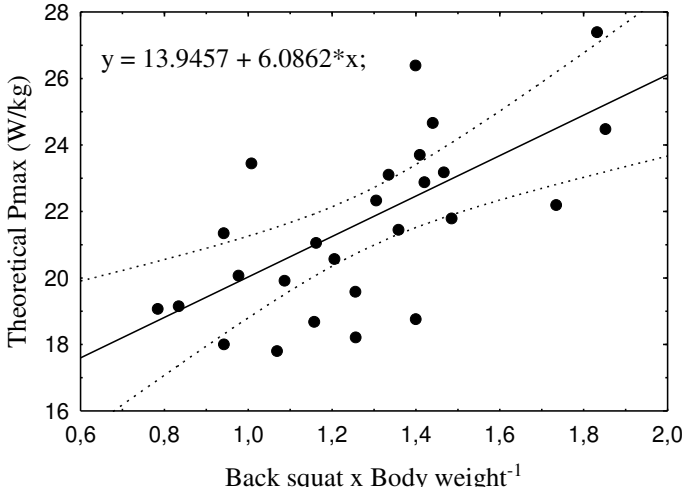


Figure 2

Correlation between the theoretical maximal Power and the relative values of one repetition maximum back squat. Dotted lines represent 95% CI.

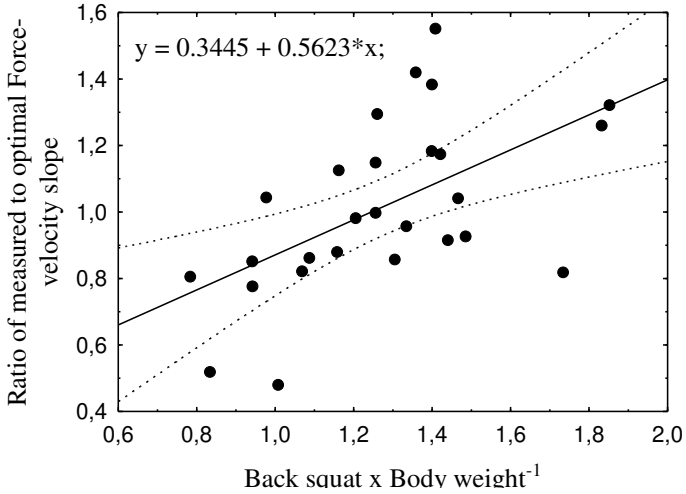


Figure 3

Correlation between ratio of the measured to the optimal Force-velocity slope and the relative values of one repetition maximum back squat. Dotted lines represent 95% CI.

4 Discussion

The aim of this study was to investigate the relationship between maximal strength abilities and the *F-v* profile in resistance-trained female athletes, attempting to determine the minimum strength requirements for an optimal profile. *F-v* profiling has gained attention as a diagnostic measurement by separating the strength and velocity abilities of an athlete and describing training directions according to the individual needs.

Previous studies with female athletes have reported *F-v* profile of mainly force-deficit [13; 14]. Force-deficit indicates the need for strength-oriented training, which targets the force-end of the force-velocity spectrum [19]. This is supported by studies showing that the type of strength training has a direct effect on the force-velocity qualities and that the *F-v* profile reflects adaptations in strength training [12; 23]. Generally, resistance-trained athletes usually demonstrate a force-oriented profile, while athletes not used to heavy loads training have a velocity-oriented profile [10; 12; 24]. Being used to resistance training with heavy loads the proportion of the athletes in our sample, who had a low to high force deficit was 48%, which is considerably lower than the same proportion found in our previous study with elite handball and volleyball players (83.3%) [14]. The typical force-oriented profiles of resistance-trained athletes were demonstrated also in the study of Jimenez-Reyes et al. [25] with athletes from different sports and competition levels. Olympic weightlifters had significantly higher F_0 values than elite athletes from team sports (soccer, handball), indicating the differences in the mechanical attributes of the lower limbs in sports of different movement patterns and physical requirements [25]. The weightlifters in our study had similar F_0 and v_0 values to high-level weightlifters and high-level track & field athletes reported by Jimenez-Reyes et al. [25]. While F_0 values (38.0 ± 7.3 N/kg) were higher in weightlifters in both studies, v_0 values (2.6 ± 0.1 m/s) seem to be within the average range consisting of several sports as represented in the study of Jimenez-Reyes et al. [25]. CrossFit athletes had lower F_0 values (35.4 ± 5.7 N/kg) with a lower v_0 (2.5 ± 0.4 m/s) than weightlifters. Both groups produced greater forces, than athletes from non-strength-based sports, confirming the force-oriented profiles resulting from resistance training.

The larger prevalence of force deficit in previous studies highlights the importance of maximal strength in power training. In relation to maximal strength, an observation and a question emerge. As expected, and confirming existing knowledge [16], maximal strength is a determinant factor in performance during ballistic movements. This was well reflected in the correlation analysis. Relative values in 1RM back squat positively correlated with the performance in vertical jumping (Figure 1) and with the theoretical maximal power output (Figure 2). Athletes with higher relative to bodyweight back squat achieved higher jump height and had a higher theoretical maximal power. These results clearly reinforce the significance of maximal strength training to increase vertical jump performance.

However, the question, which is of great interest for coaches and practitioners appertains to the levels of maximal strength which may result in an optimal profile. How strong an athlete should be in order to prevent force deficit? Haff and Nimphius [19] have suggested that strong athletes must be able to squat with two times their bodyweight. However, this is a general recommendation and does not discriminate between men and women or in relation to different bodyweight categories. With regard to the *F-v* profile, these values need to be examined experimentally separately for male and female athletes. In our sample with female athletes and according to our expectations, the statistical analysis revealed a positive, large correlation between relative 1RM back squat and the ratio of optimal to the measured *F-v* slope. The findings suggest that the optimal range of the *F-v* profile is associated with a relative 1RM back squat around 1.2* of body weight. Athletes with greater relative values show force-oriented, while lower values show mostly velocity-oriented profiles. This value is significantly lower than the recommendations of a back squat with 2.0* of body weight. Although significant, the magnitude of the correlation raises several questions concerning the conclusive use of *F-v* profile results from 1RM in the back squat.

Two aspects should be considered in the interpretation of these results. First, the robustness of the measurements. The 1RM measurement is a robust measurement showing little variations within a microcycle [26]. Meaningful changes in 1RM are usually observed after greater time periods, a minimum of four weeks of strength training has been reported in the literature [27]. In the contrary, we are unaware of any studies examining the consistency of *F-v* profiling across consecutive days. However, based on our experience, it seems that *F-v* profiling is more sensitive to variations within a training microcycle and may depend more on the training load on the days preceding the measurements. While acknowledging the need to confirm this experimentally, the sensitivity of these measurements to changes in the instantaneous training status of the athletes may have affected the correlation statistical analysis.

A second aspect to be considered relates to the limitations of the extrapolated *F-v* relationship. Measured points on the *F-v* curve cover only about 30% of the velocity values [15]. Mean velocity at the heaviest load during vertical jumping was on average 1.46 m/s. Conversely, the velocity usually reached during 1RM exercise in female athletes is around ~0.2 m/s, which is less by about 700%, than during *F-v* profile assessment [28]. Thus, exercise at 1RM is a much slower movement than during vertical jumping. In this sense, these two measurements demonstrate different kinematic profiles and indicate different qualities in strength training. Moreover, and according to Riviere et al. [15], the force developed during the 1RM measurement is about 16 ±4% higher than the force during the jump against the heaviest load. Given, that Riviere et al. [15] used a half squat 1RM measurement we assume that the force developed during a full squat as applied in our study is higher and therefore less close to the force measured during the jump against the heaviest load. In other words, relatively large differences

exist between the maximum forces developed during the two measurements. This suggests that when examining the relationship between 1RM and *F-v* profile during vertical jumps maximum loads should exceed 100% of BW, hence being closer to the 1RM force exertion.

Conclusions

While force-velocity profiling is a well-known tool in performance diagnostics, our study contributes to the existing literature by examining the relationship of the *F-v* profile during vertical jumps with the relative 1RM in the back squat. Such an approach aims at determining the minimal requirements in strength in order to achieve an optimal profile. The results (1) confirm the force-oriented profiles of resistance-trained athletes and (2) suggest that female athletes with a ratio of 1.2 back squat * bodyweight^{-1} can achieve an optimal *F-v* profile. Such data can be used by coaches and researchers as benchmark values in the evaluation of the athletes' strength qualities. They can provide a reference point to design training interventions based on the one-repetition maximum values by determining the main directions in strength training.

Acknowledgements

The authors would like to thank all athletes who participated in our study. We are grateful to Dr. Tamás Szabó (University of Physical Education) for his substantial contribution in the design and formation of our laboratory allowing the conduction of such measurements. Also, to Balázs Farkas, head coach of CrossFit Mayfly Budapest, who provided the venue to conduct part of the measurements.

References

[1] A. Hill, The heat of shortening and the dynamic constants of muscle, P. Roy. Soc. B-Biol. Sci., 1938, pp. 136-195

[2] P. Jimenez-Reyes, P. Samozino, V. Cuadrado-Penafiel, F. Conceicao, J. Gonzalez-Badillo, J. Morin, Effect of countermovement on power-force-velocity profile. European Journal of Applied Physiology 114 (2014) 2281-2288

[3] P. Jimenez-Reyes, P. Samozino, F. Pareja-Blanco, F. Conceicao, V. Cuadrado-Penafiel, J. Gonzalez-Badillo, J. Morin, Validity of a Simple Method for Measuring Force-Velocity-Power Profile in Countermovement Jump. International Journal of Sports Physiology and Performance 12 (2017) 36-43

[4] J. Morin, P. Samozino, Interpreting Power-Force-Velocity Profiles for Individualized and Specific Training. International Journal of Sports Physiology and Performance 11 (2016) 267-272

[5] P. Samozino, J. Morin, F. Hintzy, A. Belli, A simple method for measuring force, velocity and power output during squat jump. Journal of Biomechanics 41 (2008) 2940-2945

[6] D. Behm, J. Young, J. Whitten, J. Reid, P. Quigley, J. Low, Y. Li, C. Lima, D. Hodgson, A. Chaouachi, O. Prieske, U. Granacher, Effectiveness of Traditional Strength vs. Power Training on Muscle Strength, Power and Speed with Youth: A Systematic Review and Meta-Analysis. Frontiers in Physiology 8:423 (2017)

[7] N. Kawamori, R. Newton, Velocity specificity of resistance training: Actual movement velocity versus intention to move explosively. Strength and Conditioning Journal 28 (2006) 86-91

[8] G. Duthie, W. Young, D. Aitken, The acute effects of heavy loads on jump squat performance: An evaluation of the complex and contrast methods of power development. Journal of Strength and Conditioning Research 16 (2002) 530-538

[9] N. Elbadry, A. Hamza, P. Pietraszewski, D. Alexe, G. Lupu, Effect of the French Contrast Method on Explosive Strength and Kinematic Parameters of the Triple Jump Among Female College Athletes. Journal of Human Kinetics 69 (2019) 225-230

[10] P. Samozino, E. Rejc, P. Di Prampero, A. Belli, J. Morin, Optimal Force-Velocity Profile in Ballistic Movements-Altius: Citius or Fortius? Medicine and Science in Sports and Exercise 44 (2012) 313-322

[11] J. Morin, P. Jimenez-Reyes, M. Brughelli, P. Samozino, When Jump Height is not a Good Indicator of Lower Limb Maximal Power Output: Theoretical Demonstration, Experimental Evidence and Practical Solutions. Sports Medicine 49 (2019) 999-1006

[12] P. Jimenez-Reyes, P. Samozino, J. Morin, Optimized training for jumping performance using the force-velocity imbalance: Individual adaptation kinetics. Plos One 14 (2019)

[13] R. Marcote-Pequeno, A. Garcia-Ramos, V. Cuadrado-Penafiel, J. Gonzalez-Hernandez, M. Gomez, P. Jimenez-Reyes, Association Between the Force-Velocity Profile and Performance Variables Obtained in Jumping and Sprinting in Elite Female Soccer Players. International Journal of Sports Physiology and Performance 14 (2019) 209-215

[14] L. Petridis, G. Pálinkás, Z. Tróznai, B. Béres, K. Utczás, Determining strength training needs using the force-velocity profile of elite female handball and volleyball players. International Journal of Sport Science & Coaching (2021) 123-130

[15] J. Riviere, J. Rossi, P. Jimenez-Reyes, J. Morin, P. Samozino, Where does the One-Repetition Maximum Exist on the Force-Velocity Relationship in Squat? International Journal of Sports Medicine 38 (2017) 1035-1043

[16] K. Wirth, H. Hartmann, A. Sander, C. Mickel, E. Szilvas, M. Keiner, The Impact of Back Squat and Leg-Press Exercises on Maximal Strength and

Speed-Strength Parameters. Journal of Strength and Conditioning Research 30 (2016) 1205-1212

[17] M. Hammami, N. Gaamouri, G. Aloui, R. Shephard, M. Chelly, Effects of a Complex Strength-Training Program on Athletic Performance of Junior Female Handball Players. International Journal of Sports Physiology and Performance 14 (2019) 163-169

[18] L. Rodriguez-Lorenzo, M. Fernandez-Del-Olmo, R. Martin-Acero, Strength and Kicking Performance in Soccer: A Review. Strength and Conditioning Journal 38 (2016) 106-116

[19] G. Haff, S. Nimphius, Training Principles for Power. Strength and Conditioning Journal 34 (2012) 2-12

[20] N. Linthorne, Analysis of standing vertical jumps using a force platform. American Journal of Physics 69 (2001) 1198-1204

[21] H. Vandewalle, G. Peres, H. Monod, STANDARD ANAEROBIC EXERCISE TESTS. Sports Medicine 4 (1987) 268-289

[22] W. Hopkins, S. Marshall, A. Batterham, J. Hanin, Progressive Statistics for Studies in Sports Medicine and Exercise Science. Medicine and Science in Sports and Exercise 41 (2009) 3-12

[23] P. Jimenez-Reyes, P. Samozino, M. Brughelli, J. Morin, Effectiveness of an Individualized Training Based on Force-Velocity Profiling during Jumping. Frontiers in Physiology 7 (2017)

[24] P. Jimenez-Reyes, M. Cross, A. Ross, P. Samozino, M. Brughelli, N. Gill, J. Morin, Changes in mechanical properties of sprinting during repeated sprint in elite rugby sevens athletes. European Journal of Sport Science 19 (2019) 585-594

[25] P. Jimenez-Reyes, P. Sarnozino, A. Garcia-Ramos, V. Cuadrado-Penafiel, M. Brughelli, J. Morin, Relationship between vertical and horizontal force-velocity-power profiles in various sports and levels of practice. Peerj 6 (2018)

[26] P. Comfort, J. McMahon, RELIABILITY OF MAXIMAL BACK SQUAT AND POWER CLEAN PERFORMANCES IN INEXPERIENCED ATHLETES. Journal of Strength and Conditioning Research 29 (2015) 3089-3096

[27] T. Freitas, A. Martinez-Rodriguez, J. Calleja-Gonzalez, P. Alcaraz, Short-term adaptations following Complex Training in team-sports: A meta-analysis. Plos One 12 (2017)

[28] A. Torrejon, C. Balsalobre-Fernandez, G. Haff, A. Garcia-Ramos, The load-velocity profile differs more between men and women than between individuals with different strength levels. Sports Biomechanics 18 (2019) 245-255

Reading Volume Datasets from Storage – Using Segmentation Metadata, for an Enhanced User Experience

Branislav Madoš, Norbert Ádám

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9/A, 042 00 Košice
e-mail: branislav.mados@tuke.sk, norbert.adam@tuke.sk

Abstract: This paper deals with the issues of volume dataset representation as an important part of data storage and processing in many fields including science, research and development, medicine or industry. Due to the significant amount of data included in volume datasets, operations performed on them are often, time- and space-consuming. One of those operations – loading data from secondary storage into the operating memory of computer or memory of graphics card – can be time-consuming and lead to a bad user experience and significantly delay the subsequent processes. Therefore, the main contribution hereof is the design and introduction of an algorithm to generate volume dataset segmentation metadata. It allows (with a small data overhead, as a trade-off) to prepare metadata about splitting the particular volume dataset into segments with different priority levels. Subsequently, it is possible to reorganize the volume dataset according to the priority of the data segments, in descending order. The algorithm proposed herein allows to start the visualization of the volume dataset in its final quality (resembling visualization of the complete volume dataset, although only a part of the data was loaded from the secondary storage), within a fraction of the total load time of the volume dataset. The remaining data are continually read in the background during data visualization, without affecting volume data visualization quality. The first section herein, contains an introduction to the proposed algorithms. Results of tests, performed with different parameter setups on non-invasive medical imaging volume datasets, obtained by computed tomography and magnetic resonance imaging, are included in the second part of the paper. Conclusions, drawn from test results, are summarized in the last part of the paper.

Keywords: volume dataset; three-dimensional image; 3D image, image segmentation algorithm; metadata; user experience; computed tomography; CT; magnetic resonance imaging; MRI

1 Introduction

Volume datasets are often represented as regular three-dimensional grids of scalar values or vectors of scalar values. Volume dataset representation, pre-processing, visualization and other forms of processing are important in the field of science, research and development [1] [2], medicine [3], industry [4] [5], etc. The amount of the space that volume datasets need for their representation in the operating memory of computer or memory of graphics card and even the secondary storage has always been a challenge. Although the capacities of computers, in terms of system memory, secondary storage size and data throughput are constantly and significantly growing, as do the storage space requirements of volume datasets – in terms of their geometrical resolutions and the size of their binary representations (the number of bits) of voxels.

Computed tomography, invented by Sir Godfrey Hounsfield in 1967, was first used to scan a patient in 1971 [6]. Back then, computed tomography produced volume datasets of 64^3 voxels (262144 voxels), taking 2.5 hours to compute. Nowadays, 512^3 voxel computed tomography scans are common: when using 12b/vox, without any compression, they need 192 MB of secondary storage space (256MB when using 16b/vox). Raising the resolution of common volume datasets to $2K^3$ $(2048 \times 2048 \times 2048)$ voxels and using 16b/vox, their size on the secondary storage and in system memory will rise to 16 GB. While in 2007, 70 million CT scans were performed in the USA alone, this number raised to 80 million in 2015 [7].

Not only the size of volume datasets itself can be significant challenges (in terms of both time and space), but also operations performed on them during their pre-processing, visualization and other forms of processing. Modern approaches to volume dataset visualization, in combination with virtual reality, augmented reality [8] [9] or computer vision [10] are even more demanding. Spatiotemporal volume datasets are even more demanding to process. One of those time-consuming operations – loading the volume dataset from secondary storage into system or graphics memory – can result even in bad user experience and can significantly delay the subsequent processes.

That is why the authors decided to work on an algorithm that can enhance the user experience, concerning the loading of the volume dataset, from secondary storage, into operating memory of computer. In [11], we designed an algorithm decomposing the volume dataset into two segments, creating a three-dimensional image – segmentation metadata – and rearranging the volume dataset to allow reading of the important segment of voxels from secondary storage preceding the unimportant segment. In comparison to the original volume dataset, the produced segmentation metadata represent a significant amount of data and that is why we applied lossless compression to the volume dataset metadata (see: Related works section of the paper).

In this work, we build on this previous research, proposing an algorithm to create volume dataset segmentation metadata, to rearrange the volume dataset and to enhance reading the volume dataset from storage.

The contribution lies in the following:

- An algorithm splitting the volume dataset into data blocks, assigning a level of priority (importance) to each data block of the volume dataset, creating volume dataset segmentation metadata to represent the information stored in the dataset and reorganizing the volume dataset (linearizing the segments, ordered by the level of priority in descending order). The segmentation metadata allow reconstruction of the original location of the voxels for each segment of the volume dataset.
- An algorithm that allows to start the visualization of the volume dataset in its final quality (resembling visualization of the complete dataset, although only a corresponding fraction of the dataset was loaded from secondary storage), within a fraction of the total load time of the dataset. The remaining data are continually read in the background, during data processing, without affecting data visualization quality.

The structure of the remainder of this paper is as follows:

Section 2 presents the related works concerning multi-dimensional data linearization, volume dataset segmentation metadata and lossless compression of those segmentation metadata using domain-specific hierarchical data structures based on octant trees and directed acyclic graphs and other lossless compression algorithms including Run-Length Encoding and ΔRLE.

Section 3 introduces the proposed algorithms: one performing volume dataset segmentation and creating the volume dataset segmentation metadata and the other, improving the user experience concerning reading the volume dataset from storage. The inputs of these algorithms, their particular steps and outputs are described in detail.

Section 4 represents the test results of the algorithms described in the previous section using various medical imaging volume datasets using CT and MRI and various parameter setups.

The *Conclusions* section summarizes the conclusions based on the tests, described in Section 4.

2 Related Works

This section mentions only very close related works, related to the linearization of the multi-dimensional data, to the enhancement of the user experience concerning reading volume datasets from secondary storage and compression of volume datasets using hierarchical data structures.

Linearization of multi-dimensional data. Space-Filling Curves (SFC), introduced by Peano and Hilbert at the end of the 19th Century [12] [13], are used for linearization not only of two- or three- but in general of multi-dimensional data. The Morton order is an SFC popular in computer graphics for its better addressing abilities [14] and Hilbert Space Filling-Curve (HSBC) is used in computer science for better locality preserving [15].

Volume dataset segmentation metadata. In [11], we designed an algorithm assigning 1b of metadata to each voxel of the volume dataset – this allows including that voxel into the background (the unimportant voxel segment) or into the region of interest (the important voxel segment) of the volume dataset. It allows rearranging the volume dataset in the manner that all voxels from the important segment are linearized in the first part of the dataset, to be read before the unimportant voxels, stored in the second part of the dataset. In each segment of the rearranged dataset, the order of voxels is the same as it is in linearized form of the original dataset. The size of metadata is 1b per voxel – if 16b are used as the size of the binary representation of voxels, metadata represent 1/16 of the volume dataset size. That is too much, which is why lossless compression of the volume dataset metadata is applied. The 3D metadata image of the volume dataset allows reconstruction of the original location of voxels, for each segment of the volume dataset.

Lossless compression of volume segmentation metadata. In connection with the above-mentioned algorithm, lossless compression of the image segmentation metadata was proposed and different Run-Length Encoding (RLE) schemes were used [16]. Then, ΔRLE – a new compression algorithm based on the combination of Delta encoding and Run-Length Encoding – was tested on the image segmentation metadata of the volume datasets.

Suitable solutions for compressing volume dataset segmentation metadata – the volume dataset itself – are octree-based Hierarchical Data Structures (HDS), also in their pointerless versions [17] [18] (these are suitable for dense volume datasets and can encode multi-bit-value voxels), and sparse, octree-derived, hierarchical data structures - directed acyclic graphs (DAGs) – e.g. Sparse Voxel Directed Acyclic Graphs (SVDAGs) [19], Symmetry-aware Sparse Voxel Directed Acyclic Graphs (SSVDAGs) [20] and Pointerless Sparse Voxel Directed Acyclic Graphs (PSVDAGs) [21]. The latter are suitable for compressing metadata if the volume dataset has only two segments e.g. one bit per voxel can be used for the geometry representation.

3 Proposed Algorithms

This section introduces the main contribution of the paper: an algorithm to create the volume dataset segmentation metadata (described in subsection 3.1) and an

algorithm to enhance the user's experience concerning reading the volume dataset from the secondary storage (described in subsection 3.2).

3.1 An Algorithm to Generate Volume Dataset Segmentation Metadata

This section describes the input, the steps of algorithm and the outputs of the proposed algorithm.

3.1.1 Input

The inputs of the proposed algorithm are as follows:

1) The volume dataset $R[X,Y,Z]$, organized as a regular three-dimensional grid of voxels, with $X,Y,Z \in \mathbb{N}$ grid dimensions. Each voxel $v[x,y,z] \in R: x \in < 0; X-1 >;\ y \in < 0; Y-1 >;\ z \in < 0; Z-1 >; x,y,z \in \mathbb{N}_0$ is represented by a scalar value $val \in < 0; v_{max} >$, where v_{max} is the maximal value.
2) The size of the volume dataset data block that is represented by its dimensions $B_x \in < 1; X >$, $B_y \in < 1; Y >$, $B_z \in < 1; Z >$; $B_x, B_y, B_z \in \mathbb{N}$.
3) Number of priority levels $\rho_{max} \in \mathbb{N}$.

3.1.2 Steps

The following five steps of the algorithm are performed consecutively:

Step 1 of the algorithm divides the volume dataset $R[X,Y,Z]$ into a regular three-dimensional grid $R'[X',Y',Z']$ of voxel data blocks (from the volume dataset); $X' = \left\lceil \frac{X}{B_x} \right\rceil, Y' = \left\lceil \frac{Y}{B_y} \right\rceil, Z' = \left\lceil \frac{Z}{B_z} \right\rceil; X',Y',Z' \in \mathbb{N}$ are the dimensions of the grid. Each data block $[x',y',z'] \in R'$: $x' \in < 0; X'-1 >;\ y' \in < 0; Y'-1 >;\ z' \in < 0; Z'-1 >; x',y',z' \in \mathbb{N}_0$.

Each volume dataset data block has a size B_{size} determined by its dimensions B_x, B_y and B_z, expressed in number of voxels. This can be calculated using the following formula:

$$B_{size} = B_x * B_y * B_z \tag{1}$$

The data block count N (i.e. the number of data blocks, to which the volume dataset is divided), can be calculated as follows:

$$N = \left\lceil \frac{X}{B_x} \right\rceil * \left\lceil \frac{Y}{B_y} \right\rceil * \left\lceil \frac{Z}{B_z} \right\rceil \tag{2}$$

Step 2 To each data block $\delta[x',y',z'] \in R'[X',Y',Z']$ of the volume dataset, its priority level is assigned from the set P with the cardinality $|\rho_{max}|$, when this set is defined as

$$\mathrm{P} = \{0,1,2,\ldots,\rho_{max}-3,\rho_{max}-2,\rho_{max}-1\} \tag{3}$$

$R''[X'',Y'',Z'']$, where $X'' = X', Y'' = Y', Z'' = Z'; X'', Y'', Z'' \in \mathbb{N}$, is the regular three-dimensional grid of scalar values, each of these represents the priority level of the particular data block in the volume dataset. Priority level 0 is assigned to the data blocks having the highest priority, while priority level $\rho_{max}-1$ is assigned to the data blocks having the lowest priority in the volume dataset.

The above-mentioned assignment is performed using function f (this function can be designed according to the specific needs of the particular volume dataset segmentation):

$$R''[X'',Y'',Z''] = f:(R'[X',Y',Z']) \rightarrow \mathrm{P} \tag{4}$$

The priority level of each volume dataset data block is encoded in binary. If fixed-length encoding in the form of unsigned integer values is used, the encoding requires $\lceil \log_2 \rho_{max} \rceil$ bits per data block. The overall number of volume dataset data blocks is quantified using (2) as N, so the overall size of $R''[X'',Y'',Z'']$, the R''_{size} value, in bits, is:

$$R''_{size} = N \times \lceil \log_2 \rho_{max} \rceil \, [b] \tag{5}$$

All data blocks constituting R' and having the same particular priority level $p \in P; p \in <0; \rho_{max}-1>; p \in \mathbb{N}_0$ assigned in R'', constitute segment S_p of that particular priority in the volume dataset. The set of segments Σ with the cardinality $|\rho_{max}|$ may be described as follows:

$$\Sigma = \{S_0, S_1, S_2, \ldots, S_{\rho_{max}-3}, S_{\rho_{max}-2}, S_{\rho_{max}-1}\} \tag{6}$$

Each data block from R' belongs to one of the above segments and therefore:

$$R'[X',Y',Z'] = \bigcup_{r=0}^{\rho_{max}-1} S_r \tag{7}$$

Step 3 Volume dataset R' is linearized, when all of its data blocks are ordered in a one-dimensional stream according to the selected linearization.

Step 4 Volume dataset R' is rearranged, when its data segments are ordered according to the priority in the linearized one-dimensional stream of data blocks, in descending order. In each volume dataset segment, the data blocks are ordered according their order in the linearized volume dataset obtained in ***Step 3*** of the algorithm.

Step 5 Voxels of volume dataset R' are linearized separately in each volume dataset data block according to selected linearization.

3.1.3 Outputs

The output of the algorithm is represented by the volume dataset segmentation metadata stored in R'' and by the linearized rearranged volume dataset R' obtained in ***Step 5***.

3.1.4 Metadata Generation and Volume Dataset Rearrangement – an Example

Figure 1 shows the steps of the proposed algorithm – for the sake of simplicity, using two-dimensions. Figure 1a contains a grid R' of 8×8 pixels.

Figure 1
A two-dimensional (for the sake of the simplicity) example of the steps of the algorithm generating volume metadata and rearranging dataset

In ***Step 1*****,** as shown in Figure 1b, the grid is divided into 16 data blocks forming the 4×4 grid R''. Each data block has 2×2 pixels.

In ***Step 2***, each data block is assigned a priority from the set $P = \{p_0, p_1, p_2, p_3\}$. Priority values are encoded in binary as the vectors

$$p_0 = \text{"00"}, p_1 = \text{"01"}, p_2 = \text{"10"}, p_3 = \text{"11"}$$

and also in color – see the color coding table in Figure 1.

Data blocks constitute the segments $\Sigma = \{S_0, S_1, S_2, S_3\}$ of the volume dataset.

In ***Step 3,*** the data blocks are linearized as shown in Figure 1c. The linearized sequence of the binary priority codes shown in Figure 1c constitutes the linearized form of the segmentation metadata in the algorithm output.

In ***Step 4,*** the data blocks are rearranged (Figure 1d), when all data blocks from the same segment $S_r : r \in < 0; 3 >$ are located, within the segment, in the order of appearance in the linearized form (Figure 1c). The segments are ordered according to their priority levels, in descending order.

Step 5, in which each data block is linearized according the selected linearization, is missing from Figure 1. The rearranged form of the data blocks is the output of the algorithm.

It is possible to reconstruct the original position of each voxel in the volume dataset for each data block loaded from the secondary storage, using rearranged volume dataset and metadata (volume segmentation metadata R''), even if the volume dataset is loaded only partially, i.e. if only some segments of the dataset are loaded from the secondary storage.

In ***Step 1***, volume segmentation metadata is transformed from the linearized form into the three-dimensional grid R'', respecting selected linearization.

In ***Step 2***, volume dataset data blocks forming particular priority segment are placed into the locations of particular priority codes respecting order of volume dataset data blocks in the priority segment.

The number of data blocks in each segment can be obtained as the result of analysis of segmentation metadata R'' in time of original volume dataset reconstruction from rearranged volume dataset and therefore can be omitted from the metadata representation.

3.2 An Algorithm for Improved Reading of Volume Datasets from Storage

The algorithm is based on the above-mentioned algorithm, when all its steps have been performed, using a specific function f for the assignment of the priority level to volume dataset data blocks, on a pre-processed volume dataset.

3.2.1 Input

The proposed algorithm has the following inputs:

1) The volume dataset $R[X, Y, Z]$, organized as a regular three-dimensional grid of voxels, with $X, Y, Z \in \mathbb{N}$ grid dimensions. Each voxel $v[x, y, z] \in R: x \in < 0; X - 1 >;\ y \in < 0; Y - 1 >;\ z \in < 0; Z - 1 >; x, y, z \in \mathbb{N}_0$ is represented by a scalar value $val \in < 0; v_{max} >$ where v_{max} is the maximal value.
2) The size of the volume dataset data block that is represented by its dimensions $B_x \in < 1; X >$, $B_y \in < 1; Y >$, $B_z \in < 1; Z >$; $B_x, B_y, B_z \in \mathbb{N}$
3) Number of priority levels $\rho_{max} = 4$
4) Threshold $\tau \in < 0; v_{max} >$

3.2.2 Steps

Steps P1 and P2 are the pre-processing steps of the algorithm; after these, all five steps of the volume dataset segmentation and metadata generation algorithm are performed, applying function f designed specifically for this algorithm.

Step P1. This step decomposes the volume dataset $R[X,Y,Z]$ (a slice of example volume dataset is shown in Figure 2a) into two segments (unimportant, i.e. background and important, i.e. foreground voxels) and creates a three-dimensional grid $R_s[X,Y,Z]$. As a proof of concept, the authors used a segmentation threshold value τ (that is why τ is one of the inputs of the algorithm):

$$R_s[x,y,z] = \begin{cases} 0 \; if \; R[x,y,z] < \tau \\ 1 \; if \; R[x,y,z] \geq \tau \end{cases} \tag{8}$$

Two segments T_0 and T_1 are formed, when:

$$R[x,y,z] \in T_0 : R_s[x,y,z] = 0$$
$$R[x,y,z] \in T_1 : R_s[x,y,z] = 1 \tag{9}$$

Each volume dataset voxel from R belongs to one of those segments and therefore:

$$R[X,Y,Z] = T_0 \cup T_1 \tag{10}$$

After the ***Step P1*** of the algorithm is performed:

- Segment T_0 contains each voxel of R, having a value lower than the threshold τ (the corresponding value is set to 0 in R_s). In Figure 2b, these are displayed in grey.
- Segment T_1 contains each voxel of R, having a value equal or greater than the threshold τ (the corresponding value is set to 1 in R_s). In Figure 2b, these are displayed in green.

Step P2. The volume dataset is segmented further, into set Φ of four segments U_0, U_1, U_2 and U_3:

$$\Phi = \{U_0, U_1, U_2, U_3\} \tag{11}$$

Each volume dataset voxel from R belongs to one of those segments, therefore:

$$R[X,Y,Z] = \bigcup_{r=0}^{3} U_r \tag{12}$$

After ***Step P2***:

- Segment U_0 contains all voxels of the surface of the region of interest (foreground). In Figure 2c, these are displayed in red.

- Segment U_1 contains all important voxels within the region of interest that are not constituting its surface. In Figure 2c, these are displayed in green.
- Segment U_2 contains all unimportant voxels within the region of interest. In Figure 2c, these are displayed in grey.
- Segment U_3 contains all unimportant voxels beyond the region of interest (background voxels). In Figure 2c, these are displayed in blue.

In the proof of concept, the authors used flood fill to implement this step.

Step P2a. Voxels constituting the borders of the volume dataset grid are examined, each voxel $v[x, y, 0] \in R$, $v[x, y, Z-1] \in R$, $v[x, 0, z] \in R$, $v[x, Y-1, z] \in R$, $v[0, y, z] \in R$, $v[X-1, y, z] \in R$ is added to segment U_3, if this voxel is a part of segment T_0. In that case, this voxel is used also as the start point of the further flood fill (***Step 2b***). If this voxel belongs to segment T_1, it is added to segment U_0.

Step P2b. Flood fill is performed starting from each voxel that was determined as the starting point in ***Step 2a*** and each voxel reached and belonging to segment T_0, not yet added to segment U_3, is added to segment U_3 and used for the further flood fill (***Step 2b***). Each reached voxel from segment T_1, not a member of segment U_0, is added to this segment (it is not used in further flood fill). Flood fill ends when all reached voxels are members of the U_0 or U_3 segments and there are no more starting points for further flood fill available.

Step P2c. All voxels from the dataset R that are members of segment T_0 and were not added to segment U_3 are now added to segment U_2. All voxels from the dataset R that are members of segment T_1 and were not added to segment U_0 are now added to segment U_1.

Step 1 is identical to ***Step 1*** of above-mentioned algorithm (Subsection 3.1.4).

Step 2 is identical to ***Step 2*** of above-mentioned algorithm, with four segments: S_0, S_1, S_2 and S_3.

Function f performs the following mapping:

- Segment S_0 represents the set of surface data blocks of the region of interest. Each data block in this set contains at least one voxel belonging to the U_0 surface voxel set. The assigned priority level is 0. In Figure 2d, these data blocks are displayed in red.
- Segment S_1 represents the set of important data blocks of the region of interest. Each data block in this set contains at least one voxel belonging to the U_1 important voxel set and can contain voxels from segment U_2 but does not contain any voxels from the U_0 surface voxel set nor from the U_3 unimportant voxel set. The assigned priority level is 1. In Figure 2d, these data blocks are displayed in green.

- Segment S_2 represents the set of unimportant data blocks within the region of interest. Each data block is homogenously filled with voxels from U_2, the set of unimportant voxels of the region of interest. The assigned priority level is 2. In Figure 2d, these data blocks are displayed in grey.
- Segment S_3 represents the set of unimportant data blocks beyond the region of interest (the background). Each data block is homogenously filled with voxels from U_3, the set of unimportant voxels beyond the region of interest. The assigned priority level is 3. In Figure 2d, these data blocks are displayed in blue.

Each data block is assigned a 2b priority tag: data blocks from the S_0 segment are tagged with "00", from the segment S_1 with "01", from the S_2 with "10" and from segment S_3 with "11". The cardinality – the number of data blocks in the particular set – of each segment S_0, S_1, S_2 and S_3 is evaluated.

Step 3, ***Step 4*** and ***Step 5*** of the algorithm are identical to the corresponding steps of the algorithm described in subsection 3.1.

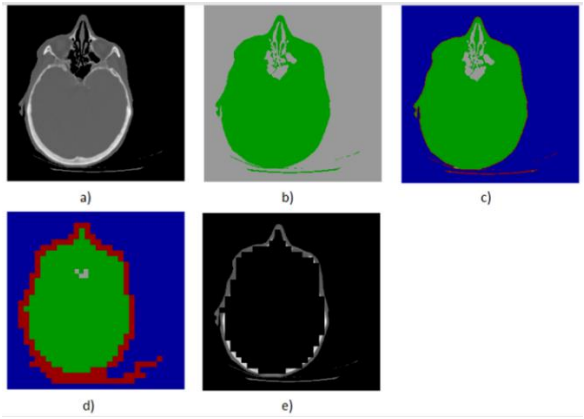


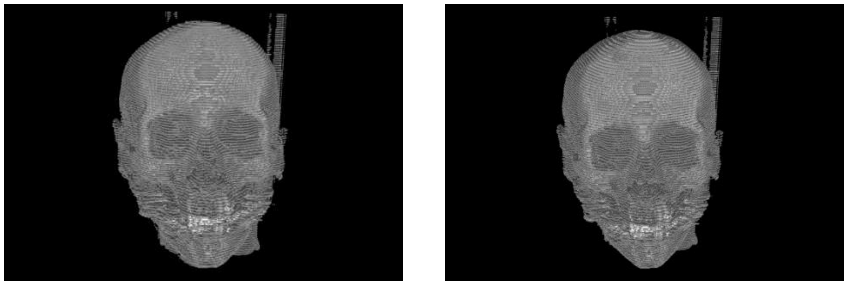
Figure 2

Visualization of algorithm steps: a) slice of the original data; b) slice of binarized data; c) slice of data with four segments of voxels; d) slice of dataset split into volume data blocks of volume segments; e) visualization of data blocks constituting segment S_0

3.2.3 Output

The output of the algorithm is represented by the volume dataset segmentation metadata stored in R'' and by the linearized rearranged volume dataset R' obtained in ***Step 5*** of the algorithm. The enhancement of reading volume datasets from storage lies in the possibility to load the volume dataset segmentation metadata along with the segment S_0 of the dataset, which allows to start the visualization of this segment in 3D immediately – with full possibility of interactions and transformations – while loading the remaining, less important segments on the

background, according to the set priority level. It allows to see surface of all parts of the region of interest in the final quality after only a fraction of the time needed to load the whole volume dataset (see Figure 3).



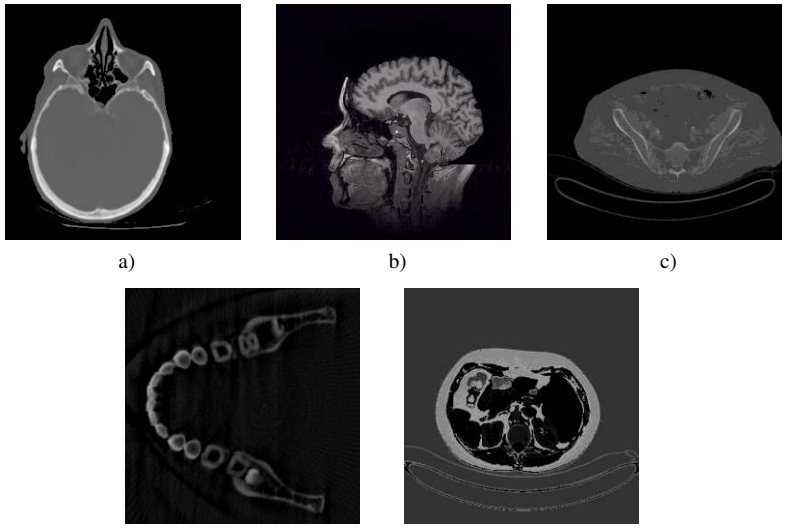
a) b)

Figure 3

Visualization of volume dataset a) only S_0 segment b) all segments

4 Test Results

Tests were performed on volume datasets obtained using medical imaging techniques, including Computed Tomography and Magnetic Resonance Imaging– see the volume dataset parameters in Table 1 and slice visualizations in Figure 4.



a) b) c)

d) e)

Figure 4

Slices of volume datasets obtained using different medical imaging techniques including Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) that were used for testing purposes: a) Head, b) Brain, c) Abdomen, d) Human Skull, e) Pancreas

Voxels of these volume datasets were represented as scalar values. Some volume datasets were encoded as 8-bit unsigned integers per voxel (range: <0;255>), others were encoded as 16-bit unsigned integers per voxel (range: <0;4095>) (using 12 bits for the value and keeping 4 bits reserved).

For test purposes, the authors used cubic data blocks of $n^3; n \in \mathbb{N}$ voxels. Four different data block sizes were selected: $2^3 = 8$, $4^3 = 64$, $8^3 = 512$ and $16^3 = 4096$ voxels, respectively.

Tests were performed on a computer with a four-core Intel® Core i5-8265U @ 1.6 GHz CPU, 8 GB system memory, an NVIDIA GeForce GTX 1050 3 GB graphics card and a 256 GB SSD as the secondary storage.

Table 1

Summary of parameters of the tested volume datasets obtained by the Computed Tomography (CT) and Magnetic Resonance Imaging (MRI)

	Dataset	Dimensions	Voxels [M]	b/vox	Size [MB]	Threshold	Active voxels [mil] / [pct]
A	**Head**	256 × 256 × 112	7.34	16	14.0	200	2.437 (33.21%)
B	**Brain**	256 × 256 × 96	6.29	16	12.0	1150	1.670 (26.54%)
C	**Abdomen**	512 × 512 × 160	41.94	16	80.0	200	21.646 (51.61%)
D	**Human skull**	256 × 256 × 256	16.78	8	16.0	25	2.199 (13.11%)
E	**Pancreas**	240 × 512 × 512	62.91	16	120.0	1150	57.114 (90.78%)

In all dataset tests, when only data blocks of the S_0 segment – the surface data blocks of the region of interest – of the volume datasets were loaded from secondary storage and the size of data block was 2^3 voxels, only a fraction of data blocks – ranging from 2.96% (in case of dataset e) to 15.47% (in case of dataset d) – had to be read. Considering also metadata, the share of data needed to load rose, ranging from 4.52% (in case of dataset e) to 18.60% (in case of dataset d) (see Table 2 and Figure 5).

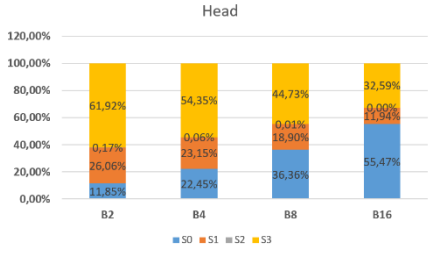
In general, an increase in the data block size generates – compared to the total voxel count – an increased share of voxels from the S_0 data block set and a decreased share of voxels from the S_1, S_2 and S_3 data block sets. For example, in dataset a, the share of S_0 raised from 11.85% (using 2^3 voxels pre data block) to 55.47% (using 16^3 voxels per data block). The total overhead (image segmentation metadata) is dependent on the data block size. In the tests, this ranged from 0.003% (using 16^3 data block size) to 1.563% (using 2^3 data block size) with datasets encoded using 16 bits per voxel and from 0.006% to 3.125% in case of datasets encoded using 8 bits per voxel.

The share of the S_0 segment of the volume dataset and the share of metadata are inversely proportionate. That is why there is a trade-off between those two parameters. Increasing the metadata share to 1.563% and/or 3.125% allows a significant reduction of the share of the S_0 dataset segment, leading to an overall improvement of the user experience.

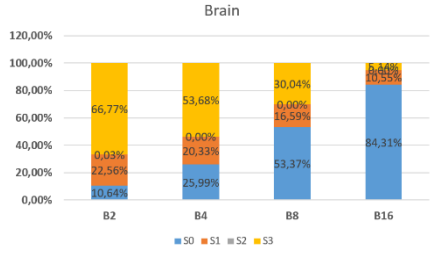
Table 2

Test results where image segmentation metadata were created for five volume datasets and four data block sizes, ranging from 2^3 to 16^3 voxels. S_0 is the segment of surface blocks, S_1 is the segment of important data blocks within the region of interest, S_2 is the segment of unimportant data blocks within the region of interest and S_3 is the segment of unimportant blocks of background data

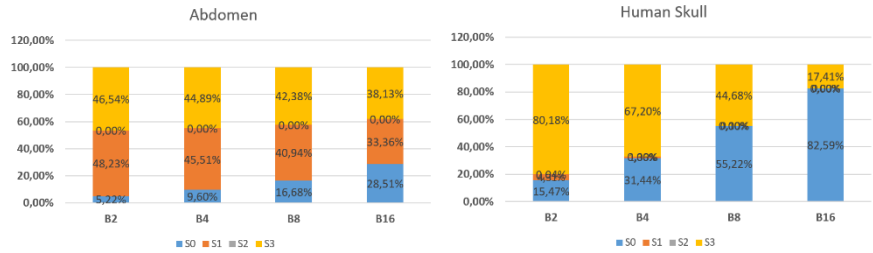
Block size	S_0		S_1		S_2		S_3		Blocks total	Metadata	
	Blocks	%	Blocks	%	Blocks	%	Blocks	%		[KB]	[%]
a – Head											
2^3	108769	11.85	239105	26.06	1536	0.17	568094	61.92	917504	224.00	1.563
4^3	25745	22.45	26546	23.15	69	0.06	62328	54.35	114688	28.00	0.195
8^3	5213	36.36	2709	18.90	2	0.01	6412	44.73	14336	3.50	0.024
16^3	994	55.47	214	11.94	0	0.00	584	32.59	1792	0.44	0.003
b – Brain											
2^3	83652	10.64	177408	22.56	261	0.03	525111	66.77	786432	192.00	1.563
4^3	25546	25.99	19989	20.33	0	0.00	52769	53.68	98304	24.00	0.195
8^3	6558	53.37	2039	16.59	0	0.00	3 691	30.04	12288	3.00	0.024
16^3	1295	84.31	162	10.55	0	0.00	79	5.14	1536	0.38	0.003
c – Abdomen											
2^3	273824	5.22	2528784	48.23	54	0.00	2440218	46.54	5242880	1280.00	1.563
4^3	62901	9.60	298245	45.51	0	0.00	294214	44.89	655360	160.00	0.195
8^3	13664	16.68	33541	40.94	0	0.00	34715	42.38	81920	20.00	0.024
16^3	2919	28.51	3416	33.36	0	0.00	3905	38.13	10240	2.50	0.003
d - Human skull											
2^3	324435	15.47	90337	4.31	835	0.04	1681545	80.18	2097152	512.00	3.125
4^3	82412	31.44	3574	1.36	10	0.00	176148	67.20	262144	64.00	0.391
8^3	18093	55.22	34	0.10	0	0.00	14641	44.68	32768	8.00	0.049
16^3	3383	82.59	0	0.00	0	0.00	713	17.41	4096	1.00	0.006
e – Pancreas											
2^3	232906	2.96	6970321	88.63	15	0.00	661078	8.41	7864320	1920.00	1.563
4^3	60037	6.11	843948	85.85	0	0.00	79055	8.04	983040	240.00	0.195
8^3	14979	12.19	98721	80.34	0	0.00	9180	7.47	122880	30.00	0.024
16^3	3616	23.54	10784	70.21	0	0.00	960	6.25	15360	3.75	0.003



a)

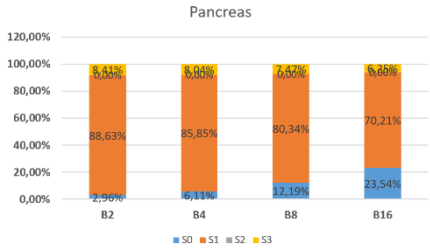


b)



c)

d)



e)

Figure 5

Shares of the respective segments of the corresponding volume datasets and block sizes, where the number of voxels in the block is: $B2 = 2^3$ voxels, $B4 = 4^3$ voxels, $B8 = 8^3$ voxels and $B16 = 16^3$ voxels

Tests show that, in most optimistic case, the user can start working with the volume dataset after only 4.52% of the total time required to load the whole volume dataset from secondary storage, when image segmentation metadata and all S_0 data blocks are loaded.

Table 3

Share of data blocks, when loading only segment S_0, both S_0 and S_1 and all of S_0, S_1 and S_2, respectively, for the corresponding volume datasets and data block sizes. Loading all segments ($S_0 \cup S_1 \cup S_2 \cup S_3$) represents 100% of the dataset

Block size	$\boldsymbol{S_0}$ [%]	$\boldsymbol{S_0 \cup S_1}$ [%]	$\boldsymbol{S_0 \cup S_1 \cup S_2}$ [%]
a – Head			
$\mathbf{2^3}$	11.85%	37.92%	38.08%
$\mathbf{4^3}$	22.45%	45.59%	45.65%
$\mathbf{8^3}$	36.36%	55.26%	55.27%
$\mathbf{16^3}$	55.47%	67.41%	67.41%
b-Brain			
$\mathbf{2^3}$	10.64%	33.20%	33.23%
$\mathbf{4^3}$	25.99%	46.32%	46.32%
$\mathbf{8^3}$	53.37%	69.96%	69.96%
$\mathbf{16^3}$	84.31%	94.86%	94.86%

Block size	S_0 [%]	$S_0 \cup S_1$ [%]	$S_0 \cup S_1 \cup S_2$ [%]
c – Abdomen			
2^3	5.22%	53.46%	53.46%
4^3	9.60%	55.11%	55.11%
8^3	16.68%	57.62%	57.62%
16^3	28.51%	61.87%	61.87%
d - Human skull			
2^3	15.47%	19.78%	19.82%
4^3	31.44%	32.80%	32.80%
8^3	55.22%	55.32%	55.32%
16^3	82.59%	82.59%	82.59%
e – Pancreas			
2^3	2.96%	91.59%	91.59%
4^3	6.11%	91.96%	91.96%
8^3	12.19%	92.53%	92.53%
16^3	23.54%	93.75%	93.75%

We may assume that using the same volume dataset at a higher resolution and using the same data block size, the share of the S_0 segment – in terms of both data blocks and voxels, compared to all data blocks and voxels – will be smaller. That will result in a shorter load time of the S_0 segment, compared to the load time of the total volume dataset.

Another possible use case allows loading of the S_0, $S_1$and S_2 segments of the volume dataset, excluding the S_3 segment (the unimportant data outside the region of interest). This is useful when the system or the graphics card is short on memory or to shorten the total load time of the volume dataset. In the performed tests, the load time of the S_0, $S_1$and S_2 segments of the volume datasets using 2^3 data block size ranged from 19.82% (in case of dataset d) to 91.59% (in case of dataset e). Considering also metadata, the values range from 22.95% (in case of dataset d) to 93.15% (in case of dataset e).

Conclusions

This paper examined the issues of volume data representation. Due to the large amount of data included in volume datasets, many operations, including loading into operating memory of computer or graphics card memory, can be time consuming and lead to a negative user experience. For this reason, the algorithm generating segmentation metadata and subsequently rearranging the volume dataset, was developed. It improves the user experience, concerning loading the data from secondary storage, into the operating memory of the computer or the graphics card memory.

In performed tests, it allowed for work to begin on the volume dataset, in a fraction (4.52% – 18.60%) of the total time required to load the whole volume

dataset into operating memory, giving the impression of loading the complete volume dataset. However, this comes with a trade-off – a 1.563% and 3.125% data overhead, respectively. The algorithm allows for the stepwise loading of the volume dataset – each step represents the loading of a less important portion (a segment having lower priority) of the volume dataset. This process can be interrupted after each step. Therefore, loading unimportant data can be omitted, decreasing the total load time. The time required, for loading the S_0, $S_1$and S_2 segments, using 2^3 data block size, ranged from 19.82% to 91.59% in the performed tests. These values rose from 22.95% to 93.15%, when also considering metadata.

In future research, we will focus on the structure and encoding of the volume dataset segmentation metadata. The constant yet relatively high ratio of metadata size and data block size, can be significantly decreased, using lossless compression. Hierarchical data structures may also be used, albeit, their potential, has yet to be investigated. Smaller metadata size can contribute to further minimization of the load time of the first segments of the volume datasets.

Acknowledgement

This research was supported by the Slovak Research and Development Agency, project number APVV-18-0214. The volume datasets are courtesy of the following: CT Cadaver Head and MR Brain – The University of North Carolina Volume Rendering Test Data Set; Abdomen – Michael Meißner, Viatronix Inc., Human skull – Siemens Medical Solutions; Pancreas – DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation.

References

[1] L. Főző, R. Andoga, L. Madarász, Mathematical model of a small Turbojet Engine MPM-20. In: Studies in Computational Intelligence Vol. 313: International Symposium of Hungarian Researchers on Computational Intelligence and Informatics. - Heidelberg: Springer, 2010, pp. 313-322 - ISBN 978-3-642-15220-7 - ISSN 1860-949X

[2] L. Nyulaszi, R. Andoga, P. Butka, et al., Fault Detection and Isolation of an Aircraft Turbojet Engine Using a Multi-Sensor Network and Multiple Model Approach, In. Acta Polytechnica Hungarica Vol. 1, No. 2, pp. 189-209, 2018, DOI: 10.12700/APH.15.1.2018.2.10

[3] P. Varga, M. Schnitzer, M. Trebuňová, R. Hudák and J. Živčák, Overview of the Current Methods for Reduction of Artifacts in CT and MR Imaging for Implants made by Additive Manufacturing, In: Acta Tecnología: International Scientific Journal about Technologies. - Šemša (Slovakia), Vol. 6, No. 2 (2020) pp. 55-58 - ISSN 2453-675X

[4] R. Andoga, L. Főző, R. Kovács, K. Beneda, T. Moravec, M. Schreiner, Robust Control of Small Turbojet Engines. Machines 2019, 7, 3, https://doi.org/10.3390/machines7010003

[5] S. Grys, L. Vokorokos and L. Borowik, Size determination of subsurface defect by active thermography – Simulation research, In: Infrared Physics & Technology. Vol. 62 (2014), pp. 147-153 - ISSN 1350-4495

[6] C. Richmond (2004) Obituary – Sir Godfrey Hounsfield, BMJ. 329 (7467): 687, doi:10.1136/bmj.329.7467.687

[7] A. Berrington de González, M. Mahesh, KP. Kim, M. Bhargavan, R. Lewis, F. Mettler and C. Land, (December 2009) Projected cancer risks from computed tomographic scans performed in the United States in 2007, Arch. Intern. Med. 169 (22): 2071-7

[8] B. Sobota and M. Guzan, Virtualization of Chua's Circuit State Space. In: Recent Advances in Chaotic Systems and Synchronization: From Theory to Real World Applications. - London, Great Britain : Elsevier Science pp. 127-164 [print] - ISBN 978-0-12-815838-8

[9] Zs. Racz, B. Sobota and M. Guzan, Parallelizing Boundary Surface Computation of Chua's Circuit - 2017. In: RADIOELEKTRONIKA 2017 - Danvers : IEEE, 2017, pp. 1-4 - ISBN 978-1-5090-4592-1

[10] L. Vokorokos, E. Chovancová, J. Radušovský and M. Chovanec, A Multicore Architecture Focused on Accelerating Computer Vision Computations - 2013. In: Acta Polytechnica Hungarica. Vol. 10, No. 5 (2013) pp. 29-43 - ISSN 1785-8860

[11] B. Madoš, A. Baláž, N. Ádám, J. Hurtuk and Z. Bilanová, Algorithm Design for User Experience Enhancement of Volume Dataset Reading from Storage Using 3D Binary Image as the Metadata - 2019. In: SAMI 2019 : IEEE 17th World Symposium on Applied Machine Intelligence and Informatics. - Danvers (USA) : Institute of Electrical and Electronics Engineers pp. 269-274 [print, online] - ISBN 978-1-7281-0249-8

[12] A. Laszloffy, J. Long and A. K. Patra, Simple data management, scheduling and solution strategies for managing the irregularities in parallel adaptive finite element simulations. Parallel Computing, 26, ISSN 1765-1788

[13] H. Sagan, Space-Filling Curves, Springer Verlag, 1994, eBook ISSN 978-1-4612-0871-6, ISBN 978-0-387-94265-0, DOI 10.1007/978-1-4612-0871-6

[14] G. M. Morton, A Computer Oriented Geodetic Data Base and a New Technique in File Sequencing, Research Report. International Business Machines Corporation (IBM), Ottawa, Canada, 20, pp. 20, March 1st, 1966. Available: https://dominoweb.draco.res.ibm.com/reports/Morton1966.pdf

[15] D. Hilbert, Via the continuous mapping of a line onto a patch of area. Mathematical annals (orig. Über die stetige Abbildung einer Linie auf ein Flächenstück. Mathematische Annalen) 38 (1891), pp. 459-460

[16] B. Madoš and N. Ádám, Evaluation of Encoding Schemas for Optimization of Bit-Level Run-Length Encoding Within Lossless Compression of Binary Images - 2019. In: Intelligent Engineering Systems. - Budapest (Hungary): IEEE Industrial Electronics Society pp. 75-80 - ISBN 978-1-7281-1212-1

[17] B. Madoš, E. Chovancová and M. Hasin, Evaluation of Pointerless Sparse Voxel Octrees Encoding Schemes Using Huffman Encoding for Dense Volume Datasets Storage, In: ICETA 2020: 18th IEEE International conference on emerging elearning technologies and applications- Denver (USA) : Institute of Electrical and Electronics Engineers pp. 424-430, ISBN 978-0-7381-2366-0

[18] B. Madoš, N. Ádám and M. Štancel, Representation of Dense Volume Datasets Using Pointerless Sparse Voxel Octrees with Variable and Fixed-Length Encoding, IEEE 19th World Symposium on Applied Machine Intelligence and Informatics, SAMI 2021, Herľany, Slovakia, Jan., 21-23, 2021, p. 6

[19] V. Kämpe, E. Sintorn, and U. Assarsson, High Resolution Sparse Voxel DAGs. ACM Transactions on Graphics. 32, 4, Article 101 (July 2013) p. 8, ISSN 0730-0301, DOI: https://doi.org/10.1145/2461912.246 2024

[20] A. J. Villanueva, F. Marton, and E. Gobbetti, SSVDAGs: Symmetry-aware Sparse Voxel DAGs. In Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '16) February 27-28 2016, Redmond, WA, USA, pp. 7-14, ACM, New York, NY, USA. ISBN: 978-1-4503-4043-4/16/03, DOI: https://doi.org/10.1145/2856400.2856420

[21] L. Vokorokos, B. Madoš and Z. Bilanová, PSVDAG: Compact Voxelized Representation of 3D Scenes Using Pointerless Sparse Voxel Directed Acyclic Graphs", In: Computing and Informatics: Computers and Artificial Intelligence. - Bratislava (Slovakia), Vol. 39, No. 3 (2020), pp. 587-616 [print] - ISSN 1335-9150

On the Irregularity Characterization of Mean Graphs

Tamás Réti, István Bárányi

Bánki Donát Faculty of Mechanical and Safety Engineering, Óbuda University, Népszínház u. 8, H-1081 Budapest, Hungary, reti.tamas@bgk.uni-obuda.hu, baranyi.istvan@bgk.uni-obuda.hu

Abstract: A connected non-regular graph G with n vertices and m edges is called a mean graph, if there exists a p ≥ 2 integer for which p=[G]=2 m/n holds. The topological index p=p(G) is called the centrality parameter of graph G. It is obvious that, if G is a mean graph, then its centrality parameter p(G) is a uniquely defined positive integer. Mean graphs represent a particular subset of connected non-regular graphs. In this note, by presenting relevant examples, some structural irregularity properties of mean graphs are studied and characterized. Comparing the degree deviations S(G) and S(H) of mean graphs G and H having equal centrality parameter p(G)=p(H) it is proved that if the only difference in the corresponding degree sequences of G and H is that the number of vertices of degree p is different, then S(G)=S(H). The smallest mean graph is the 4-vertex unicyclic graph having a degree sequence (3, 2, 2, 1). This graph is isomorphic to the 4-vertex antiregular graph A_4, for which $S(A_4)=2$ holds. Using comparative tests on preselected connected graphs it has been shown that the degree deviation S(G) is poorly suited for discriminating among non-regular graphs.

Keywords: non-regular graphs; irregularity indices; antiregular graphs; complete split graphs; balanced bidegreed graphs; degree deviation

1 Preliminary

Let G=(V, E) be a finite, simple connected graph with n vertices and m edges. For connected graph G, denote by d(v) the degree of a vertex v and by e=uv the edge connecting vertices u and v.

Let $\Delta=\Delta(G)$ and $\delta=\delta(G)$ be the maximum and minimum degrees, respectively, of vertices of G, where N_Δ and N_δ stand for the number of of vertices of degree $\Delta=\Delta(G)$ and $\delta=\delta(G)$, respectively. The average degree of a graph G denoted by [G], it is equal to [G]=2 m/n.

Using the standard terminology [1], a graph G is called R-regular if all its vertices have the same degree R. A connected graph is called non-regular if it contains at

least two vertices with different degrees. A non-regular graph G is said to be k-degreed, if its degree set contains exactly k different degrees. Consequently, a bidegreed graph G(Δ,δ) is a non-regular graph whose vertices have exactly two different degrees. An n ≥4 vertex bidegreed graph G(Δ,δ) is called a balanced graph if n is an even integer for which $N_\Delta = N_\delta = n/2$ holds. Among n-vertex connected bidegreed graphs, path P_4 with degree sequence (2,2,1,1) and the so-called diamond graph G_D with degree sequence (3,3,2,2) represent the smallest balanced bidegreed graphs.

As usual, the cyclomatic number of a connected graph with n vertices and m edges is defined as Q=Q(G)= m – n +1. A connected graph G having Q(G)=k ≥1 cycles is said to be a k-cyclic graph. As a particular case, if Q(G)=0 holds the corresponding acyclic graph is called a tree graph. A tree with n vertices has exactly n-1 edges.

For a connected graph G with n vertices and m edges, the mean degree of G denoted by [G] is defined as [G]=2 m/n. A non-regular graph G is called a *mean graph* if there exists a positive integer p for which p=[G] holds. This positive integer p is a uniquely defined graph invariant, it is said to be the centrality parameter of graph G. From the definition of mean graphs it follows that the centrality parameter is determined by the degree sequence of graphs, the value of p(G) does not depend on the distances between vertices in a graph. Consequently, the definition of mean graphs is independent of the distance-based centrality concept known from classical graph theory.

A connected graph G is called a *really mean graph* if it is a mean graph and the degree set of G contains at least one vertex of degree p. In a really mean graph a vertex of degree p is called a *mean vertex*.

The number of mean vertices in a mean graph G is denoted by $N_p(G)$. A mean graph G is said to be a minimal graph if it has no mean vertices. An edge uv in a really mean graph G is called a *mean edge* if d(u)=d(v)=p holds. The number of mean edges of G is denoted by $M_p(G)$.

By definition, a topological invariant IT(G) of a graph G is called an irregularity index if IT(G) ≥ 0 and IT(G)=0 if and only if graph G is a regular graph. The degree deviation S(G) of a graph G belongs to the family of most popular graph irregularity indices. This graph invariant was introduced by Nikiforov [2], and for a connected non-regular graph G with n vertices and m edges it is defined as

$$S(G) = \sum_{i=1}^{n} \left| d_i - \frac{2m}{n} \right| = \sum_{i=1}^{n} |d_i - [G]|. \qquad (1)$$

2 Some Fundamental Properties of Mean Graphs

It is obvious that every non-regular unicyclic graph G is a really mean graph with identical centrality parameter p(G)=2. The smallest mean graph is the 4-vertex unicyclic graph with a degree sequence (3, 2, 2, 1). It is isomorphic to the 4-vertex antiregular graph A_4, for which $S(A_4)=2$ holds.

Proposition 1 Let G be a non-regular unicyclic graph and denote by H the unicyclic graph generated by inserting into G some vertices of degree 2. Then S(G)=S(H).

Proof. Because the average degree [G]=2 m/n of a unicyclic graph G is equal to 2, it follows that S(G) and S(H) are independent of the number of vertices with degree 2, consequently S(G)=S(H) holds. This observation for 8 and 12 vertex unicyclic graphs U_8 and U_{12} is demonstrated in Fig. 1.

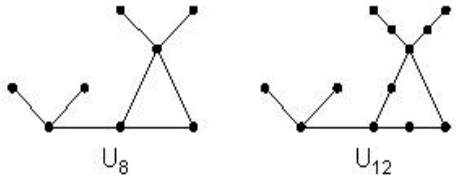


Figure 1
Four-degreed unicyclic graphs U_8 and U_{12} with centrality parameter of p=2

Graph U_8 with 8 vertices has a degree sequence $DDs(U_8)= (4^1 , 3^2 , 2^1, 1^4)$ and graph U_{12} with 12 vertices has a degree sequence $DDs(U_{12})= (4^1, 3^2 , 2^5, 1^4)$. Because the difference between degree sequences of two graphs is represented only in the different numbers of vertices of degree 2, this implies that U_8 and U_{12} have identical degree deviation $S(U_8)=S(U_{12})=8$.

Based on the above considerations the following proposition can be obtained.

Proposition 2 Let G be a mean graph having n vertices and m edges and a centrality parameter p=p(G). Denote by H the graph constructed from G by inserting into G finite number vertices of degree p. Then H will be a really mean graph, consequently p(G)=p(H) and S(G)=S(H) hold. It follows that the number $N_p(H)$ of mean vertices of degree p in H can be arbitrarily large.

Example 1 One bidegreed and two tridegreed mean graphs are depicted in Fig. 2. These graphs have different degree sequences $DDs(Y_6) = (5^3, 3^3)$, $DDs(TY_7) = (5^3, 4^1, 3^3)$, $DDs(TY_9) = (5^3, 4^3, 3^3)$ but identical degree deviation $S(Y_6) = S(TY_7) = S(TY_9) = 6$.

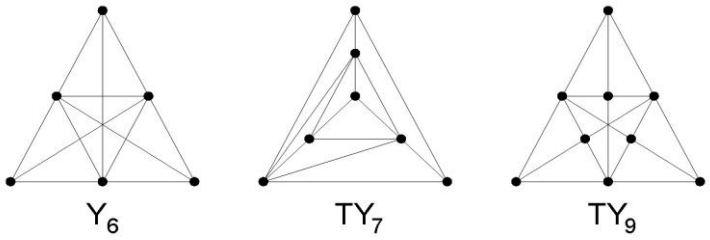


Figure 2
Graphs with identical centrality parameter p=4

Remark 1 For an n-vertex mean graph G consider the topological invariant defined by n/p(G). It is conjectured that for a mean graph G the sharp inequality $n/p(G) \geq 3/2$ holds. For example, $n/p(Y_6)=3/2$ for graph Y_6 depicted in Fig. 2.

Proposition 3 There are no mean graphs with cyclomatic number Q = 0, 2 and 3.

Proof. For an n-vertex connected graph G

$$p(G) = \frac{2(Q(\mathrm{G}) + n - 1)}{n} = \frac{2Q(\mathrm{G})}{n} + 2 - \frac{2}{n} \qquad (2)$$

It follows that if Q(G) = 0, 2 or 3, then G cannot be a mean graph.

Proposition 4 For the centrality parameter p of a mean graph the inequality $2 \leq p \leq \Delta\text{-}1$ holds.

Proof. As an example, consider the unicyclic graphs depicted in Fig. 3. For these mean graphs $\Delta = 3$, $\delta = 1$ and $2= p= \Delta - 1=2$ holds.

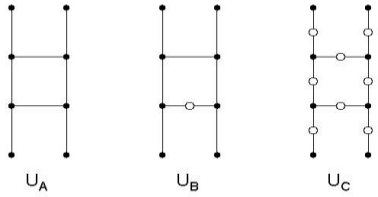


Figure 3
Tridegreed graphs U_B and U_C constructed from bidegreed unicyclic graph U_A

As can be seen, the number of mean vertices can be an arbitrary non-negative integer (i.e. $N_2 = 0, 1, 2, 3, \ldots$).

Proposition 5 Let $j \geq 3$ be an arbitrary integer. Then there exists a bidegreed balanced mean graph F_n with n=2j vertices for which $p(F_n)=3$ and $Q(F_n)=1 + n/2$ and $S(F_n)=n$ hold.

Proof. Consider the infinite sequence of n-vertex balanced bidegreed mean graphs F_n depicted in Fig. 4.

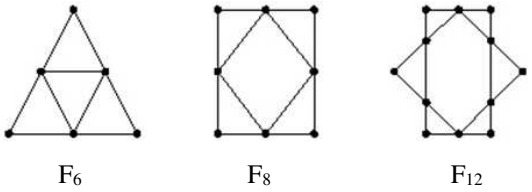


Figure 4
Bidegreed balanced mean graphs F_n having degree deviation of $S(F_n)=n$

The fundamental property of graphs F_n with $n\geq 6$ vertices is that $p(F_n)=3$, $Q(F_n)=1 + n/2$ and $S(F_n)=n$ hold for them. From bidegreed graphs F_n tridegreed really mean graphs of various type can be constructed by inserting vertices of degree 3 into F_n.

Proposition 6 Let $j\geq 3$ be an arbitrary integer. Then there exists an infinite sequence of tridegreed really mean graphs H_n with $n=2j$ vertices for which $p(H_n)=3$, $Q(H_n)=1 + n/2$ and $S(H_n)=4$ hold.

Proof. Consider the tridegreed mean graphs depicted in Fig. 5.

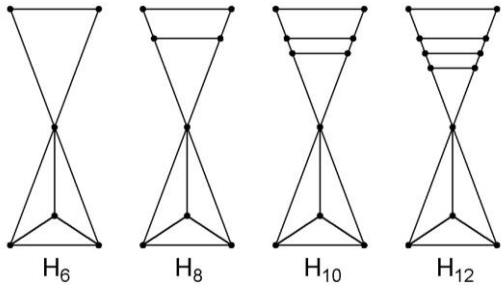


Figure 5
Tridegreed mean graphs H_n having identical degree deviation, S=4

Graphs H_6, H_8, H_{10} and H_{12} contain 3, 5, 7 and 9 mean vertices, respectively. It is easy to check that for all tridegreed graphs $p(H_n)=3$, $Q(H_n)=1 + n/2$ and $S(H_n)=4$ hold.

There exist infinitely many complete bipartite graphs belonging to the family of mean graphs. It is easy to prove the following proposition.

Proposition 7 Let $K_{\Delta,\delta}$ be a complete bipartite graph with $n= \Delta+\delta$ vertices and $m= \Delta\delta$ edges where $\delta\geq 2$ even integer and $\Delta=3\delta$. Then $K_{\Delta,\delta}$ is a mean graph with a centrality parameter $p(K_{\Delta,\delta})= \Delta/2 = 3\delta/2$ and $S(K_{\Delta,\delta}) = p(K_{\Delta,\delta})(\Delta-\delta) = m$.

Remark 2 The smallest complete bipartite mean graph is the graph $K_{2,6}$ with 8 vertices and 12 edges, for which $p(K_{2,6})=3$ and $S(K_{2,6})= m = 12$.

3 Pseudo-Antiregular Mean Graphs

A connected n-vertex graph A_n whose degree set consists of n-1 elements is called an antiregular graph [3, 4, 5]. It follows that a connected antiregular graph has exactly two vertices of the same degree. These two vertices with same degree are called exceptional vertices [5].

Lemma 1 [6]: Let G be an n-vertex connected triangle-free graph. Then for every edge uv in G the inequality $d(u) + d(v) \leq n$ holds.

Lemma 2 [4, 5]: Two vertices u and v of a connected n-vertex antiregular graph A_n are adjacent if and only if $d(u) + d(v) \geq n$.

Remark 3 There exists n-vertex connected graph G for which $d(u) + d(v) \geq n$ holds for every edge uv of G, but G does not belong to the family of antiregular graphs. For example, such graphs where $d(u) + d(v) = n$ holds for every edge uv are the n-vertex stars S_n.

Proposition 8 Let A_n be an $n \geq 4$ vertex antiregular graph where n is an even integer. Then graph A_n is a mean graph and for its average degree $p=[A_n]=n/2$ holds. It follows that A_n has exactly two mean vertices (exceptional vertices) with degree p=n/2.

Proof. It is known that the edge number $m(A_n)$ of an n-vertex antiregular graph is $m(A_n) = \lceil n/2 \rceil \lfloor n/2 \rfloor$. It follows that if n is even integer then $p=[A_n]=n/2$ and the corresponding edge number is $m(A_n)=n^2/4$.

Remark 4 One can easily determine the degree deviation of antiregular graphs A_n with $n \geq 4$ even vertex number. It is

$$S(A_n) = 2 \sum_{i=1}^{p(A_n)-1} i = 2 \sum_{i=1}^{n/2-1} i \qquad (3)$$

where $p(A_n)=n/2$.

Let uv be an edge of a connected n-vertex graph G. Edge uv is called a strong edge of G, if $d(u) + d(v) \geq n$, and edge uv is called a weak edge of G, if $d(u) + d(v) < n$ holds. It is easy to check that in a connected antiregular graph A_n every edge is strong.

The construction of *pseudo-antiregular mean graphs* is based on the following concept. Let A_n be a traditional n-vertex connected antiregular graph with n vertices where $n \geq 4$ is an arbitrary even integer and $p(A_n)=n/2$. Now, by inserting k novel vertices of degree n/2 into A_n as a result of this operation one obtains an N-vertex pseudo-antiregular mean graph $PA_N(n,k)$ with centrality parameter $p(PA_N(n,k))=n/2$ and with vertex number N=n+k. As can be observed, the vertex

sequences of graphs A_n and $PA_N(n,k)$ are different, but the degree sets of both graphs are identical.

In Fig. 6 six really mean graphs with equal centrality parameter p=3 are depicted. Graph J_1 is isomorphic to the traditional connected 6-vertex antiregular graph A_6. Graphs J_2, J_3, JA_4, JB_4 and JC_4 are pseudo-antiregular mean graphs.

Their common properties are as follows: all of them have mean vertices and mean edges, they have identical degree sets (1, 2, 3, 4, 5), and for them the corresponding degree deviation is equal, namely S=6.

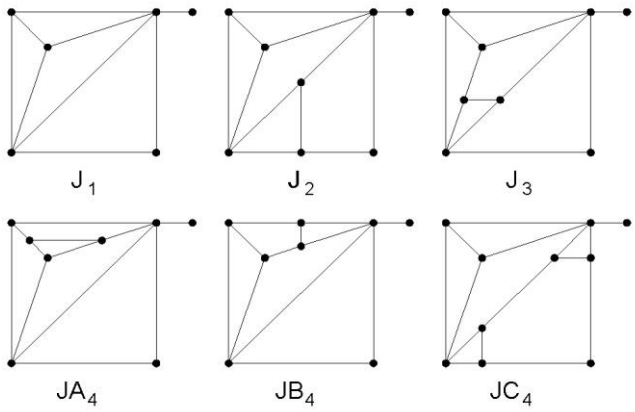


Figure 6
Non-regular graphs with the same degree set (1, 2, 3, 4, 5) and the same degree deviation S=6

Every edge of antiregular graph J_1 is a strong edge. Graphs J_2, J_3, JA_4 and JB_4 are 8-vertex graphs, they have 3 strong edges and 9 weak edges. Graphs J_2 and J_3 have exactly 2 and 3 mean edges, respectively. Graphs JA_4 and JB_4 are non-isomorphic graphs including 4 mean edges. Graph JC_4 with 10 vertices has also 4 mean edges, among its 15 edges every edge is a weak edge.

From the previous considerations the following proposition yields.

Proposition 9 Let n and k be integer numbers where $n \geq 4$ is even, and $k \geq 0$. Then, for appropriately selected n and k parameters there exist (n-1) degreed really mean graphs with centrality parameter p=n/2. Such graphs are the traditional antiregular graphs A_n with even $n \geq 4$ vertex number and k=0, moreover the corresponding pseudo-antiregular mean graphs $PA_N(n,k)$ with vertex number N=n+k where $k \geq 1$.

4 Two Conjectures

Additionally, for the structural characterization of connected mean graphs we introduce two novel graph irregularity indices formulated as

$$IRD(G) = \frac{2N_{\Delta}N_{\delta}}{N_{\Delta} + N_{\delta}}(\Delta - \delta) \tag{4}$$

$$IRR(G) = \frac{n}{2}(\Delta - \delta) \tag{5}$$

where N_{Δ} is the number of vertices of degree Δ and N_{δ} is the number of vertices of degree δ, respectively.

Let G be an arbitrary non-regular connected n-vertex and m-edge graph with maximum degree Δ and minimum degree $\delta \geq 1$. For graph G, the following conjectures are established.

Conjecture 1 It is conjectured that

$$S(G) = \sum_{i=1}^{n} \left| d_i - \frac{2m}{n} \right| \geq \frac{2N_{\Delta}N_{\delta}}{N_{\Delta} + N_{\delta}}(\Delta - \delta) = IRD(G) \tag{6}$$

Conjecture 2 It is conjectured that

$$S(G) = \sum_{i=1}^{n} \left| d_i - \frac{2m}{n} \right| \leq \frac{n}{2}(\Delta - \delta) = IRR(G) \tag{7}$$

Proposition 10 [7]: For connected bidegreed graphs G(Δ,δ) with n vertices and m edges it has been proved that

$$S(G(\Delta, \delta)) = \sum_{i=1}^{n} \left| d_i - \frac{2m}{n} \right| = \frac{2N_{\Delta}N_{\delta}}{N_{\Delta} + N_{\delta}}(\Delta - \delta) \tag{8}$$

where $n = N_{\Delta} + N_{\delta}$.

Proposition 11 Let G be a connected bidegreed graph G(Δ,δ) with n-vertices and m-edges. Then

$$S\big(G(\Delta, \delta)\big) = \sum_{i=1}^{n} \left| d_i - \frac{2m}{n} \right| = \frac{2N_{\Delta}N_{\delta}}{N_{\Delta} + N_{\delta}}(\Delta - \delta) =$$

$$= IRD(G(\Delta, \delta)) \leq \frac{n}{2}(\Delta - \delta) = IRR(G(\Delta, \delta)) \tag{9}$$

where equality holds if G is a balanced bidegreed graph, (where $n \geq 4$ is an even integer, and $N_\Delta = N_\delta = n/2$ holds.)

Proof. Because G is a bidegreed graph where $n = N_\Delta + N_\delta$, one obtains that

$$S(G(\Delta,\delta)) = \frac{2N_\Delta N_\delta}{N_\Delta + N_\delta}(\Delta - \delta) = \frac{2N_\Delta(n - N_\Delta)}{n}(\Delta - \delta) \tag{10}$$

Consider the monotonically increasing function defined by

$$g(N_\Delta) = \frac{2}{n}N_\Delta(n - N_\Delta) \tag{11}$$

Its maximum value with respect to N_Δ can be computed from

$$\frac{\partial g(N_\Delta)}{\partial N_\Delta} = \frac{2}{n}(n - 2N_\Delta) = 2 - \frac{4}{n}N_\Delta = 0 \tag{12}$$

As can be observed, function $g(N_\Delta)$ has a maximum value if $N_\Delta = N_\delta = n/2$ is fulfilled.

Consequently, for n-vertex connected bidegreed graphs

$$S\big(G(\Delta,\delta)\big) \leq \frac{n}{2}(\Delta - \delta) = IRR\big(G(\Delta,\delta)\big), \tag{13}$$

and equality holds if $n \geq 4$ is an even integer and $N_\Delta = N_\delta = n/2$.

Remark 5 If n is an odd integer, then for any n-vertex bidegreed graph $S(G) < IRR(G)$ holds. Moreover, if n is even integer, but $N_\Delta \neq N_\delta$ for a bidegreed graph G, then IRR(G) will always be larger than S(G). For example, if G is the 4-vertex star $K_{1,3}$ then $S(K_{1,3})=3 < 4 = IRR(K_{1,3})$.

Concerning the validity of Conjecture 1 and Conjecture 2, it can be shown that equality in formulas (6) and (7) is satisfied for a broad class of tridegreed mean graphs.

Proposition 12 There exist infinitely many n-vertex really mean tridegreed graphs H_n for which $S(H_n) = IRD(H_n) = 4$ holds.

Proof. The result follows from the properties of mean graphs H_n depicted in Fig. 5.

Proposition 13 As it is demonstrated in Fig. 4, if $n \geq 4$ is an even integer, then there exist infinitely many n-vertex balanced bidegreed mean graphs F_n for which $S(F_n) = IRD(F_n) = IRR(F_n) = n$ holds.

Proof. Because bidegreed graphs F_n are balanced mean graphs the result follows from Proposition 11.

Remark 6 Consider a connected graph G with n vertices and m edges. Let A(G) and L(G) be the corresponding adjacency and Laplacian matrices of graph G,

respectively. Denote by λ_k ($1 \le k \le n$) and μ_k ($1 \le k \le n$) the eigenvalues of matrices A(G) and L(G).

Because for an n-vertex and m-edge connected graph G

$$\sum_{k=1}^{n} \lambda_k^2 = \sum_{k=1}^{n-1} \mu_k = 2m \tag{14}$$

holds, it follows that if G is a mean graph with centrality parameter p(G), then

$$p(G) = \frac{1}{n}\sum_{k=1}^{n} \lambda_k^2 = \frac{1}{n}\sum_{k=1}^{n} \mu_k \tag{15}$$

In Ref. [8] several non-isomorphic 10-vertex graphs with their Laplacian eigenvalues are presented. All of them are really mean and Laplacian equienergetic graphs having identical centrality parameter p=5. For these mean graphs the corresponding edge number is equal to $m=n^2/4=25$.

5 Additional Considerations

The discriminating power of various graph irregularity indices have been tested and compared in several publications [9-21]. We have seen that the degree deviation S(G) is poorly suited for discriminating among mean graphs. Its low-level discriminating performance was demonstrated primarily on unicyclic graphs having identical centrality parameter p=2. This means that degree deviation measure S(G) is unable to classify (order) mean graphs according to their structural irregularity.

In what follows it will be shown that the discriminatory power of S(G) is poor not only for mean graphs but for balanced bidegreed graphs as well. Starting with Proposition 11, balanced bidegreed graphs are characterized by the following property:

Proposition 14 Let $G(\Delta,\delta)$ be an n-vertex and m-edge balanced bidegreed graph where $n \ge 4$ is an even integer and the equality $N_\Delta = N_\delta = n/2$ holds. Then

$$S(G(\Delta,\delta)) = \sum_{i=1}^{n} \left| d_i - \frac{2m}{n} \right| = \frac{n}{2}(\Delta - \delta) = IRR(G(\Delta,\delta)) \tag{16}$$

Proof. Since $N_\Delta = N_\delta = n/2$, this implies that

$$2m = \Delta N_\Delta + \delta N_\delta = \frac{\Delta n}{2} + \frac{\delta n}{2} = \frac{n}{2}(\Delta + \delta) \tag{17}$$

Consequently, one obtains

$$S\big(G(\Delta,\delta)\big) = \sum_{i=1}^{n}\left|d_i - \frac{2m}{n}\right| = N_\Delta\left|\Delta - \frac{\Delta+\delta}{2}\right| + N_\delta\left|\delta - \frac{\Delta+\delta}{2}\right| = \frac{n}{2}(\Delta-\delta) \tag{18}$$

There are infinitely many balanced bidegreed graphs characterized with the above property. Such graphs can be generated by the so-called partial edge-subdivision operation (PES transformation) performed on R≥ 3 regular graphs. By using PES transformation on the R≥3 regular graph G_R with n_R-vertices we can insert n/2 new vertices of degree δ=2 into the parent graph G_R. As a result of this transformation, one obtains a balanced bidegreed graph G(R,2) with vertex number $n(G(R,2))=2n_R$ and edge number $m(G(R,2))=n_R(R+2)/2$.

Example 2 The concept of PES transformation is illustrated by graphs depicted in Fig. 7. The 3-regular graph G_T is the graph of the 6-vertex trigonal prism. Balanced bidegreed graphs G_j (j = 1, 2, 3) constructed from G_T are 12-vertex non-isomorphic graphs containing 6 vertices of degree 3 and 6 vertices of degree 2.

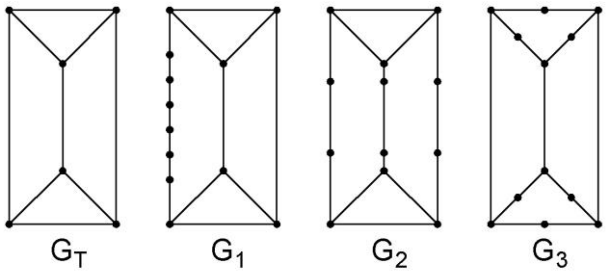


Figure 7
Balanced bidegreed graphs constructed from the 3-regular graph G_T

Non-isomorphic balanced bidegreed graphs G_j (j = 1, 2, 3) do not belong to the family of mean graphs, however, they have identical degree deviation given by

$$S(G_j) = \sum_{i=1}^{12}\left|d_i - \frac{5}{2}\right| = IRD(G_j) = IRR(G_j) = 6 \tag{19}$$

As can be observed, balanced bidegreed graphs depicted in Fig. 7 cannot be discriminated by irregularity indices S(G), IRD(G) and IRR(G).

It is interesting to note that the so-called Albertson irregularity index [9] defined by

$$AL(G) = \sum_{uv \in E} |d(u) - d(v)| \tag{20}$$

seems to be more efficient for graph irregularity characterization, i.e. it possesses a better discriminatory performance. Computing the Albertson indices for graphs Gj (j = 1, 2, 3), we have $AL(G_1)=2$, $AL(G_2)=6$ and $AL(G_3)=12$.

Pokoradi demonstrated in [23] that the comparative evaluation of the discriminating power of irregularity indices is problematic in many cases.

For example, consider balanced graphs J_1, JA_4 and JB_4 depicted in Fig. 6. For their irregularity indices one obtains: $S(J_1)=S(JA_4)=S(JB_4)=6$, and $AL(J_1) = AL(JA_4) = AL(JB_4) =16$.

Denote by M_1 the first Zagreb index of a connected graph with n vertices and m edges [17, 19]. It is interesting to note that the irregularity index defined by

$$IRM_1(G) = M_1(G) - \frac{4m^2}{n} = \sum_{i=1}^{n} \left(d_i - \frac{2m}{n}\right)^2 \tag{21}$$

has an equivalent discriminating performance with degree deviation S(G) if G is a mean graph. This observation is based on the following considerations: If connected graphs G and H are mean graphs with identical centrality parameter p, and the only difference between the degree sequences of G and H is that number N_p of vertices of degree p is different, then $IRM_1(G)= IRM_1(H)$ holds.

We end our study by pointing out a recently published paper [22] containing new results on extremal graphs having maximal degree deviation. Ghalavand and Ashrafi proved [22] that among all n-vertex connected graphs the maximal degree deviation is attained for a particular set of complete split graphs. An n-vertex complete split graph denoted by Cs(n,k) is a connected bidegreed graph consisting of an independent set of n-k vertices and a clique of k vertices, such that each vertex of the independent set is adjacent to each vertex of the clique [24]. According to Ref. [22] the corresponding degree deviation can be calculated as

$$S(Cs(n,k)) = \frac{2k}{n}(n-k)(n-k-1). \tag{22}$$

Among n-vertex connected graphs the maximal degree deviation belongs to the complete split graphs Cs(n,k) listed below, where k is defined as follows

- k=n/3, if n is divisible by 3,

- k=(n-1)/3, if n-1 is divisible by 3,

- k=(n-2)/3 if n-2 is divisible by 3 or

- k=(n+1)/3 if n+1 is divisible by 3.

From the previous considerations, we can conclude that among 12-vertex connected graphs the maximal degree deviation belongs to the complete split graph Cs(n=12, k=4). In this case, the corresponding degree deviation is

$$S(Cs(12,4)) = \frac{2 \cdot 4}{12}(12-4)(12-4-1) = \frac{112}{3} = 37.333 \quad (23)$$

As can be observed, this degree deviation is considerably larger than that of balanced bidegreed graphs depicted in Fig. 7.

Acknowledgements

The authors would like to thank Dr. László Horváth for his extensive help with computer graphics.

References

[1] C. Godsil, G. Royle, Algebraic Graph Theory, Springer-Verlag, New York, 2001

[2] V. Nikiforov, Eigenvalues and degree deviation in graphs, Linear Algebra Appl., 414 (2006) 347-360

[3] N. V. R. Mahadev, U. N. Peled, Threshold Graphs and related Topics, in: Annals of Discrete Mathematics, Vol. 57, Elsevier Science B. V., Amsterdam, 2004

[4] L. Nebeskỳ, On connected graphs containing exactly two points of the same degree, Časopis, Pést. Mat, 98 (1973) 305-306

[5] A. Ali, A survey of antiregular graphs, Contrib. Math. 1 (2020) 67-79

[6] B. Bollobás, Modern Graph Theory, Graduate Texts in Mathematics, Springer Verlag, New York, 1998

[7] T. Réti, I. Milovanović, E. Milovanović, M. Matejić, On graph irregularity indices with particular regard to degree deviation, paper under publication in FILOMAT

[8] D. Stevanović, Large sets of noncospectral graphs with equal Laplacian energy, MATCH Commun. Math. Comput. Chem. 61 (2009) 463-470

[9] M. O. Albertson, The irregularity of a graph, Ars Comb. 46 (1997) 219-225

[10] A. T. Balaban, Highly discriminating distance-based topological index, Chem. Phys. Lett. 89 (1982) 399-404

[11] I. Gutman, On Discriminativity of Vertex-Degree-Based Indices, Iranian Journal of Mathematical Chemistry, 3, (2012) 95-101

[12] T. Došlić, On Discriminativity of Zagreb Indices, Iranian Journal of Mathematical Chemistry, 3 (2012) 25-34

[13] T. Došlić, T. Reti, Novel degree-based molecular descriptors with increased discriminating power, Acta Polytech. Hung. 9 (2012) 17-30

[14] I. Gutman, Topological indices and irregularities measures, Bull. Inter. Math. Virt. Inst. 8 (2018) 469-475

[15] B. Furtula, I. Gutman, M. Dehmer, On structure-sensitivity of degree-based topological indices, Appl. Math. Comput. 219 (2013) 8973-8978

[16] T. Reti, D. Dimitrov, On Irregularity of Bidegreed Graphs, Acta Polytech. Hung. 10 (2013) 117-134

[17] A. Hamzeh, T. Reti, An Analogue of Zagreb Index Inequality Obtained from Graph Irregularity Measures, MATCH Commun. Math. Comput. Chem.72 (2014) 669-683

[18] I. Gutman, Irregularity of molecular graphs, Kragujevac J. Sci. 38 (2016) 71-81

[19] T. Reti, E. Tóth-Laufer, On the construction and comparison of graph irregularity indices, Kragujevac J. Sci. 39 (2017) 53-75

[20] M. Matejić, B. Mitić, E. Milovanović, I. Milovanović, On Albertson irregularity measure of graphs, Sci. Publ. State Univ. Novi Pazar Ser. A: Apple. Math. Inform. Mech. 11 (2019) 97-106

[21] A. Ali, E. Milovanović, M. Matejić, I. Milovanović, On the upper bounds for the degree-deviation of graphs, J. Appl. Math. Comp. 62 (2020) 179-187

[22] A. Ghalavand, A. R. Ashrafi, On a conjecture about degree deviation measure of graphs, Trans. Comb. 1 (2021) 1-8

[23] L. Pokoradi, Graph model-based analysis of technical systems, IOP Conf. Series: Materials Science and Engineering 393 (2018) 012007, doi: 10.1088/1757-899X/393/1/012007

[24] D. Stevanović, Spectral Radius of Graphs, Academic Press, Amsterdam, 2015