

Lecture Notes in Networks and Systems 710

Cezary Biele · Janusz Kacprzyk ·
Wiesław Kopeć · Jan W. Owsinski ·
Andrzej Romanowski ·
Marcin Sikorski *Editors*

Digital Interaction and Machine Intelligence

Proceedings of MIDI'2022 –
10th Machine Intelligence and
Digital Interaction – Conference,
December 12–15, 2022, Warsaw,
Poland (Online)

OPEN ACCESS

 Springer

Series Editor

Janusz Kacprzyk, *Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Advisory Editors

Fernando Gomide, *Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil*

Okyay Kaynak, *Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Türkiye*

Derong Liu, *Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA*

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, *Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada*

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, *Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus*

Imre J. Rudas, *Óbuda University, Budapest, Hungary*

Jun Wang, *Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong*

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Cezary Biele · Janusz Kacprzyk ·
Wiesław Kopeć · Jan W. Owsinski ·
Andrzej Romanowski · Marcin Sikorski
Editors

Digital Interaction and Machine Intelligence

Proceedings of MIDI'2022 – 10th Machine
Intelligence and Digital
Interaction – Conference, December 12–15,
2022, Warsaw, Poland (Online)

Editors

Cezary Biele
National Information Processing Institute
Warsaw, Poland

Wiesław Kopeć
Polish-Japanese Academy of Information
Technology
Warsaw, Poland

Andrzej Romanowski
Institute of Applied Computer Science
Łódź University of Technology
Łódź, Poland

Janusz Kacprzyk
Polish Academy of Sciences
Systems Research Institute
Warsaw, Poland

Jan W. Owsiański
Polish Academy of Sciences
Systems Research Institute
Warsaw, Poland

Marcin Sikorski
Department of Informatics in Management,
Faculty of Management and Economics
Gdańsk University of Technology
Gdańsk, Poland



ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-3-031-37648-1

ISBN 978-3-031-37649-8 (eBook)

<https://doi.org/10.1007/978-3-031-37649-8>

This work was supported by NiPI PL525 000 91 40.

© The Editor(s) (if applicable) and The Author(s) 2023. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Artificial intelligence (AI) is rapidly affecting more aspects of our lives thanks to significant advancements in its research and the widespread usage of interactive goods. This is leading to the birth of several new social phenomena.

Many nations have been working to comprehend these phenomena and discover solutions for moving artificial intelligence development in the proper direction to benefit individuals and communities at large. These initiatives necessitate multidisciplinary approaches that span not only the scientific fields involved in the creation of artificial intelligence and human-computer interaction but also strong collaboration between researchers and practitioners.

Because of this, the major objective of the MIDI conference is to combine two, up until recently distinct, areas of computer science research: artificial intelligence and human-technology interaction. Beginning in 2020, topics discussed at the MIDI conference will include artificial intelligence-related challenges in addition to interface design and user experience.

There is no denying that society is becoming more and more conscious of issues associated to the use of artificial intelligence in solutions. However, the development of artificial intelligence technology is advancing far more quickly than the search for solutions to the moral, social, and economic problems of the present. Both social research and AI technology expertise are required for the discussion of them to provide acceptable answers. As the conference's organizers, we are certain that the expanded format will make it an even more engaging platform for experts in the fields of artificial intelligence and human-technology interaction to exchange experiences.

As the conference hosted two Special Sessions, in 2022 the volume has been divided into four parts: (1) Machine Intelligence, (2) Digital Interaction, (3) Special Session: Advances in Collaborative Robotics, and (4) Special Session: Interacting with Virtual Reality Applications. As a result, the chapters deal with such topics as reinforcement learning, music prediction, medical usage of AI (prostate MRI analysis, chest scans analysis), co-designing immersive environments, gestural interaction with 3D data, collaborative robotics (i.e., electronic skin), and interaction with VR apps.

We believe that all readers interested in emerging trends as well as creators of end-user IT products and services will find inspiration and useful theoretical and practical information in this book. The eventual success or failure of a newly generated product depends on the focus on meeting the requirements of people in a future where technical solutions based on artificial intelligence are developed by people for people. No matter how cutting-edge a technology solution may be, if it is not sufficiently connected to the lifestyle or other variables affecting the target users' social behaviors, they will reject it. Underestimating the value of technology is a mistake, as evidenced by the history of technical progress and examples of technological solutions developed even by the biggest tycoons in the world.

During this year's edition of the conference, two papers were awarded Best Paper Award in memory of professor Krzysztof Marasek. The professor was one of the initiators of the first MIDI conference focused on the user's perspective in computer science and the co-organizer of subsequent editions. Being an outstanding scientist and engineer, an excellent specialist in the field of linguistics, voice user interfaces, and voice-based interaction, he understood the importance of research on technological solutions conducted from the user's perspective. We hope that this year's and subsequent editions of the MIDI conference will step up to the mark and continue the work of professor Krzysztof Marasek.

Cezary Biele
Janusz Kacprzyk
Wiesław Kopeć
Jan W. Owsński
Andrzej Romanowski
Marcin Sikorski

Contents

Machine Intelligence

Light Fixtures Position Detection Using a Camera	3
<i>Vojtěch Leischner</i>	
Improved Vehicle Logo Detection and Recognition for Complex Traffic Environments Using Deep Learning Based Unwarping of Extracted Logo Regions in Varying Angles	12
<i>Zamra Sultan, Muhammad Umar Farooq, and Rana Hammad Raza</i>	
Predicting Music Using Machine Learning	26
<i>Aishwarya Asesh</i>	
A Novel Process of Shoe Pairing Using Computer Vision and Deep Learning Methods	35
<i>Marek Koźłowski, Przemysław Buczkowski, and Piotr Brzezinski</i>	
Representation of Observations in Reinforcement Learning for Playing Arcade Fighting Game	45
<i>Huaiyu Du and Rafał Józwiak</i>	
AI4U: Modular Framework for AI Application Design	56
<i>Kamil Wołoszyn, K. Turchan, M. Rąpała, and K. Piotrowski</i>	
A Competent Deep Learning Model to Detect COVID-19 Using Chest CT Images	67
<i>Somenath Chakraborty and Beddhu Murali</i>	
AI in Prostate MRI Analysis: A Short, Subjective Review of Potential, Status, Urgent Challenges, and Future Directions	76
<i>Rafał Józwiak, Ihor Mykhalevych, Iryna Gorbenko, Piotr Sobecki, Jakub Mitura, Tomasz Lorenc, and Krzysztof Tupikowski</i>	
Performance of Deep CNN and Radiologists in Prostate Cancer Classification: A Comparative Pilot Study	85
<i>Piotr Sobecki, Rafał Józwiak, and Ihor Mykhalevych</i>	
Assessing GAN-Based Generative Modeling on Skin Lesions Images	93
<i>Sandra Carrasco Limeros, Sylwia Majchrowska, Mohamad Khir Zoubi, Anna Rosén, Juulia Suvilehto, Lisa Sjöblom, and Magnus Kjellberg</i>	

Prostate Cancer Detection Using a Transformer-Based Architecture and Radiomic-Based Postprocessing	103
<i>Jakub Mitura, Rafał Józwiak, Ihor Mykhalevych, Iryna Gorbenko, Piotr Sobecki, Tomasz Lorenc, and Krzysztof Tupikowski</i>	
Sales Forecasting During the COVID-19 Pandemic for Stock Management	111
<i>Enver Yildirim, Veli Cam, Fatih Balki, and Salih Sarp</i>	
Digital Interaction	
Seeking Emotion Labels for Bodily Reactions: An Experimental Study in Simulated Interviews	127
<i>Debora C. Firmino De Souza, Pia Tikka, and Ighoyota Ben Ajenaghughrure</i>	
“NAO Says”: Designing and Evaluating Multimodal Playful Interactions with the Humanoid Robot NAO	139
<i>Ilona Buchem, Lukas Brömmeling, and Niklas Bäcker</i>	
Representation of Air Pollution in Augmented Reality: Tools for Population-Wide Behavioral Change	150
<i>Grzegorz Pochwatko, Zbigniew Jędrzejewski, Wiesław Kopeć, Kinga Skorupska, Rafał Maszyk, Anna Jaskulska, and Justyna Świdrak</i>	
Ukrainian Version of the Copresence Scale	159
<i>Lyubov Naydonova, Grzegorz Pochwatko, Mykhaylo Naydonov, and Justyna Świdrak</i>	
Modular Platform for Teaching Robotics	167
<i>Damian Nagajek, Michał Rapata, Kamil Wołoszyn, Krzysztof Turchan, and Krzysztof Piotrowski</i>	
A Method for Co-designing Immersive VR Environments with Users Excluded from the Main Technological Discourse	176
<i>Wiesław Kopeć, Anna Jaskulska, Barbara Karpowicz, Grzegorz Pochwatko, Monika Kornacka, and Kinga Skorupska</i>	
Improving the Usability of Requests for Consent to Use Cookies	191
<i>Kristina Lapin and Laima Volungevičiūtė</i>	
Transdisciplinary Approach to Virtual Narratives - Towards Reliable Measurement Methods	202
<i>Grzegorz Pochwatko, Daniel Cnotkowski, Paweł Kobyliński, Paulina Borkiewicz, Michał Pabiś-Orzeszyna, Mariusz Wierzbowski, and Laura Oseka</i>	

Towards Gestural Interaction with 3D Industrial Measurement Data Using HMD AR	213
<i>Natalia Walczak, Franciszek Sobiech, Aleksandra Buczek, Mathias Jeanty, Kamil Kupiński, Zbigniew Chaniecki, Andrzej Romanowski, and Krzysztof Grudzień</i>	
Polish Adaptation of the Cybersickness Susceptibility Questionnaire (CSSQ-PL)	222
<i>Laura Osęka, Grzegorz Pochwatko, and Justyna Świdrak</i>	
Special Session: Advances in Collaborative Robotics	
NARX Recurrent Neural Network Model of the Graphene-Based Electronic Skin Sensors with Hysteretic Behaviour	233
<i>Jakub Możaryn</i>	
Proximity Estimation for Electronic Skin Placed on Collaborative Robot Conductive Case	242
<i>Jan Klimaszewski and Przemysław Białorucki</i>	
Finite Element Method Based Toolchain for Simulation of Proximity Estimation Using Electronic Skin	250
<i>Anna Ostaszewska-Liżewska and Jan Klimaszewski</i>	
Collaborative Robotics. Safety and Ethical Considerations	260
<i>Monika Różańska-Walczuk</i>	
Versatile Robotic Workstation for Electronic Skin - Problems and Solutions	270
<i>Jan Klimaszewski</i>	
Special Session: Interacting with Virtual Reality Applications	
VR Game for Powerlifting Training	281
<i>Krzysztof Popielski, Katarzyna Matys-Popielska, and Anna Sibilska-Mroziejewicz</i>	
A Case for VR Briefings: Comparing Communication in Daily Audio and VR Mission Control in a Simulated Lunar Mission	287
<i>Kinga Skorupska, Maciej Grzeszczuk, Anna Jaskulska, Monika Kornacka, Grzegorz Pochwatko, and Wiesław Kopeć</i>	
Construction and Evaluation of a Laboratory Stand for Testing MV Switchgears Using VR Technology in the Power Industry	298
<i>Tadeusz Daszczyński, Kacper Berdek, and Dariusz Naruszewicz</i>	

Virtual Reality Simulations of the Snake Robot 307
*Anna Sibilska-Mroziewicz, Ayesha Hameed, Jakub Możaryn,
and Andrzej Ordys*


**Prototype of Virtual Reality Game to Support Post-stroke Recovery
in Patients with Spatial Neglect Syndrome** 314
*Katarzyna Matys-Popielska, Krzysztof Popielski,
and Anna Sibilska-Mroziewicz*

Author Index 321

Machine Intelligence



Light Fixtures Position Detection Using a Camera

Vojtěch Leischner^(✉) 

Czech Technical University in Prague, Technická 2, 160 00 Prague 6, Czech Republic
leiscvoj@fel.cvut.cz
<https://dcgi.fel.cvut.cz>

Abstract. Interdisciplinary research combines computer vision with stage light design to automatically detect light fixtures' positions to create light animations. Multiple programmable light fixtures are often used in theaters, the event industry, and interactive installations. When creating complex animations such as a wave traveling from one side to the other through multiple light fixtures array, all lights' positions must be known beforehand. Traditionally the position of the light is marked in the technical plan. However, technicians make mistakes during the installation and sometimes install the light in a different position. In such a case, time-consuming troubleshooting is needed to determine which light is misplaced and either correct the position in the software or manually move the light to the correct position. Our system saves time during installation and produces a light id and position pairs that users can use in various lighting control software. As a result, users can improvise and change the light positions more intuitively without needing a technical plan. Our system saves installation costs and enables rapid prototyping of light shows to create previously impossible organic designs. We verified the system in a controlled experiment and measured the influence of camera resolution on accuracy.

Keywords: DMX512 · light fixture · computer vision · position detection · camera

1 Introduction

1.1 Use Case

We focus on large-scale dynamic light installations. Consider the light installation at 131 South Dearborn, Chicago, the USA [22] as an example of an ideal use case. The installation consists of 925 glass bubbles, and each bubble has its light source. Install 925 individually programmable light sources and ensure that the wiring is right pose a considerable challenge and is prone to mistakes during installation. We aim to automate the light position mapping process with the proposed system.

Another use case we tested was mapping multiple programmable led rings and strips. Instead of manually marking each fixture's position, we have automated

the process. As a result, the setup can be changed frequently to help designers find the ideal configuration and immediately test it. Please watch the experiment video [15] documenting the process.

1.2 Programming Light Show

Without knowing the light positions, we can create only simple light shows, such as changing the light parameters uniformly or creating animations based on noise. To create a more complex light show, we need to distinguish individual lights and know their position.

We can then create a simulation with light sources represented as points in space. See Figure 1. We add virtual objects and animate their position. We can achieve various light effects by detecting the collision of light sources' positions with a volume of moving virtual objects. We turn on the light source when the light is inside the virtual object and turn it off when it is outside. We can also use multiple virtual objects to create more elaborate animations or map them interactively based on sensor inputs. For example, one can map people's movement to light intensity in different sectors.



Fig. 1. Schema of light scene animation: Light sources represented as points in space collide with the animated virtual 3D object. Lights inside the object will be turned on.

1.3 Light Network Control

Light fixtures are often controlled using DMX512 [10] protocol. Traditionally DMX address (a unique id of the light) is selected using a hardware DIP switch directly on the light fixture. Another option is to use Remote Device Management (RDM [9]) commands from the control software [6]. However, RDM is only available with some DMX light fixtures, so we often rely on manual selection, making changing DMX addresses difficult. According to the technical plan, each light with the appropriate DMX address must be installed in the correct position.

To control the lights, most often, we use the Artnet [2,22] or sACN [1,11] protocol - an extension of DMX standard that enables us to control more lights and use network topology and devices such as Ethernet switches to send DMX packets over the network.

In essence, we can control individual lights from a single networked computer. We need an ethernet adapter and Artnet/sACN node to convert the signal to DMX. DMX signal is then sent to DMX driver that maps the DMX values to voltage and current to dim individual lights. See Figure 2.

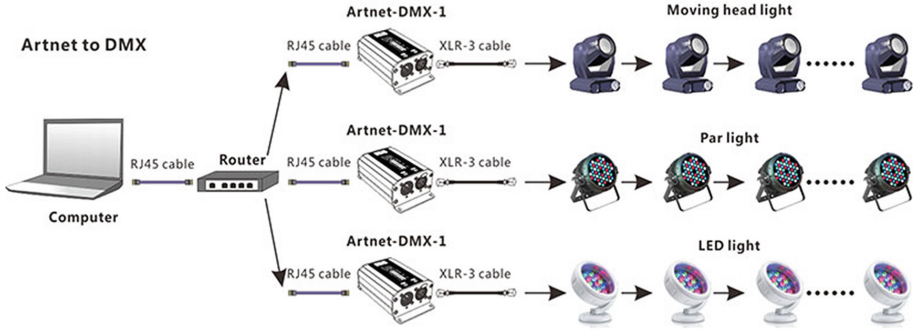


Fig. 2. Artnet light network diagram (<https://www.ledlightinghut.com/artnet-dmx512-converter.html>).

2 State of the Art

Various systems using a camera to control lights exist. Most of them deal with finding a user or controlling the light to track the user. In most cases, available solutions rely on knowing the light fixtures' position in space. We are trying to solve a different problem, localizing light sources relative to the camera. Still, some principles can be used for both problems.

Luxapose [14] system used a mobile phone as a camera and modified light fixtures to track a mobile phone's position. Light sources are modulated to produce pulses of light with encoded position and ID information that can be captured in a single frame exploiting the rolling shutter effect. While we are trying to determine the position of the light fixture, Luxapose is tracking a user. Theoretically, we could repurpose the system and use the known camera position to determine the light position. Unfortunately, the Luxapose system requires light source modification which is not practical in our use case. Moreover, the camera position relative to the light source would have to be known. While feasible, it would require precise measurement on site.

Similar to the Luxapose, a paper by Hossan and Tanvir [13] uses multiple known light fixtures to localize the camera position. We can not use triangulation methods to determine the camera position when using a single camera in our system. In practice finding the precise camera position relative to the light fixture would have to be repeated for all lights as we can not guarantee their position relative to each other. Such manual measurements would completely negate any benefits of automatization. The authors use found blob pixel count to measure

the distance from the camera to the light fixture. A similar approach could be adopted while moving the camera back and forth to determine the distance from the camera to the light source. Other approaches [24] use a photo-diode as a sensor installed in the light fixture to track objects based on changes in light propagation. Multiple light fixtures have to be taken into account, or a quadrant photo-diode is used [7]. The camera is no longer needed in these cases, but the light fixtures have to be modified to include the appropriate sensor and communication unit. For example, BlackTrax [8], a commercial solution for motorized lights, uses multiple cameras and an infrared beacon to track the moving target. Craig Hiller developed a system to detect light’s positions and even classify them as a fluorescent bulb, tube, or LED light [12]. The project relies on a Tango tablet [17] aligned with a DSLR camera and IMU in one package. The user walks under the lights to map their position and create a spatial map. Tracking is only possible with sensors being perpendicular to the light fixture, which is fine when detecting ceiling-mounted lights but would fail with vertical or organically shaped installations. The reported error is also not suitable for the light show scenario, with some lights being false positives and others missed during detection.

3 Software

3.1 Programming Environment

The main program is written in Java Processing framework [4]. To acquire the camera stream, we used Gstreamer [23] based library [11]. To perform computer vision tasks, we have used the Java wrapper of OpenCV library [5].

3.2 Light Fixture Detection

To find the light position, we create a frame difference between a frame with all lights turned off and a frame with a single light turned on. Then we threshold the resulting absolute difference to a binary image. We find the contours of the light blobs. Finally, we sort found blobs based on their area. We select the blob with the biggest area and find its centroid. The resulting x and y coordinates are saved and associated with the light DMX address. When we have processed all the lights, we normalize their coordinates and save all the information into a JSON file. Later this file can be loaded into light control software such as Touchdesigner [21], Madrix [3], OpenFrameWorks [20], Processing [4], VVVV [18], and similar software used to create light shows.

3.3 GUI and User Input

We have also developed a GUI for easier usage. We provide multiple options, such as selecting a target IP, setting custom binary thresholds, selecting from multiple cameras, and setting minimal and maximal blob size. Users can also

select whether to cycle through all lights automatically or one by one by clicking the button. The user can manually adjust every detected light position with a mouse if needed. Users can also create custom masks to select the area in the camera image that should be ignored during detection. Furthermore, we have also enabled perspective corner pin transformation to be applied to the camera image. Acquiring undistorted camera images is essential to measure uniform distances between light fixtures correctly.

4 Experiment

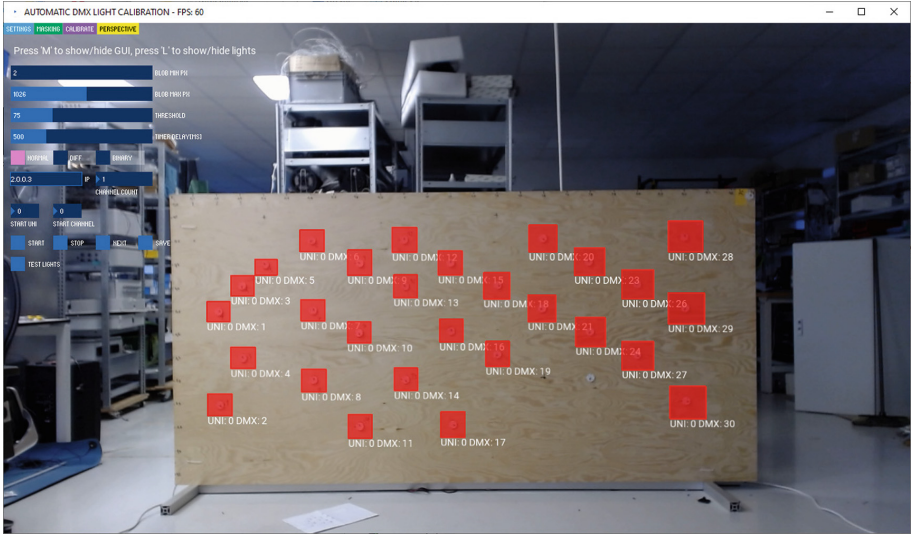


Fig. 3. Experiment setup - LEDs mounted on the wooden plate

4.1 Lights Setup

We have tested our system on a setup with 30 individually programmable LED light sources controlled via the ArtNet network. Lights were installed on 250cm by 125cm wooden plate, all facing the same direction. See Figure 3. Each light is 3.5cm in diameter with 6 LEDs controlled as a single symmetrical light source.

4.2 Camera

We have used the Logitech C922 camera, a widely available and affordable standard web camera with a USB interface. The camera was positioned perpendicular to the wooden plate 272cm away. The camera's Diagonal Field of View (FOV) is 78° , horizontal FOV is 70.42° , vertical FOV is 43.3° , and Focal Length is 3.67mm. We tested the setup, once using 640×480 px camera resolution and once 1920×1080 px, so we could determine if the used resolution correlates with accuracy.

Table 1. Mean error in the position detection and standard deviation

resolution	pixel pitch	error X	error Y	STDEV X	STDEV Y
640*480 px	4.53 mm	2.912 mm	2.426 mm	0.678 px	0.507 px
1920*1080 px	2.02 mm	2.885 mm	1.947 mm	1.069 px	1.318 px

5 Results

Experiment data is available at Zenodo 6814223 [16]. All installed lights were detected correctly. See the results in Table 1. In the case of 640*480px camera resolution, 1 pixel corresponded to 4.53mm in the wooden plate plane, and the average position error on the horizontal axis was 2.912mm and 2.426mm on the vertical axis. The maximum error on the horizontal axis was 2 pixels, and 1 pixel on the vertical axis with standard deviation 0.678px and 0.507px respectively.

In the case of 1920*1080px camera resolution, 1 pixel corresponded to 2.02mm in the wooden plate plane, and the average position error on the horizontal axis was 2.885mm and 1.947mm on the vertical axis. The maximum error on the horizontal axis was 4 pixels, and 6 pixels on the vertical axis with standard deviation 1.069px and 1.318px respectively.

6 Discussion

We assumed that the higher resolution of the camera and shorter distance to the lights would produce more accurate results as 1cm would be represented by more pixels in the camera image. Resolution does not play a crucial role, as we have not observed a significant accuracy increase when using 1920* 1080px over 640* 480px.

We need to know if particular light is left or right relative to the other light to propagate the animated wave through them correctly. We have successfully obtained correct relative relationships between lights irrespective of the used camera resolution. High absolute accuracy can be beneficial but unnecessary for the light-show creation.

The light source shines directly at the camera and produces a lens flare. Lens flare does not worsen the detection as long as it is uniform in all directions, so we can assume that the light source is in the center.

7 Conclusion

We offer a robust way to automatically detect installed light fixtures positions with a single monocular camera. Furthermore, we save the information with the respective light DMX address in both machine and human-readable formats to be used in various light control software. Our proposed automatic light detection method can be thought of as a proof of concept with clear time and cost-saving

benefits. More importantly, it opens the way for new stage light design possibilities. More testing in real-world scenarios is needed to further verify the presented system's viability.

8 Limitations

An unobstructed view of all light sources is needed to detect their position correctly. Position mapping is possible only if we find a view in which the light fixtures do not overlap. For example, it would be challenging to automatically get the positions of light fixtures organized in a 3D helix shape.

The best use case for our system is the light sources positioned in a single layer. For example, lights are hanging from the ceiling.

The problem occurs when lights are organized in multiple layers. Such as multiple lights on a single string or rod beside each other. It might not be practical to use our method in such a case.

Reflections cause another limitation. When reflective surfaces surround the light sources, it can create several false-positive hot spots in-camera images that might be hard to distinguish from the actual light source. For example, a chrome-plated ceiling has high reflectivity. In this case, reflections have to be eliminated before detection.

9 Future Research

We can improve the tool to enable 3D position mapping. We can use the binocular stereoscopic camera to calculate the distance from the camera to the light. Alternatively, we can reposition a single monocular camera to acquire two points of view to achieve the same result. We could further improve the usability by merging multiple cameras to cover a larger area to map large-scale installations. Another approach would be to enable sequential mapping. After mapping one place, the user would physically move the camera to cover another neighboring area. The relative position change of camera origin could be calculated by the standard SLAM method [19].

10 Declarations

Acknowledgements. The research was consulted with doc. Ing. Zdeněk Míkovec, Ph.D.

Funding and Competing interests. The author works commercially in the interactive installations industry that can benefit from the proposed automatic light position detection system. This research has been supported by the project funded by a grant SGS22/172/OHK3/3T/13 and by RCI (CZ.02.1.01/0.0/0.0/16 019/0000765).

Availability of data and materials. Experiment data is available at DOI: 10.5281/zenodo.6814223 [16].

References

1. SACN (2019). <https://artisticlicenceintegration.com/technology-brief/technology-resource/sacn-and-art-net/>
2. Art-net (2020). <https://art-net.org.uk/>
3. Madrix lighting control (2022). <https://www.madrix.com/>
4. Ben, F., Casey, R.: (2004). <https://processing.org/>
5. Bradski, G.: The openCV library. Dr. Dobb's J. Softw. Tools Prof. Programmer **25**(11), 120–123 (2000)
6. Choi, S.-I., Lee, S., Koh, S.-J., Lim, S.-K., Kim, I., Kang, T.-G.: Reliable transmission for remote device management (RDM) protocol in lighting control networks. In: Jeong, Y.-S., Park, Y.-H., Hsu, C.-H.R., Park, J.J.J.H. (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 280, pp. 51–58. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-41671-2_8
7. Cincotta, S., Neild, A., He, C., Armstrong, J.: Visible light positioning using an aperture and a quadrant photodiode. In: 2017 IEEE Globecom Workshops (GC Wkshps), pp. 1–6 (2017). <https://doi.org/10.1109/GLOCOMW.2017.8269150>
8. Eichel, J.A., Clausi, D.A., Fieguth, P.: Precise high speed multi-target multi-sensor local positioning system. In: 2011 Canadian Conference on Computer and Robot Vision, pp. 109–116 (2011). <https://doi.org/10.1109/CRV.2011.59>
9. ESTA: American national standard ANSI e1.20 - 2006 entertainment technology RDM remote device management over dmx512 networks. Technical report, 875 Sixth Avenue, Suite 1005, New York, NY 10001, USA (2006). https://webstore.ansi.org/preview-pages/ESTA/preview_ANSI+E1.20-2006.pdf
10. ESTA: American national standard ANSI e1.11 - 2008 (r2018) entertainment technology-usitt dmx512-a asynchronous serial digital data transmission standard for controlling lighting equipment and accessories. Technical report, 630 Ninth Avenue, Suite 609, New York, NY 10036 USA (2018). https://tsp.esta.org/tsp/documents/docs/ANSI-ESTA_E1-11_2008R2018.pdf
11. Gottfried, H., Ben, F., Reas, C., et al.: Codeanticode (2022). <https://github.com/processing/processing-video>
12. Hiller, C., Zakhor, A.: Fast, automated indoor light detection, classification, and measurement. *Electron. Imaging* **2018**(15), 2711–2714 (2018)
13. Hossan, M.T., Chowdhury, M.Z., Islam, A., Jang, Y.M.: A novel indoor mobile localization system based on optical camera communication. *Wirel. Commun. Mob. Comput.* **2018**, 9353428 (2018). <https://doi.org/10.1155/2018/9353428>
14. Kuo, Y.S., Pannuto, P., Hsiao, K.J., Dutta, P.: Luxapose: indoor positioning with mobile phones and visible light. In: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, pp. 447–458. MobiCom 2014, Association for Computing Machinery, New York, NY, USA (2014). <https://doi.org/10.1145/2639108.2639109>
15. Leischner, V.: Automatic light position detection prototype v2 (2022). <https://youtu.be/xAghkKOFq-g>
16. Leischner, V.: Light camera position detection - experiment data (2022). <https://doi.org/10.5281/zenodo.6814223>
17. Marder-Eppstein, E.: Project tango. In: ACM SIGGRAPH 2016 Real-Time Live!, pp. 25–25 (2016)
18. McDirmid, S.: Usable live programming. In: Proceedings of the 2013 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software, pp. 53–62 (2013)

19. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular slam system. *IEEE Trans. Rob.* **31**(5), 1147–1163 (2015)
20. Noble, J.: *Programming Interactivity: A Designer’s Guide to Processing, Arduino, and Openframeworks*. O’Reilly Media Inc, California (2009)
21. Rousset, I.: *Touchdesigner* (2022). <https://derivative.ca/>
22. Růžičková, J.: *Walk on clouds* (2019). <https://www.lasvit.com/project/131-south-dearborn/intro>
23. Taymans, W., Baker, S., Wingo, A., Bultje, R.S., Kost, S.: *Gstreamer application development manual* (1.2. 3). Publicado en la Web 72 (2013)
24. Wang, W., Wang, Q., Zhang, J., Zuniga, M.: *PassiveVLP: leveraging smart lights for passive positioning*. *ACM Trans. Internet Things* 1(1), 1–24 (2020). <https://doi.org/10.1145/3362123>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Improved Vehicle Logo Detection and Recognition for Complex Traffic Environments Using Deep Learning Based Unwarping of Extracted Logo Regions in Varying Angles

Zamra Sultan, Muhammad Umar Farooq^(✉), and Rana Hammad Raza

Pakistan Navy Engineering College (PNEC), National University of Sciences and Technology
(NUST), Karachi, Pakistan

zsultan.beee15pnec@student.nust.edu.pk, {umar.farooq,
hammad}@pnec.nust.edu.pk

Abstract. Intelligent Traffic Monitoring and Management System (TMMS) is a growing research area as cities infrastructure continues to evolve. Traffic situation is demanding innovative solutions for effective monitoring and management given the complex nature of the urban scenario. A major focus of this research domain is fine-grained vehicles classification that requires detection and recognition of distinct features of vehicles. Some of these features are semantic based while others are appearance based. One such appearance-based feature of a vehicle is its logo. Logo detection helps with identification of a vehicle's make during fine-grained classification process. There are various deep learning methods which give good performance for such object detection tasks. However, it is challenging to exploit these methods due to smaller size of logo especially in a surveillance environment. This work firstly presents a deep learning-based approach for detection of vehicles' logos in camera video feeds. Due to small size of logos, a unique pipeline using three different deep learning models is designed. Firstly, a modified Improved Warped Planar Object Detection Network (IWPOD-NET) selects a Region of Interest (ROI) and adjusts the orientation of vehicle logo. Then YOLO (You Only Look Once) v5 is used to detect the logo part in the selected ROI and finally, EfficientNet is used to further classify logo into different classes. This pipeline is tested on four surveillance environments namely toll control, law enforcement, dashcam, and parking lot access control. Comparative analysis shows accuracy improvement with this proposed approach in each testing case. A pose variance analysis is also performed to determine the orientation limits to which this approach can work. Secondly, a custom dataset, VL-10 (Vehicle Logos) is presented which provided further insights into the challenges w.r.t local environment settings. The whole approach improved the overall performance of the logo detection and recognition system.

Keywords: Video surveillance · Vehicle detection · Vehicle logo detection · Fine-grained vehicle logo classification · Deep learning · IWPOD-NET · EfficientNet · YOLOv5 · Image unwarping

1 Introduction

Intelligent Traffic Monitoring and Management System (TMMS) has an increasing significance as the traffic infrastructure is advancing. With the advent of technology, intelligent TMMS has taken over conventional TMMS and has become an emerging research area. It includes detection, recognition and fine-grained classification of vehicles, automatic license plate recognition and various kinds of traffic analytics. Its applications vary from commercial domains to national security levels e.g., surveillance, security system, traffic congestion avoidance, accident prevention, advanced driver assistance systems, access control systems, intelligent parking systems, electronic toll collection, etc. These applications necessitate the development of more robust techniques for Intelligent TMMS.

The fine-grained classification of vehicles is the process of categorizing a vehicle into its make and model, generally known as Vehicles Make and Model Recognition (VMMR). Many techniques are used for this classification; some by considering the whole vehicle or others through a certain area of vehicles (e.g., frontal view, grill, or logo). A vehicle logo is a distinct appearance-based feature which is an important part of this whole process. Vehicle logo detection/recognition is required for many applications of TMMS. It can be used for estimation of brand reputation and traffic monitoring etc. Its recognition also plays an important role in the scenarios where the authenticity of vehicle's make is doubtful. For example, the scenario where a stolen vehicle's license plate is replaced with another one, or when a license plate is obscured or illegible, which affects the readability process in any license plate recognition application. The logo detection can be used as an authenticity validation step for this purpose. However, identification through vehicle logo is a challenging task due to its smaller size. Detection and recognition of small objects is a fundamental challenge in Computer Vision and is a growing research area these days. Many researchers have achieved high performance for small objects detection using Convolutional Neural Networks (CNNs). CNN is able to learn various image-based features from large-scale data without the need of any human intervention. Compared to different kinds of small objects, the vehicle logo has more complex rich content information at times due to their unusual designs. These small sized logos are also affected by background clutter and noise. Similarly, most of the times, logo is blurred and not properly illuminated, and even not visible due to rainy or snowy weather which causes difficulties in features identification. An overview of some challenges associated with this logo detection and recognition task are provided below:

- *Logo Size and Resolution.* Logos account for only ~1% of the frame obtained from camera. Due to camera distance, the resolution is low at times. These issues affect the feature extraction process.
- *Illumination Variance.* Colors may vary under different illumination conditions especially in low-light conditions, where identified features may not be that effective. Similarly, vehicle logos usually have high reflectance luminous which causes issues in features extraction.
- *Logo Location.* The locations for vehicle logos vary among various vehicle manufacturers. Some manufacturers place their vehicle logos on radiator grilles, while some place them on the vehicle's front hood. This causes difficulties in the identification of logo boundaries.

- *Background Clutter and Noise.* Logo is a comparatively smaller part of vehicle due to which there are many sources of background interference including radiator grille, bumper and other small objects. These objects distract the feature extraction and results in increase of false detections.

This research has modified a state-of-the-art license plate detection model IWPOD-NET, to extract the desired ROI and then further unwarp it to fix the distortions in varying angles [32]. The used surveillance recordings are low in resolution and cause the accuracy to decrease. Therefore, a combination of different CNN models are used for detection and recognition purposes. The results show a significant improvement in overall detection and recognition of vehicle logos. The main contributions of this paper are highlighted as followed.

- An offline vehicle logo detection and recognition approach that uses video feeds acquired through low-resolution surveillance cameras for authenticity validation of vehicles.
- Improved logo detection and classification based on a modified state-of-the-art license plate detection network.
- Pose variance analysis to identify the varying angles on which the proposed approach can effectively work.
- A custom vehicle logo dataset, VL-10, containing logos of 10 commonly used vehicle makes and models.

The rest of the paper is organized as followed: Sect. 2 provides a concise overview of the related research work. Section 3 provides a detailed working methodology of the proposed approaches and gives an overview of the proposed dataset. Section 4 discusses the test cases used to conduct this research and presents a comparative analysis of the results obtained from the experimentation. Finally, Sect. 5 concludes the papers and provides a future direction of this research.

2 Related Works

Work in different domains of intelligent traffic monitoring and management system has been contributed by many researchers. Setchell discussed monitoring and management of road traffic using computer vision in details [1]. The paper presented two vision-based traffic monitoring systems including license plate recognition system and a road traffic monitoring system to track vehicles. Different ideas for vision-based traffic monitoring and management systems have been proposed [2-6]. VMMR is also an important part of traffic monitoring and management systems.

There have been many related approaches using conventional algorithms such as Scale-invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) and Sequential Pattern Mining (SPM) etc. [7-9]. Usually, logo is detected through the reference of license plate [10-12]. After the detection of license plate, different techniques are then implemented to segment the logo [13-15]. Psyllos *et al.* introduced an enhanced SIFT-based features extraction mechanism for vehicle logo recognition. The experiment results showed a significant improvement of recognition accuracy compared to conventional SIFT algorithm [16]. Mao *et al.* proposed a novel vehicle logo detection method

which highlighted logo regions using direction filters and saliency map. Information entropy was then used to choose the binary image and for precise location of logo region [17]. Yu *et al.* proposed a system for vehicle logo recognition based on Bag-of-Words (BoW) [18]. The vehicle logo images were represented as histograms of visual words and were classified by SVM in three steps. These steps involved SIFT feature extraction, quantization of features into visual words and histograms of visual words with spatial information. Wang *et al.* presented a vehicle logo detection/recognition system which uses edge features to find logo in the rough logo region detected through prior knowledge [19]. Then, a combination of template matching and HOG were used to identify the category of logo.

These conventional techniques are usually dependent on hand-crafted features, which lack robustness in certain conditions. These conventional algorithms are relatively easier to implement, however, these algorithms are computationally expensive. Compared to these conventional approaches, various deep learning-based methods are also explored by researchers. The deep learning-based algorithms have significantly better performances in detection of objects under different complex environments as discussed in Sect. 1. Tong, K *et al.* presented an in-depth analysis of different deep learning-based small objects detection approaches [20]. The authors showed with experimentation that techniques such as multi-scale feature learning, data augmentation and different training strategies can significantly improve the small objects detection. Similarly, these state-of-the-art techniques have also been tried and tested for the task of small objects such as vehicle logo.

Yang *et al.* proposed a modified YOLOv3 model for vehicle logo detection in complex scenes [21]. The approach showed good accuracy on the proposed vehicle logo detection dataset. Zhang *et al.* proposed a real-time lightweight vehicle logo detection system based on deep convolutional networks [22]. The experimentation showed a significant improvement in accuracy of vehicle logo detection task. Jiang *et al.* proposed an improved YOLOv4 model which enhanced the backbone feature extraction network for efficient extraction of small vehicle logos [23]. It further used convolutional transformer to reduce influence of complex backgrounds. Yang *et al.* proposed an improved YOLOv2 network which provided fast and accurate vehicle logo detection [24]. The experimental results show that this approach significantly improved accuracy and speed of logo detection.

The surveillance recordings usually have low resolutions which makes the logo detection challenging, particularly at varying angles. Various techniques are used to improve the quality of video prior to the detection of objects. Some researchers used Super Resolution algorithms to overcome these limitations and to convert these low-resolution images to high-resolution ones [25-29]. Others used near-field and far-field video enhancement techniques [30]. Some researchers suggested the use of ROI rate control scheme for surveillance videos [31]. However, this proposed research work adjusts the orientation of logos at varying angles, which improves the overall vehicle detection and recognition task.

3 Working Methodology

3.1 Selected Models

A unique pipeline using a combination of different deep learning models is designed in this study for improved vehicle detection and recognition. The specific details of the selected models are as follows:

Unwarping Model (IWPOD-NET). Improved Warped Planar Object Detection Network (IWPOD-NET) [32] is able to detect four corners of vehicle license plate in a variety of unconstrained scenarios. It then unwarps the license plate to a fronto-parallel view and eliminates perspective related issues. This network was introduced to tackle license plate detection with varying viewpoints. However, this research modified the model to serve as basis for ROI localization and orientation adjustment of logos. The localized area above the license plate was extended by a variable (assuming that the logo placement is always on the top of license plate) to include the logo part as well in complete ROI, which was then unwarped. For example, if the height of license plate is 3 units and variable is set to be 2 units, then the ROI will be extended by 2×3 units vertically upwards. Workflows for both original and modified network are shown in the following Fig. 1. The characters on the license plate of the vehicles utilized in this study have been concealed in order to maintain the privacy of the vehicle owners.

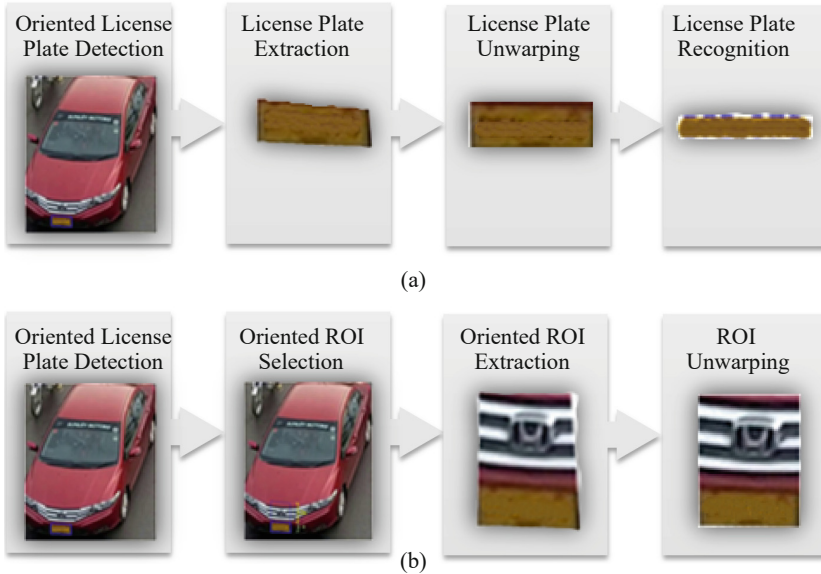


Fig. 1. Workflows for both the Original IWPOD-NET (a) and Modified IWPOD-NET (b) models

Logo Detection Model (YOLOv5m). Advent of YOLO models improved the real-time object detection. Many versions of YOLO have been introduced so far, which have significantly improved object detection. This research utilizes YOLOv5, which has different

architecture series including P5 and P6. These series are further divided into nano (n), small (s), medium (m), large (l) and extra-large (x) that vary on the basis of parameter sizes and provide optimizations for particular image sizes. For example, P5 is optimized for training images of sizes 640×640 . For the desired task, YOLO-v5-P5 medium sized model is used in this system of logo detection. Due to low classification accuracies, the model was used to detect the presence of logo only. For a comparative analysis, results are also presented in Table 2 for the approach when YOLO was used for both detection and classification purpose. The unwarped image from previous step is then fed to YOLO-v5m model for the detection of logo. The workflow is shown in following Fig. 2. As shown in the figure, logo is successfully extracted from the complete ROI.

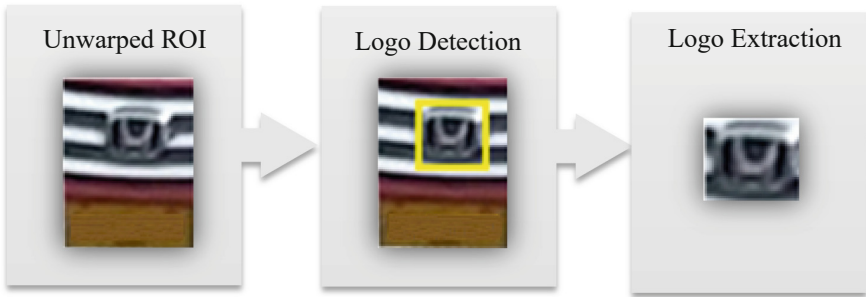


Fig. 2. Logo Detection using YOLO-v5m

Logo Classification Model (EfficientNet). EfficientNet [29] introduced by Google is a commonly used CNN architecture for image classification purposes. The model achieves state-of-the-art performance, in addition to being faster and lightweight. This research work trained EfficientNet on logos in the proposed dataset. Firstly, all images were cropped to the logo part, based on dimensions from the annotated images that were fed to the object detection model. Then these cropped logo images were used for training the classification model. In testing phase, once the logo is detected in ROI, it is then fed to the classification model as an input. It then classifies the logo into its respective class. The addition of EfficientNet has significantly improved the results as discussed in Sect. 4.

3.2 Dataset

A high-quality dataset is essentially required for tackling various computer vision tasks involving object detection and recognition etc. The lack of a local dataset can become a huge challenge when performing these tasks for any particular region with a specific complex traffic environment. To the best of our knowledge, there is no standard local dataset available. Therefore, a local dataset of vehicles, VL-10, is developed where logos are either prominently visible or partly visible due to varying angles as in surveillance environment as shown in Fig. 3.

All the logos are self-collected from various local and online sources (e.g., different car selling websites with publicly available images). The logos are categorized into ten different vehicle classes, including Honda, Suzuki, Toyota, Kia, Hyundai, Mitsubishi, Daihatsu, Faw, Hino and Nissan. A total of 500 and 50 images per class were selected to form the training and validation sets, respectively. The data augmentation was then carried out by adding blur and noise to each image. Therefore, the total images per class became 1500 and 150 for training and validation sets, respectively.



Fig. 3. Some Images from VL-10 Dataset

4 Results and Discussions

4.1 Test Cases

Different surveillance scenarios have been chosen for test/evaluation of the proposed model pipeline. These include video streams from toll control (i.e., CCTV video from height), law enforcement (road-level cameras), parking lot access validation (cameras installed at various entry/exit points), and dashboard cameras. The highlights of each test case are provided in Table 1. Similarly, example frames from each test case are shown in Fig. 4.

4.2 Comparative Analysis

Three different approaches were tested, and their comparative analysis is accordingly presented in this section.

Approach 1 - Without IWPOD-NET Using YOLOv5 for Both Detection and Classification. The YOLOv5 model was trained on VL-10 dataset for both object detection and classification tasks. All test scenarios were first passed to this model to get baseline accuracies. The mAP for detection and classification is presented

in Table 2 for each test case. IWPOD-NET was not used in this approach so that results could be compared, and the impact of IWPOD-NET could be observed.

Approach 2 - With IWPOD-NET Using YOLOv5 for Both Detection and Classification. A modified IWPOD-NET was introduced. The detected vehicles were passed to IWPOD-NET for the ROI extraction and unwarping. After extraction of ROI, it is then passed to the YOLOv5 model trained on VL-10 dataset. Here, YOLOv5 detects and classifies the logos into their respective classes. Results obtained in the form of mAP are shown in Table 2. It can be seen that results were improved to a certain level. The highest improvement shown is in the Law enforcement test scenario. This is due to the fact that this test set has side camera viewpoints and IWPOD-NET helped to unwarped these side views really well. On the other hand, the lowest improvement is shown by parking lot access validation test set, which is apparently because camera viewpoint was already frontal, and the license plates were not very oriented in this case.

Approach 3 - With IWPOD-NET Using YOLOv5 for Detection and EfficientNet for Classification. In the third approach, while using IWPOD-NET and YOLOv5, another model, EfficientNet was introduced. YOLOv5 was trained to detect the presence of logo only. All the classes of VL-10 dataset were trained on EfficientNet which was then used for logo classification. Firstly, each test dataset was passed to modified IWPOD-NET for the extraction and unwarping of ROI. The ROI was then passed to YOLOv5 model for the detection of logo. After the logo was detected, it was then passed to the EfficientNet for classification of the detected logo. The mAP for the detection by YOLOv5 and the accuracy of classification by EfficientNet is shown in Table 2. Considerable improvements in the results were observed. The highest improvement in the detection were observed in the Dashcam testing case. The margin left was due to the motion blur present in this case since both the objects (vehicle/logo) and the camera were in motion. Highest classification accuracy was produced in the case of parking lot access validation. The margin left in this case was due to the inclusion of night images in dataset where the logos were not perfectly visible, and the use of streetlights create the issue of reflection. Approach 3 stood out from the other two approaches. Therefore, this pipeline could be used as an offline tool for logo detection and classification in various surveillance environments.

4.3 Pose Variation Calculation

Pose variation analysis was conducted to determine the capacity of the proposed model to detect logos accurately at varying angles. For this purpose, some images were captured in the daylight similar to the settings in parking lot access validation dataset. Different angles of each vehicle were captured as shown in Fig. 5(a). Poses were categorized by comparing the oriented bounding boxes and rectangular bounding boxes of license plate as shown in Fig. 5(b). IWPOD-NET detected the four corners of license plate (green dots). It then made an oriented bounding box joining these four corner points (blue bounding box). Another algorithm was used to make rectangular bounding boxes (red bounding boxes). The upper lines for both bounding boxes were used to identify the poses. This was done by calculating the angle “ θ ” subtended between both lines.

According to the observations, the proposed module was able to detect logo in the range of 0–30° while it was mostly undetectable beyond this range. In the range 0–10°, the module’s logo detection mean Average Precision (mAP) was 0.91 while the classification accuracy was 0.97. In the range 10–20°, the module’s logo detection mAP was 0.85 while the classification accuracy was 0.90. On the other hand, in the range 20–30°, the module’s logo detection mAP was 0.80 while the classification accuracy dropped to 0.75. Angles from 0–40° are depicted in Fig. 6 along with the visual appearance of logos.

Table 1. Highlights of video data for different test cases

	Toll Control	Law Enforcement	Dashcam	Parking Lot Access Validation
Sequence Type	Outdoor	Outdoor	Outdoor	Outdoor
Environment	Sunny	Sunny	Day (Sunny), Night	Day (Sunny), Night
Object Class	Logo	Logo	Logo	Logo
Object Size	Small	Small to Medium	Medium to Small	Medium
Object Speed	Medium to Fast	Medium to Fast	Medium to Fast	Slow
Camera View	Top frontal and back	Side Frontal and Side back	Frontal and Back	Frontal
Camera Motion	Static	In motion	In Motion	Static
Resolution	3840 × 2160	1920 × 1080	1920 × 1080	1920 × 1080
Noise Level	Low	Medium to high (motion blur)	High (motion blur)	Low

Table 2. Results for each testcase using all the three approaches

Test Case	Approach 1	Approach 2	Approach 3	
			Detector	Classifier
	mAP	mAP	mAP	Accuracy
Toll Control	0.470	0.589	0.73	0.77
Law Enforcement	0.322	0.453	0.624	0.64
Dashcam	0.391	0.486	0.685	0.71
Parking Lot Access Validation	0.528	0.61	0.79	0.82



Fig. 4. (a) Single frame from each test case; (b)–(d) Individual vehicles in respective datasets of each test case along with their original logo containing ROIs and their respective unwrapped ROIs

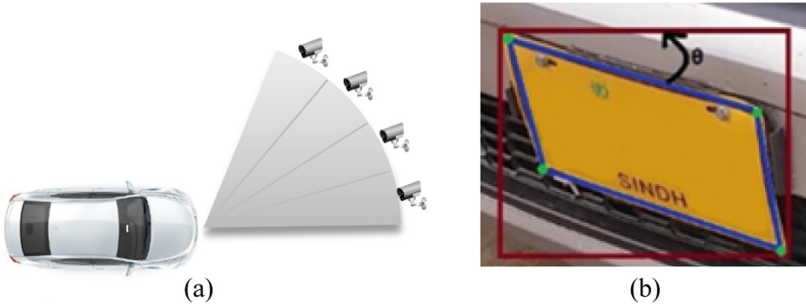


Fig. 5. (a): Photos capturing from different angles; (b): angle calculation for pose variance analysis

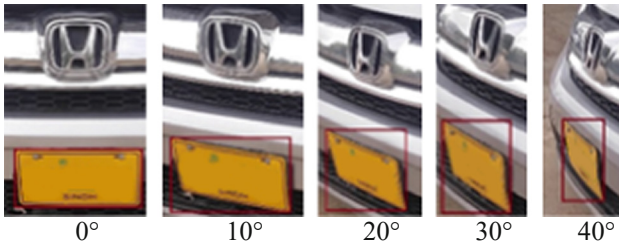


Fig. 6. Pose variance analysis on different angles along with respective logo appearance

5 Conclusion and Future Work

Intelligent TMMS is grasping researchers' attention due to the evolving infrastructure of cities. This is not only becoming complex day-by-day but also demanding more innovative solutions. Vehicle classification is now shifting towards more fine-grained classification. For this, all such features of vehicles are being considered which are distinct and specific. One such feature is the vehicle's logo. Though, it stands out as a unique feature depicting the information regarding vehicle make/model, yet its small size is a complex challenge for its detection and recognition. The detection of small sized objects like a logo in a surveillance environment where vehicles themselves appear to be very small is nearly impossible. Video enhancement techniques increase computational cost and time. This paper specifically targeted surveillance environment and presented a novel method of detection and recognition of vehicle's logo. A modified IWPOD-NET was introduced to overcome the orientation issues. Then YOLOv5 was used to detect logos and EfficientNet was used to classify logos. This pipeline was tested on four surveillance environments namely toll control, law enforcement, dashcam, and parking lot access control. Comparative analysis showed accuracy improvement with this proposed approach in each testing case. Pose variance analysis was also performed to determine the orientation limits to which this approach can work. It was observed that the proposed approach worked best in the 0–30° range, while it was mostly undetectable beyond this range. Secondly, a new dataset, VL-10 was presented which focused on local environment settings. The proposed technique in this research work improved the results for logo detection and can therefore be used in offline surveillance environment.

As part of a future work, the deep learning architecture can be optimized to detect and recognize vehicle logos in real-time. Similarly, a completely new deep learning architecture can also be designed particularly for this vehicle logo task.

Acknowledgement. We acknowledge support from National Center of Big Data and Cloud Computing (NCBC) and Higher Education Commission (HEC) of Pakistan for conducting this research.

Data Availability. The custom dataset utilized in this study is available on request from the corresponding author.

References

1. Setchell, C.J.: Applications of computer vision to road-traffic monitoring. Doctoral dissertation. University of Bristol (1998)
2. Huang, M.C., Yen, S.H.: A real-time and color-based computer vision for traffic monitoring system. In: 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No. 04TH8763), June 27, vol. 3, pp. 2119–2122. IEEE (2004)
3. Kun, A.J., Vamossy, Z.: Traffic monitoring with computer vision. In: 2009 7th International Symposium on Applied Machine Intelligence and Informatics, 30 January, pp. 131–134. IEEE (2009)
4. Poddar, M., Giridhar, M.K., Prabhu, A.S., Umadevi, V.: Automated traffic monitoring system using computer vision. In: 2016 International Conference on ICT in Business Industry & Government (ICTBIG), 18 November, pp. 1–5. IEEE (2016)
5. Osman, T., Psyche, S.S., Ferdous, J.S., Zaman, H.U.: Intelligent traffic management system for cross section of roads using computer vision. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), 9 January, pp. 1–7. IEEE (2017)
6. Malhi, M.H., Aslam, M.H., Saeed, F., Javed, O., Fraz, M.: Vision based intelligent traffic management system. In: 2011 Frontiers of Information Technology, 19 December, pp. 137–141. IEEE (2011)
7. Peng, Y., Yan, Y., Zhu, W., Zhao, J.: Vehicle classification using sparse coding and spatial pyramid matching. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 8 October, pp. 259–263. IEEE (2014)
8. Narhe, M.C., Nagmode, M.S.: Vehicle classification using SIFT. *Int. J. Eng. Res. Technol.* **03**(06), 1735–1738 (2014). ESRSA Publications
9. Llorca, D.F., Arroyo, R., Sotelo, M.A.: Vehicle logo recognition in traffic images using HOG features and SVM. In: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), 6 October, pp. 2229–2234. IEEE (2013)
10. Peng, H., Wang, X., Wang, H., Yang, W.: Recognition of low-resolution logos in vehicle images based on statistical random sparse distribution. *IEEE Trans. Intell. Transp. Syst.* **16**(2), 681–691 (2014)
11. Huang, Y., Wu, R., Sun, Y., Wang, W., Ding, X.: Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy. *IEEE Trans. Intell. Transp. Syst.* **16**(4), 1951–1960 (2015)
12. Li, W., Li, L.: A novel approach for vehicle logo location based on edge detection and morphological filter. In: Proceedings of the 2009 Second International Symposium on Electronic Commerce and Security, Nanchang, China, 22–24 May 2009, pp. 343–345 (2009)

13. Liu, Y., Li, S.: A vehicle-logo location approach based on edge detection and projection. In: Proceedings of the IEEE International Conference on Vehicular Electronics and Safety, Beijing, China, 10–12 July 2011, pp. 165–168 (2011)
14. Sulehria, H., Zhang, Y.: Vehicle logo recognition using mathematical morphology. In: Proceedings of the Wseas International Conference on Telecommunications and Informatics, Dallas, TX, USA, 22–24 March 2007 (2007)
15. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, 20 September, vol. 2, pp. 1150–1157. IEEE (1999)
16. Psyllos, A.P., Anagnostopoulos, C.-N.E., Kayafas, E.: Vehicle logo recognition using a SIFT-based enhanced matching scheme. *IEEE Trans. Intell. Transp. Syst.* **11**, 322–328 (2010). <https://doi.org/10.1109/TITS.2010.2042714>
17. Mao, S., Ye, M., Li, X., et al.: Rapid vehicle logo region detection based on information theory. *Comput. Electr. Eng.* **39**, 863–872 (2013). <https://doi.org/10.1016/j.compeleceng.2013.03.004>
18. Yu, S., Zheng, S., Yang, H., Liang, L.: Vehicle logo recognition based on Bag-of-Words. In: 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 353–358 (2013). <https://doi.org/10.1109/AVSS.2013.6636665>
19. Wang, Y., Liu, Z., Xiao, F.: A fast coarse-to-fine vehicle logo detection and recognition method. In: 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 691–696 (2007). <https://doi.org/10.1109/ROBIO.2007.4522246>
20. Tong, K., Wu, Y.: Deep learning-based detection from the perspective of small or tiny objects: a survey. *Image Vis. Comput.* **123**, 104471 (2022). <https://doi.org/10.1016/j.imavis.2022.104471>
21. Yang, S., Zhang, J., Bo, C., et al.: Fast vehicle logo detection in complex scenes. *Opt. Laser Technol.* **110**, 196–201 (2019). <https://doi.org/10.1016/j.optlastec.2018.08.007>
22. Zhang, J., Yang, S., Bo, C., Zhang, Z.: Vehicle logo detection based on deep convolutional networks. *Comput. Electr. Eng.* **90**, 107004 (2021). <https://doi.org/10.1016/j.compeleceng.2021.107004>
23. Jiang, X., Sun, K., Ma, L., Qu, Z., Ren, C.: Vehicle logo detection method based on improved YOLOv4. *Electronics* **11**(20), 3400 (2022). <https://doi.org/10.3390/electronics11203400>
24. Yang, S., Bo, C., Zhang, J., Wang, M.: Vehicle logo detection based on modified YOLOv2. In: Lu, H., Yujie, L. (eds.) 2nd EAI International Conference on Robotic Sensor Networks. EICC, pp. 75–86. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-17763-8_8
25. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
26. Huang, J., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5197–5206 (2015)
27. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: Proceedings IEEE International Conference on Computer Vision, pp. 349–356 (2009)
28. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2015)
29. Wang, X., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11133, pp. 63–79. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11021-5_5
30. Jia, Z., et al.: A two-step approach to see-through bad weather for surveillance video quality enhancement. *Mach. Vis. Appl.* **23**(6), 1059–1082 (2012)

31. Wu, C.-Y., Su, P.-C.: A region of interest rate-control scheme for encoding traffic surveillance videos. In: 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE (2009)
32. Silva, S.M., Jung, C.R.: A flexible approach for automatic license plate recognition in unconstrained scenarios. *IEEE Trans. Intell. Transp. Syst.* **23**, 5693–5703 (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Predicting Music Using Machine Learning

Aishwarya Asesh^(✉)

University of Utah, Salt Lake City, USA

a.asesh@gmail.com

Abstract. The intricate temporally prolonged sequences seen in music make it a perfect environment for the study of prediction. Melody, harmony, and rhythm are three examples of the structural elements found in music. This research incorporates music excerpts prediction by understanding structural details using Markov chain and LSTM models. The novel approach compares to state-of-the-art algorithms by predicting how a musical excerpt would continue after being given as input. To compare the variations in prediction and learning, different learning models with different input feature representations were utilized. This algorithm envisions multitude of usage including next generation music recommendation system using intra-sequence matching, pitch-tone correction, amongst others by integrating with recent advances in deep learning, computer vision, and speech techniques.

Keywords: music excerpt · markov chain · LSTM · sequence prediction

1 Introduction

Even though the western music scale only has twelve notes per octave, it encompasses the vast majority of the music one is familiar with, from Bach and Beethoven to Elton John and Beyonce. The past ten years have seen an increase in interest and a diversity of methods for algorithmically creating music using a variety of learning models [1]. Can deep learning models be used to study and utilize the genius of Mozart or Beethoven? How simple is it to recreate particular styles of music that the generation is accustomed to hearing?

While music has attracted a lot of attention, research on music prediction has received less attention. In this research, using a musical snippet, predictions are made for following N musical occurrences (for instance, 10 quarter-note beats). The actual musical occurrences are then contrasted with the predicted music, and the prediction is graded using a scoring system. Experiments leaves one wondering on whether music prediction or music generation is trickier.

2 Data

The Patterns for Prediction Development Dataset, was used. This data was produced by analysing a randomly chosen portion of the Lakh MIDI Dataset¹, a

¹ https://www.music-ir.org/mirex/wiki/2020:Patterns_for_Prediction.

dataset made up of one million popular music tracks. For both monophonic and polyphonic songs, the symbolic MIDI format was applied. Each input, or prime, corresponds to about 35s of music, and each output, or continuation, includes the following 10 quarter-note beats.

3 Feature Representation

3.1 Basic Music Notation

The fundamental component of music and the foundation from which all chords and melodies are built are notes. Each note has a pitch and a duration. Pitch is essentially the note’s sound frequency. The MIDI music file system can store 128 different pitches. Quarter-note lengths are used to measure duration, which is the length of the note. When several notes are played simultaneously, they form chords. Since only one note is played at a time, monophonic music lacks chord structure. Polyphonic music has chords.

3.2 Note and Chord Representation

Each note was encoded as an integer ranging from 0 to 128 due to the fact that notes have 128 pitches in the MIDI file system. The number 129 was used to symbolize a rest, or the amount of time between notes. In string format, a chord was shown as a collection of pitches. For instance, “68.70.75” might be used to denote the chord with the pitches 68, 70, and 75. The training set’s chords were all sampled, and a dictionary was built that converted the uncommon pairings into integer values. This corresponded to 101,039 distinct combinations and consequently input data dimensions for the medium-sized dataset.

How to handle inputs (chords) that are present in the testing set but absent from the training set is one of the difficulties in music prediction. This issue is avoided while creating music since the model only generates chords that are present in the training set. There will invariably be note combinations in the testing data that the model hasn’t “seen” previously while making a forecast. The initial strategy was replacing the unfamiliar chord with the most prevalent chord in the practice set.

A more sophisticated approach is to substitute the most frequent chord found inside the prime sequence for the unknown chord as the next sequence item. Combinatorial proliferation of chord possibilities in polyphonic music makes the process a bit challenging.

Duration Representation: The representation of the durations is a quarter-note length. The number of durations may be insurmountable, in principle. The duration length is typically spread among a limited set of choices. All of the training data’s durations were sampled, and a dictionary was built that converted each distinct duration length to an integer. The dictionary’s length varied depending on the dataset, although it was often less than 100 words long.

Combination Representation: The combination of note pitch and durations as a tuple was represented initially as a Pre-process step. For instance, the tuple was used to represent a note with a pitch of 60 and a duration of 1 quarter-note length (60, 1). Each distinct tuple was mapped to an integer value, just how the chords and durations were done.

Interval Representation: Note intervals were used in the final input representation that was assessed. Pitch difference between two notes is an interval. Each interval denotes a semitone of difference between the pitches of two notes that are right adjacent to each other. For instance, the pitch difference between the notes 60 and 65 is 5 semitones. Negative intervals are used to illustrate drops in pitch. There are 256 total possible interval values because there are 128 total pitches. Instead of using pitch sequences for training and prediction, interval sequences were used for this representation. For instance, the sequence of intervals for a prime consisting of the four notes 65, 70, 45, and 60 would be 5, -25, 15. To find the note predictions using the list of interval changes, the last pitch of the prime was preserved in a list and utilized as the beginning point.

4 Methods and Modeling

4.1 Markov Chain Model

One of the earliest models utilized for music production was the Markov chain. A first-order and a second-order Markov chain was trained and evaluated. Separate training sessions were held for notes and durations. This model could be utilized with inference methods like restricted inference and beam search [2–4].

4.2 LSTM Model

A Forest of Long-Short-Term Memory (LSTM) network was used for next set of experiments. A single network with multivariate input made up one model (both notes and durations). For the notes and durations, respectively, there were two different LSTM networks in the other model. It was established that notes and durations have a significant relationship. Input layer, 64-dimensional embedding layer, 512-hidden unit LSTM layer, batch normalization layer, and dropout layer with a dropout rate of 0.3 were the layers present in both models. The multivariate model then has two dense layers employing a categorical cross-entropy loss and an rmsprop optimizer for gradient descent after concatenating the two inputs. The batch size was 64, and there were typically 100 epochs. Experiments were performed using more epochs, but the outcomes didn't change, over-fitting was also observed in some cases. Figure 1 below displays the multivariate LSTM model's graph.

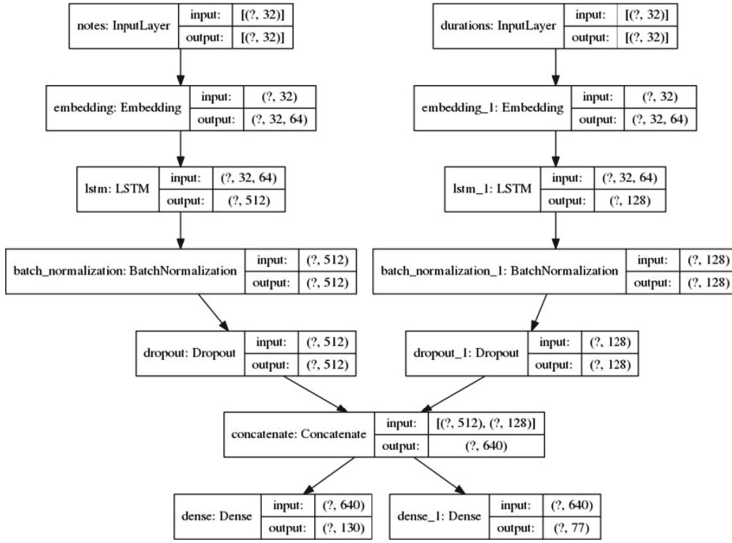


Fig. 1. LSTM Multivariate Model

4.3 LSTM Encoder-Decoder Model

Two different models for the notes and durations, as well as a single multivariate model, were trained. The parameters matched those of the LSTM model aforementioned. The longest prime sequence in the training set served as the input sequence length, and the longest target sequence in the training set served as the output sequence length. The graph of the multivariate encoder-decoder model is shown in Fig. 2.

4.4 Inference Modeling

Prediction was performed using the common beam search algorithm with beam sizes ranging from $k=2$ to $k=4$. Only the notes from the prime part were used for the constrained inference [5]. The following note with the highest score, for instance, would be the new option for prediction of song's prime sequence.

4.5 Other Techniques

Transposition was a method used to increase the uniformity of the input data. Transposition in music refers to the process of raising or lowering notes and chords by a fixed distance. The spaces between succeeding notes remain constant. Six major scales and six minor scales make up the 12 different scales in Western music. One of these scales is used in the majority of harmonic music. For instance, changing a song from the key of C major to the key of E major would require adding four semitones or four intervals to each pitch. Depending on whether the

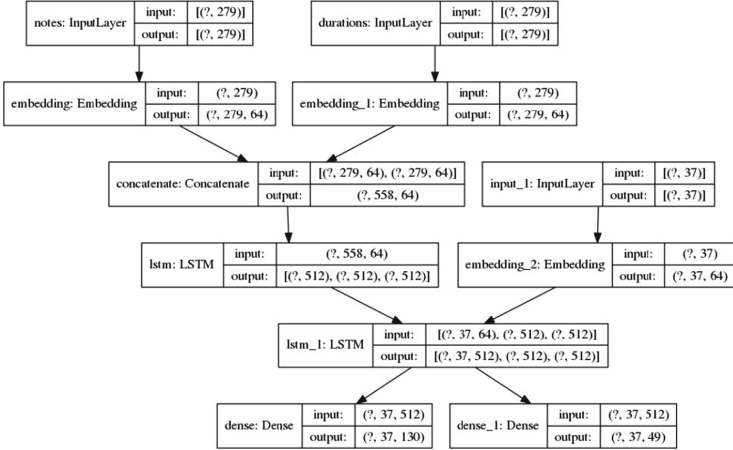


Fig. 2. LSTM Encode-Decoder Multivariate Model

music was originally in a major or minor key, each input song was transposed to either C major or C minor. The training then took place using the transposed pitches. The prediction notes were created by transposing the prime sequence into either C major or C minor before testing. The anticipated target notes were then recomposed back into their original key before being contrasted with the actual target notes.

5 Results and Observations

Both the training timeframes and the outcomes were adversely affected by the exponential growth of the input data for the polyphonic music. Scoring was done based on cardinality score and pitch score.

Cardinality Score: Minimizing a general supermodular monotone non-increasing function $f()$ with respect to a cardinality constraint. Let $x_1^* \in_x f(\emptyset) - f(x)$ and use the greedy algorithm on the set function

$$g(S) := f(x_1^*) - f(x_1^* \cup S),$$

which is a monotone non-decreasing submodular set function with $g(\emptyset) = 0$. Thus, the greedy algorithm maximizes the function with respect to a cardinality constraint within an approximation factor of $1 - 1/e$. However, there are two caveats. First, the greedy algorithm uses a budget of $k - 1$ instead of k (budget of one is spent on identifying x_1^*). Second, and most importantly, observe that the approximation factor is obtained on the function $g(S)$ and not $f(S)$, they get a set S of size $k - 1$ such that

$$f(x_1^*) - f(x_1^* \cup S) \geq 1 - \frac{k}{(k - 1)e} \cdot (f(x_1^*) - f(x_1^* \cup S^*)),$$

where S^* is the optimal set of size $k - 1$ to be added to x_1^* with the goal of minimizing $f()$.

The Pitch Score. The right “shape” or separation from each note is rewarded by the cardinality of continuations. However, it rewards a continuation that is transposed one semitone lower in most cases. The pitches are additionally scored because this isn’t entirely accurate. Two normalized histograms of the pitches from the prediction and the ground truth are made to achieve this. The amount by which the histograms are overlapped determines the score. The operation is carried out while ignoring octaves as well. The cardinality score and the pitch score are combined linearly to produce the final score.

Baseline. The first baseline estimate included selecting notes from the training set at random. The second technique involved selecting notes from the training set based on their frequency. The last approach used the same steps as the first, but added restrictions required that the sampled note be in the song’s expected prime sequence. The third method generated the highest baseline for both the polyphonic and monophonic music sets.

Discussions. Polyphonic and monophonic datasets were evaluated individually. 90% of the songs were used for training and 10% for testing in the first division of the datasets into training and testing groups. Figure 3 and Fig. 4 displays the outcomes of numerous training and testing runs for the monophonic and polyphonic dataset. The LSTM model produced the greatest results when it used constrained inference, transposition on the input data, and a longer sequence length of 64. Compared to the best baseline score, the model’s accuracy increased by over two times. The accuracy of the LSTM model was improved by transposing the input and lengthening the sequence. Transposition improved all of the models’ accuracy, suggesting that more reliable input data improves learning. There were no appreciable improvements in accuracy brought about by the beam search method of inference. Prediction accuracy was also improved by constrained inference, particularly for the Markov model.

The notes and duration feature representations for the input data were far more accurate than the interval feature representation. This suggests that compared to pitches, intervals may have less structure and greater unpredictability. Unsurprisingly, the results were significantly worse for the input with larger dimensions (i.e., the tuple combinations of pitches and durations). There were some minor differences between employing a multivariate LSTM model and training two distinct models for notes and durations. When all other variables remained constant, the multivariate model consistently beat the two independent models by a little margin. This suggests that there is some relationship between the anticipated notes and durations. In addition to having a slightly higher accuracy, the multivariate models also required less time to train than two individual models.

Polyphonic Music Datasets									
Model Type	Description	Feature Representation	Transpose	Sequence Length	Beam Search	Constrained Inference	Score	Cardinality Score	Pitch Score
Markov	First Order	Chords and durations as integers				✓	0.2664	0.0981	0.1683
Markov	First Order	Chords and durations as integers					0.2505	0.1020	0.1480
Baseline	Sampled with Constraints	Chords and durations as integers		N/A		✓	0.2359	0.0570	0.2789
Baseline	Sampled	Chords and durations as integers		N/A			0.2160	0.0440	0.1720
Baseline	Random	Chords and durations as integers		N/A			0.1830	0.0602	0.1228
Markov	First Order	Chords and durations as integers	✓			✓	0.1625	0.0435	0.1186
LSTM	Multivariate Input	Chords and durations as integers		32			0.0742	0.0273	0.0469

Fig. 3. Results - Model Predictions for Polyphonic Music Data

Across all models, the pitch score was far more variable than the cardinality score. When the findings were examined more closely, it became clear that many of the projected sequences only had two or three different pitches. As a result, the model could determine which pitch or pitches appeared in the target sequence the most frequently, but not which order they were played. Additionally, no predictions were ever made for notes that in the target sequence seemed to be more “random” (i.e., not present in the prime sequence). While music might have a strong structural element, it also has randomness or surprises that provide appeal.

Overall, the LSTM model performed better than every other model. The Markov models had the highest variance of all the models, indicating that they would benefit from further restrictions or domain expertise [6-8].

The baseline values for the polyphonic music were significantly higher than those for the monophonic music, neither set of results should be used to judge the other. Due to the method used to determine the score, the baseline scores for the two forms of music are different.

Monophonic Music Datasets									
Model Type	Description	Feature Representation	Transpose	Sequence Length	Beam Search	Constrained Inference	Score	Cardinality Score	Pitch Score
LSTM	Multivariate Input	Notes and durations as integers	✓	64		✓	0.3503	0.1116	0.2387
LSTM	Multivariate Input	Notes and durations as integers	✓	64	✓ (k=2)		0.3290	0.1070	0.2220
LSTM	Multivariate Input	Notes and durations as integers	✓	64			0.3280	0.1068	0.2214
LSTM	Multivariate Input	Notes and durations as integers	✓	64	✓ (k=3)		0.3280	0.1068	0.2214
Markov Model	First Order	Notes and durations as integers	✓	N/A		✓	0.3183	0.0900	0.2273
LSTM	Two separate models	Notes and durations as integers	✓	64	No		0.3140	0.0940	0.2196
LSTM	Multivariate Input	Notes and durations as integers	✓	32			0.3034	0.1055	0.1979
LSTM	Multivariate Input	Notes and durations as integers	✓	32	✓ (k=3)		0.3027	0.1031	0.1906
LSTM	Multivariate Input	Notes and durations as integers	✓	32	✓ (k=2)		0.3020	0.1028	0.1992
LSTM	Multivariate Input	Notes and durations as integers	✓	32			0.3003	0.1033	0.1970
LSTM	Multivariate Input	Notes and durations as integers	✓	64	✓ (k=3)		0.2830	0.1040	0.1790
LSTM	Multivariate Input	Notes and durations as integers		64			0.2810	0.0990	0.1820
LSTM	Multivariate Input	Notes and durations as integers		32			0.2760	0.1018	0.1741
Markov Model	First Order	Notes and durations as integers		N/A		✓	0.2317	0.0746	0.1570
LSTM	Two separate models	Notes and durations as integers		32			0.2286	0.0867	0.1420
LSTM	Multivariate Input	Intervals and durations as integers	✓	32			0.2248	0.1111	0.1132
Markov Model	Second Order	Notes and durations as integers	✓	N/A		✓	0.2130	0.0839	0.1290
LSTM	Single model	Tuple of notes and duration combinations		32			0.1930	0.1012	0.0923
Markov Model	First Order	Notes and durations as integers	✓	N/A	✓ (k=3)		0.1904	0.1164	0.0740
LSTM Encoder/Decoder	Two separate models	Notes and durations as integers		32			0.1800	0.0940	0.0860
LSTM Encoder/Decoder	Multivariate Input	Notes and durations as integers		32			0.1760	0.0897	0.0863
Markov Model	Second Order	Notes and durations as integers		N/A			0.1730	0.0750	0.0980
Baseline	Sampled with Constraints	Notes and durations as integers		N/A		✓	0.1649	0.0780	0.0869
Markov Model	First Order	Notes and durations as integers	✓	N/A			0.1490	0.0679	0.0816
Markov Model	First Order	Notes and durations as integers		N/A	✓ (k=2)		0.1184	0.0907	0.0277
Markov Model	First Order	Notes and durations as integers		N/A			0.1144	0.0660	0.0484
Baseline	Sampled	Notes and durations as integers		N/A			0.1037	0.0596	0.0441
Baseline	Random	Notes and durations as integers		N/A			0.0652	0.0541	0.0111

Fig. 4. Results - Model Predictions on Monophonic Datasets

6 Conclusion and Future Work

Because music has both structure and randomness, it presents an intriguing and difficult prediction problem. Although feature representation is significant, feature size is even more crucial. Training is substantially more challenging for inputs with very high dimensions. Constraints can significantly shrink the search space and improve model accuracy when used in inference. Meaningful learning and inference depend on limiting the input space and the search space in ways that nonetheless capture pertinent relationships and significant features.

Random pop tracks were employed in this research. Pop music is just one of many different musical styles. Jazz music is more free-form than classical music, which has a lot more structure. A fascinating expansion of this research would be to compare the results of other musical genres. Models, such as Hidden Markov Models and LSTM with attention mechanisms, leave room to wonder [9, 10].

Acknowledgments. All that I am, or ever hope to be, I owe to my angel mother.

References

1. Briot, J.-P., Hadjeres, G., Pachet, F.-D.: *Deep Learning Techniques for Music Generation*, vol. 1. Springer, Heidelberg (2020)
2. Lisena, P., Meroño-Peñuela, A., Troncy, R.: MIDI2vec: learning MIDI embeddings for reliable prediction of symbolic music metadata. *Semant. Web Preprint* **13**, 1-21 (2022)
3. Asesh, A.: SentiSeries: a trilogy of customer reviews, sentiment analysis and time series. In: Shakya, S., Balas, V.E., Kamolphiwong, S., Du, K.-L. (eds.) *Sentimental Analysis and Deep Learning*. AISC, vol. 1408, pp. 31–45. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-5157-1_3
4. Kim, S.T., Oh, J.H.: Music intelligence: granular data and prediction of top ten hit songs. *Decis. Support Syst.* **145**, 113535 (2021)
5. Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S., Lakshminarayanan, B.: Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **22**(57), 1–64 (2021)
6. Savage, P.E.: Music as a coevolved system for social bonding. *Behav. Brain Sci.* **44**, e59 (2021)
7. Gronauer, S., Diepold, K.: Multi-agent deep reinforcement learning: a survey. *Artif. Intell. Rev.* **55**(2), 895–943 (2022)
8. Guzmán, C., Mehta, N., Mortazavi, A.: Best-case bounds in online learning. *Adv. Neural Inf. Syst.* **34**, 21923–21934 (2021)
9. Chen, L., et al.: Decision transformer: reinforcement learning via sequence modeling. *Adv. Neural Inf. Process. Syst.* **34**, 15084–15097 (2021)
10. Schrittwieser, J., Hubert, T., Mandhane, A., Barekatin, M., Antonoglou, I., Silver, D.: Online and offline reinforcement learning by planning with a learned model. *Adv. Neural Inf. Process. Syst.* **34**, 27580–27591 (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





A Novel Process of Shoe Pairing Using Computer Vision and Deep Learning Methods

Marek Kozłowski¹(✉), Przemysław Buczkowski¹, and Piotr Brzezinski²

¹ National Information Processing Institute, Warsaw, Poland
markozlow@gmail.com

² Vive Textile Recycling, Warsaw, Poland

Abstract. The industrialisation of the footwear recycling processes is a major issue in the European Union—particularly in view of the fact that at least 90% of shoes consumed in western economies are ultimately sent to landfill. This requires new AI-empowered technologies that enable detection, classification, pairing, and quality assessment in a viable automatic process. This article discusses automatic shoe pairing, which comprises two sequential stages: a) deep multiview shoe embedding (compact representation of multiview data); and b) clustering of shoes' embeddings with a fixed similarity threshold to return sets of possible pairs. Each shoe in our pipeline is represented by multiple images that are collected in industrial darkrooms. We present various approaches to shoe pairing—from fully unsupervised ones based on image descriptors to supervised ones that rely on deep neural networks—to identify the most effective one for this highly specific industrial task. The article also explains how the selected method can be improved by hyperparameter tuning, massive increases in training data, and data augmentation.

Keywords: clustering · deep learning · multiview embedding · representation learning

1 Introduction

The increased availability of affordable mass-produced goods, coupled with rapidly changing consumer fashion trends, has resulted in a sharp increase in the consumption of products in many industrial sectors. The production of textiles for clothing and footwear is expected to increase by 2.7% each year until 2030¹. In [8], the authors claim that approximately 5% of the twenty billion pairs of shoes produced worldwide every year are recycled or reused. In the European Union, it is estimated that the waste that results from postconsumer shoes could reach several million tonnes per year. Politicians and members of civil society

¹ <https://www.businessoffashion.com/reports/news-analysis/the-state-of-fashion-2021-industry-report-bof-mckinsey/>.

are starting to implement the Zero Waste to Landfill policy to address one of the major challenges of the twenty-first century for the footwear sector. This ambitious goal requires extensive support from intelligent information systems.

The problem considered in this article results from practical need. VIVE Textile Recycling is the leader of the textile recycling industry in Poland and Europe. The firm is regularly required to process massive deliveries of shoes. A conveyor belt transports the shoes to a darkroom, where their profile photos are taken (of the vamp, sole, and heel). Next, a management system, ShoeSelector consumes information from many external intelligent systems, and decides, for example, where the shoes should be thrown off of the conveyor belt, and whether they should be paired with their counterparts or remain single shoes. Shoe pairing is crucial for the shoes to be resold.

The primary goal of this article is to present a novel shoe-pairing method that can be used effectively in the process of analysing large quantities of shoes that are transported on a conveyor belt. We have verified various well-known classical approaches, including image descriptors like Central/Hu moments, HOG, and SSIM-based approaches, as well as a novel deep-learning-based method that outperforms others.

In our deep learning approach to shoe pairing, we introduced two sequential stages: a) multi-image shoe embedding that uses a deep neural network; and b) clustering of the shoes' embeddings with a fixed similarity threshold to return pairs (clusters in which size = 2) and singles (unpaired shoes). Each shoe in our pipeline is represented by three images (vamp, sole, and heel) that are collected in the darkrooms. Sample images are presented in Fig. 1. Our approach is evaluated on diverse custom industrial datasets, as provided by Vive Textile Recycling.

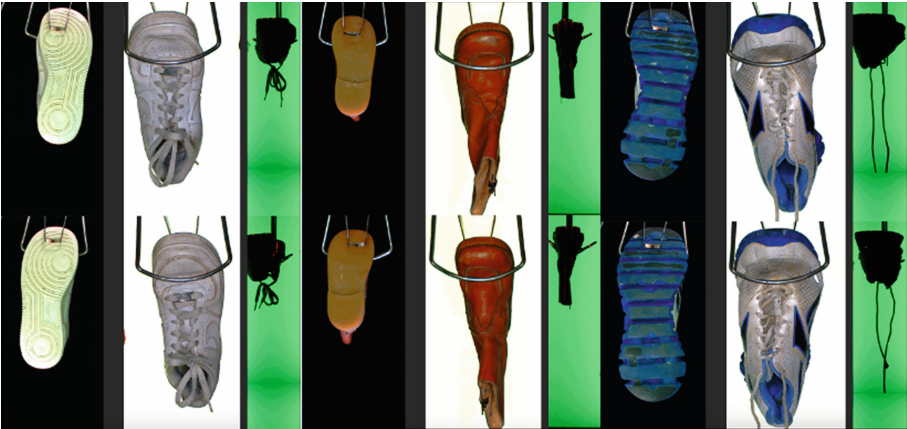


Fig. 1. Sample shoe images. The top row presents all three images (sole, vamp, and heel) for three distinct shoes. The bottom row presents the corresponding shoes found by the system.

2 Related Work

Shoe pairing can be modelled as clustering of shoes' compact representations within a cluster (maximum size = 2) and a minimum similarity threshold to discriminate correct pairs (two-element clusters) against single shoes without counterparts.

Clustering is a fundamental task in the machine learning paradigm whose objective is to divide subjects into a number of groups in such a way that subjects in the same groups are more similar to each other and less similar to those in other groups. Traditional methods cluster subjects on the basis of a single set of features per subject [11].

Multiview clustering (MVC) is a variant of clustering in which each subject is represented by multiple sets of features [2]. MVC has been applied successfully to various applications, including computer vision, natural language processing, healthcare, and social media. Given that our project involves only images assigned to shoes, our focus lies in computer vision.

MVC has been used widely in image clustering [7] and motion segmentation [5] tasks. Chi et al. [3] conducted MVC for web image retrieval ranking. Xin et al. [14] successfully applied MVC for person reidentification. Typically, several feature types, such as HOG [4], LBP [10], and SIFT [9], can be extracted from images prior to cluster analysis.

Since 2012, deep learning has proved outstandingly efficient in a variety of applications, such as image classification, speech recognition, object detection, and natural language processing. Multiview deep clustering methods demonstrated better performance than traditional multiview shallow clustering methods. Most of the literature presents multiview deep clustering as the process of clustering on the representations obtained from multiview representation models built in a supervised manner [2].

Multiview representation learning (MVRL) has recently become popular by its exploitation of complementary information of multiple features or modalities. Recently, due to the remarkable performance of deep models, deep MVRL has been adopted in many domains, including computer vision and signal processing. One article [15] presents a comprehensive review of deep MVRL in two perspectives: (1) deep MVRL extensions of traditional methods; (2) MVRL methods in full deep learning scope. The first group of methods introduce the advancements of deep learning models into traditional MVRL methods, such as multiview canonical correlation analysis and matrix factorisation. The second group represents pure deep learning MVRL methods, such as multiview autoencoders, conventional neural networks, and deep belief networks [15].

The deep neural networks used to learn deep representations adopted in multiview clustering methods have superior expression ability to reflect multiview data comprehensively. Multiview deep representation learning is a method of transforming a collection of inputs (in our case, profile images) into compact, fixed-size, floating-point representations that exhibit the desired properties. Such embedding can be subject to further processing, such as clustering. However, the authors of [2] claim that the separate processes of representation learning and

clustering entails limitations, such as representation learning being unaware of clustering’s goal. Our approach resolves this problem by using learning representations that are tailored to the purpose of our clustering, which is pairing.

The outstanding efficiency of multiview deep-based clustering and its ability to make representation learning aware of the clustering’s goal inspired us to develop a novel shoe-pairing process based on the clustering of representations obtained from a trained deep multiview representation model (hereinafter referred to as deep multiview embedding-based clustering, DMVEC).

3 The Proposed Approach

The solution is a web service that is capable of responding to a variety of shoe-related requests via REST API. This article describes only the pairing aspect of the system. Answering pairing-related questions requires the system to perform a series of processing steps—some of which are shared with other tasks. The first (1) is input preprocessing (e.g. image decoding from base64 encoding, image scaling, and normalisation). Later (2), shoe detection is performed (empty hangers occasionally get photographed). Next, (3) deep multiview embedding of a shoe is performed (this step assumes that the embedding model is already trained and available for inference). Last (4), clustering of the shoes’ embeddings into pairs can be requested. In the two sections below, we describe (3) and (4) in greater detail.

3.1 Deep Multiview Representation Learning

The deep multiview embedding approach is a supervised method of changing the representation of a shoe into a single fixed-size vector based on its three images (shoe embedding). The process is based on a deep neural network, which is responsible for transforming the three images of a shoe into a compact vector of floating numbers. This transformation is trained in such a way that each shoe in a pair should be located close to the other, but away from unrelated shoes. The Euclidean distances between the embeddings can be used by the clustering method as a distance metric.

This method is a multibranch combination of convolutional layers and dense layers. The input comprises three photographs of a shoe (vamp, sole, and heel). The output is a single 64-dimensional unit floating vector. In the initial part of the proposed deep neural network, we used the convolutional base of the VGG-16 network [12], pretrained on the ImageNet dataset². The results of the convolutional base are globally pooled by channels. The use of the convolutional base of the VGG-16 network in our method is an example of transfer learning. The nonconvolutional part of the network is trained from scratch by freezing the VGG-16 weights. The vector obtained as a result of global average pooling of the output of the convolutional base is processed using a fully connected layer. Such

² <https://image-net.org>.

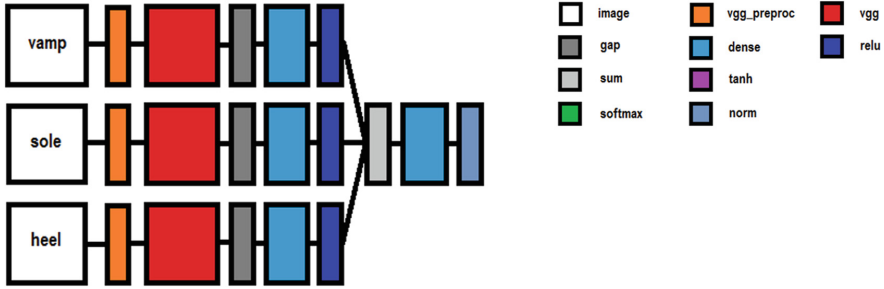


Fig. 2. The deep multiview shoe embedding model. vgg_preproc – a preprocessing layer in which images are converted from RGB to BGR, before each colour channel is zero-centered relative to the ImageNet dataset; vgg – the convolutional base of the original VGG-16 network pretrained on ImageNet; gap – global average pooling, the layer used for dimensionality reduction; dense - a dense or otherwise fully-connected layer; relu – an activation layer in which all negative values are replaced with zero and the remaining values are passed to the output; sum – a layer for aggregating the output by adding tensors from multiple layers; norm – a layer that normalises the input vector to a unit vector.

a network fragment, called a branch, is repeated between one and three times, depending on the number of image types desired. Then, all branches are aggregated over the corresponding indices (element-wise) and transformed once more using the fully connected layer. Last, the 64-dimensional output is normalised so that it lies on a multidimensional sphere with a unit radius. Figure 2 presents the entire architecture of the deep neural network. To train our embedding model, we used triplet learning (namely, the TripletSemiHardLoss and TripletHardLoss loss functions), in which each reference (anchor) shoe requires a positive example (a corresponding shoe from a pair) and a negative example (a nonsimilar shoe). The deep neural network is trained in a regime that forces similarity (e.g. Euclidean-distance-based similarity) between the representations of the reference and the compatible shoe, while reducing the similarity of the representation of the reference example and the negative example.

3.2 Clustering

Given the trained deep multiview representation model (the shoe embedding model), a visually represented batch of shoes can be transformed into a set of embeddings. The most similar embeddings can then be identified to pair the shoes. The desired pairing should connect the elements so that the distance—and, therefore, the differences in appearance—is as short as possible (or alternatively, similarity is as high as possible).

Although we evaluated many methods, considering the heavy deep multiview embedding model and specific nonfunctional requirements (such as the number of shoes in the clustered population being limited by the capacity of the conveyor belt, which is around 1,000–2,000 hangers), the most robust and comprehensive

was the greedy method, which is based on agglomeration clustering with an additional termination condition. It works as follows:

1. Create a collection of unpaired shoes L ;
2. Create an empty collection of result pairs P ;
3. Find two shoes l_a and l_b in L that are separated by the shortest Euclidean distance between their embeddings, which does not exceed the threshold t ;
4. Remove them from L and add a pair (l_a, l_b) to P ;
5. If at least two items remain in L , go to step 3;
6. Return pairing P plus the remaining unpaired shoes as *singles*.

4 Results

In the initial phase of our experiments, we had limited access to labelled data. Approximately 1,000 hand-labelled pairs of shoes were split into training (80%) and test (20%) sets. Using these relatively small datasets, we evaluated various methods to identify the most promising one. At this point, we suspected that we using too little data to fully unlock the benefits of deep learning. We opted to use transfer learning, which is helpful in cases of scarce data.

We verified four unsupervised classical methods: Central and Hu moments, HOG, and SSIM. We used a special evaluation measure called pairing accuracy, which is the percentage of shoes that are correctly paired (or unpaired). We define pairing accuracy as the number of shoes with correct assignment (properly paired or properly unpaired) divided by the number of shoes considered.

Table 1. The quality evaluation of the shoe-pairing methods on the test set measured by pairing accuracy; VAMP/SOLE/BOTH indicates from which images the shoe-pairing model was inferring.

Method	VAMP	SOLE	BOTH
DMVEC	0.73	0.95	0.99
Central moments	0.02	0.01	0.01
Hu moments	0.01	0.01	0.02
HoG	0.03	0.05	0.08
SSIM	0.05	0.09	0.10

Table 1 presents the results of the shoe-pairing experiments on our initial test set. The first row presents the proposed DMVEC approach (deep multiview embedding-based clustering). The subsequent rows present the results obtained by our unsupervised classical baselines:

1. Central moments – the shoe images, separately or joined horizontally in one image, are described by image descriptors called central moments [6] before clustering is applied;

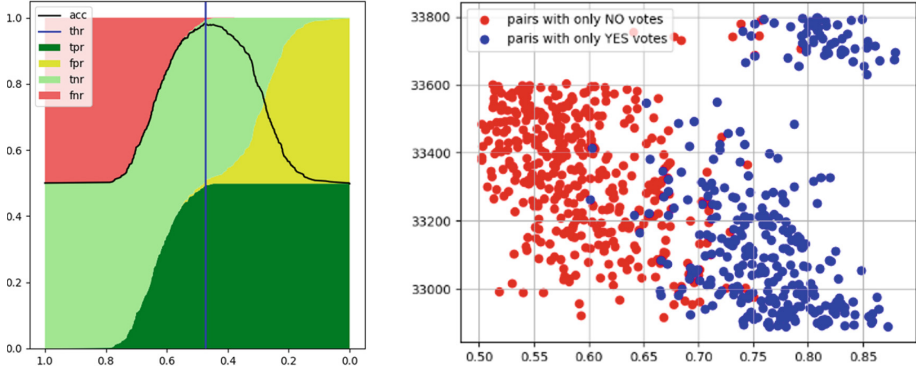


Fig. 3. Adjustment of the similarity threshold is a crucial task in shoe pairing. The image on the right shows the similarity between shoes in proper and incorrect pairs in a set of 34,000 pairs returned by a prototype system and next labelled by human taggers. The image on the left demonstrates how we tuned the hyper-param-similarity threshold using the 1,000-shoe validation set. We counted partial errors and accuracy (acc) for each threshold setting. The legend presents the errors as a ratio of TP/FP/TN/FN to all shoes in the paired set. *Thr* is represented by the vertical line that corresponds to the threshold value that reports the highest accuracy on the 1,000-shoe validation set.

2. Hu moments – the shoe images, separately or joined horizontally in one image, are described by image descriptors called Hu moments [6] before clustering is applied;
3. HoG – the shoe images, separately or joined horizontally in one image, are described by image descriptors called histogram of oriented gradients - (HoG) [13] before clustering is applied;
4. SSIM – the shoe images, separately or joined horizontally in one image, are compared one vs one using a structural similarity index measure (SSIM) [1]. SSIM distances are used for clustering.

The initial experiment revealed that DMVEC obtained the best results; however, the test set consisted solely of pairs i.e. all shoes in the test set had their counterparts within the test set. This situation is a borderline one; usually, the ratio of paired shoes in a real-world population of shoes under pairing is between 50% and 90%. The appearance of singles in a population demands the introduction of a threshold, which prevents further pairing and returns the remaining shoes as singles. We received a 1,000-shoe validation set from Vive Textile Recycling that represented the most common distribution of pairs in a population. In Fig. 3, we present how we tuned the similarity threshold value based on the validation set.

The next phases of the project involved conducting experiments on diverse test populations that contained significant proportions of singles. The results of the experiments revealed that accuracy reported a huge drop from an initial 99% to around 80–90% (lower scores when more singles were in the population;

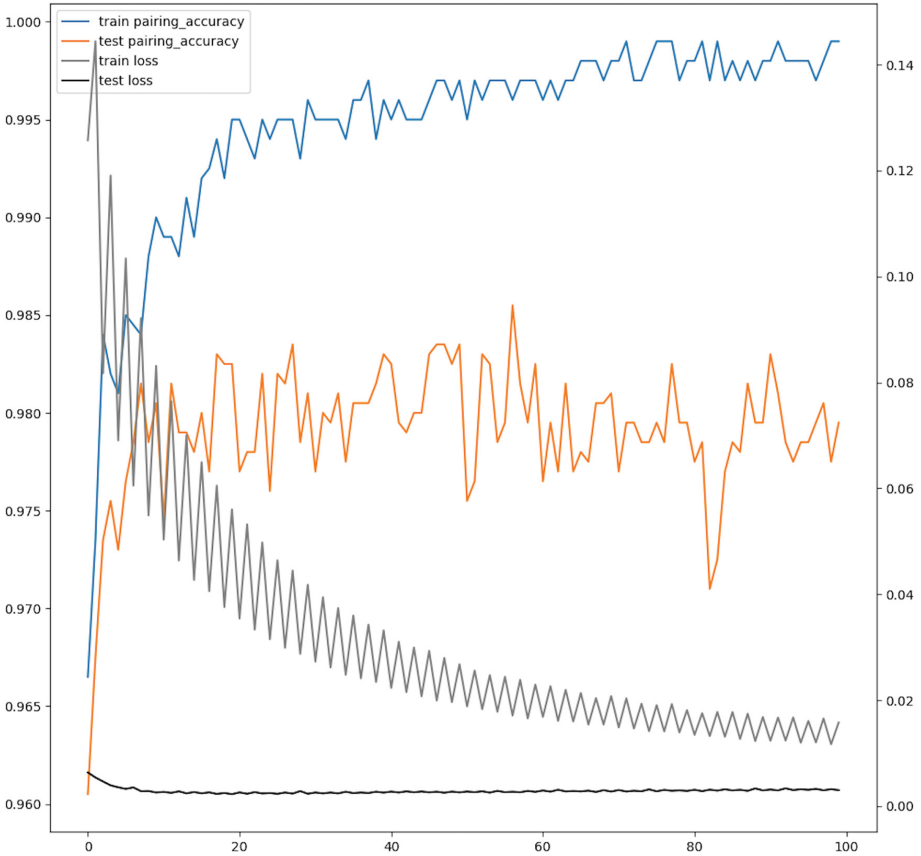


Fig. 4. Training and validation (on the test set) metrics concerning our best pairing model. The model was trained using the maximum number of true pairs collected over two years of labelling efforts. We accumulated 108,000 shoes in pairs, and validated/tested after each epoch using 2,000 shoes, of which 1,000 were paired and 1,000 were singles (the x-axis represents epochs; the y-axis represents accuracy - left axis and loss score - right axis).

80% was reported when we had equal numbers of paired shoes and singles). After investigating the drop in pairing accuracy, we reached two conclusions. (1) Finding pairs in a population that does not contain singles is a significantly easier task than pairing a general population. The threshold is challenging to discover; embeddings are typically not separable by any single threshold and this introduces a tradeoff between precision and recall. (2) The difficulty of shoe pairing depends inherently on the size of the population. This differs from other tasks, like classification, in which scores are independent of validation set size. This happens because in pairing, examples in the dataset are not independent

of each other. This can be also seen from the perspective of a solution space, in which a combinatorial explosion appears in the number of possible pairings.

We decided that our training dataset was insufficiently representative for deep multiview representation learning, and that a more robust embedding model was demanded. We manually labelled tens of thousands of pairs returned by a prototype system in a preindustrial environment. After more than a year of labelling, we had collected approximately 54,000 proper pairs. In Fig. 4, we present the training and validation process using our largest training dataset and data augmentation (cropping, saturation, and hue disturbance). The trained model is highly robust and reports high accuracy scores during clustering on the test set—even when it was tested on a very difficult set (1,000 shoes in pairs and 1,000 singles), it presents accuracy above 97% after the 10th epoch.

5 Conclusions

This article presents a novel approach to shoe pairing, a dual-stage approach that comprises deep multiview shoe embedding and clustering. We evaluated different approaches to shoe pairing, from classical unsupervised ones based on image descriptors to the proposed supervised one that applies deep neural networks. The best-performing model in this task is the proposed supervised method, DMVEC. It reports almost 100% accuracy on the test sets that exclusively comprise shoes in pairs, and at least 97% when singles cover almost half of the test population. Over a broad number of tests, we evaluated different test sets (with different size and various distributions of singles). We also demonstrated how our selected method can be improved by hyperparameter tuning (similarity threshold tuning), massive increases in training data, or data augmentation.

In the future, we plan to conduct research on further optimisations for DMVEC. We suspect that multiple factors can impact the pairing accuracy of our model, such as the margin hyperparameter in triplet learning, the schedule of `TripletHardLoss` and `TripletSemiHardLoss` during training, or the GAP layer discarding too much information. There are also other necessary tasks apart from pairing, which will be addressed soon.

References

1. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. *IEEE Trans. Image Process.* **21**(4), 1488–1499 (2011)
2. Chao, G., Sun, S., Bi, J.: A survey on multiview clustering. *IEEE Trans. Artif. Intell.* **2**(2), 146–168 (2021)
3. Chi, M., Zhang, P., Zhao, Y., Feng, R., Xue, X.: Web image retrieval reranking with multi-view clustering. In: *Proceedings of the 18th International Conference on World Wide Web*, pp. 1189–1190 (2009)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893. IEEE (2005)

5. Djelouah, A., Franco, J.S., Boyer, E., Le Clerc, F., Pérez, P.: Multi-view object segmentation in space and time. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2640–2647 (2013)
6. Flusser, J., Zitova, B., Suk, T.: Moments and Moment Invariants in Pattern Recognition. John Wiley & Sons, Hoboken (2009)
7. Jin, C., Mao, W., Zhang, R., Zhang, Y., Xue, X.: Cross-modal image clustering via canonical correlation analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29 (2015)
8. Lee, M.J., Rahimifard, S.: An air-based automated material recycling system for postconsumer footwear products. *Resour. Conserv. Recycl.* **69**, 90–99 (2012)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
10. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002)
11. Schütze, H., Manning, C.D., Raghavan, P.: Introduction to Information Retrieval, vol. 39. Cambridge University Press, Cambridge (2008)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
13. Suard, F., Rakotomamonjy, A., Bensrhair, A., Broggi, A.: Pedestrian detection using infrared images and histograms of oriented gradients. In: 2006 IEEE Intelligent Vehicles Symposium, pp. 206–212. IEEE (2006)
14. Xin, X., Wang, J., Xie, R., Zhou, S., Huang, W., Zheng, N.: Semi-supervised person re-identification using multi-view clustering. *Pattern Recognit.* **88**, 285–297 (2019)
15. Yan, X., Hu, S., Mao, Y., Ye, Y., Yu, H.: Deep multi-view learning methods: a review. *Neurocomputing* **448**, 106–129 (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Representation of Observations in Reinforcement Learning for Playing Arcade Fighting Game

Huaiyu Du¹ and Rafał Józwiak^{1,2}

¹ Faculty of Mathematics and Information Science,
Warsaw University of Technology, Warsaw, Poland
Rafal.Jozwiak@opi.org.pl

² Applied Artificial Intelligence Laboratory, National Information Processing
Institute, Warsaw, Poland

Abstract. Reinforcement learning (RL) is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning. Reinforcement learning algorithms have become very popular in simple computer games and games like chess and GO. However, playing classical arcade fighting games would be challenging because of the complexity of the command system (the character makes moves according to the sequence of input) and combo system. In this paper, a creation of a game environment of The King of Fighters '97 (KOF '97), which implements the open gym env interface, is described. Based on the characteristics of the game, an innovative approach to represent the observations from the last few steps has been proposed, which guarantees the preservation of Markov's property. The observations are coded using the "one-hot encoding" technique to form a binary vector, while the sequence of stacked vectors from successive steps creates a binary image. This image encodes the character's input and behavioural pattern, which are then retrieved and recognized by the CNN network. A network structure based on the Advantage Actor-Critic network was proposed. In the experimental verification, the RL agent performing basic combos and complex moves (including the so-called "desperation moves") was able to defeat characters using the highest level of AI built into the game.

Keywords: Reinforcement learning · KOF '97 · Arcade fighting game · Python · pattern recognition · Artificial Intelligence · Neural Network · CNN · graphical representation · agent's reward system · transfer learning

1 Introduction

Since DeepMind's AlphaGo defeated the world champion in the ancient GO game in 2015, with the rise of deep learning, more and more researchers have been dedicated to studying reinforcement learning (RL). One of the favorite RL

study field is playing video games. Since “Street Fighter 2” was published on the arcade platform in 1991, fighting games have taken the world by storm, and the fighting genre has become an essential category in video games. Many researchers have studied applying RL techniques to fighting games such as the 2D game “Super Smash Bros” [1], 2.5D game “Little Fighter 2” [3], and 3D game “Blade & Soul” [5]. In [1], researchers used information such as player’s location, velocity, and action state read from game memory as observations instead of raw pixels. They used two main classes of model-free RL algorithms: Q-learning and policy gradients, and both approaches were successful even though the two algorithms found qualitatively different policies from each other. After training the agent to fight against in-game AI, a self-play curriculum was utilized to boost the agent’s performance, making the agent competitive against human professionals. In [3], since in 2.5D games, 3D objects in a scene are orthographically projected to the 2D screen, the main challenge is the ambiguity between distance and height. To tackle this problem, in addition to an array of 4-stacked, gray-scaled successive screenshots, the location information of both the agent and the opponent read from the environment was used as an observation. A CNN was used to extract visual features from the screenshots, and other information features were concatenated with the CNN’s output. They also utilized LSTM in the network in order to enable the agent to learn proper game related and action-order related features. In [5], a novel self-play curriculum was utilized. By reward shaping, three different styles (Aggressive, Balanced, and Defensive) of agents were created and trained by competing against each other for robust performance. By introducing a range of different battle styles, diversity in the agent’s strategies were enforced, and the agents were capable of handling a variety of opponents.

Compared with classical fighting games on the arched platform, such as the “Street Fighter” series and “The Kind of Fighters” series, the command system of games mentioned above is relatively simple: special moves are triggered either by combinations of direction keys and action keys or by dedicated keys. While in the game of KOF ’97, the movement system is more complex. Except for basic moves such as a punch or a kick, special and desperation moves’ commands consist of a sequence of joystick movements and button presses. A desperation moves’ command is longer than special moves’ and using a desperation move consumes a power stock indicated by a green flashing orb at the bottom of the screen. The movement and combo systems require from the RL agent to master the ability to complete relatively long sequences of accurate inputs.

To the best of our knowledge, till now, there is no RL research dedicated to arcade fighting games like KOF ’97, which typically feature special moves and desperation moves that are triggered using rapid sequences of carefully timed button presses and joystick movements. The aim of the study described in this paper was to develop a RL agent which can not only win the game KOF ’97 but also masters a complex game movement system. In order to achieve this goal, special attention was paid to the effective representation of observations in the RL environment.

2 Environment Setup for KOF '97

2.1 Interaction with Arched Emulator

An arched game emulator called MAME is used to set up the python game environment. It is possible to drive MAME via Lua scripts externally, and the event hook allows interaction with the game at every step. An open source Python library called MAMEToolkit [4] encapsulates Lua scripts operation into Python classes. The KOF '97 game environment, which implements the Gym env interface, is built based on this python lib. Figure 1 shows a simplified communication process between the game environment and MAME emulator. Whole source code, including game environment, training, and evaluating, is available on GitHub¹

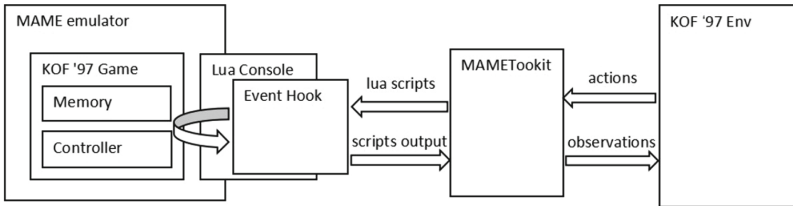


Fig. 1. Simplified communication process

2.2 Action Space

There are eight basic input signals: UP, DOWN, LEFT, RIGHT, A, B, C, and D. The action space has 49 discrete actions, including all meaningful combinations of the basic input signals. One of the problem is that direction inputs in the commands are based on the character's orientation, i.e., if the character changes its orientation, all direction inputs in the commands should be flipped horizontally. Since the agent operates 1P's (Player 1) character, most of the time, the character is facing to the right side. It makes the training samples unbalanced. Moreover, the experience learned from the default orientation could not be applied to the other side. A trick of action flipping is used to tackle these problems. In the action space, FORWARD and BACKWARD commands are used instead of LEFT and RIGHT, and the environment will transfer FORWARD and BACKWARD to the LEFT or RIGHT according to the character's orientation. With the help of this trick, the agent's win rate increased to over 90% from 70% after 30M timesteps' training.

¹ Source code address: <https://github.com/duhuaiyu/AIbotForKof97>.

2.3 Observation

The observation vector from a single step consists of several parts. The first part contains basic information (such as characters’ HP, location, POW state, and combo count). All scalar features are normalized. The second part is the characters’ ACT code, which represents the character’s action. ACT is a concept in the game implementation mechanism. The character’s ACT code and corresponding action can be checked in debug mode of the game. A character’s move consists of one or more ACTs, and each ACT consists of several animation frames. As Figs. 2 shows, the character’s “100 Shiki Oni Yaki” Move consists of 3 ACTs (Codes are 129, 131, and 133, respectively) which denote the start, middle, and end of the move. Each character can have a maximum of 512 ACTs. Although some of the ACT codes are not used, for compatibility, a 512 bits one-hot encoding vector is used to represent the ACT code feature. The basic information and characters’ ACT code can be read from the game memory. The last part is the agent’s input, which is also a one-hot encoding vector.

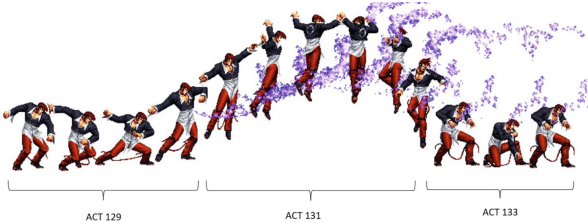


Fig. 2. “100 Shiki Oni Yaki” Move’s ACTs

2.4 Graphical Representation of the Input and ACT Sequences

For readers can have a better understanding of the approach proposed below, it is worth to mention about the command judgment system of the game. As mentioned, to make a special move or desperation move, the player needs to finish a sequence of input in a short period. The game system would determine which move the character is going to do according to the input history. As human input varies in rhythm and precision, the judgment of input sequence has some flexibility. Take the move “hopping back” for example. The command is “← ←”. As Fig. 3 shows, as long as the player can finish such an input pattern in 8 frames, the character can make the move successfully. In case 1 and case 2, although activated keys last for a different time, the input pattern “← ←” is finished in 8 frames, so the character can make the hopping back move. In case 3, such an input pattern is not finished in the time window, so the character failed to make the move.

The observation features from a single step are not enough to capture all important information since, on the one hand, the character’s move highly

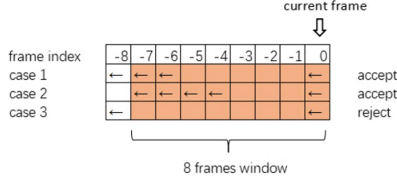


Fig. 3. Hop back command's input judgment

depends on the recent input history; on the other hand, according to the sequence of actions the characters have already made and for how long these actions have lasted, the player can make more accurate predictions and make an optimal response. In order to preserve the Markov's property, it is necessary to maintain the character's input sequence and ACT sequence within a specific time window. But simply concatenating features from several steps is not a good idea. Firstly, the feature vector for each step is long (over 1000 per step), and at least 30 frames' feature is needed since desperation move's commands are pretty long. It means more neurons for the network are needed, and training such a network would be slow and difficult. Besides, because of the command judgment system's flexibility, the input sequence can vary for the same move, and it is hard for an MPL network to find such a pattern. Our idea is to stack the 1P's input sequence, 1P's ACT sequence, and 2P's ACT sequence as three binary images. Figure 4 shows our approach to the representation of the feature sequences in the form of binary images, and these images encode the character's input and behavioural pattern, which are then retrieved and recognized by the CNN network. Although input sequences can differ for the same move, the input sequences will form a similar pattern on the graphic representation, and CNN is good at image pattern recognition.

2.5 Reward System

Based on the goal, the following reward system inspired by the reward system described in [1] is proposed:

$$R^{total} = R^{1P_damage} - P^{2P_damage} - P^{distance} - P^{time} - P^{POW}$$

where

$$R^{1P_damage} = d^{1P} * c^{1P} * \gamma^{1P}$$

$$P^{2P_damage} = d^{2P} * c^{2P} * \gamma^{2P}$$

$$P^{distance} = Max(0, g - \beta) * \gamma^{distance}$$

$$P^{time} = 1 * \gamma^{time}$$

$$P^{POW} = \begin{cases} 5 & \text{if a POW orb consumed} \\ 0 & \text{otherwise} \end{cases}$$

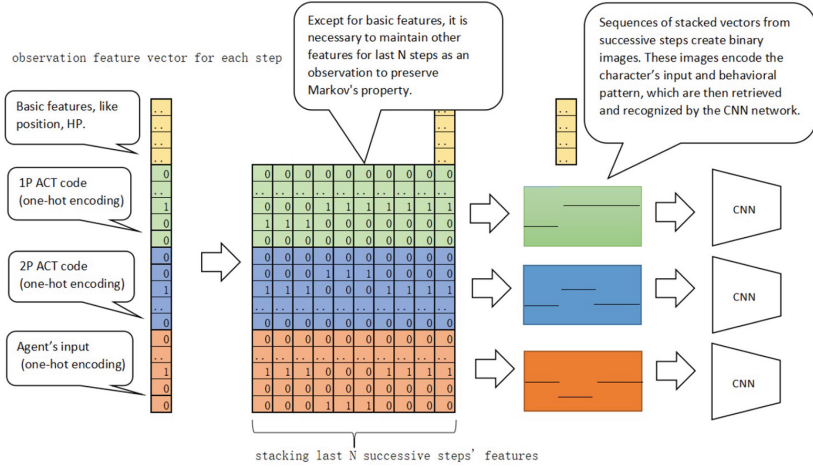


Fig. 4. Graphical representation of the input sequence and ACT sequences

The first two terms are a damage reward and a damage penalty (R^{1P_damage} , P^{2P_damage}). The priority is to win the game, so it is quite natural that if the character deals damage (d^{1P}), the agent gets a reward. In contrast, if the opponent deals damage (d^{2P}), the agent gets a penalty. In order to encourage the agent to perform more combo, the damage reward and penalty will be amplified linearly with the combo numbers (c^{1P} , c^{2P}). On top of that, two factors: γ^{1P} , and γ^{2P} , are introduced to balance the reward and penalty caused by damage. To prevent the agent from behaving too aggressively, c^{2P} is set to 1.3, and c^{1P} is set to 1.

Except for the damage reward and penalty, some other terms are also introduced to adjust the agent's behavior. The distance penalty term $P^{distance}$ is used to make the agent keep a reasonable distance between the two characters. If the distance exceeds a certain threshold β , the agent will get a small penalty controlled by the factor $\gamma^{distance}$. The term P^{time} gives the agent a small penalty at every step. It is used to encourage the agent to deal damage and win the game as soon as possible. The last term in the formula is P^{POW} . It gives a moderate penalty (specifically 5) when the POW orb is consumed, and if the desperation move can hit the opponent later, the reward will counter-weight this penalty. Otherwise, the agent will be punished for using the POW orb for nothing.

2.6 Proposed Network Structure

Figure 5 shows the overall structure of our proposed network. The network's input consists of four parts: basic features as a vector, stacked input sequence (last 40 frames), stacked 1P characters' ACTs (last 36 steps), and stacked 2P characters' ACTs (last 36 steps). There are three CNNs to extract features from these images, and the results of these CNN extractors are concatenated with the

basic features as the input of our Action Network and Value Network. Figure 5b shows the structure of CNN network. The proposed network is referred to as Multi-CNN in the following content.

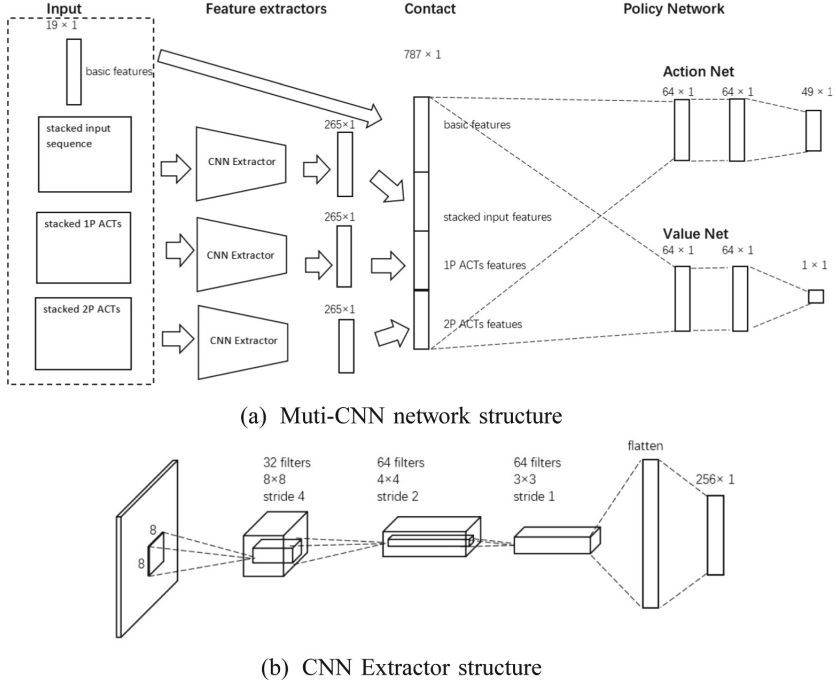


Fig. 5. Proposed network structure

3 Experiment

3.1 Training Process

The game’s difficulty level (Built-in AI level) was set to 8, which was the hardest setting. Built-in AI refers to a programmed action strategy that is used to fight against players in this game. Even for intermediate-level gamers, winning a level 8 built-in AI is quite challenging. On top of that, to simplify the training process, the game model was changed from team-play to single-play. The chosen RL Algorithm was Proximal Policy Optimization [6]. The model’s training was carried out on a personal computer with a 12-cores CUP and an RTX3060 GPU with 12 environments running in parallel.

3.2 Results

Multi-CNN Model and Transfer Learning. Each model was evaluated for 50 episodes. The first model’s agent operated the character Iori to fight against Kyo Kusanagi. As Table 1 shows, after 54 h of training, the first Multi-CNN model achieved a 100% win rate. A random agent was made as a negative baseline, and the win rate of the random agent is only 18%. Based on the first Multi-CNN model, using transfer learning, three more models were trained to fight with another three arbitrary chosen characters: Goro Daimon, Mai Shiranui, and Billy. After 30 M timesteps’ training, as Table 1 shows, the transfer learning models also achieved similar performance. As Fig. 6 shows, the transfer learning models’ performance increases more rapidly than the one learned from scratch, meaning that the experience learned from the previous model can be transferred to new models. The agents can defeat their counterparts in a short time and have lots of HP remaining. Besides, our agents are able to perform more combos and desperation moves than the random agent. Overall the Multi-CNN models are more than capable of winning the game by a considerable margin. Sample videos of the Multi-CNN model agents are available on Youtube².

Table 1. Test Result

Model	Transfer Learning	Opponent	Win Rate	Total Timesteps	Training Time	Mean Reward	Std Reward
Multi-CNN	No	Kyo Kusanagi	100%	50 M	54 h	157.4	48.2
Multi-CNN	Yes	Goro Daimon	100%	30 M	31 h	156.6	44.6
Multi-CNN	Yes	Mai Shiranui	100%	30 M	30 h	111.0	76.0
Multi-CNN	Yes	Billy	100%	30 M	33 h	151.0	86.0
LSTM	No	Kyo Kusanagi	46%	50 M	43 h	-27.6	105.5
Random	-	Kyo Kusanagi	18%	-	-	-140.2	83.4

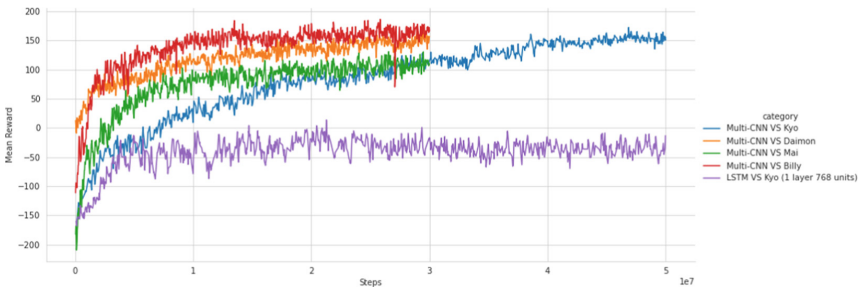


Fig. 6. Mean rewards of Multi-CNN models and LSTM model

² Sample videos: <https://youtu.be/v22Le-c9Uak>.

LSTM Network for Comparison. The LSTM network is well-suited to tackle tasks based on time series data since the cells can remember values over arbitrary time intervals. In order to illustrate our proposed network’s validity, an LSTM network was constructed for comparison. As Fig. 7 shows, the structure is similar to the Multi-CNN’s. The only difference is that the multiple CNN extractors are replaced with an LSTM module. As Table 1 shows, given the same budget (50 M timesteps), the LSTM model can only achieve a 46% win rate. As Fig. 6 shows (the purple line), at the first 5 M steps of training, the performance increases steadily, then it begins to fluctuate without apparent improvement. Whereas in the Multi-CNN model, the reward gradually increases as the step increases. Figure 8 shows that more LSTM models with different numbers of hidden layers and hidden units were trained for 10 M timesteps, but none of these models was promising. Another observation is that the training process dramatically slowed as the number of hidden layers and units increased.

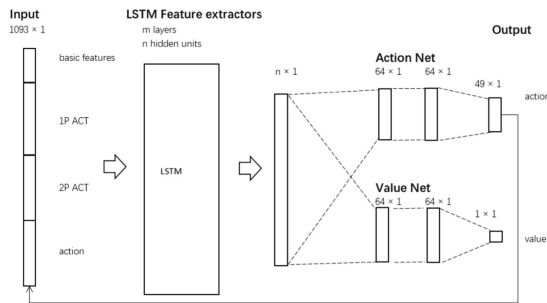


Fig. 7. LSTM network structure

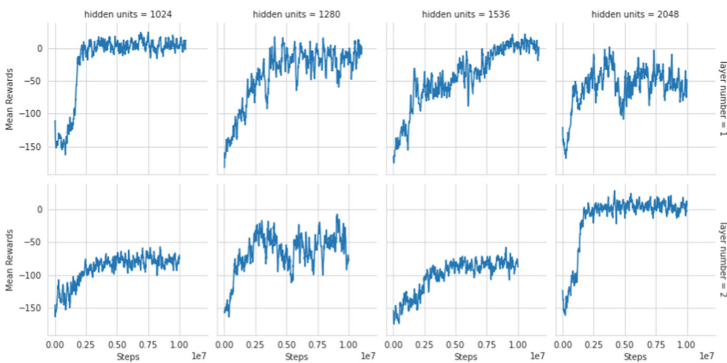


Fig. 8. Mean reward over timesteps of more LSTM models

3.3 Discussion

Even though our agent can perform some simple combos and defeat the built-in AI by a large margin, the agent failed to learn advanced combo skills. It is because performing advanced combos needs a long sequence of accurate inputs at the right time. Such opportunities are rare unless the player intentionally creates them. Furthermore, at each step, our agent has many actions to choose from, so purely by random exploration, it is almost impossible for our agents to get such an experience. Some studies, such as [2, 7], combined supervised learning and reinforcement learning to make the training process more effective. In the first stage, they used supervised learning to make the model imitate human operation, and in the second stage, they used RL to enhance their model. It would be a promising direction for further improvement.

Atari games are usually used as a testbed for RL algorithms, and the operation skill of such games is relatively low. Whereas in reality, manipulating things is often more complex and involves a sequence of operations with some patterns. The Recurrent Neural Network (RNN) is commonly used to process time series data. However, in such models, more recent data usually impact the results more, so it is not quite efficient for detecting behavioral patterns with some flexibility. The proposed graphical representation of sequences and network structure provides a new idea for detecting and simulating complex action patterns that are required by more practical tasks.

4 Conclusions

In this paper, a lot of valuable experience from the previous research is applied to our task: playing the arched fighting game KOF '97. Based on the features of this game and our goal, a novel graphical representation of the input sequence and characters' action sequence is proposed. The input pattern and characters' behavior pattern are well extracted by the proposed Multi-CNN network. Besides, a contribution to the RL community has been made by adding a new game environment of KOF '97, which is one of the most iconic arcade fighting game.

The experiments show that the agent can not only win the game by a large margin but also learns to perform some combos and desperation moves according to the situation. They also show that the experience learned from fighting against one character can be transferred to fighting against other characters via transfer learning.

References

1. Firoiu, V., Whitney, W.F., Tenenbaum, J.B.: Beating the world's best at super smash bros. with deep reinforcement learning. arXiv preprint [arXiv:1702.06230](https://arxiv.org/abs/1702.06230) (2017)

2. Kim, D.W., Park, S., Yang, S.: Mastering fighting game using deep reinforcement learning with self-play. In: 2020 IEEE Conference on Games (CoG), pp. 576–583. IEEE (2020)
3. Li, Y.J., Chang, H.Y., Lin, Y.J., Wu, P.W., Wang, Y.C.F.: Deep reinforcement learning for playing 2.5 d fighting games. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 3778–3782. IEEE (2018)
4. M-J-Murray: M-j-murray/mametoolkit: A python toolkit used to train reinforcement learning algorithms against arcade games. <https://github.com/M-J-Murray/MAMEToolkit>
5. Oh, I., Rho, S., Moon, S., Son, S., Lee, H., Chung, J.: Creating pro-level AI for a real-time fighting game using deep reinforcement learning. *IEEE Trans. Games* **14**, 212–220 (2021)
6. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint [arXiv:1707.06347](https://arxiv.org/abs/1707.06347) (2017)
7. Vinyals, O., et al.: Grandmaster level in starcraft II using multi-agent reinforcement learning. *Nature* **575**(7782), 350–354 (2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





AI4U: Modular Framework for AI Application Design

Kamil Wołoszyn¹ , K. Turchan¹ , M. Rapala² , and K. Piotrowski¹  

¹ IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder), Germany
{woloszyn,turchan,piotrowski}@ihp-microelectronics.com

² CBK PAN - Centrum Badan Kosmicznych PAN, Zielona Gora, Poland
mrapala@cbk.waw.pl

<https://www.ihp-microelectronics.com> , <https://cbkpan.pl>

Abstract. This paper presents the concepts of a universal, modular framework that shall enable rapid development of AI applications. The goal of the conceptual tool is the versatility in terms of changing the environment, as well as integration of different types of sensors to perform a specific task. The framework follows the Sens4U approach that will facilitate the entire process of building the AI applications and simplify the testing process.

Keywords: AI · Artificial Intelligence · AI Application · Framework

1 Motivation

Artificial intelligence (AI) has grown in popularity in recent years. AI has been applied in many scenarios, from intelligent assistants to advanced real-time vision systems. AI enables devices to make decisions based on data collected from cameras or various sensors. This data is processed and the device responds correspondingly. Artificial intelligence (AI) is a part of computer science domain which builds systems or machines that in specific tasks can imitate human intelligence. They are also able to constantly improve their performance according to collected data. Artificial intelligence is often associated with humanoid robots that are believed to replace humans, but their real purpose is to increase human capabilities instead [1]. Artificial intelligence is divided by the mechanism of complexity.

Artificial Weak/ Narrow Intelligence is the type we are using nowadays. Most of current AI systems can be classified as narrow artificial intelligence. All software using technologies such as natural language processing, Machine learning, pattern recognition, data mining is considered as narrow artificial intelligence. Narrow artificial intelligence focuses mainly on tasks in which human beings can be outperformed.

Artificial Strong/General Intelligence is the concept of an intelligent system with a great deal of knowledge and cognitive ability, which will be able to

independently perform tasks even those not yet known. It should think similarly to a person and with comparable or even better efficiency. At present, there is no machine that can understand the world as a human being can.

Artificial Super Intelligence is the most popular term circulating in the computer science world in recent times. An artificial super intelligence will hypothetically be able to interpret and to understand human behaviour and/or intelligence. It will be able to surpass human intelligence in mathematical tasks, science, medicine or applying solutions to different types of problems. What is more, artificial super intelligence is assumed to be able to evoke understanding, emotions or to be capable of having its own desires.

Artificial intelligence is an idea that integrates human intelligence with machines, as shown in Fig. 1. An issue related to artificial intelligence is ML (Machine Learning), which is one of the subcategories, that belongs to the area of artificial intelligence. Its main goal is to allow computers to learn using specific algorithms.

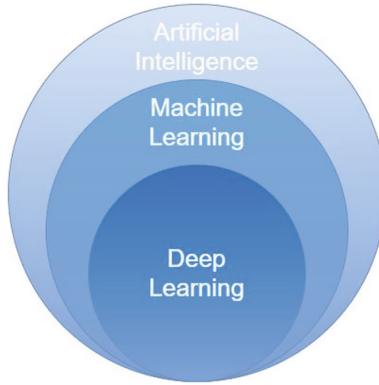


Fig. 1. Artificial intelligence structure.

This paper presents the concept of a universal, modular tool that will enable rapid construction of AI applications. The goal of this conceptual tool is its versatility in changing environment and different sensor types integration needed to perform a specific task. The concept of the proposed tool using the Sens4U approach [2], will improve the construction process, testability and reliability of the application. The idea of Sens4U is based on accelerating the application construction for non-experts in the field by using the modularity of available libraries.

Currently, there are many tools available for training deep neural networks, and many examples of software codes that support the sensors needed to create an intelligent robot or implementation of trained models. However, putting these together and making it fully function requires a large time investment.

When working on the SpaceRegion project, the initial focus was to build a stationary robot with no mobility capability. The robot is equipped with video cameras and servos to control the cameras, and its task was to observe and track detected objects. In the course of the work, it was decided that a mobile robot would be additionally built for the project. Initially the mobile robot was supposed to have the same functionality as the stationary robot, but during the code implementation to the mobile robot it turned out that most of the software code was the same, and only the code responsible for controlling the robot chassis has changed significantly.

This gave rise to the idea of building a universal tool that would speed up the process of developing AI applications. This approach can be also used to move one application from a particular scenario to another, for example a consumer solution to a space solution, without re-implementing the application from scratch. A universal tool must define application areas and requirements that simplify the migration process between scenarios. According to the Sens4U approach, the application area of the proposed tool will be expandable, by adding support for new sensors or updating those already implemented in the conceptual tool.

This paper is structured as follows. Section 2 discusses the related work, while Sect. 3 describes the proposed approach, followed by the example scenarios given in Sect. 4. Section 5 concludes the paper.

2 Related Work

At present, one can find systems that provide users with libraries and tools to help building robot applications, such as Robot Operating System (ROS) [3] or CubeSystem [4]. ROS is a meta operating system focused on building robots, which main goal is to support users in recycling of previously written code for robotics research or development purposes. ROS has a Gazebo [5] robot simulator that allows applications to be developed virtually, prior to the real robot, saving time and cost of modifications that might arise during construction. Another advantage of the ROS system is language versatility, thanks to which software can be written in such programming languages as Python [6], C++ [7], Lisp [8]. The big disadvantage of this system is the lack of multi-platform. The stable version work only on Unix [9] systems, but there are also experimental versions for Windows system.

ROS is open source software which allows the community to create new solutions. CubeSystem on the other hand is a collection of hardware and programmable components that enable rapid prototyping of robots. CubeSystem in combination with the RobLib library [10] allows customization and linking to a distinct set of libraries that are tailored to a specific application. For example, the RobLib library contains general functions for controlling robots based on a two-wheeled differential drive. It is generally difficult to find systems or sets of libraries that have standardized functionality for sensors of the same type. The operation complexity level of the application is raised by a number of additional

software, libraries and an adequate operation system that have to be installed. Fulling of such requirements can be problematic for a user unfamiliar with the field. Given the existing software and libraries, it seems reasonable to present the concept of a tool that is functional, modular, and cross-platform to support the construction of a complete application.

3 Proposed Approach

The concept of the tool for fast and modular construction of the application is shown in Fig. 2. The fact is almost every available sensor on the market has ready-made hardware drivers as well as dedicated software. The aim of the proposed tool is to create software, which will be a set of dedicated libraries, but any introduced modification in the conceptual tool will be applicable on different sensors. As a result, it will always return the same output, assuming that all the requirements specified during the construction of the conceptual tool are met.

The idea for the proposed tool emerged during working on the development of a computer vision application. The programming language used to build the application is Python version 3. Python is a high-level programming language, and thanks to its simplicity and versatility, it has become an ideal tool for building AI systems.

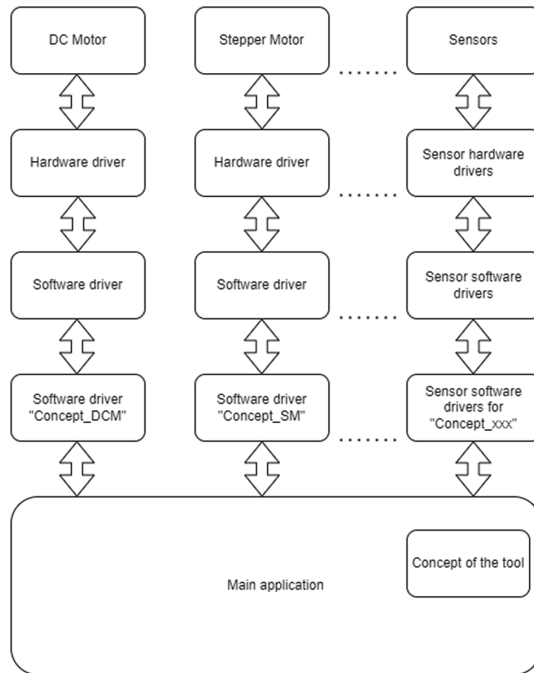


Fig. 2. Concept of the proposed tool.

The base AI application that was created for the project had to observe the environment, recognise people in the same room and check from time to time whether the person is still in the same place. Based on the people present in the room, the robot estimated in which office it was located. The basic application was initially built on a stationary robot, which is the QBO ONE [11] produced by Thecorpora as shown in Fig. 3.



Fig. 3. Robot QBO.

Further, the plan was also to investigate the possibility of a smooth transfer of the application samples to more demanding environments (like space). In order to support that, the application should fulfil its tasks when the original sensors and actors are replaced by devices that fit the new environmental conditions. The application should thus use generic application programming interface (API) to apply the functionalities of the devices and for instance adapt to their physical features (video resolution of a camera module or the resolution of a motor turn).

Replacing a device with another one does not change the application implementation although it can change its internal behaviour and features. The API equips the application with the control over the device and provides information about its features. The application developer who uses the conceptual tool will be able to obtain a list of available functions for a given sensor. When building

the tool, it would be necessary for the user to define the communication interface and to enter the parameters so it could work correctly.

For example, a user is building a robot with built-in DC motors. Using the API functionality of the conceptual tool, the user would be able to call, for example, the `getFeaturesDCmotor()` method. This method will return information to the user about parameters that should be entered and about available functions for DC motors e.g., `turnForward()`, `turnBackward()`, `stop()`.

Another functionality of the tool concept is its modularity, which aims to support the seamless modification of the project. As an example lets take the modification of the mentioned robot. The user replaces the standard DC motors with stepper motors, which brings a number of changes to the created main application. However, when using the proposed conceptual tool, the only modification that the user should provide, would be possibly a name change of the used motors. The basic functionality would be consistent for both motors and the program operation after the motors replacement would not be disturbed.

What's more, after modifications, calling the `getFeaturesSmotor()` method for stepper motors will display additional functions, besides `turnForward()`, `turnBackward()`, `stop()` stepper motors can also have functions like `turnForwardAboutAngle(x_angle)`, `turnBackwardAboutAngle(x_angle)` and `turnLeftAboutAngle(x_angle)`, `turnRightAboutAngle(x_angle)`, `setStartPosition()` while maintaining the requirements related to the motor mounting direction specified during the tool construction.

The proposed tool will allow expansion with new libraries depending on the users' needs. The main advantage of the conceptual tool would be reliability to environmental changes, and unification of functionality for sensors of the same type, what would result in rapid modifications in the design. During the construction of the tool, it will be possible to extend its operation to support different types of sensors. Moreover, sample functionalities such as object detection on the provided video stream or real-time object tracking from trained deep learning models can be implemented in the tool for artificial intelligence purposes.

For devices with low processing power, the proposed conceptual tool will be able to implement functionality to support AI accelerators. The AI accelerators will enable the satisfactory performance of artificial neural networks combined with computer vision. AI accelerator is, for example, the Intel Neural Compute Stick 2 [12]. One advantage of the proposed tool will be its cross-platform capability, which would allow the tool to run on both Microsoft Windows and Linux platforms.

4 The Application Area

This section presents examples of scenarios from which requirements for the correct operation of the proposed tool may arise. Some of the scenarios have been implemented, some are in the process of being implemented and some may be created in the future for development, experimental or research purposes during the development of the proposed conceptual tool. While creating the application,

scenarios based on terrestrial solutions were used for testing reasons and finding available solutions. However, the conceptual tool also envisions applications in extraterrestrial scenarios.

4.1 Watchman Scenario

The first developed scenario is the Watchman scenario. The main task of this scenario is to monitor a specific room and control people authorized to be in it. The system based on optical sensor and machine learning must be able to detect, recognize and perform verification of a person present in the room. The next step that the created system must take is to react during the detection of a possible intruder. During the detection of an intruder, the task of the system is to take a picture of a person who is not authorized to be in the room and record the event in the database for further verification by the administrator. The system must be adapted to expand the database of people authorized to stay in the room.

4.2 Parkmonitor Scenario

The second scenario is the Parkmonitor scenario. The main task of this scenario is to monitor a specific area and count free and occupied parking spaces. System using optical sensor and machine learning must be able to detect and return information about number of free and occupied parking spaces.

4.3 Tracker Scenario

The third scenario is the Tracker scenario. The main task in this scenario is to track a specific object by an unmanned craft (drone). For example, such an object can be a ball, a cyclist, a person or a car. In general, any model of object or objects that can be trained using neural networks suit this scenario. The system, based on the optical sensor and machine learning, after activating the function, must be able to follow a specific object until it disappears from the frame or the function is disabled. The drone operator must be able to activate and deactivate the tracking function.

4.4 Mapping the Environment Scenario

The fourth scenario is the Mapping the Environment scenario. The main task in this scenario is to build a map of the area on which the mobile robot moves. Created system has to enable real preview of the built map and must have the ability to export the created map of space/area. Moreover, the system using an optical sensor and machine learning has to be able to mark certain objects on the built map. Additionally, the system must have the possibility to choose the option of controlling the vehicle. The first option is to drive autonomously to build a map of the room. The second option after building the map is to drive

to a specific point on the map and the last required option is manual control of the vehicle. A prototype of the mobile vehicle under development is shown in Fig. 4.

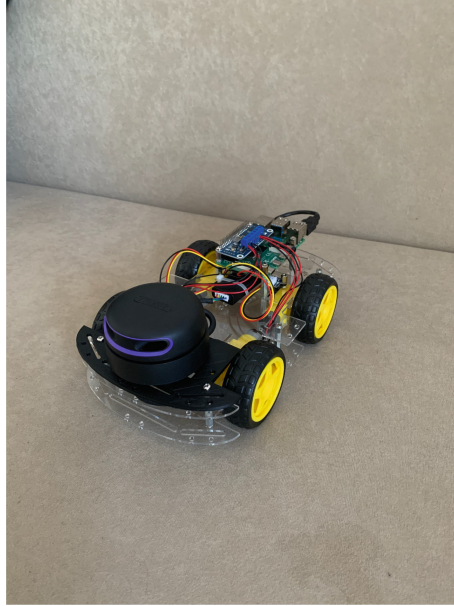


Fig. 4. Mobile vehicle prototype.

4.5 Vehicle Counting

The fifth is the scenario of Vehicle counting. The main task of this scenario is to build a system capable of counting vehicles. Vehicles that will be counted are wheeled vehicles such as: cars, trucks and motorcycles. Counting of vehicles has to be done separately in two directions for vehicles approaching and moving away from the device. Additionally, the device on which the created system will operate must be resistant to external weather conditions and it should be as small as possible. Prototype of vehicle counting device under development is shown in Fig. 5. After installing and testing the device, the system will transmit the vehicle counting results via a network to the middleware. Vehicle counting device prototype consists of Raspberry Pi 4, Raspberry Pi High Quality Camera, 6mm 3MP Wide-Angle Lens and Intel Neural Compute Stick 2.

4.6 Space Surveyer

The sixth scenario is the space scenario of the Space Surveyer. The main task in this scenario is for the mobile robot to explore the terrain and determine the landing site of the lander. The created system must allow communication



Fig. 5. Vehicle counting device prototype.

between the lander and the mobile robot to exchange information while exploring the terrain. The mobile robot after exploring the terrain must be able to determine the best landing place and send information about it to the lander. The site selection will be determined by the detected obstacles, hills or hardness of the ground. In case the communication between the platforms fails, the mobile robot will be equipped with a light system which will allow the lander to start the landing procedure after detection. The light signal from the mobile robot will determine the center of the selected landing area.

4.7 Mission to Mars

The seventh scenario is the space scenario of Mission to Mars. The main task in this scenario is the autopilot function for the spacecraft. The developed system using optical sensors and machine learning has to be able to track the planet Mars and enable and disable the autopilot function. When the autopilot function is enabled, the system will be able to control and correct the flight to the designated planet.

4.8 First Conclusions

The above scenarios, while each different, have something in common. Namely, each of them contains optical sensors and machine learning. So, building one system with the previously described modular approach will reduce the time needed to build the next system. A user or even a developer involved in building AI systems, after introducing a conceptual tool will not have to write a consecutive system from scratch. The gained time can be spent on the process of building and training the model, which is time consuming.

In the first scenario, the previously mentioned QBO platform was used to build a watchman application capable of monitoring a room. This platform, thanks to the installed servos, allowed searching for objects around the room.

The built-in VGA cameras were used for object detection. For image processing, a programming function library was used, which is OpenCV [13]. It is a library mainly oriented to work with computer vision in real time. This library has the advantage of supporting hardware acceleration and the ability to load machine learning models trained on various available development platforms such as Tensorflow [14], PyTorch [15], Cafe [16] and many others. However, for proper real-time operation, the QBO platform was modified. The Raspberry Pi 3 computer platform was replaced with a Raspberry Pi 4 and due to limited computing capabilities, an AI gas pedal such as Intel Neural Compute Stick 2 was added to the computer platform.

After the modifications, the AI application created in conjunction with the QBO platform enabled seamless real-time monitoring of the room and verification of people from the created photo database. Having built a system from the first scenario, it is easy to build another system, for example for the fifth scenario. Thanks to the modular approach to building applications, the module responsible for the detection of objects from the first scenario can be fully applied in the fifth scenario.

It is enough to change the trained model from people to vehicles and replace part of the code responsible for controlling the camera with a new code created to count the occurring vehicles. The same is true for second scenario. A minor modification of the code and a change of the trained model will allow building a parking lot monitoring application without much effort. In the first scenario, after replacing the module responsible for controlling servos with a module for controlling DC motors, the functionality can be extended to run on a mobile platform without much effort.

5 Conclusions and Further Work

The paper presents the concept of the tool, which is still under development and may change during the implementation process when new problems and challenges are discovered within the SpaceRegion project. Different scenarios based on terrestrial and space solutions are currently developed and implemented. Based on the presented scenarios, the capabilities and requirements of the proposed tool will be adjusted. We will also analyse and compare implemented code for different scenarios in order to find common parts, which will be eventually transferred to the proposed tool. A library will be created consisting of several sensors of the same type in order to test the behaviour of the application during its modification.

Furthermore, sample applications related to artificial intelligence, robotics and computer vision using the conceptual tool will be created. In addition, an analysis of the application implementing modifications under scenario change, will be performed. The main goal of the proposed conceptual tool is to simplify the construction of AI applications.

Acknowledgments. This work was supported by the European Regional Development Fund within the BB-PL INTERREG V A 2014–2020 Programme, “reducing barriers - using the common strengths”, project SpaceRegion, grant number 85038043. The

funding institutions had no role in the design of the study, the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Jajal, T.D.: Distinguishing between Narrow AI, General AI and Super AI. <https://medium.com/mapping-out-2050/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22>. Accessed 21 Apr 2022
2. Piotrowski, K., Peter, St.: Sens4U: wireless sensor network applications for environment monitoring made easy. In: Proceedings of the 4th International Workshop on Software (2013)
3. ROS. ROS-Robot Operating System. <http://wiki.ros.org/ROS/Introduction>. Accessed 22 Mar 2022
4. Birk, A.: Fast robot prototyping with the cubesystem. In: Proceedings of the IEEE International Conference on Robotics and Automation, ICRA 2004 (2004)
5. GAZEBO. <http://gazebo.org/>. Accessed 22 Apr 2022
6. Python. Python™. <https://www.python.org/>. Accessed 22 Mar 2022
7. C++. <https://www.cplusplus.com/>. Accessed 22 Apr 2022
8. Lisp. <https://lisp-lang.org/wiki/>. Accessed 22 Apr 2022
9. Unix. <https://www.hpc.iastate.edu/guides/unix-introducti-on>. Accessed 22 Apr 2022
10. RobLib. <http://robotics.jacobsuniversity.de/-CubeSystem-em>. Accessed 22 Apr 2022
11. Thecorpora Robotic Company, QBO ONE. <https://thecorp-ora.com>. Accessed 07 Apr 2022
12. Intel Corporation. Intel®Neural Compute Stick 2 (Intel®NCS2). <https://www.intel.com/content/www/us/en/dev-eloper/tools/neural-compute-stick/overview.html>. Accessed 22 Mar 2022
13. OpenCV. <https://opencv.org/>. Accessed 04 Apr 2022
14. Tensorflow. <https://www.tensorflow.org/>. Accessed 21 Apr 2022
15. PyTorch. <https://pytorch.org/>. Accessed 21 Apr 2022
16. Café. <https://caffe.berkeleyvision.org/>. Accessed 21 Apr 2021


Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





A Competent Deep Learning Model to Detect COVID-19 Using Chest CT Images

Somenath Chakraborty¹ (✉)  and Beddhu Murali²

¹ West Virginia University Institute of Technology, Beckley, WV 25801, USA
Somenath@ieee.org

² University of North Texas, Denton, TX 76203, USA
Beddhu.murali@unt.edu

Abstract. The ongoing terrifying wave of COVID-19 which is caused by the Severe Acute Respiratory Syndrome Coronavirus 2 or SARS-CoV-2 continues to affect human life and disrupt healthcare systems. Chest computed tomography (CT) is an effective clinical tool for estimating the patient's severity levels and deciding appropriate treatment regimes. In this paper, we use a deep learning method for detecting COVID-19 using chest CT images with the more advanced balanced dataset. We used a dataset of 8054 real patient CT scans, of which 5427 had COVID-19 and 4223 were Non-COVID-19 patient images. Our model had an average detection accuracy of 91.96% on the test dataset. In conclusion, Automated Deep Learning (DL) methodologies allow for speedy evaluation of CT images to detect COVID-19.

Keywords: COVID-19 · Severe Acute Respiratory Syndrome Coronavirus 2 · SARS-CoV-2 · computed tomography (CT) · Deep Learning (DL) · healthcare system

1 Introduction

The ubiquitous COVID-19 pandemic is the distinguishing tragedy of the present century thus far. To combat this battle, clinicians and radiologists work together and invented a couple of different diagnostic tools. COVID-19 reportedly started in Wuhan City of China in December 2019 [1–3]. According to the Centers for Disease Control and Prevention (CDC), the total number of cases and deaths as of 8th May 2022 is 517 million and 6.25 million, respectively worldwide. That is why this research has a significant value from the social perspective as it helps faster the processes and automate the tasks rapidly and build more confidence in the medical fraternity. The numbers continue to increase with the spread of the Omicron coronavirus variant (B.1.1.529). COVID-19 is a disease that initially affects the respiratory organs, like the lungs [1]. As it is a highly transmittable viral infection without a cure, people are dying all over the world and many severely suffering patients do not find hospital accommodation due to overloading with the existing COVID-19 patients. The Critical Care facility is also overloaded and the situation is overwhelming.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 67–75, 2023.

https://doi.org/10.1007/978-3-031-37649-8_7

The diagnosis of COVID-19 is one of the crucial aspects of the fight against this disease. In the early days of the pandemic, the detection was very limited and error-prone due to a lack of understanding of the virus genome, but researchers came up with the Reverse Transcription Polymerase Chain Reaction (RT-PCR) [4] pathological method, which is still a gold standard for the detection of COVID-19. However, many challenges still remain in the COVID-19 detection research. The RT-PCR sensitivity rate is 60%–70%, which is very low. Rapid antigen testing is also a fast method to detect COVID-19 patients but due to the different types of variants coming up every now and then it is also not so effective as lots of True negatives coming in the test due to the mutation of the virus genome.

Recently, many automated machine learning approaches have been introduced in the literature for Covid-19 detection such as Chest CXR, Chest CT, etc.

In this paper, we present a very efficient deep learning model for analyzing CT images.

2 Literature Review

As the situation in early 2020 had already started deteriorating and new cases were growing exponentially, The World Health Organization (WHO), declared it a pandemic on 11th March 2020 [5]. WHO distinguishes the variants of coronavirus into two categories, one is a ‘variant of interest’ and the other one is a ‘variant of concern’. According to the transmissibility and detrimental effects, the variant of concern is more dangerous compared to the variant of interest.

He, X. et al. [6] present a deep learning diagnosis method using chest CT images. As their work was reported in early April 2020, their COVID-19 chest CT image dataset was limited. Zhang, K. et al. [7] proposed an AI System for diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. As their research was published in early June of 2020 they used the idea of image slices to increase the size of sample data as only 752 COVID-19 positive patient data was included in their work. Gunraj, H. et al. [8] proposed a tailored deep convolutional neural network (CNN) using chest CT images. They generated a dataset of 104,009 chest CT slices across 1,489 patient cases. Harmon, S. A. et al. [9] present a deep neural network using a multinational dataset. M. Khushi et al. [10] and Maillo, Triguero and Herrera [11] show that having limited positive samples in a dataset could lead to data inconsistency and data imbalance issues. Panwar, H. et al. [12] proposes a grad-CAM-based color visualization in their deep learning approach using CT images.

3 Our Proposed Methodology

Deep Convolutional Neural Network (CNN) is one of the most widely used neural networks for image classification problems [13]. CNN in a deep learning setting has been widely applied in the literature, for example, [14–18]. With the availability of different deep learning networks such as ResNet18, Res-Net50 [19], AlexNet [20], DenseNet [21], VGG16 [22], EfficientNet [24], etc., one can design many automated machine learning models that can detect COVID-19 from CXR, CT images, or a combination of both. In

our approach, we designed our own classification layer but leveraged the advantages of EfficientNet in multi-feature generation layers through the transfer learning technique. It helps to generate lower-level and higher-level features in a rapid way. Iterative multiple training helps to optimize parameters and hyperparameters with rapid succession due to this integrated deep multi-layer approach.

3.1 Dataset Description and Preprocessing

The data consisted of chest CT images divided into two categories. COVID-19 infected patients and non-COVID-19 patients. We used multiple sources for data acquisition [9, 12–15] as the opensource COVID-19 data increases over time. These images are labeled by licensed radiologists and freely available in open-source platforms like GITHUB repositories and Kaggle platforms. Once we obtained the CT images, our next step was to preprocess those images with proper annotation and labeling, including normalization and augmentation using machine learning models. We use Pytorch vision models [23], to flip, rotate and include those images to increase the size of the dataset, which basically helps the deep learning model to be more robust and accurate. The deep neural network models derives feature information in a very different way they human radiologist visually inspect the images. Initially, some sort of histogram equalizing and contrast-enhancing.

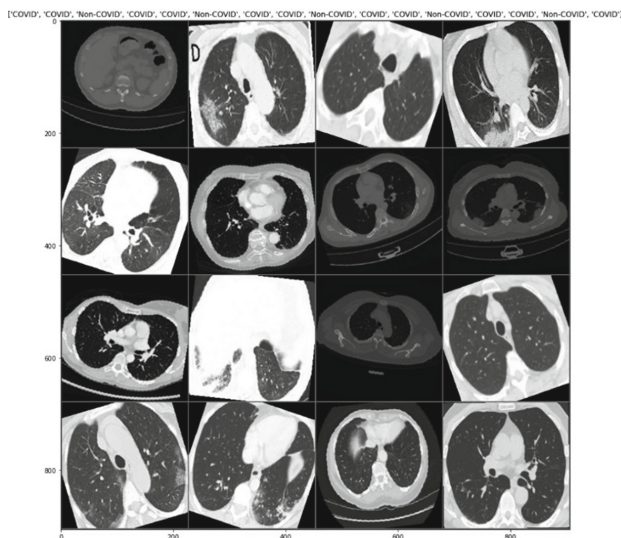


Fig. 1. Chest CT images after preprocessing.

Our proposed model used a dataset of 8054 real patient CT scans, of which 5427 had COVID-19 and 4223 were Non-COVID-19 patient images. The dataset is further divided into train, validation, and test sets with the ratio of 7:2:1. Figure 1 shows the chest CT images after preprocessing. Our dataset is very well balanced thus does not

incurred data imbalance issues. Data imbalance is a very common problem, especially happened in medical data and image analysis if the model used fewer number of positive sample cases than the negative sample cases.

3.2 Methodology

The Fig. 2. It shows the block diagram representation of our proposed model. EfficientNet [24], pretrained weightages helps to reduce the time of our training process as due to transfer learning technique of EfficientNet we used the hyperparameter values in our model.

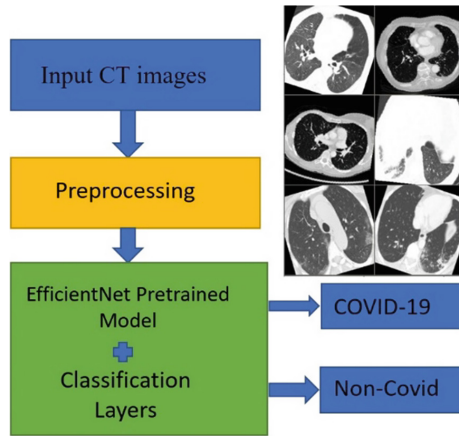


Fig. 2. Abstract Block Diagram Representation of our Proposed Model

After preprocessing the images we used EfficientNet [24], a pretrained Model, for the feature generation tasks and then we add a deep classification layer to classify the CT images into COVID-19 position images and COVID-19 negative images, i.e. Non-COVID-19 images. After execution of the feature generation tasks, the total number of parameters of our dataset is 18,830,011, out of which Trainable parameters are 1,281,395 and Non-trainable parameters are 17,548,616.

The architecture of EfficientNet baseline Neural Network is shown in Fig. 3.

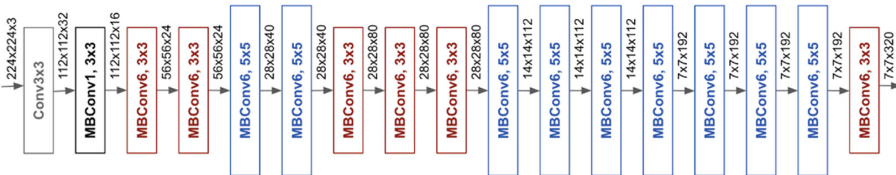


Fig. 3. EfficientNet B0 Baseline neural network Architecture

EfficientNet is a very efficient high-performance convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth, width and resolution using a compound coefficient. Where benchmark deep neural models arbitrarily scale these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients. The effectiveness of model scaling is based on the baseline network which is playing a key role in high performance. The baseline network builds up with a deep neural architecture search using the AutoML MNAS framework [25], which optimizes both accuracy and efficiency. After leveraging the potential of transfer learning by using the deep learning model EfficientNet in the feature generation architecture we design our deep learning classification architecture. The elements of the architecture is given in the Fig. 4. We use piecewise linear function rectified linear unit in our model as an activation function instead of Softmax or other activation function as we are interested in getting probabilistic value from 0 to 1.

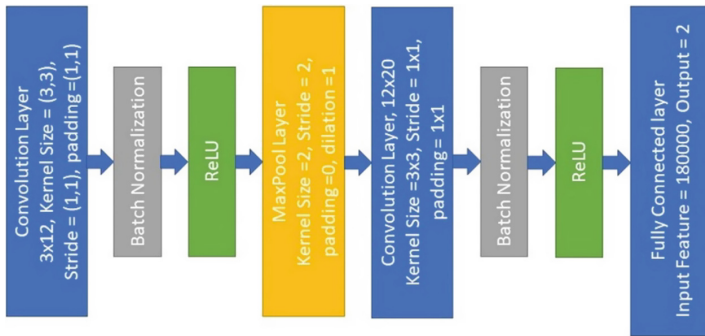


Fig. 4. Classification Architecture of our Deep Learning Model

The classification layers provide the percentage of the score for COVID-19 patients and Non-COVID-19 patients.

4 Results

The proposed deep learning model had an average detection accuracy of 91.96% on the test CT images. Sensitivity is 92.24, Specificity is 93.01 and F1 Score is 0.90. Figure 5 shows the Loss Curve. We used three fold cross validation technique and after the 20 epoch as the values converge to the optimum values so no further training of the model needed, We heuristically checked this is the optimum level of the loss in our experiment while running with many epoch heuristically to understand the optimum level. Hence, we only included the optimum value figures upto 20 epoch. From the graph, in Fig. 5a, it is evident the loss decreased over the epoch. Figure 5.b shows the accuracy curve. From the graph, it is evident that the accuracy improves over the epoch.

Figure 6a, b, c are the results coming from an unknown test sample and then matching their value for the predicted score obtained using our proposed model. It is clearly evident from the Fig. 6 that the predicted score clearly confirmed the type of the class with very high accuracy.

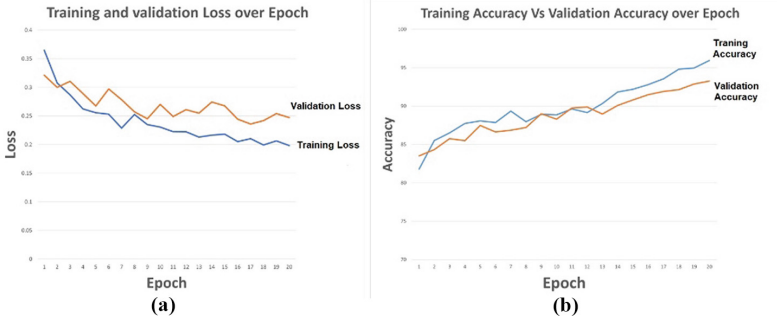


Fig. 5. a) Training and Validation Loss b) Training Accuracy Vs Validation Accuracy

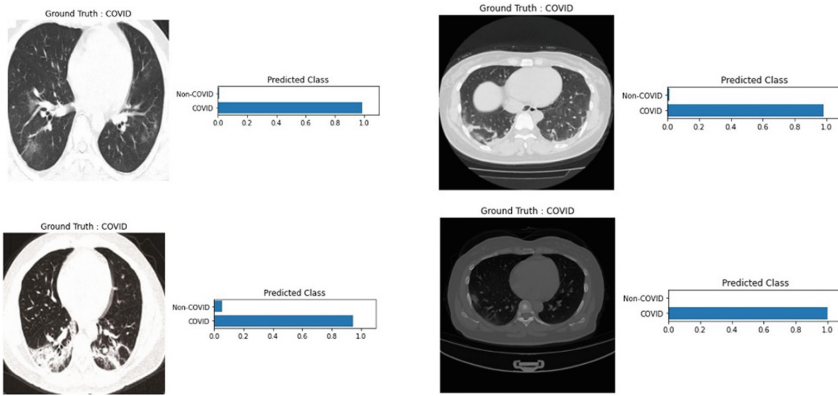


Fig. 6. Predicted score matching with the ground truth values for different test images.

The following Table 1 presents the comparison of the proposed model with other existing models using our dataset with the model they described in their papers.

Table 1. Comparison Analysis with Existing Models

Model name	Accuracy	Sensitivity	Specificity	F1 score
Proposed Model	91.96	92.24	93.01	0.90
He, X. et al. [6]	86.45	87.02	87.54	0.86
Gunraj, H. et al. [7]	87.21	88.82	87.98	0.87
Zhang, K. et al. [10]	88.67	89.52	89.87	0.88
Harmon, S. A. et al. [11]	87.36	89.54	89.98	0.89
Panwar, H. et al. [12]	90.14	89.85	90.51	0.88

5 Future Work and Conclusion

The Deep learning Model is not intended to replace the RT-PCR or other bio-chemist methods rather it would help to give extra confidence to the doctors to detect COVID-19 with more accuracy, hence helps to reduce wrong diagnosis. Many earlier models reported in the literature utilized a limited number of COVID-19 positive images due to the lack of availability of such data during the primary phases of the pandemic. In the present work, 8054 CT images were used with balanced positive and negative cases. The results show that this work which is based on EfficientNet is capable of detecting COVID-19 based on CT images with high accuracy. This adds to the growing body of evidence that deep learning techniques may be used in a clinical setting with confidence for detecting COVID-19. One of the recommendations of WHO is to increase testing for COVID-19. Usage of deep learning-based methods would bring down the cost and accelerate the diagnosis process which will enable rapid testing on a global scale. In our future work, we would investigate vision transformer models with this dataset. Transformer architecture performs very well in Natural Language Processing (NLP) tasks and is widely used in NLP domains. But recently many research works comes up with vision applications as well, though Deep Convolutional Neural Network is mostly preferred by most of the researchers in computer vision applications, especially in image processing domains, object detection domains, etc. We also like to extend our work on our classification layers with more variety of activation functions and loss functions.

References

1. Huang, C., et al.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**(10223), 497–506 (2020)
2. Wu, F., et al.: A new coronavirus associated with human respiratory disease in China. *Nature* **579**(7798), 265–269 (2020)
3. McIntosh, K.: Coronavirus Disease 2019 (COVID-19): Epidemiology, Virology, Clinical Features, Diagnosis, and Prevention (2020)
4. Wang, W., et al.: Detection of SARS-cov-2 in different types of clinical specimens. *JAMA* **323**(18), 1843–1844 (2020)
5. World Health Organization: WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19, 11 March 2020 (2020)
6. He, X., et al.: Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. *Health Informatics* (2020, Preprint). <https://doi.org/10.1101/2020.04.13.20063941>
7. Zhang, K., et al.: Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **181**, 1423–1433 (2020)
8. Gunraj, H., Wang, L., Wong, A.: COVIDNet-CT: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images. *Front. Med. (Lausanne)* **7**, 608525 (2020). <https://doi.org/10.3389/fmed.2020.608525>
9. Harmon, S.A., et al.: Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat. Commun.* **11**, 1–7 (2020)
10. Khushi, M., et al.: A comparative performance analysis of data resampling methods on imbalance medical data. *IEEE Access* **9**, 109960–109975 (2021). <https://doi.org/10.1109/ACCESS.2021.3102399>

11. Maillou, J., Triguero, I., Herrera, F.: Redundancy and complexity metrics for big data classification: towards smart data. *IEEE Access* **8**, 87918–87928 (2020). <https://doi.org/10.1109/ACCESS.2020.2991800>
12. Panwar, H., et al.: A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. *Chaos Solitons Fractals* **140**, 110190 (2020). <https://doi.org/10.1016/j.chaos.2020.110190>
13. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*, pp. 253–256. [5537907] (ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems) (2010). <https://doi.org/10.1109/ISCAS.2010.5537907>
14. Chakraborty, S., Zhang, C.: Survival prediction model of renal transplantation using deep neural network. In: *2020 IEEE 1st International Conference for Convergence in Engineering (ICCE)*, pp. 180–183 (2020). <https://doi.org/10.1109/ICCE50343.2020.9290695>
15. Chakraborty, S., Murali, B.: A novel medical prognosis system for breast cancer. In: Mandal, J.K., Buyya, R., De, D. (eds.) *Proceedings of International Conference on Advanced Computing Applications*. AISC, vol. 1406, pp. 403–413. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-5207-3_34
16. Chakraborty, S.: Category identification technique by a semantic feature generation algorithm. In: *Deep Learning for Internet of Things Infrastructure*, pp. 129–144. CRC Press
17. Chakraborty, S., Bandyopadhyay, S.K.: Scene text detection using modified histogram of oriented gradient approach. *IJAR* **2**(7), 795–798 (2016)
18. Chakraborty, S., Murali, B.: Investigate the correlation of breast cancer dataset using different clustering technique. *ArXiv abs/2109.01538*, n. pag. (2021)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Pattern Recognition (CVPR), pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017). <https://doi.org/10.1145/3065386>
21. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269 (2017). <https://doi.org/10.1109/CVPR.2017.243>
22. Zhang, X., Zou, J., He, K., Sun, J.: Accelerating very deep convolutional networks for classification and detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(10), 1943–1955 (2016). <https://doi.org/10.1109/TPAMI.2015.2502579>
23. PyTorch Vision Models. <https://pytorch.org/vision/stable/models.html>
24. Tan, M., Le, Q.V.: EfficientNet: rethinking model scaling for convolutional neural networks. *ArXiv, abs/1905.11946* (2019)
25. Tan, M., Chen, B., Pang, R., Vasudevan, V., Le, Q.V.: MnasNet: platform-aware neural architecture search for mobile. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2815–2823 (2019)
26. Chakraborty, S., Murali, B., Mitra, A.K.: An efficient deep learning model to detect COVID-19 using chest X-ray images. *Int. J. Environ. Res. Public Health* **19**, 2013 (2022). <https://doi.org/10.3390/ijerph19042013>
27. Chakraborty, S., Ali, D., Murali, B.: A novel distributed database architectural model for mobile cloud computing. In: Mandal, J.K., Hsiung, P.A., Sankar Dhar, R. (eds.) *ICCTA 2021. LNNS*, vol. 426, pp. 155–161. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-0745-6_17








Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





AI in Prostate MRI Analysis: A Short, Subjective Review of Potential, Status, Urgent Challenges, and Future Directions

Rafał Józwiak^{1,2} , Ihor Mykhalevych^{1,2} , Iryna Gorbenko¹ ,
Piotr Sobecki¹ , Jakub Mitura¹ , Tomasz Lorenc³ ,
and Krzysztof Tupikowski⁴ 

¹ Laboratory of Applied Artificial Intelligence, National Information Processing Institute, Warsaw, Poland
Rafał.Jozwiak@opi.org.pl

² Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

³ I Department of Clinical Radiology, Medical University of Warsaw, Warsaw, Poland

⁴ Lower Silesian Oncology, Pulmonology and Hematology Center, Wrocław, Poland

Abstract. Artificial intelligence (AI) in prostate MRI analysis shows great promise and impressive performance. A large number of studies present the usefulness of AI models in tasks such as prostate segmentation, lesion detection, and the classification and stratification of a cancer’s aggressiveness. This article presents a subjective critical review of AI in prostate MRI analysis. It discusses both the technology’s current state and its most recent advances, as well as its challenges. The article then presents opportunities in the context of ongoing research, which possesses the potential to reduce bias and to be applied in clinical settings.

Keywords: artificial intelligence · deep learning · image analysis · magnetic resonance imaging · prostate cancer

1 Introduction

Prostate cancer (PCa) is the second most common cancer in men, with almost 1.4 million new cases diagnosed per year worldwide [22]. With the acceleration of the industrialization process and the impact of environmental pollution, the incidence of PCa—caused by enriched foods, smoking, and excessive alcohol use—continues to increase at a rate of 6.63% per year. Early detection of PCa can improve patients’ prognoses considerably. Recent advances in MRI technology that allow both anatomical and functional imaging to be performed simultaneously, mpMRI, have improved our ability to detect and characterise prostate tumors [16]. According to patient management guidelines, noninvasive diagnostic tools such as mpMRI play an important role in the referral of patients to

active surveillance, watchful waiting, and active treatment [15, 24]. The Prostate Imaging Reporting and Data System (PI-RADS) was introduced by the European Society of Urogenital Radiology (ESUR) in 2012 to standardize prostate mpMRI examination protocols and the reporting of suspicious lesions (providing standardized terminology and sector map-based locations). The PI-RADS system categorizes prostate lesions based on the likelihood of cancer according to a five-point scale. PI-RADS was developed by a consensus-based process that uses a combination of published data, and expert observations and opinions [20]. The clinical utility of PI-RADS scoring is growing, and several studies have confirmed that the system improves the diagnostic accuracy of mpMRI [12].

The definitive diagnosis of PCa depends on the recognition of cancer cells in a tissue biopsy. Based on histological tumor architecture, a Gleason classification system was proposed. The Gleason score of biopsy-detected PCa comprises the Gleason grade of the most extensive (primary) pattern, plus that of the second most common (secondary) pattern, and ranges from one to five [11]. The “clinically significant” (csPCa) descriptor is used widely to differentiate PCa types that cause morbidity or death from those that do not. Defining what is clinically significant and what is insignificant PCa (iPCa) is challenging. According to the literature, iPCas do not typically cause harm and are at high risk of being overtreated, with the treatment itself risking harmful side effects to patients [4].

In recent years, a large number of review articles that concern the application of artificial intelligence (AI) in prostate cancer diagnosis have been published [7–9, 17, 21]. They discuss various aspects of AI application in PCa, which concern not only to mpMRI image analysis, but also ultrasound image analysis, histopathology image analysis, MRI-ultrasound fusion, MRI-histopathology registration, and clinical outcome predictions. In 2022 alone various different review articles were published. Li et al. [13] presented an extensive review over a long period that studied the applications of machine learning (ML) and deep learning (DL) in prostate MRI segmentation, registration, lesion detection and scoring, and treatment decision support. Sushentsev et al. [19] analyzed two classes of AI method: DL and traditional machine learning (TML), demonstrating their comparable performance in the differentiation of csPCa/iPCa, as well as discovering common methodological limitations. According to the authors, consensus on datasets, segmentation, ground truth assessment, and model evaluation remains to be established. The narrative review of Belue and Turkbey [5] introduced emerging medical imaging AI paper quality metrics, such as the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) and Field-Weighted Citation Impact (FWCI), and applied those analyses to the top AI models for segmentation, detection, and classification of PCa—including potential areas of impact in radiologists’ workflow. Although those methods are commonly reported in the literature with promising results, the authors concluded that prospective multicenter studies are necessary to determine the impact of AI on improving radiologists’ performance. Sunoqrot et al. [18] provided an interesting review that focuses on open datasets, commercially/publicly available AI, and grand challenges. The authors concluded that well curated public datasets are avail-

able, but are relatively small and vary in quality. Computational AI challenges are needed to deliver independent validation and to build trust in AI for prostate MRI.

This article presents a subjective critical review of AI in prostate MRI analysis. It considers the most recent advances, challenges, and opportunities presented in the context of the ongoing project, *AI-augmented radiology - detection, reporting and clinical decision making in prostate cancer diagnosis* (INFOS-TRATEG) being conducted at the Laboratory of Applied Artificial Intelligence of the National Information Processing Institute in Poland.

2 The Potential of Artificial Intelligence in MpMRI Analysis

Current clinical practice and guidelines utilize mpMRI prior to biopsies to identify potentially suspicious lesions. Radiological interpretation, together with relevant clinical information, supports clinicians in proper patient management, which remains crucial in light of the high prevalence of PCa and its low mortality rate. Many patients with no cancer or with indolent cancer can benefit from long-term active surveillance and lower numbers of unnecessary biopsies; this, in turn, minimizes the occurrence of unnecessary side-effects, such as pain, bleeding, and infection. Despite continuous improvement in MRI technique, interpretation of prostate MRI remains challenging and is generally recognized to present a steep learning curve [13]. Low specificity and high interobserver variability remain problematic disadvantages of MRI—particularly for nondedicated or less experienced radiologists, who have received only short term training in prostate MRI [23]. At present the processing and interpretation of prostate mpMRI data in clinical routine is performed chiefly by human experts; it remains highly subjective and strongly dependent on experience and training. Improving the PCa diagnostic pathway (and potentially reducing overdiagnosis of iPCa and underdiagnosis of csPCa) is a key challenge. AI techniques may support the radiological workflow of PCa diagnosis and reduce interobserver variability among radiologists. This enables more consistent diagnoses by clinicians, which can result in improved patient outcomes. AI models utilize the quantitative nature of imaging data to construct a more robust feature space based on mpMRI representation. AI models can help in cancer diagnosis by facilitating ancillary tasks in cancer detection that are labor- and experience intensive, such as prostate gland segmentation, PCa detection on mpMRI images, and characterization of a cancer’s advancement and aggressiveness [6].

2.1 Prostate Segmentation

Prostate gland segmentation aims to outline the whole gland boundary, as well as its zonal division. This is critical for the calculation of the entire prostate volume and of the serum prostate-specific antigen density, which are important

PcA biomarkers. Manual segmentation of the prostate and its zones is a time-consuming and tedious task. It is also highly subjective and dependent on the experience of the radiologist. Prostate gland segmentation is used commonly in clinical practise estimation of prostate volume based on the ellipsoid formula, as it is easy to apply, highly time-efficient, and is characterized by high interobserver agreement; however, it offers only an approximation of a reality, which, in many cases, is much more complex. AI/DL methods have high potential to reduce the time and variability associated with prostate gland segmentation on MRI.

In [3], the authors propose a segmentation pipeline that comprises three convolutional neural networks. The first localizes the prostate by creating a bounding box. The second completes prostate gland segmentation by classifying each pixel as belonging either to the prostate or to the background. The third differentiates between the transition zone and the peripheral zone by classifying every voxel in the image as one of these two classes. Each of the convolutional neural networks was implemented using a customized hybrid three-dimensional/two-dimensional (3D/2D) U-Net architecture. The model achieved mean Dice scores for segmentation of 0.940 for the whole prostate, 0.914 for the transition zone, and 0.776 for the peripheral zone. Recently a comparison of three standard DL architectures for prostate segmentation was proposed in [10]. UNet, an efficient neural network (ENet), and an efficient residual factorized ConvNet (ERFNet) were trained and tuned on the PROSTATEx public dataset to segment the whole gland and the transition zone separately (the peripheral zone masks were obtained by subtraction). The top result was achieved by ENet: 91% for the whole gland, 87% for the transition zone, and 71% for the peripheral zone.

2.2 Prostate Lesion Detection and Characterization

Identifying and characterizing csPCa is a crucial component of proper treatment planning. The probability of csPCa can be assessed radiologically based on PI-RADS (although even when using the current version, v2.1, considerable inter-/intrareader disagreement is observed frequently. AI/DL methods have the potential to become common tools for differentiation between csPCa and icsPCa, and for assessment of the locations and extents of aggressive cancers.

Typically, AI models are subdivided into two groups of methods, with regard to the nature of the input data and the expected result of the analysis. The first group focuses on lesion detection, uses whole MRI images for analysis and provides pixel-level output, as well as extracting regions with probable csPCa and icsPCa. The pixel-level analysis provides pixel-level probability maps of cancer distribution. This produces patient-level predictions of suspicious areas automatically. Although AI-based detection models are typically of the two-class variety (csPCa versus non-csPCa), multiclass lesion detection models—in which detection results relate to different grading systems like histopathological International Society of Urological Pathology (ISUP) score to express the aggressiveness of csPCa, or radiological PI-RADS score—are also viable. The second group of models, dedicated purely to lesion classification, assumes radiologist-outlined

lesions (regions of interest) as inputs, which are then sorted into different categories. Some two-class or multiclass models aim to automate stratification at lesion level.

Studies reported in the literature demonstrate detection models that range between 75% to 85% accuracy; however, the methodology and study population are highly diverse. Nevertheless, it can be observed that the results reported fall within the range of reported radiologist performance [8]. A key challenge for fully automated detection algorithms is the number of false positive lesions. Mehralivand et al. [14] presented a new, fully automated, DL-based PCa detection system for prostate MRI using a large scale, diverse, expert annotated training dataset. Although the approach achieved reasonable performance metrics, an average of 0.8–1.9 false positive lesions per patient were reported. Multiclass lesion classifications are more varied due to their diverse cohort sizes and methodology. When classifying according to ISUP grade, mean AUC per classification category can range significantly—even for the same category. In a review study conducted by Twilt et al. [21], ISUP 3 category mean AUC ranged between 0.379 and 0.96.

3 Urgent Challenges

3.1 Datasets

The AI community continues to wait for extensive and well annotated datasets. In PCa research most datasets are small (often deriving from a single institution) that are homogeneous in acquisition protocols and scanner manufacturers. The development of robust and unbiased AI models requires large, heterogeneous, and reliably annotated datasets, which reflects the variability of cancer’s appearance and the diversity of equipment manufacturers. Labels are typically created by a single expert; however, intraobserver variability exists even between experienced radiologists in the assessment of cancer extent on different mpMRI modalities and in the selection of individual lesions features. Such differences continue to be observed, despite the introduction of the PI-RADS standard. Dataset labeling should be performed during multireader studies, including interdisciplinary discussions and panels, to reduce bias in labeling.

3.2 Defining Ground Truth

“Ground truth” refers to the labels that are assigned to expert annotations. Radiological delineations without histopathology confirmations are severely limited due to the high risk of missing cancer foci that were not identified by a radiologist. In PCa the most reliable validation is based on retrospective identification of cancerous regions on whole-mount radical prostatectomy specimens, which can be projected onto an mpMRI scan. Although histopathology information is evident, this method requires advanced MRI-histopathology registration. Moreover, manual pathologists’ annotations are even more time consuming than

radiological delineations, which limits the possibility of creating large datasets. Prostatectomy is typically performed on cancer at advanced stages, which limits the possibility of including cases of low risk or indolent cancer. The alternative approach assumes pathological information from biopsies. The use of systematic biopsy is limited due to the random sampling procedure and the range being limited mainly to the peripheral zone, which might entail the missing or undersampling of some PCa foci, and lead to underdiagnosis. The best solution assumes pathological confirmation from target biopsies, implemented under a fusion of MRI and TRUS-ultrasonography. PCa candidates are typically selected by a prebiopsy MRI in which a radiologist highlights PI-RADS 3 or above lesions. Much attention should be paid to accurate mpMRI analysis, which, in the case of database creation might relate to multireader verification of visible lesions and verification with biopsy history.

3.3 Different Evaluation Criteria

Comparison of different AI models is quite problematic—not only due to the diversity of the databases used and the various cohort sizes, but also in context of the lack of standardized evaluation criteria. In the case of PCa, which progresses gradually, using clinical endpoints in the form of patient outcomes like death or recurrence is mostly unviable. This underlines the importance of defining suitable benchmarks and verification criteria for model evaluation. Organization of grand challenges and open databases might improve the validation and benchmarking of models.

3.4 Limited Multireader Studies and Prospective Evaluation

As well as comparisons between AI models, attention should be directed toward multireader studies that compare the efficiency of AI models with that of radiologists or clinicians. Such comparisons may better indicate the potential of AI models—particularly in the support or training of less experienced specialists. Finally, a prospective evaluation in a controlled medical environment should be performed to open the door for clinical deployment, preceded by clinical trials.

4 Future Directions

We observe the development of AI models in PCa diagnosis and their increasing effectiveness every day. We are moving, incrementally, toward personalized medicine, and exploring the potential of radiomics and domain knowledge, while attempting to overcome the existing limitations and challenges.

A research group at the Laboratory of Applied Artificial Intelligence of the National Information Processing Institute is conducting the *AI4AR PCa*[1] project, which is dedicated to the analysis of mpMRI images for PCa diagnosis. During the initial phase reliable, well annotated databases of mpMRI examinations are being compiled. The researchers aim to collect between 400 and 600

cases with full clinical descriptions. All cases are annotated by three experienced radiologists who have over five years of experience in describing mpMRI examinations and who possess practical knowledge of the PI-RADS standard. Radiologists analyze mpMRI independently, without access to historical information. All cases are later analyzed by an interdisciplinary team of researchers, radiologists, and clinicians to reduce bias and to confirm lesions' locations and extent. Ground truth is based on MRI-ultrasound fusion guided targeted prostate biopsy. Only cases confirmed histopathologically, with verified lesion locations that correlate with historical biopsy data, are treated as reliable and are included in the database.

Simultaneously a novel structured reporting system is under development as a flexible environment for the structured and standardized reporting of radiological image data. The system possesses a modular architecture and is integrated with the XNAT[2] imaging informatics platform, which facilitates common management, storage and quality assurance tasks for imaging and associated data. A dedicated module for prostate mpMRI structured reporting was proposed, which is standardized according to the PI-RADS radiological lexicon. The structured report scheme is also used in the data labeling process; all cases in the database possess not only visual labels, but also structured ones, which contributes to the uniqueness of the proposed dataset. In the second phase, the database will be extended by inclusion of cases from other medical centers to increase its heterogeneity regarding different equipment manufacturers and acquisition protocols.

The final implementation of the system is planned in the form of a radiological educational platform that is dedicated to learning the structural reporting of prostate mpMRI examinations, and considers various forms of support by AI algorithms. Verification of the platform in the third phase of the project assumes the conductance of multireader studies to assess the effectiveness and impact of AI models on the quality and accuracy of the reporting process. A study of the behavior of platform users will allow us to assess the potential of AI for less experienced radiologists, and will standardize or improve human reader performance.

5 Conclusions

AI in prostate MRI analysis shows great promise and impressive performance that is comparable to that of human experts. Overcoming the technology's limitations and demonstrating its clinical effectiveness will unlock opportunities for clinical deployment in the form of educational systems, reporting and patient management support systems, and "second reader" or patient triage systems.

Acknowledgements. This work has been funded by the Polish National Centre for Research and Development under the program INFOSTRATEG I, project INFOSTRATEG-I/0036/2021 AI-augmented radiology - detection, reporting and clinical decision making in prostate cancer diagnosis.

References

1. Ai-enhanced radiology detection, reporting and clinical decision-making in prostate cancer diagnosis. <https://ai4ar.opi.org.pl/en/>
2. The extensible neuroimaging archive toolkit (XNAT). <https://www.xnat.org/>
3. Bardis, M., et al.: Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning. *Radiol.: Imaging Cancer* **3**(3), e200024 (2021)
4. Bell, K.J., Del Mar, C., Wright, G., Dickinson, J., Glasziou, P.: Prevalence of incidental prostate cancer: a systematic review of autopsy studies. *Int. J. Cancer* **137**(7), 1749–1757 (2015)
5. Belue, M.J., Turkbey, B.: Tasks for artificial intelligence in prostate MRI. *Eur. Radiol. Exp.* **6**(1), 1–9 (2022)
6. Bhattacharya, I., et al.: A review of artificial intelligence in prostate cancer detection on imaging. *Thera. Adv. Urol.* **14**, 17562872221128792 (2022)
7. Castillo, T.J.M., Arif, M., Niessen, W.J., Schoots, I.G., Veenland, J.F.: Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. *Cancers* **12**(6), 1606 (2020)
8. Chaddad, A., et al.: Magnetic resonance imaging based radiomic models of prostate cancer: a narrative review. *Cancers* **13**(3), 552 (2021)
9. Cuocolo, R., et al.: Machine learning applications in prostate cancer magnetic resonance imaging. *Eur. Radiol. Exp.* **3**(1), 1–8 (2019)
10. Cuocolo, R., et al.: Deep learning whole-gland and zonal prostate segmentation on a public MRI dataset. *J. Magn. Reson. Imaging* **54**(2), 452–459 (2021)
11. Epstein, J.I., Egevad, L., Amin, M.B., Delahunt, B., Srigley, J.R., Humphrey, P.A.: The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *Am. J. Surg. Pathol.* **40**(2), 244–252 (2016)
12. Hamm, B., Asbach, P.: Magnetic resonance imaging of the prostate in the PI-RADS era. In: Hodler, J., Kubik-Huch, R.A., von Schulthess, G.K. (eds.) *Diseases of the Abdomen and Pelvis 2018-2021. ISS*, pp. 99–115. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75019-4_11
13. Li, H., Lee, C.H., Chia, D., Lin, Z., Huang, W., Tan, C.H.: Machine learning in prostate MRI for prostate cancer: current status and future opportunities. *Diagnostics* **12**(2), 289 (2022)
14. Mehralivand, S., et al.: Deep learning-based artificial intelligence for prostate cancer detection at biparametric MRI. *Abdom. Radiol.* **47**(4), 1425–1434 (2022)
15. Mottet, N., et al.: EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. part 1: screening, diagnosis, and local treatment with curative intent. *Eur. urol.* **79**(2), 243–262 (2021)
16. de Rooij, M., Hamoen, E.H., Witjes, J.A., Barentsz, J.O., Rovers, M.M.: Accuracy of magnetic resonance imaging for local staging of prostate cancer: a diagnostic meta-analysis. *Eur. Urol.* **70**(2), 233–245 (2016)
17. Suarez-Ibarrola, R., et al.: Artificial intelligence in magnetic resonance imaging-based prostate cancer diagnosis: where do we stand in 2021? *Eur. Urol. Focus* **8**(2), 409–417 (2021)
18. Sunoqrot, M.R., Saha, A., Hosseinzadeh, M., Elschot, M., Huisman, H.: Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges. *Eur. Radiol. Exp.* **6**(1), 1–13 (2022)
19. Sushentsev, N., et al.: Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review. *Insights Imaging* **13**(1), 1–17 (2022)

20. Turkbey, B., et al.: Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur. Urol.* **76**, 340–351 (2019). <https://doi.org/10.1016/j.eururo.2019.02.033>
21. Twilt, J.J., van Leeuwen, K.G., Huisman, H.J., Fütterer, J.J., de Rooij, M.: Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review. *Diagnostics* **11**(6), 959 (2021)
22. Urakami, A., et al.: Stratification of prostate cancer patients into low-and high-grade groups using multiparametric magnetic resonance radiomics with dynamic contrast-enhanced image joint histograms. *Prostate* **82**(3), 330–344 (2022)
23. Westphalen, A.C., et al.: Variability of the positive predictive value of PI-RADS for prostate MRI across 26 centers: experience of the society of abdominal radiology prostate cancer disease-focused panel. *Radiology* **296**(1), 76 (2020)
24. Witherspoon, L., Breau, R.H., Lavallée, L.T.: Evidence-based approach to active surveillance of prostate cancer. *World J. Urol.* **38**(3), 555–562 (2020)




Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Performance of Deep CNN and Radiologists in Prostate Cancer Classification: A Comparative Pilot Study

Piotr Sobecki^(✉) , Rafał Józwiak , and Ihor Mykhalevych 

National Information Processing Institute, Warsaw, Poland
{piotr.sobecki, rafal.jozwiak, ihor.mykhalevych}@opi.org.pl

Abstract. In recent years multiple deep-learning solutions have emerged that aim to assist radiologists in prostate cancer (PCa) diagnosis. Most of the studies however do not compare the diagnostic accuracy of the developed models to that of radiology specialists but simply report the model performance on the reference datasets. This makes it hard to infer the potential benefits and applicability of proposed methods in diagnostic workflows. In this paper, we investigate the effects of using pre-trained models in the differentiation of clinically significant PCa (csPCa) on mpMRI and report the results of conducted multi-reader multi-case pilot study involving human experts. The study aims to compare the performance of deep learning models with six radiologists varying in diagnostic experience. A subset of the ProstateX Challenge dataset counting 32 prostate lesions was used to evaluate the diagnostic accuracy of models and human raters using ROC analysis. Deep neural networks were found to achieve comparable performance to experienced readers in the diagnosis of csPCa. Results confirm the potential of deep neural networks in enhancing the cognitive abilities of radiologists in PCa assessment.

Keywords: Deep learning · Prostate Cancer · Computer Aided Diagnosis

1 Introduction

In light of the increasing incidence rate of prostate cancer (PCa) over the previous years [2], there is a global focus on providing modern solutions that can address this growing health issue. Noninvasive diagnostics based on multiparametric magnetic resonance imaging (mpMRI) became essential in clinical decision-making as it enables more accurate risk stratification and therefore plays an important role in selecting patients for biopsy and direct targeting of lesions [6, 11].

Radiological assessment of the prostate gland involves the interpretation and reporting of mpMRI examinations according to the established global standards. The current version of the standardized prostate MRI assessment Prostate Imaging-Reporting and Data System (PI-RADS v2.1) [8], provides an approach

to the interpretation and reporting of PCa examinations. The system assumes the evaluation of each mpMRI sequence separately. Each lesion assessment category is established on a 5-point scale according to the assessment algorithm involving previously scored sequences. Introducing the standard in diagnostic practice improved the diagnostic accuracy of performed examinations and improved the availability of the method. Low specificity, however, remains a considerable aspect of MRI assessment and clinically significant (cs) PCa differentiation, potentially leading to unnecessary biopsies. Because of the assessment complexity and steep learning curve, it is mainly the case for radiologists with low experience in prostate MRI reporting [10].

Recently, solutions based on machine learning have achieved promising results in applications in PCa diagnostics. A PCa classification challenge held in 2017 (ProstateX) provided a way of comparing tools of automatic PCa differentiation based on mpMRI [1]. 71 competing methods were evaluated on the lesion classification task. The area under the receiver operating characteristic curve (AUC) of submitted models ranged between 0.45 to 0.87 AUC. The top three scoring teams achieved results of $AUC = 0.84$ and 0.87 . We used the ProstateX dataset to develop and validate the deep convolutional neural network (CNN) model that achieved $AUC = 0.84$ on the test ProstateX dataset [7].

Narrative Review by Twilt et. al. [9] presents an overview of recently proposed tools (between 2018 and 2022) that have been suggested to aid in the diagnosis of PCa. Overall deep learning (DL) solutions achieve the highest performance on PCa detection and diagnosis tasks. Computational models show the potential in enhancing the diagnostic processes and increasing the specificity of mpMRI assessment. However, only a limited number of studies validated the results in clinical workflows - 85% of them report only stand-alone model diagnostic accuracy [9]. It remains a question of how the diagnostic accuracy of DL solutions relates to that of radiology experts and what could be expected from the integration of computational models in diagnostic workflows.

The objective of our study was to evaluate the diagnostic accuracy of radiologists with various levels of diagnostic experience in comparison to the proposed DL solution for csPCa differentiation.

2 Methods

The retrospective study design involved the assessment of 32 suspicious lesions by six radiologists and the deep CNN model in a multi-case multi-reader (MCMR) setting.

2.1 Dataset

A group of cases from a publicly available database of annotated mpMRI data were used in the study [3]. Complete mpMRI data (T2W, DCE, DWI, and ADC sequences) were included for all cases in the database. We have selected a thirty-two lesion dataset diversified according to its clinical significance based on the results of a histopathological evaluation.

The selected dataset contained:

- 14 PZ lesions (7 cs and 7 not cs),
- 11 TZ lesions (5 cs and 6 not cs),
- 7 AS lesions (4 cs and 3 not cs).

2.2 Radiological Assessment Study

The study was carried out by a group of specialists with diversified expertise. Six radiologists involved in the study had practical experience in PCa diagnosis based on the mpMRI (three specialists with diagnostic experience of one to five years; three specialists with more than ten years of diagnostic experience, and at least five years of experience using the PI-RADS standard). Those groups of experts are referred to in the paper as experienced and inexperienced raters. The participating experts did not interact with each other during the assessment phase.

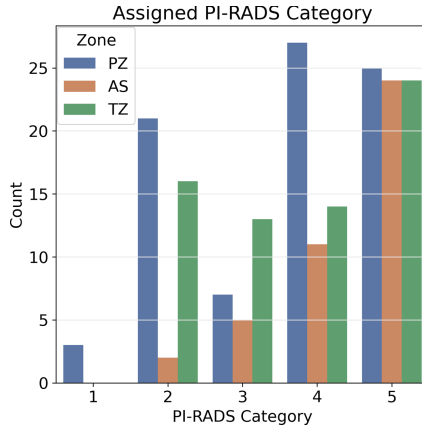


Fig. 1. PI-RADS score evaluations for lesions in the dataset. Only three assessments assigned lesions to the PI-RADS 1 category (only PZ lesions).

Experts participating in the study were not involved in the dataset selection, development of the study methodology, and experiment results analysis.

The results of the assessments are presented in the Fig. 1. Even though the dataset was balanced (close to an equal rate of cs and non-cs lesions) the distribution of assigned scores did not reflect that.

2.3 Deep CNN Model

The model evaluated in this study was a multi-modal deep CNN network of VGG-inspired architecture, adapted to the input sequence resolution and

problem complexity. Introduced modifications reduced the number of trainable parameters. The developed model architecture design reflected a PI-RADS category assessment algorithm based on lesion zonal location. This was done by integration of output routing and using complex loss function for optimization. Resulting predictions were assigned using two subnetworks designed to base predictions on T2W and DWI (subnetwork for TZ, AS, and SV lesions) and on DWI/ADC and DCE (subnetwork for PZ lesions). The model has been evaluated on the test reference dataset and resulted in a score corresponding to the state-of-the-art results (AUC = 0.84) [1]. Model architecture and analysis of achieved diagnostic accuracy have been described in the previously published study[7].

CNN predictions were made on unseen samples using 5-fold cross-validation and collecting validation split classification results.

2.4 Probability Mapping

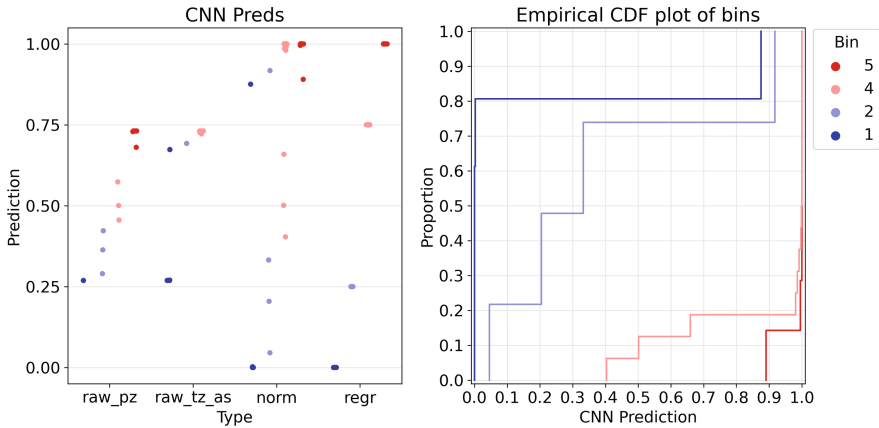


Fig. 2. The left figure shows the results of a mapping of raw CNN predictions to bins corresponding to PI-RADS categories. Separate raw predictions for PZ (raw_pz) and TZ and AS lesions (raw_tz_as) show different response characteristics resulting from separate network optimization processes. The right figure presents the empirical cumulative distribution function (ECDF) of bins in relation to normalized CNN output.

It could be argued that continuous predictions resulting from softmax output layers can produce superior AUC results in comparison to the ordinal estimations made by human experts using the Likert scale. Therefore, to further evaluate the diagnostic characteristics of the proposed model, we have mapped the raw continuous CNN predictions to bins corresponding to PI-RADS scores (Fig. 2). Bin discretization involved several steps that mapped the raw CNN predictions to ordinal categories. We have used the mode of PI-RADS assessments for lesions resulting from a radiological assessment study as ground truth for labels.

First, continuous outputs resulting from PZ and TZ/AS sub-networks were normalized to the range of [0,1]. Leave-one-out cross-validated estimates of bins were obtained for each lesion using two ordinal regression models[4] (separate each subnetwork predictions). This resulted in the mapping of continuous network output to the Likert scale that reflected the PI-RADS category characteristics based on the CNN prediction (Fig. 2).

We have mapped the 5-point Likert scale of manually assigned PI-RADS categories and automatically estimated bins to the [0, 0.25, 0.5, 0.75, and 1] probability values to perform the ROC analysis.

2.5 Statistical Analysis

We compare the model performance with that of experienced and inexperienced radiology specialists using assessments collected during the retrospective study. To evaluate the differences, we have used the Area under the Receiver operating characteristic Curve (AUC) as a measure of diagnostic accuracy. Extensive simulations using bootstrap re-sampling (1000 tests) [5] were conducted to construct 95% confidence intervals and perform hypothesis testing in various scenarios. We have conducted separate experiments to compare the diagnostic characteristics in relation to the combinations of assessment methods (CNN, human raters), lesion location (PZ, TZ, and AS), and examiner experience. An alpha of 0.05 was used as the cutoff for statistical significance (we additionally report test results with an alpha of 0.1 due to the small dataset sample size).

3 Results

The following section presents the results of the comparison of diagnostic accuracy between inexperienced, experienced radiologists and model predictions. Additionally, we report the change in diagnostic accuracy resulting from the integration of human and CNN assessments.

3.1 Results of Raw CNN Predictions

The results achieved by the CNN model (AUC = 0.83, CI [0.80, 0.88]) demonstrated superior diagnostic accuracy in comparison with both:

experienced (AUC = 0.80, $p > .1$, CI [0.74, 0.86]) and inexperienced (AUC = 0.71, $p < .1$, CI [0.63, 0.80])

specialists in the evaluation of lesions' clinical significance using the PI-RADS v2.1 standard.

The lowest diagnostic accuracy has been observed for AS lesions, where CNN solution provides higher quality estimations in comparison both to experienced and inexperienced radiologists (AUC = 0.79 vs. AUC = 0.64 vs. AUC = 0.59). In the case of other lesion locations, the differences between the neural network and the experienced radiology specialists were less pronounced: PZ (AUC = 0.88 vs.

AUC = 0.85 vs. AUC = 0.72) and TZ (AUC = 0.89 vs. AUC = 0.84 vs. AUC = 0.78). Differences in diagnostic accuracy dependent on lesion location were however not statistically significant.

Following analyses were performed using binned CNN predictions.

3.2 CNN Performance Compared to Human Raters

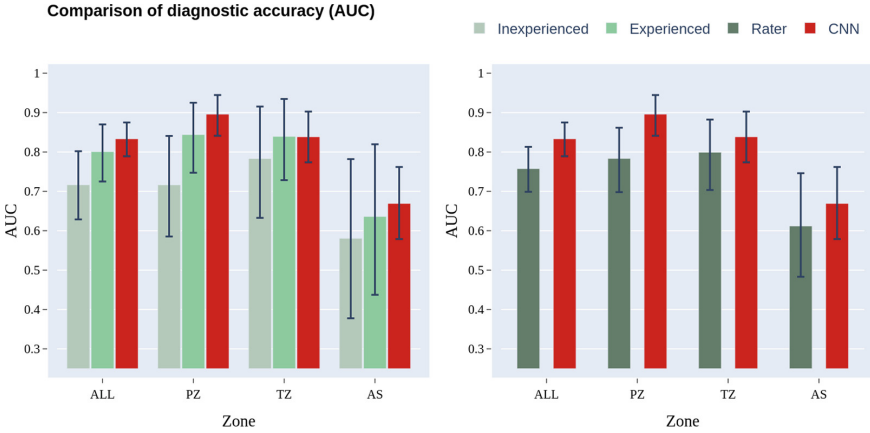


Fig. 3. Diagnostic accuracy of inexperienced, experienced (left), and all (right) assessments in comparison to the CNN predictions, expressed in established AUC values and 95% confidence intervals.

The Fig. 3 presents the results of a comparison of diagnostic accuracy measured in AUC between inexperienced and experienced raters in comparison to model predictions restricted to ordinal categories. Overall, CNN achieved superior ($p < .1$) diagnostic accuracy (AUC = 0.83, CI [0.79, 0.87]) in comparison to all (AUC = 0.76, CI [0.70, 0.81]) and inexperienced (AUC = 0.72, CI [0.63, 0.80]) rater assessments.

Differences were statistically significant for PZ lesion evaluation when considering assessments of all (CNN AUC = 0.90 vs AUC = 0.78, $p < .05$) and inexperienced raters (AUC = 0.71, $p < .05$). There were no statistically significant differences found in diagnostic accuracy between assessments of experienced raters and CNN predictions.

3.3 Diagnostic Accuracy of Combined Assessment

To investigate the potential change in diagnostic accuracy by integration of computer-aided assessment we analyzed the potential results of combining human and automatic predictions. Integrated predictions were obtained by computing average expert and binned CNN predictions on the lesion level and mapping those back to the Likert scale. This allowed investigation of the potential gain in diagnostic accuracy in computer-aided PCa diagnosis.

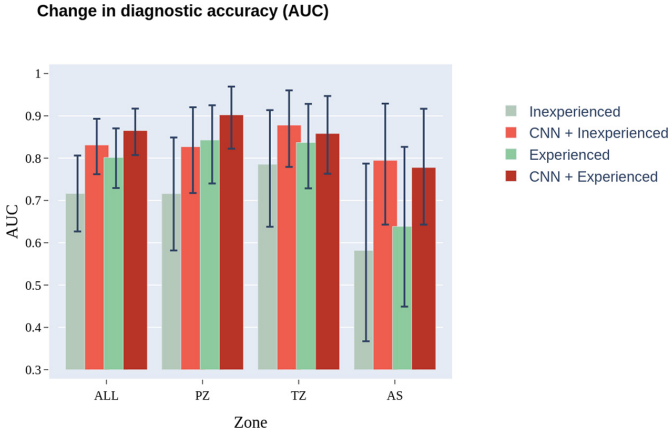


Fig. 4. Diagnostic accuracy for assessments of inexperienced and experienced radiologists compared to combined predictions expressed in established AUC values and 95% confidence intervals.

A positive diagnostic accuracy change has been observed after combining the model predictions with expert assessments in all tested settings (Fig. 4). Integration of CNN with rater predictions resulted in an overall increase of diagnostic accuracy by 0.09 AUC (CNN+rater AUC = 0.85, $p < .05$, CI [0.80, 0.89])

4 Discussion and Conclusion

In this study, we investigated the performance of the deep CNN model for PCa diagnosis on mpMRI data in comparison to human raters in an MRMC study setting on a subset of the reference dataset.

The results suggest that the proposed model outperformed inexperienced radiologists and achieved diagnostic accuracy similar to that of experienced raters. The achieved results are promising, yet decisive conclusions cannot be drawn confidently given the study design and small sample size used for validation.

Our study had several limitations. First of all, the dataset size used for validation in the conducted study was limited by the availability of readers. The study involved a substantial number of readers, however, the analysis has been performed in subgroups defined based on radiologist experience, which limited the number of assessments considered in hypotheses testing. The modest sample size resulted in wide 95% CIs constructed using bootstrap simulations and therefore affected the power of performed statistical tests. Furthermore, the study design was far from the clinical setting and based on the evaluation of selected single lesions. Finally, we could not evaluate the stability of the model performance on external data.

Although promising, results need confirmation in further, more extensive studies.

References

1. Armato, S.G., et al.: PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *J. Med. Imaging* **5**(4), 044501 (2018)
2. Carioli, G., et al.: European cancer mortality predictions for the year 2020 with a focus on prostate cancer. *Ann. Oncol.* **31**(5), 650–658 (2020)
3. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H.: SPIE-AAPM PROSTATEx challenge data (2017). <https://doi.org/10.7937/K9TCIA.2017.MURS5CL>, <https://wiki.cancerimagingarchive.net/x/iIFpAQ>
4. Liu, Q., Shepherd, B.E., Li, C., Harrell, F.E., Jr.: Modeling continuous response variables using ordinal regression. *Stat. Med.* **36**(27), 4316–4335 (2017)
5. MacKinnon, J.G.: Bootstrap hypothesis testing. *Handb. Comput. Econometrics* **183**, 213 (2009)
6. Mottet, N., et al.: EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. part 1: screening, diagnosis, and local treatment with curative intent. *Eur. Urol.* **79**(2), 243–262 (2021)
7. Sobeci, P., Józwiak, R., Sklinda, K., Przelaskowski, A.: Effect of domain knowledge encoding in CNN model architecture-a prostate cancer study using mpMRI images. *PeerJ* **9**, e11006 (2021)
8. Turkbey, B., et al.: Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur. Urol.* **76**(3), 340–351 (2019)
9. Twilt, J.J., van Leeuwen, K.G., Huisman, H.J., Fütterer, J.J., de Rooij, M.: Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review. *Diagnostics* **11**(6), 959 (2021)
10. Westphalen, A.C., et al.: Variability of the positive predictive value of PI-RADS for prostate MRI across 26 centers: experience of the society of abdominal radiology prostate cancer disease-focused panel. *Radiology* **296**(1), 76 (2020)
11. Witherspoon, L., Breau, R.H., Lavallée, L.T.: Evidence-based approach to active surveillance of prostate cancer. *World J. Urol.* **38**(3), 555–562 (2020)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Assessing GAN-Based Generative Modeling on Skin Lesions Images

Sandra Carrasco Limeros^{1,2}, Sylwia Majchrowska^{1,2}(✉), Mohamad Khir Zoubi¹, Anna Rosén¹, Juulia Suvilehto¹, Lisa Sjöblom¹, and Magnus Kjellberg¹

¹ Sahlgrenska University Hospital, Blå stråket 5, 413 45 Göteborg, Sweden

² AI Sweden, Lindholmspiren 3-5, 402 78 Göteborg, Sweden
{sandra.carrasco, sylwia.majchrowska}@ai.se

Abstract. We explored unconditional and conditional Generative Adversarial Networks (GANs) in centralized and decentralized settings. The centralized setting imitates studies on large but highly unbalanced skin lesion dataset, while the decentralized one simulates a more realistic hospital scenario with three institutions. We evaluated models' performance in terms of fidelity, diversity, speed of training, and predictive ability of classifiers trained on the generated synthetic data. In addition, we provided explainability focused on both global and local features. Calculated distance between real images and their projections in the latent space proved the authenticity of generated samples, which is one of the main concerns in this type of applications. The code for studies is publicly available (<https://github.com/aidotse/stylegan2-ada-pytorch>).

Keywords: GAN · federated learning · skin lesion classification · XAI

1 Introduction

In recent years, the use of neural networks has become a very popular and attractive topic for many medical researches [7, 9, 17], as one of the key promises of using Artificial Intelligence (AI) in healthcare is its potential to improve diagnosis. However, to create reliable deep learning (DL) algorithms that can identify complex patterns of medical conditions, they must be trained on a large amount of data. In addition, it is desirable for the model to have a diverse range of cases, as data from a single source may be biased by the acquisition protocol or the population [6, 20].

Unfortunately, preparation and annotation of medical data is a costly procedure that demands the assistance of medical specialists. Additionally, access to medical data requires a lengthy approval process due to patient privacy concerns. This makes it almost impossible for different institutions to share data and thus expertise with one another. Although there are some high quality open access dataset initiatives [9, 17], there is still a great need for much more diverse and complex databases to effectively apply DL.

Synthetic data appears to be a good solution to mitigate the issues with privacy policies. It can be used in two ways - firstly as extensions of small and unbalanced datasets

S. C. Limeros and S. Majchrowska—These authors contributed equally to this work.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 93–102, 2023.

https://doi.org/10.1007/978-3-031-37649-8_10

(e.g., of rare diseases) and secondly for anonymization purposes (to replace instead of augment real samples). In both scenarios, synthetic medical data must accomplish two competing goals. The data should accurately reflect the real data and simultaneously offer strong privacy protection for the individuals whose records were used to create it.

In this work we perform a detailed study of GAN-based artificial data generation in the case of International Skin Imaging Collaboration (ISIC) 2020 [17] database using StyleGAN2-ADA [11] in conditional and unconditional settings. All trained models are evaluated in terms of both fidelity and diversity. Furthermore, we conduct an extensive latent space analysis of the generated images to better understand the structure of the real and synthetic images for the subsequent binary classification task (benign and malignant). Performed evaluations base on image editing in latent space, local and global explanations of trained classifiers. As far as we know, such detailed analysis has not been attempted before.

Moreover, to deal with a more realistic scenario where a single hospital does not have a sufficiently large dataset to generate artificial data, we simulate a scenario with three hospitals with a different amount of data each. We propose to use Federated Learning (FL) [16] with the aim to synthesise a more complex, fair and diverse dataset through the collaboration of multiple medical institutions without exchanging local data samples.

2 Materials and Methods

2.1 International Skin Imaging Collaboration Database

In our experiments, the reference dataset for real images is based on the training set of the ISIC 2020 challenge [17] extended by malignant cases from previous years' competitions [15]. The database consists of the 37 648 images – the whole ISIC 2020 dataset, adding 4522 malignant samples from ISIC 2019 – where 20% were used for validation in first phase of central trainings. Later, we splitted the training subset based on patient ID attributes. To make the FL setup more appropriate, we ensured that the data from an individual patient would not be present on more than one client. For this setup, we created 3 clients and for them, data subsets with 2k, 12k, 20k images respectively. For each client the proportion of malignant and benign was roughly the same as in the whole dataset. In all experiments, we resized the input images to 256×256 pixels.

2.2 Training Details

We investigated StyleGAN2-ADA performance using an original implementation from NVIDIA Research group¹. We trained StyleGAN2-ADA models with each of the two classes of training set as input, as well as in a conditional setting with and without augmentations. To select the best model, we considered both the Fréchet Inception Distance (FID) [8] and Kernel Inception Distance (KID) [5] metrics, along with training speed, similarly as proposed in [4]. The classification task was performed using EfficientNet-B2 model [19], pretrained on ImageNet, with Ross Wightman's implementation². During training, we used the Adam method for optimizing the network weights with an

¹ <https://github.com/NVlabs/stylegan2-ada-pytorch>.

² <https://github.com/rwightman/efficientdet-pytorch>.

adaptive learning rate initialized to 5×10^{-4} . We trained the models for a maximum of 20 epochs and an early stopping with a patience of 3 epochs. We applied standard data augmentation techniques, such as random rotation, horizontal and vertical flip, during the training phase for all experiments. For the experiments in a FL setup, we used Flower framework [3]. In our simulated setup, we created a network with 3 clients with different amounts of data and a server, where the weights of the trained model were exchanged every 100 iterations. We used the Federated Average (FedAvg) algorithm [14] as it is an effective and simple method that is commonly used for federated aggregation.

2.3 Evaluation Protocol

Various dimensions should be considered when evaluating GANs [2]. Firstly, fidelity as a measure of reliability, and diversity as a measure of fairness. FID and KID metrics evaluate these two characteristics, but rely on a preexisting classifier trained on ImageNet, and are insensitive to the global structure of the data distribution. Also Precision (P) and Recall (R) scores measure, respectively, the fraction of synthetic samples that look realistic (fidelity) and the fraction of real samples that the model can synthesize (diversity). Perceptual Path Length (PPL) [12] estimates whether and how much latent space is entangled or regularized, ultimately being able to capture the coherence of images. Another dimension to look at is predictive performance, referring to the fact that samples should be as useful as real data when used for the same predictive purpose. Here, we built a melanoma classifier using synthetic data for training and real data for testing. Since privacy is the most important factor in medical study, we evaluated the generalization or authenticity of the generative process [2], which measures the model capability to creation of new samples. Additionally, a survey was conducted in which experts assessed whether each of the 200 tested images is real or generated artificially by cGAN. Finally, we investigated whether it is possible to edit the image by manipulating the latent input of the trained GAN. The semantic factorization (SeFa) [18] method, as it do not need a large sample of latent vectors and auxiliary classifier, was tested to see if we could obtain directions in latent space, where the influence of one feature could be controlled while preserving the rest of the image.

3 Results

3.1 GANs Trainings

In the first phase of our experiments, we established the best model in terms of fidelity and diversity using well-known metrics such as KID, FID, P, R, and PPL (see Table 1). It is worth noting that the GAN responsible only for malignant melanoma generation (mal-GAN) had around 6 times less data than for benign cases (ben-GAN). In general, the unconditional models have lower PPL scores, showing better regularity of latent space due to the fact that they model only the distribution of one class. Additionally, the vast majority of malignant melanoma examples in ISIC 2020 and ISIC 2019 show a black dermatoscope frame, which leads to the generation of darker images.

The conditional setting was used to provide the model with a wider variety of images, since there are a lot of characteristics that are common for both classes. Achieved higher FID and KID scores confirmed that this is beneficial for the minority class (malignant). For this setting, we used ADA mechanism with and without ($w/o\ col$) color augmentation, and achieved the best scores for second one. Color augmentation leads to leakage of color hue (unnatural red or violet) to the generated examples. Subjective assessment based on four responses in a qualitative survey, from two dermatologists and two deep learning experts, achieved an overall average accuracy of 54% for participants (at level 58% for dermatologists and 50% – deep learning experts). There was no feature in any image that clearly suggested to the participants that the image is either real or synthetic.

Table 1. Calculated metrics for each of the generative models tested in the centralized setting.

Scenario	KID (%)	FID	P	R	PPL
ben-GAN	0.42	7.99	0.77	0.45	60
mal-GAN	0.47	15.46	0.62	0.40	51
cGAN	0.32	7.33	0.75	0.42	193
cGAN $_{w/o\ col}$	0.24	7.02	0.75	0.44	101

In case of simulated hospital scenario in FL setup, we observed faster convergence (1.6 times) and improved quality of the generated images mainly for the client with the smallest data resources. As the data distributions between different clients only differed in size, we put more emphasis on the classification task with centrally trained models.

3.2 Predictive Performance with Classifier

After the evaluation with general metrics, we performed a study on predictive performance to measure how useful the synthetic data is for the subsequent task, i.e. malignant melanoma diagnosis. As a baseline for the experiments, we first train the classifier on training subset of the real images of ISIC dataset, and then tested it on the validation set. Secondly, GAN-based augmentation was performed using two types of GANs models with two scenarios: training on balanced synthetic dataset with 55k images (*syn*) and testing on real validation subset (the same as in baseline experiment) and training on real images adding 22k synthetic melanoma samples (*aug*) to balance the dataset. The introduction of highly underrepresented malignant melanoma cases improves the classification accuracy roughly of few pp. in both scenarios, as summarised in Table 2. Overall GAN-based augmentation technique does not provide reliable improvements in case of classification using the whole ISIC 2020 and malignant samples from ISIC 2019.

3.3 Explanations of the Predictions

To measure the authenticity we projected 12k samples from the real dataset into the latent space of the generator. This gave us the latent codes that caused our generator to synthesize the most similar output to the input image. To optimize for a latent

Table 2. Calculated metrics for each of the classification scenarios with EfficientNet-B2: trained on real (real-baseline), only synthetic samples (*syn*), and augmented balanced dataset with additional 22k fake malignant lesions images (*aug*) from conditional (cGAN) and unconditional (GAN) models. In all scenarios the models were tested on the same real images validation set.

Scenario	Acc (%)	AUC (%)	Scenario	Acc (%)	AUC (%)	Scenario	Acc (%)	AUC (%)	Scenario	Acc (%)	AUC (%)
real baseline	97.8	98.8	syn-GAN	94.1	94.2	syn-cGAN	94.7	96.7	syn-cGAN _{w/o col}	92.6	92.7
			aug-GAN	97.8	98.6	aug-cGAN	97.8	98.6	aug-cGAN _{w/o col}	97.9	98.8

code for the given input images, we followed [1]. We used a VGG16 model as a feature extractor, computed the loss on the difference of the extracted features for both the target image and the generated output, and performed backpropagation. Next, we extracted the features of both the real and their projected images using the last convolutional layer of our classifier trained on real and synthetic data (aug-cGAN_{w/o col}). These embeddings were visualized in a 3D space using t-distributed stochastic neighbor embedding (t-SNE) method [13]. This allows visually exploring the closest near neighbors of each real image using cosine distances. Figure 1 shows examples of real images projections in the latent space of the generator (with benign marked on red, malignant – blue) and projected embeddings of real and synthetic data. In both cases there is visible separation between two clusters created by two examined skin lesion classes. However, there are still plenty of the cases in the middle between two clusters and mixed with improper class, what is visible in Fig. 1(a). Additionally, we spotted some clusters inside classes, which are associated with instrumental bias, such as ruler and black dermatoscope frame.

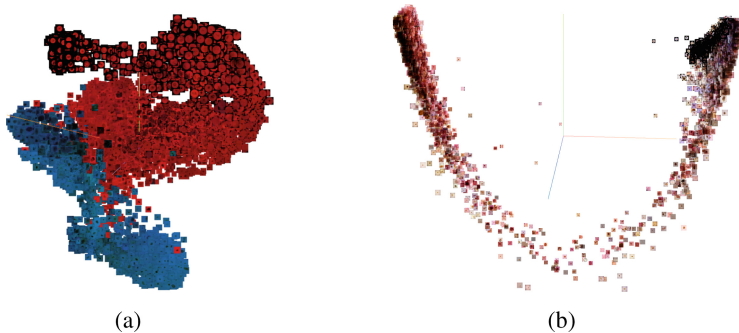


Fig. 1. Real images projections in the latent space of the generator (a). Projected embeddings of real and synthetic data coming from the classifier trained on synthetic data (b).

For a more systematic inspection, we computed the cosine distances between the different pairs of real images and their projected samples. The mean distance was equal to 0.1444 and the median 0.00283 with only two projections being *too close* in terms

of $Q1 = 0.013$ (range of $1e-5$) to the real images. Only in these two cases the closest neighbor was the projection of the target image, meaning that the generative model could have memorized that sample. We treated this as a measure of the authenticity of the generated samples. We also spotted that some of the images were very distant from their projections (around 2) but still resembled the target image (Fig. 2(a)).

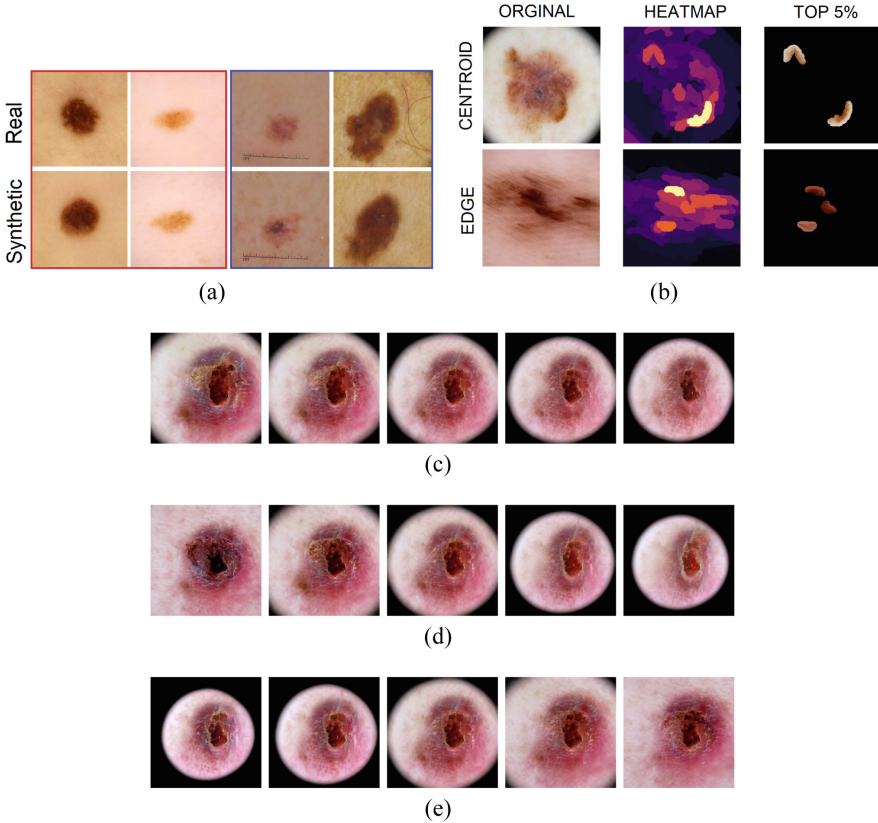


Fig. 2. A few examples of the closest (red frame) and the most distant (blue frame) pairs real-synthetic in terms of cosine distance (a). Two examples from the malignant class, which were found in the center, and in the boundary between two clusters respectively, examined using XRAI heatmaps (b). Examples of image editing using the SeFa framework shifted along the 2nd (c), 4th (d) and 6th (e) eigenvectors.

The images in the center and boundary between the two clusters (Fig. 1(b)) were studied using local explanations with the XRAI method [10]. For images of malignant lesions that belong to the centroid of the embeddings, we found that the mole itself is the most important part of the image for the final prediction. In the sample image, the network focuses on boundary pixels which represent asymmetry in the mole, one of the main clues for detecting malignant melanomas. On the other hand, in edge cases the

results are not as evident due to image distortions or poorly centred moles (Fig. 2(b)). We selected all the misclassifications and edge cases and generated N neighbors using a distance of 0.1 to augment the dataset with more complex examples with the aim of making it more robust. First experiments showed an improvement in performance in those edge cases.

Finally, we edited the latent input in an attempt to eliminate the dermoscopic frame in malignant melanoma images using the SeFa method [18]. The latent w -vector corresponding to the image in Fig. 2(c)–2(e) is shifted along the 2nd, 4th and 6th eigenvectors. The image in the middle in all three rows (3rd column) is the original image. Left and right of the original image are positive and negative directions along these eigenvectors. The eigenvectors displayed were chosen from a larger qualitative evaluation of 100 images along the first 10 largest eigenvectors. Applying SeFa image editing suggests less entangled features from visual inspection of the images of different directions – the black frame was removed leaving the other features (such as shape, size, color) almost intact.

To assess the quality of the edited images, we first generated a large sample size of images all containing frames. After acquiring the images we removed the frames by shifting the latent vectors along the direction where the presence of the dermoscopic frame was minimized. Finally, we trained a classifier on these images for the malignant melanoma with a training set of 10k images per class and a test set consisting of real images, which result in accuracy equalled 87%.

4 Discussion

In our study, we explored the state of the art DL-based techniques to generate, classify, and explain computed results for skin lesion diagnosis. Our experiments are based on ISIC 2020 and ISIC 2019 datasets, which are one of the largest but very unbalanced open access database.

Samples generated using different types of GANs and settings exhibits slightly different appearance, as evidenced by the calculated metrics shown in the results (see Sect. 3). The PPL measure, which is capable of capturing the consistency of the images, is the lowest for generated malignant melanoma samples by unconditional GAN. However, this is not connected with the lowest KID and FID scores indicating the dissimilarity between two probability distributions (real and fake) using samples drawn independently from each distribution. Lower PPL score is related to the smallest amount of malignant data, and in result more regularized and narrow distribution of latent space. The second observation may be connected with the fact, that KID and FID rely on a pre-existing classifier (InceptionNet) trained on ImageNet that consists of different images rather than skin samples. The results also indicates that the cGAN model is prone to generating more realistic looking melanoma (using some features from benign samples) than the mal-GAN. No statistical conclusion can be drawn from the small sample size in the survey where cGAN generated images were used. However, the results do suggest that subjectively, experts are unable to tell an artificial lesion from that of a real patient. There was no specific feature that the experts picked up on in the generated data as an artifact of the model. Therefore, qualitatively the synthetic data pass for real in the eyes of experts.

In the case of classification, we have not observed a large improvement of the performance of the classification network based on synthetic data generated by StyleGAN2-ADA. Actually, the results achieved in different scenarios do not differ much. This may be affected by the large size of the real dataset, but also by the fact that some features coupled with, for example, methods of collecting data (existence of black dermoscopic frame) may be entangled with a specific class.

Performed exploration of the latent space showed that there is a clear separation between the projections of the real and generated samples. Measured distance between the projections of real and the closest synthetic image proved the authenticity of the generated samples. Our main interest in the explanation of classification results focused on the edge cases, as the dermatologists are paying special attention to those cases that lie in the boundary and are not so obvious. We noticed that the network output is often biased by acquisition protocols, as well as some patient-related features. The main issue seems to be the area covered by the mole on the image. However, this topic requires closer examination. Editing images using latent directions could be a useful tool in removing unwanted artifacts from images. Nevertheless, dermoscopic frames were present mostly in images of malignant melanoma, thus the characterization of class labels was entangled with dermoscopic frames. This entanglement resulted in changes in separate features when removing the frame artifact and did not leave the malignant melanoma data intact. For future steps, using this technique may show promising results in data normalization and generalization in different domains.

On the other hand, as GAN training requires a large investment in computing and data resources, the FL setup may be a solution for smaller institutions with a lack of access to sufficient data resources. Achieved results confirmed that generation of skin lesions in a distributed setup can lead to similar performance with respect to the quality and diversity of generated samples, with a significant faster convergence. However, to reach a final verdict on this matter, it is necessary to conduct further research into different aggregation algorithms, privacy preserving techniques, and even defense mechanisms against adversarial attacks.

5 Conclusions

GAN-based augmentation is an extensively explored technique for medical imaging applications, especially in the case of very rare diseases. First of all, it helps in the creation of larger and more balanced datasets. Secondly, it creates non-real data, which can be more easily shared amongst the medical community. However, the results achieved with the addition of synthetic data reported in literature show an improvement in accuracy of only a few percents without clearly explaining the reason. On the other hand, GAN-based anonymization suffers from an unset gold standard in measuring its performance.

To utilize GANs in generating synthetic healthcare data, a number of considerations need to be made. First, one should consider the architecture. In our case, we chose between central unconditional GANs per class, conditional GAN and FL setup. The usefulness of chosen architectures mainly depends on computational resources and time - unconditional GAN can be good option with small amount of classes due to long

duration of training of single GAN. If a massive, annotated dataset exists, training the GAN centrally is preferable but in case of a more realistic scenario of data being siloed in an institution, the benefit from FL is noticeable particularly for smaller institutions.

Second, the created synthetic data should be inspected from multiple different points of view. Common features to emphasise are fidelity and diversity, which are important to understand how well the synthetic data represents the underlying real data. Importantly, as the goal in healthcare is to avoid sharing data, it is also crucial to inspect the authenticity of the synthetic examples to make sure they are not simply copying the training data. Additionally, the synthetic data should be as useful as the real data for the subsequent task (e.g. classification) and not allow inferences based on features that are not related to the case, but, for example, to the way the data were collected (e.g., linking a black dermatoscope to malignant melanoma).

Acknowledgements. This work has been carried out during the *Eye for AI* and *Master Thesis* programs thanks to the support of Sahlgrenska University Hospital, Chalmers University of Technology, and AI Sweden.

References

1. Abdal, R., Qin, Y., Wonka, P.: Image2StyleGAN: how to embed images into the StyleGAN latent space? In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4431–4440 (2019). <https://doi.org/10.1109/ICCV.2019.00453>
2. Alaa, A.M., van Breugel, B., Saveliev, E., van der Schaar, M.: How faithful is your synthetic data? Sample-level metrics for evaluating and auditing generative models. arXiv preprint [arXiv:2102.08921](https://arxiv.org/abs/2102.08921) (2021)
3. Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., Lane, N.D.: Flower: a friendly federated learning research framework. arXiv preprint [arXiv:2007.14390](https://arxiv.org/abs/2007.14390) (2020)
4. Bissoto, A., Valle, E., Avila, S.: GAN-based data augmentation and anonymization for skin-lesion analysis: a critical review. In: 2021 IEEE/CVF CVPRW, pp. 1847–1856 (2021)
5. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying MMD GANs. arXiv preprint [arXiv:1801.01401](https://arxiv.org/abs/1801.01401) (2021)
6. Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., Yap, M.H.: Analysis of the ISIC image datasets: usage, benchmarks and recommendations. *Med. Image Anal.* **75**, 102305 (2022)
7. Gillstedt, M., Hedlund, E., Paoli, J., Polesie, S.: Discrimination between invasive and in situ melanomas using a convolutional neural network. *J. Am. Acad. Dermatol.* **86**(3), 647–649 (2022)
8. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Guyon, I., et al. (eds.) *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc. (2017)
9. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L.A., Roger, M.: MIMIC-IV. *PhysioNet* (2020)
10. Kapishnikov, A., Bolukbasi, T., Vi’egas, F., Terry, M.: XRAI: better attributions through regions. In: 2019 IEEE/CVF ICCV, pp. 4947–4956 (2019)
11. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. arXiv preprint [arXiv:2006.06676](https://arxiv.org/abs/2006.06676) (2020)
12. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8107–8116 (2020)

13. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(86), 2579–2605 (2008)
14. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.Y.: Communication-efficient learning of deep networks from decentralized data. In: Singh, A., Zhu, J. (eds.) *Proceedings of the 20th AISTATS. Proceedings of Machine Learning Research*, vol. 54, pp. 1273–1282. PMLR (2017)
15. Nozdrin, R.: *Melanoma external malignant 256* (2020)
16. Rajotte, J.F., et al.: Reducing bias and increasing utility by federated generative modeling of medical images using a centralized adversary. arXiv preprint [arXiv:2101.07235](https://arxiv.org/abs/2101.07235) (2021)
17. Rotemberg, V., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Sci. Data* **8**(1), 34 (2021)
18. Shen, Y., Zhou, B.: Closed-form factorization of latent semantics in GANs. In: *CVPR*, pp. 1532–1540 (2021)
19. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) *Proceedings of the 36th ICML. Proceedings of Machine Learning Research*, vol. 97, pp. 6105–6114. PMLR (2019)
20. Wachinger, C., Rieckmann, A., Pölsterl, S.: Detect and correct bias in multi-site neuroimaging datasets. *Med. Image Anal.* **67**, 101879 (2021)









Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Prostate Cancer Detection Using a Transformer-Based Architecture and Radiomic-Based Postprocessing

Jakub Mitura^{1,5} , Rafał Józwiak^{1,2}  , Ihor Mykhalevych^{1,2} ,
Iryna Gorbenko¹ , Piotr Sobecki¹ , Tomasz Lorenc³ ,
and Krzysztof Tupikowski⁴ 

¹ Laboratory of Applied Artificial Intelligence, National Information Processing Institute, Warsaw, Poland

² Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

Rafał.Jozwiak@opi.org.pl

³ I Department of Clinical Radiology, Medical University of Warsaw, Warsaw, Poland

⁴ Lower Silesian Oncology, Pulmonology and Hematology Center, Wrocław, Poland

⁵ Medical University Lublin, Lublin, Poland

Abstract. The detection of prostate cancer is an important challenge for medical personnel. To improve the medical system's ability to process increasing numbers of oncological patients, demand for automation systems is growing. At the National Information Processing Institute, such systems are undergoing active development. In this work, the authors present the results of a pilot study whose goal is to analyze possible directions in the development of new, advanced deep learning systems using a high quality dataset that is currently in development.

Keywords: Prostate cancer · Magnetic resonance imaging · Medical image segmentation

1 Introduction

Prostate cancer is one of the most common neoplasms in men [6]. This indicates the importance of developing systems for its efficient detection, treatment, and monitoring. The gold standard of cancer diagnosis is the study of histopathology; however, due to high variability in the structure of the prostate gland, particularly among older patients, the selection of optimal sites for biopsy remains challenging. This explains the necessity of medical imaging. The most established imaging modality for prostate cancer detection is multimodal magnetic resonance imaging (MRI). However, the interpretation of the multimodal 3D images requires time and expertise from radiologists. The increasing average age of patients and the rising prevalence of cancers place intense pressure on medical organizations to supply enough skilled personnel to meet growing demand.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 103–110, 2023.

https://doi.org/10.1007/978-3-031-37649-8_11

One possible solution for alleviating this problem lies in the design of automated systems for cancer detection. This, in turn, has led to growing demand for high quality datasets and deep learning algorithms. Both solutions are undergoing active development at the National Information Processing Institute.

The selection of architecture is one of the most crucial decisions that influences a model's performance. Until recently, most of the research conducted in computer vision was based on convolutional neural networks, during a time when natural language processing tasks witnessed an explosion of transformer-based architectures. However, according to new research in computer vision, transformer-based architectures promise performance that is consistently better than that of convolutional neural networks [8]. One of the main characteristics of convolutional neural networks is the enforcement of models to include information on the local co-occurrence of image features, which have been proven to be a significant inductive bias. Pure transformers do not share this characteristic; they learn the spatial correlations between image features via attention mechanisms. This adds a number of degrees of freedom to the models that enable them to learn the nonlocal, long-range dependencies in images, at the cost of requiring larger datasets to achieve the same performance. Moreover, the newest research [19] tackles the high memory requirements of nonmodified transformer architectures and the technical problems in training larger models on graphical processing units. One solution involves fusing convolutional and transformer-based architectures to take advantage of both using a hybrid transformer. This can be achieved by inserting a transformer into different layers of a U-shaped architecture, composing architectures, and using attention mechanisms on features calculated by convolutional neural networks [8]. The authors of this article concentrated on the first type of hybrid architecture, as they have already proven to be efficient in multimodal MRI settings [7] and specifically in prostate cancer detection [17]. At the time of writing, no consensus exists on the best available transformer-based architecture for prostate cancer detection and segmentation. This points to the necessity of further research and experimentation, the preliminary results of which are presented below.

2 Material and Methods

The data used to train and validate the model was accessed from *Artificial Intelligence and Radiologists at Prostate. Cancer Detection in MRI: The PI-CAI Challenge* [1]. The data encompasses 1,500 partially labelled cases of prostate parametric MRI (bpMRI). The labels, when present, indicate the locations of prostate cancer. The algorithm described below utilized T2-weighted imaging (T2W), axial-computed high value (≥ 1400 s/mm²) diffusion-weighted imaging (DWI), and axial-apparent diffusion coefficient maps (ADCs). The labels were annotated manually by human experts, and at least two changes were considered significant for the International Society of Urological Pathology (ISUP). The main library used in the work was Monai [4], which is a PyTorch-based [14] framework for deep learning in the medical imaging domain. To improve the

code structure and training time, the code was refactored for use with Pytorch Lightning [5]. Image preprocessing was completed using the proposed algorithm from the PI-CAI Challenge [1], based on the nnUnet [9] architecture. All preprocessing steps were implemented as Monai transforms. Image augmentations were performed using the batchgenerators library [10]. To improve the reproducibility of the algorithm, training and inference were conducted using Docker containers [13]. All experiments were performed in the Google Cloud cluster using a server with NVIDIA A100 40 GB RAM GPU.

2.1 Preprocessing

The MRI data was normalized in each channel using z-score normalization. The image shape was set to (256, 256,32) for it to be a multiple of the sixteen in each axis, as the chosen architecture required. The spacing of the dataset was highly inhomogeneous; for this reason, all images were resampled to achieve (0.5,0.5,3.0) voxel size. Image augmentations were performed using the batchgenerators library [10] and encompassed Gaussian noise, elastic deformations, Gaussian blur, brightness modifications, contrast augmentations, simulations of low resolution, and mirroring. All of the labels were converted to binary masks and included in augmentations that led to spatial deformations of the original images.

2.2 Deep Learning Architecture

We selected Swin UNETR [7] as the architecture because it demonstrates characteristics that are crucial for the further development and finetuning of the algorithm on the new dataset in development. The neural network architecture is based on transformers. This has multiple advantages over traditional, convolution-based architectures. Primarily, it increases the receptive field, which enables the learning of long-range image dependencies. It partially avoids translation invariance of convolutions, which, in the context of medical imaging, can lead to the loss of relevant location-based information. Transformer-based architectures also have generally higher expressive power due to their less pronounced inductive bias. However, such architectures also cause difficulties due to their high memory footprint and relatively poor performance on small datasets (because of reduced inductive bias). The architecture is summarised in Fig. 1.

For the current work and the dataset in development, the Swin UNETR architecture has additional crucial characteristics that are well suited to modelling multimodal images. As a transformer architecture, it is possible to extend Swin UNETR to incorporate clinical data in tabular form.

2.3 Optimization

The model's optimization was implemented using the PyTorch AdamW [12] optimizer. Cosine annealing with warm restarts [11] was used for the Learning Rate Scheduler, and the initial learning rate was established by the Learning Rate Finder [18] implemented in PyTorch Lightning.

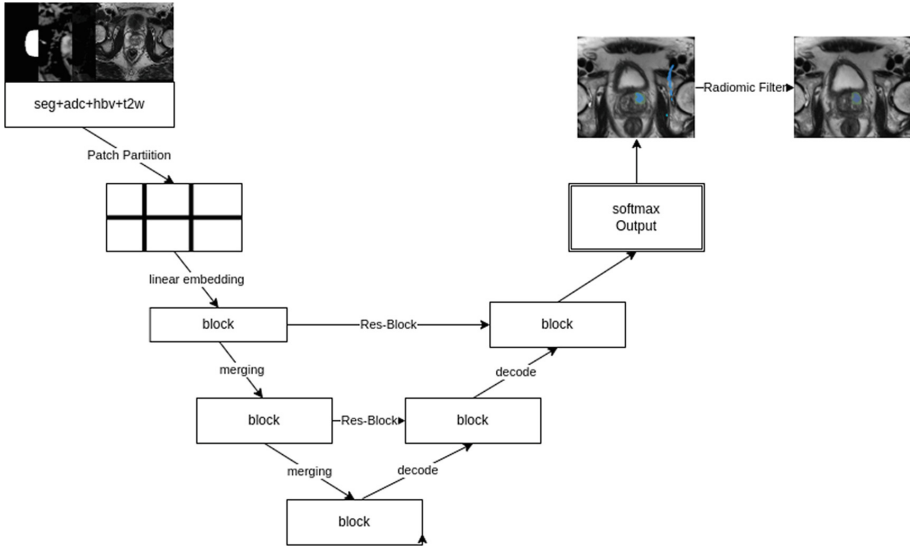


Fig. 1. A simplified schematic diagram of Swin UNETR on the basis of Fig. 1 from Hatamizadeh et al. The input comprises four channels with output of segmentation of whole gland, ADC, HBV, and T2W values [7]

2.4 Hyperparameter Selection

Hyperparameter tuning was achieved using a genetic algorithm implemented in the Optuna [2] library. Hyperparameter tuning was used in the selection of the optimizer, architecture, and optimizer-related decisions like the Learning Rate Scheduler.

2.5 Postprocessing

The training was conducted as a five-fold cross-validation using splits provided by the contest organizers, and the outputs of each fold were combined by a mean ensemble algorithm. The model's output was passed through a sigmoid activation function before lesion candidates were extracted using the report guided annotation library [3]. The proposed lesions were analyzed further by assessment of simple radiomic characteristics that are important for the task at hand; this can help increase the model's precision by filtering out some false positive results. Proposed lesions were assessed for their:

- size, where too big and small lesions were filtered out;
- elongation and roundness, where highly elongated changes were filtered out, as they typically represented the obturator internus muscle or some of the large vessels in the pelvis;
- the hypointensity of the ADC map and the hyperintensity of a high b-value DW image, defined as the difference of the mean value of complementary

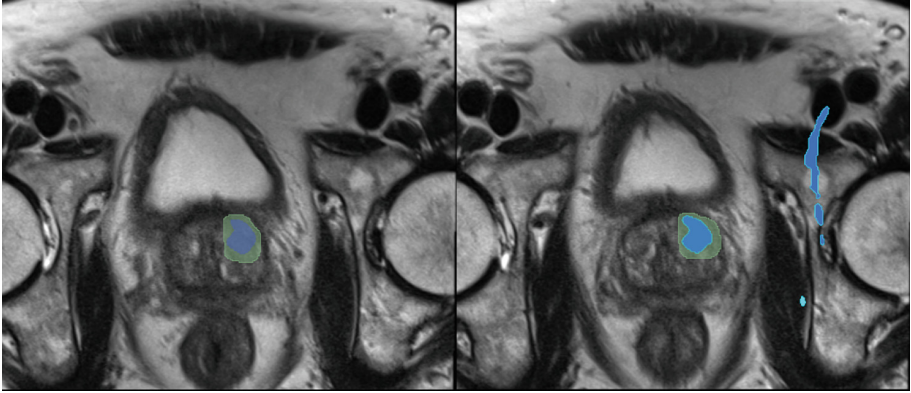


Fig. 2. A transverse T2W image of the prostate. In green, the gold standard label indicates prostate cancer; in blue, the changes detected by the neural network (and on the left, after being filtered by post-processing).

modalities concerning a lesion’s neighborhood. As the presence of hyperintense lesions on a high b-value DW image with related hypointense signal intensity on the ADC map is typical for prostate cancer, lesions that failed to meet this criterion were filtered out.

Figure 2 presents an example of the algorithm output, before and after the changes are filtered out by their radiomic features.

Table 1. A summary of the simple shape statistics of segmented instances

		elongation	physical size	roundness
TP	Min	1.0	18.3	0.4
FP		0.0	0.4	0.1
TP	Max	3.2	40726.5	1.2
FP		7.8	75598.7	1.7
TP	Median	1.4	1512.1	0.8
FP		1.9	1402.5	0.9
TP	STDEV	0.4	5227.1	0.1
FP		0.8	3789.3	0.2

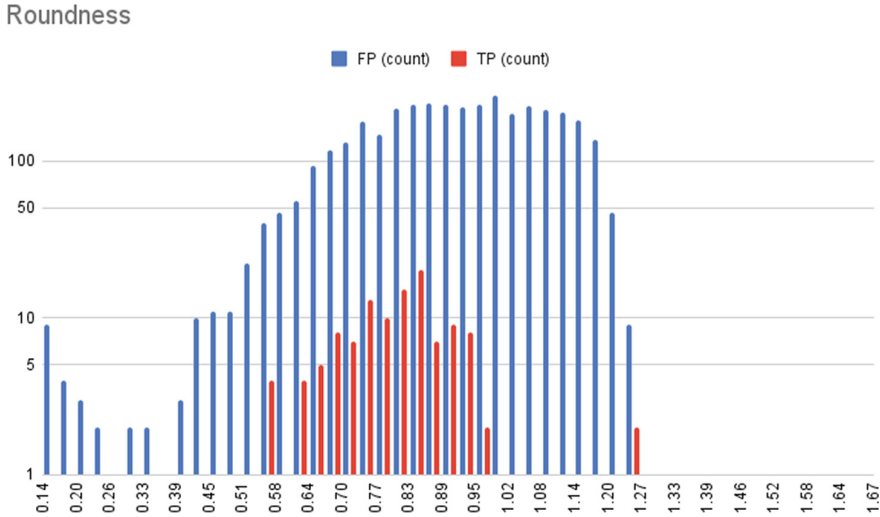


Fig. 3. A histogram of the distribution of roundness values: in blue, false positives; in red, true positives. The number of samples per histogram bin is scaled logarithmically.

3 Results and Discussion

Validation of the algorithm was performed using the Picai evaluation library [15] on the validation dataset provided by the contest organizers. Preliminary results for the model give a Ranking Score of 0.531, Area Under the Receiver Operating Characteristic curve of 0.686, and Average Precision of 0.376. An analysis of simple radiomic characteristics was performed and is summarized in Table 1. For each measured quantity—elongation, physical size, and roundness—the incorrect segmented instances presented approximately two times higher standard deviations, which indicates far higher variability. This also suggests a far wider distribution of the aforementioned quantities and the possibility of identifying suitable thresholds that define some of the segmented instances as false positives with high probability. As an example, in Fig. 3, one can observe that in the dataset, all segmented instances with roundness lower than 0.4 were false positives. A similar analysis can be performed for all other quantities. However, final conclusions regarding increases in model specificity using radiomic-based postprocessing require further study.

The results suggest that the model performs comparably to the state-of-the-art non-transformer-based baseline architectures provided by the contest organizers. However, a significant number of the top-ranking results that are presented on the contest leaderboard are based on transformer architectures. This demonstrates their impressive ability to learn the presented task and the presence of further opportunities for optimization.

4 Conclusions

This study indicates the usefulness of new transformer-based architectures in multimodal three-dimensional medical imaging. An additional feature considered necessary for analyzing the dataset is the proven ability of transformer-based architectures to incorporate data from different sources [16]. This provides a strong base for incorporating clinical data directly into the neural network architecture. Radiomic analysis performed in the postprocessing step proved helpful in the study by increasing the model’s specificity; work on more advanced radiomic analysis is fully justified. The use of model PyTorch-based libraries enabled efficient training, which supplies further proof of its efficiency. Such tools can serve as the basis for additional work on the algorithm’s development.

Acknowledgements. This work has been funded by the Polish National Centre for Research and Development as part of the program, INFOSTRATEG I, project INFOSTRATEG-I/0036/2021 “AI-augmented radiology - detection, reporting and clinical decision making in prostate cancer diagnosis”.

References

1. Artificial intelligence and radiologists at prostate cancer detection in MRI: The PI-CAI challenge. <https://pi-cai.grand-challenge.org/PI-CAI/>
2. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
3. Bosma, J., Saha, A., Hosseinzadeh, M., Sloopweg, I., de Rooij, M., Huisman, H.: Report-guided automatic lesion annotation for deep learning-based prostate cancer detection in bpMRI (2021)
4. Diaz-Pinto, A., et al.: Monai label: A framework for AI-assisted interactive labeling of 3d medical images (2022)
5. Falcon, W., et al.: Pytorchlightning/pytorch-lightning: 0.7.6 release (2020). <https://doi.org/10.5281/zenodo.3828935>
6. Grönberg, H.: Prostate cancer epidemiology. *Lancet* **361**(9360), 859–864 (2003). [https://doi.org/10.1016/S0140-6736\(03\)12713-4](https://doi.org/10.1016/S0140-6736(03)12713-4), <https://www.sciencedirect.com/science/article/pii/S0140673603127134>
7. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.: Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images. In: Crimi, A., Bakas, S. (eds.) BrainLes 2021. LNCS, vol. 12962, pp. 272–284. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08999-2_22
8. He, K., et al.: Transformers in medical image analysis: a review (2022). <https://doi.org/10.48550/ARXIV.2202.12165>, <https://arxiv.org/abs/2202.12165>
9. Isensee, F., Jaeger, P., Kohl, S., Petersen, J., Maier-Hein, K.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 1–9 (2021). <https://doi.org/10.1038/s41592-020-01008-z>
10. Isensee, F., et al.: batchgenerators - a python framework for data augmentation (2020). <https://doi.org/10.5281/zenodo.3632567>

11. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts (2016). <https://doi.org/10.48550/ARXIV.1608.03983>, <https://arxiv.org/abs/1608.03983>
12. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization (2017). <https://doi.org/10.48550/ARXIV.1711.05101>, <https://arxiv.org/abs/1711.05101>
13. Merkel, D.: Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* **2014**(239), 2 (2014)
14. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024–8035. Curran Associates, Inc. (2019). <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
15. Saha, A., et al.: Artificial intelligence and radiologists at prostate cancer detection in MRI: the PI-CAI challenge (study protocol) (2022). <https://doi.org/10.5281/zenodo.6667655>
16. Serdyuk, D., Braga, O., Siohan, O.: Transformer-based video front-ends for audio-visual speech recognition for single and multi-person video (2022). <https://doi.org/10.48550/ARXIV.2201.10439>, <https://arxiv.org/abs/2201.10439>
17. Singla, D., Cimen, F., Aluganti, C.: Novel artificial intelligent transformer u-net for better identification and management of prostate cancer. *Mol. Cell. Biochem.* **478**, 1439–1445 (2022). <https://doi.org/10.1007/s11010-022-04600-3>
18. Smith, L.N.: Cyclical learning rates for training neural networks (2015). <https://doi.org/10.48550/ARXIV.1506.01186>, <https://arxiv.org/abs/1506.01186>
19. Wu, C.Y., et al.: Memvit: memory-augmented multiscale vision transformer for efficient long-term video recognition, pp. 13577–13587 (2022). <https://doi.org/10.1109/CVPR52688.2022.01322>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Sales Forecasting During the COVID-19 Pandemic for Stock Management

Enver Yildirim¹, Veli Cam¹, Fatih Balki¹, and Salih Sarp²(✉)

¹ Aurora Bilisim, Istanbul, Turkey

{enver.yildirim,veli.cam,fatih.balki}@aurorabilisim.com

² Virginia Commonwealth University, Richmond, USA

sarps@vcu.edu

Abstract. Stock management is very important for the companies to supply necessary demand for the products they sell, to pricing the products and to the aspect of storage cost. In stock management, the products to be sold to the customer are procured by ordering from the vendors. The orders given to the vendors are determined by estimating the sales quantities of the products. When estimating sales, if we order in large quantities, the storage and expiration dates of the products may exceed, or if we order less than demand, the customer cannot find the product in the store. With the Covid-19 pandemic entering our lives, there have been some changes in our habits. One of these changes is the change in shopping habits of people due to the isolation period. By managing this change in terms of stock management on the store side, it ensures that people can reach the products they demand in these difficult times and that companies do not create extra costs by making more stock than necessary. We made a sales forecast on 5-lt sunflower oil which is a basic food product using the data of a grocery chain with machine learning methods and developed models to use these forecasts in stock management. Our data is multivariate and contains both quantitative and qualitative features. In our study, we used the supervised learning method and the XGBoost, LGBMRegressor and Ridge models used in many machine learning projects. As a result of our studies, an improvement of approximately 25% has emerged with the features we added specifically for the pandemic.

Keywords: Machine Learning · Demand Forecasting · Decision Support System · Stock management

1 Introduction

Markets are workplaces where customers can meet their needs. It is essential that the market supplies the needs of the customers, namely the demand. If the market does not present the customer's needs to the customer, this may lead to customer dissatisfaction and loss of customers for the market [1]. The reputation of the stores in the eyes of the customers is also important, their customers want to buy the product they want at an affordable price. Inventory management plays an important role in customers' access to products and in affordable prices compared to other competing markets [2]. Effective stock management benefits the market and the customer [3].

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 111–123, 2023.

https://doi.org/10.1007/978-3-031-37649-8_12

Inventory management is to ensure that the products offered by the markets to their customers are supplied from the suppliers and that the customers can reach them on the shelves of the market at more affordable prices than the competitors of the market. There are two criteria here; It should be ensured that the customer has continuous and affordable access to the product [4]. When the criteria are met, customer satisfaction will increase the profitability rate in the market.

In order to be able to manage the stock, it is necessary to predict how much of each product will be sold [5]. If the sale of the product is underestimated and the customer cannot reach the product, it will cause loss of customer and reputation in terms of the market. If we overestimate the sales of the product, we have to store the product, which increases our warehouse cost and therefore the cost of the product. It makes it difficult to offer affordable products to our customers. In another case, when we have excess demanded product and the expiry date of the product has passed, we can no longer sell the products and the cost of the product increases [6]. However, it causes additional costs because the products in the warehouse or on the shelf have to be removed for the market.

In extraordinary times such as stock management pandemics, it becomes even more important that the demand, which consists of the needs of the people, can be supplied by the market [7]. This situation, which is out of normal, causes abnormality in sales. Especially during the pandemic period, together with the problems experienced in logistics, stock management and thus sales forecasting are of vital importance in presenting the required products to customers [8].

Sales forecasting provides continuation of sales before stocks run out and provides real-time forecasts suitable for all situations [9]. Statistics, mathematical models, machine learning and deep learning etc. on stock management. Many techniques are used [10]. Instead of a rule-based model [8], we decided to examine this issue using a machine learning technique that could extract the change in sales from the data. First of all, we made general analyzes on our data and tried to understand our data. In the light of the analyzes we made, we tried to find features that would positively affect the result while predicting sales by performing feature engineering on our data. We developed a model using our data and newly discovered features and the XGBoost machine learning model [11], which is frequently used in making predictions. In our study, with the development of the pandemic feature, an improvement of approximately 25% has emerged in sales forecasts.

2 Dataset

Our data set includes sales records of 5-lt sunflower oil product in all branches between 01-05-2019 and 01-05-2021 of a chain market consisting of 10 branches located in various parts of Turkey. Data set contains 1613 daily records. The selected time period also includes the COVID-19 pandemic period so that the pandemic effect can be examined. With the COVID-19 pandemic, there have been unusual curfews and serious difficulties related to logistics [12]. Limits have been imposed on the working hours of the markets and the number of customers inside. At the same time, restrictions were introduced for a certain segment of people to go out to the street, depending on their age. Such restrictions on both markets and customers have led to significant changes in sales [13]. The features included in our data are given in Table 1.

Table 1. Data information.

Property	Explanation
Date	Date of the sale
Sales Quantity	Amount of sales made during the day
Promotion	Promotional information on the product
Available Per Day	Existing product found at the beginning of the day before the sale starts
Sale price	Selling price of the product
Cost	The cost of the product to the market

Using the date information here, the period without a pandemic before March 2020, when the pandemic effect was seen, and the period after it were described as the pandemic period.

3 Methodology

3.1 Problem Identification

Figure 1 shows our methodology. The methods that are actively used to forecast the sales of the product in non-pandemic times have changed due to changes and limitations in shopping habits during the pandemic. When we use the machine learning algorithm, which is trained with normal data, under pandemic conditions, it makes bad predictions compared to the previous period estimates due to the change in the data. Due to this change, the previous forecasting model will have a high margin of error during the pandemic period. In this study, it is aimed to prevent this.

3.2 Data Preparation

After our data was obtained from the database. It was pre-processed as follows in order to eliminate the errors and deficiencies found in our data. Some of these errors are Identifying the missing areas in our data and removing them from our data will prevent our model from making mistakes and making biased learning during the training. Features in the dataset are typecasted in order to extract additional features. For example: date columns transformed string to datetime type.

3.3 Exploratory Data Analysis

In this study, only human exposure estimation was made without distinguishing activity. Exploratory data analysis provides the opportunity to summarize our data, discover the features of our data, understand the relationship between features, and detect abnormal data by analyzing our data with statistical methods and data visualization methods. These analyzes can also be used during feature extraction processes.

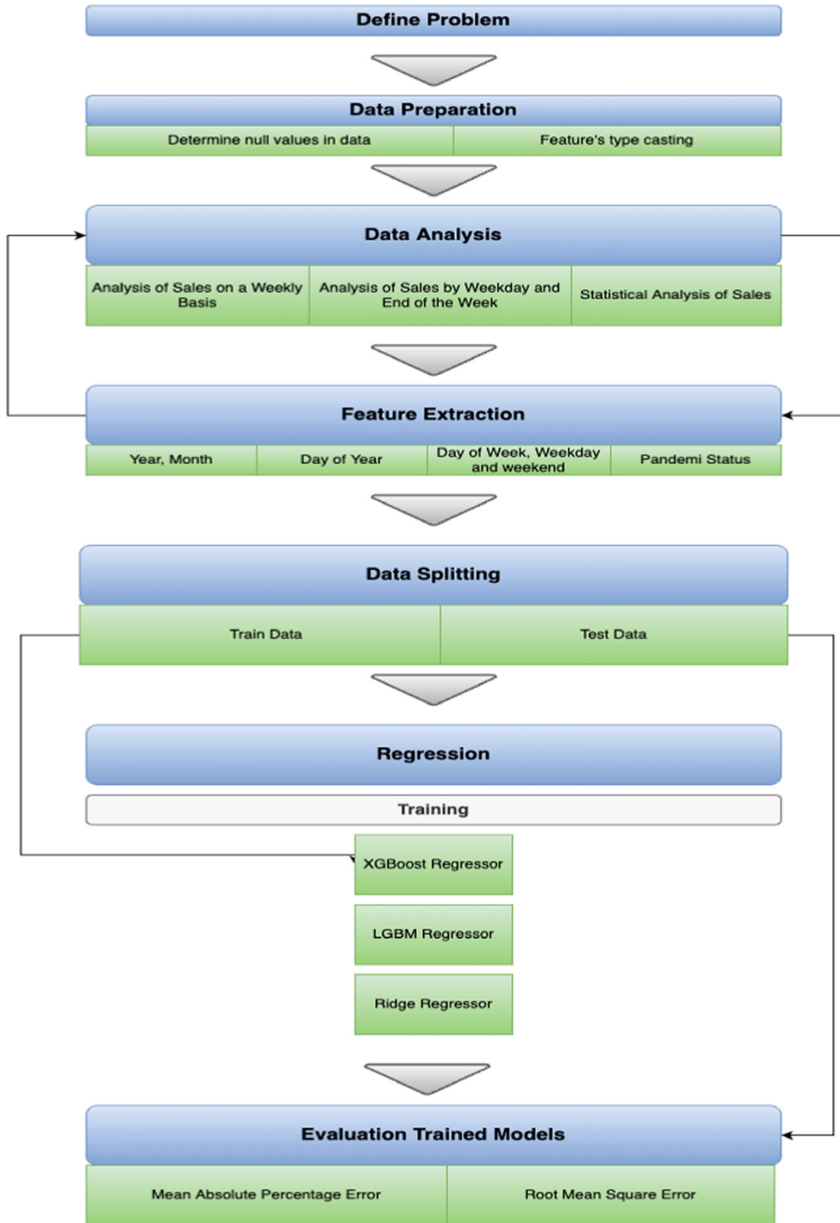


Fig. 1. Flowchart for our processes in sales forecasting

General information can be obtained by examining our data statistically first. When we examine Table 2, where the data are analyzed statistically, we can see that the sales figures of the pandemic period have decreased in general. We analyzed the sales data of our product statistically: the number of records, average sales, standard deviation,

minimum value, maximum value and data distribution of 25%, 50%, 75%. We obtained these values in Table 2 using the `print(df.describe())` function of the pandas library. As can be seen in the table, there has been a decrease of approximately 30% in the average of sales. Changes in other areas are also at this level, and sales have generally decreased as expected during the pandemic period (Fig. 2).

Table 2. Data information.

Criteria	Pandemic Era	Regular Era
Data Number	1186	427
Average sale	19	31
Minimum	0	0
25%	9	13
50%	14	20
75%	22	35
Maximum	368	730

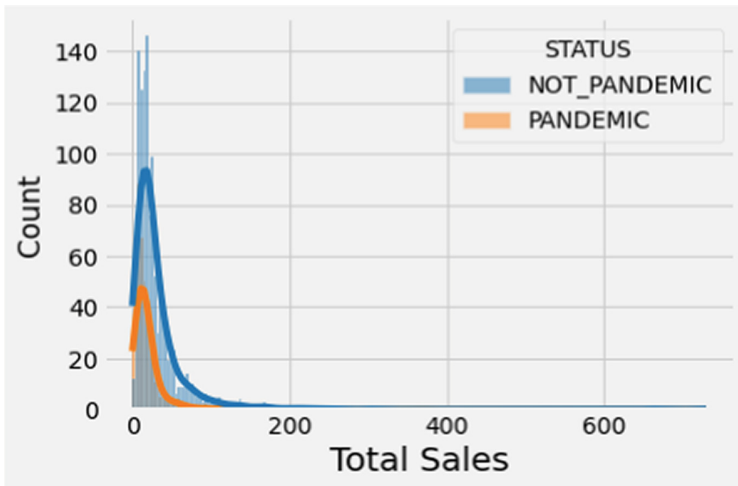


Fig. 2. Distribution of data

In order to see the changes in our sales data with and without pandemics on the basis of days of the week, we obtained Figs. 3, 4 and 5 using the seaborn library and column (bar plot) and line (line plot) graphics. Percentage changes are expressed in the bar chart.

In the graphs in Fig. 4, a column chart of the sales in the pandemic period (shown in red) and in the non-pandemic period (shown in green) is shown by weekdays. In this chart, the change in sales is observed due to the expected weekend closure. Percentage changes are given on the column. When we interpret the graph, we can see that the sales

in the normal period are high on the weekend. In the pandemic period, we can see that sales are less due to the weekend closure.

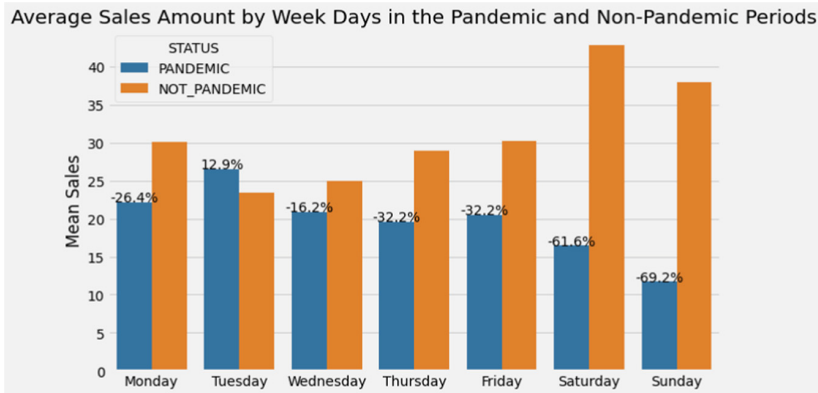


Fig. 3. Average sales by weekdays during the pandemic and non-pandemic period.

In order to understand the changes in sales on weekdays and weekends, we showed our data in the form of time with and without pandemics, as in Fig. 3, with a column chart as the average sales on weekdays and weekends. When we examine these graphics, we need to update our model that we use in normal times, because the sales character has changed during the pandemic period. In order to predict sales during the pandemic period, a model suitable for the development of our model should be put forward by adding new features.

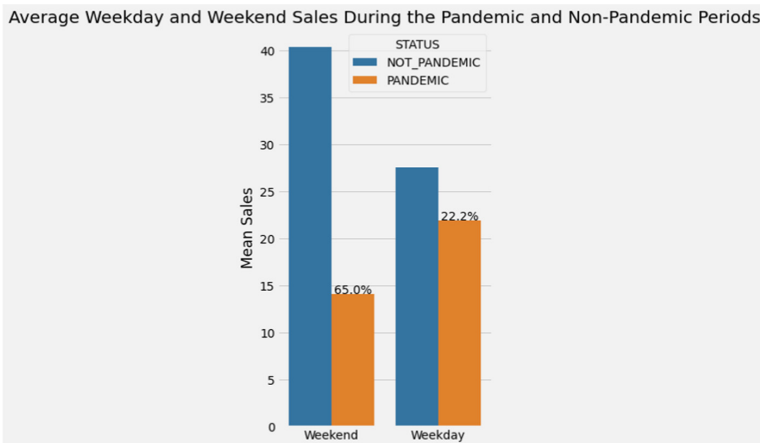


Fig. 4. Average weekday and weekend sales in the pandemic and non-pandemic period.

As seen in the line graph in Fig. 5, sales follow a decreasing trend on weekends.

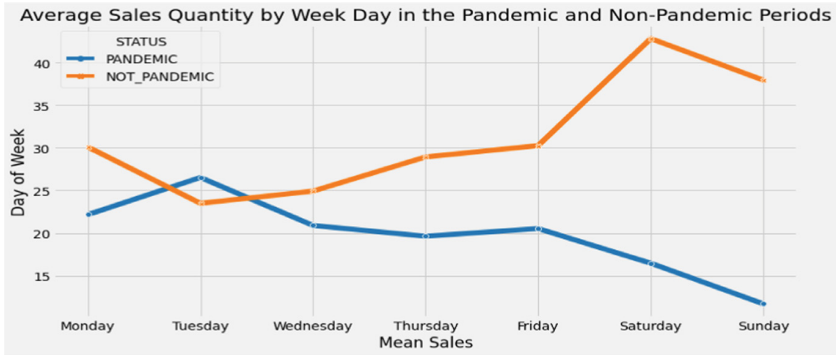


Fig. 5. Average sales per weekday in the pandemic and non-pandemic period

3.4 Feature Extraction

In the light of the graphics in the analysis section, the pattern and characteristics of our data have changed due to the change in sales habits during the pandemic period. Because of this change, when our machine learning models are trained with the data before the pandemic, the consistency in the predictions is lost because the pandemic data is different. For this, we can add new features, taking into account the information in the analysis section, so that our model can better learn the changing data and non-pandemic data. In this way, we can better predict the sales during the pandemic period and make better stock management by placing the orders accordingly. We use pandas library's ready-made functions on time data types while extracting properties.

These features were extracted from the time feature of our data in order to be able to deal with the year, month, week of the year, which day of the year, which day of the week, and when and in which situation the sale was made. Thanks to these features, we expect our model to establish a connection between sales data and time and to better learn the sales values made at the same time. We evaluate these temporal features categorically, and we want our model to evaluate it that way. Since there will be a connection between time-based sales, records with the same time feature will have similar values and this approach will enable our model to learn better.

We extract a true-false property from our time feature, whether it is weekdays or weekends. Weekday and weekend sales are different in the retail industry. It is said by retail experts that people shop more during the holidays. In order to express this situation in our data, we add this feature to our data.

The character of the sales during the pandemic period, which we observe in the graphics, is changing. We add a categorical feature to our data that takes values as pandemic and non-pandemic when it is in the pandemic period.

3.5 Dataset Separation

We divide our data into two as data to be used in training the models and then as data to be used in testing in order to measure the predicted performance of the trained model. When dividing the data, considering that our data is a time series, we determine the training and test parts by shifting. In order to see the change over time in the time series data, the part immediately after the data taken for training is taken for testing. Figure 5 illustrates this approach. Each shift is evaluated as a step, and the model is trained with the training data at each step, and evaluation metrics are obtained with the test data. Then the average of the scores obtained in all steps is taken.

Time series cross validation method is used before a certain time for training and after for testing, so we can objectively train time series data and measure the performance of our model. We did it using the time series cv algorithm of the sklearn library.

```
tscv = TimeSeriesSplit(n_splits = 3, test_size = 2).
```

Example: Size of data set = 1613.

1. Fold = 1
 - a. Train size = 1607
 - b. Test size = 2
2. Fold = 2
 - a. Train size = 1609
 - b. Test size = 2
3. Fold = 3
 - a. Train size = 1611
 - b. Test size = 2

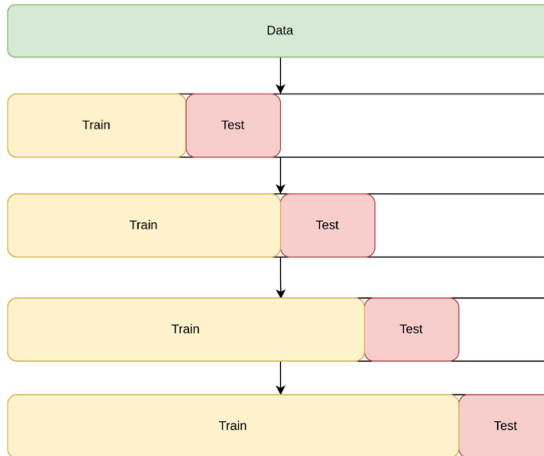


Fig. 6. Time series cross validation method.

3.6 Regression and Machine Learning Model

Regression is a statistical measurement that tries to determine the strength of the relationship between a dependent variable and other independent variables, and makes predictions according to this power. It is used as a predictive modeling method in machine learning where an algorithm is used to predict the dependent variable.

Solving regression problems is one of the most common applications for machine learning models, especially in supervised machine learning. Algorithms are trained to understand the relationship between independent variables and an outcome or dependent variable. The model can then be used to predict the outcome of new and invisible input data or to fill a gap in the missing data.

Machine learning is a sub-branch of artificial intelligence algorithms. It has a structure that can learn on data, and allows us to make predictions after this learning. This model takes an input and an output data as training data. The algorithm establishes a function by learning between the input data and the output data and learns the statistical pattern in our data. Estimates are made with the learned model.

Here, we decided to use the XGBoost and LGBMRegressor model with the tree-based gradient boosting [14] feature, which is the most popular of the machine learning algorithms, and the linear-based Ridge model. This choice was made taking into account both the overall success of the models in machine learning and the number of records of our data. Due to the scarcity of data we have, deep learning methods were not preferred.

XGBoost and LGBMRegressor algorithms are decision trees-based machine learning algorithms that use gradient boosting. XGBoost has brought some improvements over plain GBM, such as the use of regularization, pruning, and parallelization to prevent over-learning. It works faster than other algorithms with its parallel operation. It is used in many projects and competitions. Since the XGBoost library is an open source project, it is developed and supported by many users.

XGBoost, short for Extreme Gradient Boosting, is a scalable distributed gradient assisted decision tree (GBDT) machine learning library. It provides parallel tree reinforcement and is the leading machine learning library for regression, classification and sorting problems. XGBoost, a supervised machine learning method, uses algorithms to train a model to find patterns in a dataset containing tags and features, and then uses the trained model to predict tags in a new dataset's features.

LGBMRegressor has the following differences from XGBoost. Instead, it grows trees in the form of leaves. He chooses the leaf he believes will provide the greatest reduction in loss. Also, LightGBM does not use the commonly used rank-based decision tree learning algorithm, which looks for the best split point in the sorted feature values, as XGBoost or other applications do. Instead, LightGBM implements a highly optimized histogram-based decision tree learning algorithm, which provides huge advantages in both efficiency and memory consumption. The LightGBM algorithm uses two new techniques called Gradient-Based Unilateral Sampling (GOSS) and Special Feature Packing (EFB), which allow the algorithm to run faster while providing a high level of accuracy.

Ridge regression is a method of estimating the coefficients of multiple regression models in scenarios where the independent variables are highly correlated. It has been used in many fields, including econometrics, chemistry, and engineering.

3.7 Performance Evaluation

The metrics allow us to evaluate our model's predictive performance after training. It allows us to measure how successful a model trained with training data is on the test data, or how wrong predictions it makes. We will use the MAPE [15] and RMSE [16] metrics to measure our model. These metrics are frequently used in regression tasks.

The problem we are trying to solve is a regression problem. Among the most preferred metrics in regression problems are MAPE and RMSE. The reason why we use two different metrics at the same time here is to be able to look at our mistakes from two different angles and to be more objective.

MAPE; Average absolute percent error (MAPE), also known as mean absolute percent deviation (MAPD), is a measure of the error in estimation of a forecasting method in statistics. The MAPE scale is set out in Eq. 1.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{true} - y_{pred}}{y_{true}} \right| \times 100 \quad (1)$$

Here, A_t , T_t , and n represent the actual value, the predicted value, and the number of cases tested in the total data set, respectively. The MAPE metric is our error expressed as a percentage. Calculation of the error as a percentage will reflect the error more objectively, if the magnitude of the values change as a scalar, since this is calculated as a percentage to the error.

RMSE; The mean square deviation (RMSD) or mean square error (RMSE) is a commonly used measure of the difference between values predicted by a model or estimator (sample or population values) and observed values. The RMSE scale is shown in Eq. 2.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{true} - y_{pred})^2} \quad (2)$$

The abbreviations used in the RMSE metric are also used in the same sense as MAPE. The performance of the extracted feature and the operated model with these two performance scales are given in Table 3 and Fig. 6.

4 Findings and Interpretation

By training our data in a comparative way, we will make evaluations by observing the effect of the features we add on the result. Two datasets were created, the first contains features related to the pandemic and the other does not. We trained these datasets with XGBRegressor, LGBMRegressor and Ridge machine learning models, and evaluated our models with the performance metrics that occurred in the estimation they made on the test data.

The results we obtained with the test data after the training are given in Table 3 and Fig. 6. Considering these results, our model makes 25% improvement in the one step ahead sales prediction (Table 4).

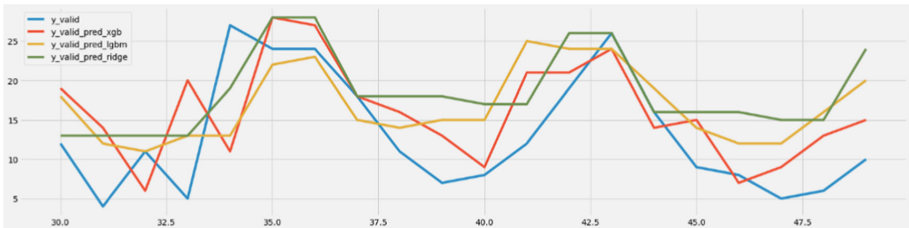
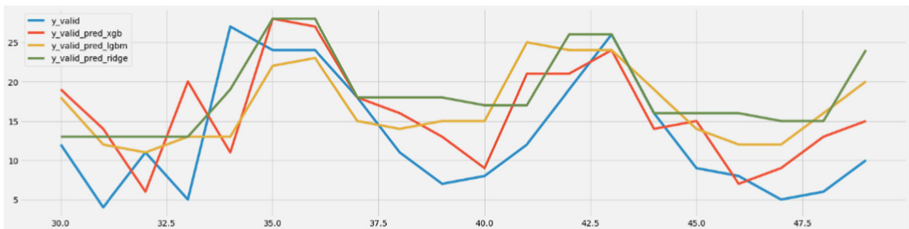
Table 3. Data information.

Era	Metrics	
	MAPE	RMSE
Using Pandemic specific model	48	6.29
Regular model	36	4.61

Table 4. Data information.

Metrics	XGBRegressor		LGBMRegressor		Ridge	
	Pandemic	No Pandemic	Pandemic	No Pandemic	Pandemic	No Pandemic
RMSE	3.11	6.09	6.52	9.94	6.69	7.01
MAPE	24.82	64.11	38.35	58.35	59.84	63.61

Figure 7 in below shows the estimation of our model on the test data graphically. As seen in this way, our model generally catches the sales trend, but makes an one step ahead prediction that sometimes less or more sales prediction will be made in sales (Fig. 8).

**Fig. 7.** Model predictions with Pandemic feature (blue model prediction, red actual test data)**Fig. 8.** Model predictions without pandemic feature

5 Conclusion

If the market chains cannot manage the stock well, they may face the risk of loss of reputation in the public, storage due to excess stock, or deterioration of the product due to insufficient supply to the incoming demand. Any market chain sells around 10000 products and the stock management of this large-volume product portfolio requires serious resources. Here, XGBoost, LGBMRegressor and Ridge machine learning models were preferred as a method where we can efficiently manage stocks through algorithms created with today's developing techniques and data obtained, update our data despite changing conditions, and provide rapid training because there are so many product types.

As a result of our work, we have developed a new feature that can be used for the pandemic period, and by adding this feature to other already existing features, we have improved the consistency of our predictions.

In the future, when feature extraction studies on our data reach the desired point and the number of data increases, we are considering using deep learning methods, feature extraction and models. The amount of data is of great importance for deep learning methods.

With this developed model, the sales forecast of the products in the portfolio of a market can be made and an adequate stock management can be realized in this way. In this way, the stock management of the entire market will be fully automated with the recommended machine learning model.

Production and consumption, which is the basis of the economy, is directly related to stock management. This study will be an example to the literature for information and guidance for each company working on stock management.

References

1. Erceg, Ž., et al.: A new model for stock management in order to rationalize costs: ABC-FUCOM-interval rough CoCoSo model. *Symmetry* **11**(12), 1527 (2019)
2. Li, R., Chiu, A., Seva, R.: A process-based dead stock management framework for retail chain store systems. In: *RSF Conference Series: Business, Management and Social Sciences*, vol. 2, no. 1 (2022)
3. van Ryzin, G., Mahajan, S.: On the relationship between inventory costs and variety benefits in retail assortments. *Manag. Sci.* **45**(11), 1496–1509 (1999)
4. Chase Jr., C.W.: What you need to know when building a sales forecasting system. *J. Bus. Forecast.* **15**(3), 2 (1996)
5. McCarthy, T.M., et al.: The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. *J. Forecast.* **25**(5), 303–324 (2006)
6. Chen, C.-Y., et al.: The study of a forecasting sales model for fresh food. *Expert Syst. Appl.* **37**(12), 7696–7702 (2010)
7. Gupta, C., Poddar, S., Ghosh, A.: The consumer demand recovery beyond the pandemic. *Int. J. Manag. Hum. Sci. (IJMHS)* **5**(3), 60–68 (2021)
8. PriyaDarshani, M., Sinha, M.P., Sinha, K.: A study on evolutionary technique to predict the sales during COVID-19. In: *Handbook of Research on Library Response to the COVID-19 Pandemic*, pp. 376–402. IGI Global (2021)
9. Kiracı, M.: Stok yönetimi ve karlılık ilişkisinin finansal oranlar aracılığıyla incelenmesi: İMKB imalat sektöründe bir araştırma (2009)

10. Cadavid, J.P.U., Lamouri, S., Grabot, B.: Trends in machine learning applied to demand & sales forecasting: a review. In: International Conference on Information Systems, Logistics and Supply Chain (2018)
11. Chen, T., et al.: XGBoost: extreme gradient boosting. R package version 0.4-2 1.4 1-4 (2015)
12. Açıkgöz, Ö., Günay, A.: The early impact of the Covid-19 pandemic on the global and Turkish economy. *Turk. J. Med. Sci.* **50**(9), 520–526 (2020)
13. Priya, S.S., Cuce, E., Sudhakar, K.: A perspective of COVID 19 impact on global economy, energy and environment. *Int. J. Sustain. Eng.* **14**(6), 1290–1305 (2021)
14. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
15. De Myttenaere, A., et al.: Mean absolute percentage error for regression models. *Neurocomputing* **192**, 38–48 (2016)
16. Botchkarev, A.: Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. arXiv preprint [arXiv:1809.03006](https://arxiv.org/abs/1809.03006) (2018)
17. <https://en.wikipedia.org/wiki/LightGBM>
18. https://en.wikipedia.org/wiki/Ridge_regression

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Digital Interaction



Seeking Emotion Labels for Bodily Reactions: An Experimental Study in Simulated Interviews

Debora C. Firmino De Souza^(✉), Pia Tikka,
and Ighoyota Ben Ajenaghurure

Tallinn University, Tallinn, Estonia

deboracs@tlu.ee, pia.tikka@tlu.ee, ighoyota@tlu.ee

Abstract. Computers rely on different methods and approaches to assess human affective information. Nevertheless, theoretically and methodologically, emotion is a challenging topic to address in Human-Computer Interaction. Exploring methods for assessing physiological responses to emotional experience and for aiding the emotion recognition features of Intelligent Virtual Agents (IVAs), this study developed an interface prototype for emotion elicitation and simultaneous acquisition of the user's physiological and self-reported emotional data. Supplementary, the study ventures to combine such data through event-related signal analysis.

Keywords: Emotion Recognition · Multimodal Signal Acquisition · Self-report methods

1 Introduction

Humans benefit from emotional interchange as a source of information to adapt and react to external stimuli and navigate their reality. Computers, on the other hand, rely on classification methods to do so. It uses models to calculate and differentiate affective information from other human inputs because of the emotional expressions that emerge through human body responses, language, and behavior changes. The present study is structured to investigate methods for identifying and interpreting variations of physiological responses related to emotional states as a way of improving emotion recognition in interactive systems. The study frame observes emotional responses during interactions with Intelligent Virtual Agents (later IVAs) in a simulated context.

In recent years, increasing investments have introduced IVAs in customer service, education, health care, entertainment, immigration settings, and social media. IVAs encapsulate the embodiment of different interactive channels and establish meaningful communication with humans by enacting emotions, empathy, and social behavior [6]. The social and emotional capabilities displayed by IVAs motivate the users to establish empathy and bonding [5]. Moreover, IVAs' real-time perception, cognition, and emotional awareness bring novel solutions for human-machine cooperative tasks [21] within social domains.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 127–138, 2023.

https://doi.org/10.1007/978-3-031-37649-8_13

1.1 Research Goal and Motivation

This study investigates methods for assessing psychophysiological responses to emotional experiences evoked in a simulated IVA interview. Closely, the project reflects on the implications of IVA's emotion recognition for applications in real-life scenarios, like easing the critical situation of migration management in Europe [7, 8]. Ultimately, the study encourages an adequate use of emotional information and thoughtful development of automated mediators to aid in contexts involving high-stakes decisions that shape people's life chances [8, 23]. Identifying the asylum seekers' narratives aspect which proves their situation of fear is essential to granting asylum-seeker protection, as the Refugee Status Determination (RSD) procedure assesses applicant's "well-founded fear of being persecuted" due to "happenings of major importance" and, for that unable to return to their home countries ([9] - Paragraphs 34, 36, 32–110). Therefore, to automate stages of the RSD assessment, it is critical to explore the design of emotion recognition methodologies to assist automated migration management processes.

1.2 Hypotheses and Research Question

Two hypotheses are formulated to guide the development of the research:

- H1: People react emotionally to a simulated interview interaction with a virtual agent, even in as-if contexts where they are assigned specific roles ("Imagine that you are -")
- H2: Individual human subjects' responses to the perceived affective information in simulated settings can be identified and related to a set of emotional states based on a combination of quantitative (psycho-physiological) and qualitative (subjective reports) measures of their behavior.

Correspondingly, three research questions are investigated:

- RQ1: How do people emotionally appraise the context of the interview?
- RQ2: Can the individual subjective self-reported emotional experience (qualitative data) be associated with features recognized from the same individual's physiological behavior (quantitative data)?
- RQ3: Within the context, is it possible to identify certain emotional reactions by looking at the patterns of physiological data?

2 Theoretical Background

Psychology researchers have waged an endless debate since William James (1884) interrogated what emotion is. Nevertheless, there is little consensus around the definitions [11], albeit the diverse answers presented. Damasio's studies described that humans use the body as a theater for emotions, supporting that the mind is embodied and the experience of emotions is a manifestation of the drives and instincts that help the organism regulate itself and respond to changes in the environment [24]. As Fontaine and colleagues [4] points out, some elements of the emotional experience are uncontroversial:

- Certain events tend to elicit specific affective responses in organisms.
- Physiological changes, behavior expressions, behavioral intention, and direction shifts characterize these organism responses.
- The response patterns produce a conscious feeling of a certain quality in the person’s experience over a certain period.
- Emotional episodes and the associated experiences are labeled by the experiencing person and by the observers with a specific word or expression.

The appraisal theory of emotion explains that affective states are elicited by people’s subjective evaluation or appraisal of the events’ significance for their well-being and goal achievement. Also, these theories acknowledge that emotions arise from comparing individual needs and external environmental demands. Brave and Nass’s investigation remarks that the knowledge of appraisal theories serves HCI researchers in modeling and predicting users’ emotional states in real-time [10,25]. Likewise, the appraisal theory of emotions can be mapped to Artificial Intelligence concepts, like belief-desire-intention models of agency [21].

By examining its users’ psycho-physiological cues, computers can sense implicit communication by assessing affective and cognitive states. Emotion recognition is based on automated classifiers trained to identify general emotional states from patterns found in the users’ biosignals, whether explicit (i.e., measurements of facial expression and gestures) or implicit (i.e., electrocardiography or electrodermal activity measurements). Furthermore, it expands the interface’s limited input modalities (i.e., mouse, keyboard, and camera), moving the limits of interaction beyond stated and visible emotion parameters [10]. Stemmler’s studies [12] demonstrated relationships between one emotion and specific physiological changes, showing that, for instance, fear can be characterized by strong cardiovascular and electrodermal activity. Although there is evidence that it is possible to identify certain emotions through measures of bodily changes, the results are confined for multiple reasons. Issues of validity and reliability are often pointed out in research [1,3,13].

In general, emotions do not display a universal signature of physiological activity for different people, as emotions arise from a unique and personal experience for each individual, and the relationship between appraisal of a situation and emotional response is, to a great extent, context-bound [10]. Nonetheless, recent studies suggest that to foster the development of the classifiers, combined measurements of psycho-physiological activity during emotional events are suitable strategies for sorting out overlapping signal problems [11,22]. Moreover, self-assessment results can be correlated to the physiological metrics to classify emotions measured from various parameters [14].

3 Methodology

This study adopts an experimental approach to examine the influence of questions raised by an IVA on participants’ emotions in the context of an automated initial interview in the RSD assessment. The researchers developed a within-subject study consisting of one session in which participants are presented to

one hundred questions, as shown in Fig. 1 below, displayed audio-visually by a VA. Precisely, in each session, participants interact with an interview simulation where they encounter a looping video of a female avatar that articulates via voice-over questions regarding practical and personal matters (stimuli). Essentially, each question represents one unique condition. The independent variable in this study is the answer provided to the emotional self-assessment, and the dependent variable is the emotional bodily experience recorded from participants.

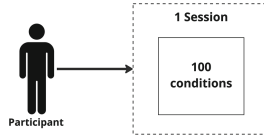


Fig. 1. In a single session, the study exposes participants to the different conditions

The display of the female VA on the screen remains constant throughout the session. Although the virtual agent has no adaptive reaction, its display on the simulation is meant to foster the enactment of an IVA presence by emulating the gaze of the computer [16].

3.1 Experiment Design

The researchers created a prototype interactive interview simulation that embeds simultaneous psycho-physiological data collection. At the experimental level, when studying participants' emotional perception via as-if situations, the narrative supports psychological immersion into the interaction context [26] and furthermore, the fictitious scenario help to situate the participants into the specific context of the research [17]. Moreover, inspired by the research on social appraisal [2, 15], it is assumed that even though participants might answer questions based on their own experience, the simulated interview context still allow them to identify with the asylum seeker's emotional situation.

In this study, context is presented as a narrative prior to the experiment and later enhanced by the simulation that situates participants into the RSD procedure and invites them to take the role of an asylum seeker. To provide stimuli, the study uses references from the list of Affective Norms for English Words (ANEW) [19] for phrasing the emotionally evocative questions. The questionnaire comprises a hundred closed questions, sixty-five percent of which are assumed to be emotionally evocative. The simulation ultimately intends to benefit from the physiological responses generated over the enunciation of the sensitive questions and does not have follow-up inquiries.

Accordingly, as shown in Fig. 2, a question is first enunciated by the virtual officer and presented as text on the screen. Once that is completed, participants are taken to the self-report screen and asked to select from the wheel the

label and intensity of the emotion they experienced. After completing the self-assessment, participants are presented with the answer options. Finally, after answering that question, participants are presented with a new one. Randomization of the questions was not included in the design of the simulation.



Fig. 2. Screenshots illustrating the loop of tasks through which participants labeled the emotional state experienced for each question presented by the virtual officer

As the experiment design establishes no fixed time for completing the simulation, every time a new question is presented to participants, a mark is created in the database to inform the start of the question. The approach is justified for not inducing pressure on participants during the simulation play.

3.2 Procedure

It is worth remarking that the experiment was piloted twice to test out a) the simulation setup and questions structure and b) data collection methods. After fine-tuning the latest adjustments, the experiments were carried out in laboratory settings under uniform conditions. Figure 3 illustrates the experimental procedure designed for the study.

Experiments were conducted during the COVID-19 pandemic; thus, the necessary safety measures stood as part of the procedure. At first, participants answered a BIS/BAS questionnaire [20] to analyze the symmetry of personality in terms of the motivational systems underlying behavior and affect responses. When conforming, participants were considered apt for the experiment. Later, information regarding the experiment, the purpose of the study, and the technical apparatus was provided. Participants were exposed to the contextual narrative upon signing the consent form. Following, the sensors were connected, as it is shown in the Data Collection section below. Before engaging in the simulation, participants were invited to perform a simple breathing exercise. It is noteworthy that, while playing the simulation, participants were alone, and monitored from a nearby room via a camera streaming setting.

3.3 Participants

Seven (7) English-speaking individuals currently residing in Tallinn, Estonia, participated in the study. Participation was voluntary, and the age range of

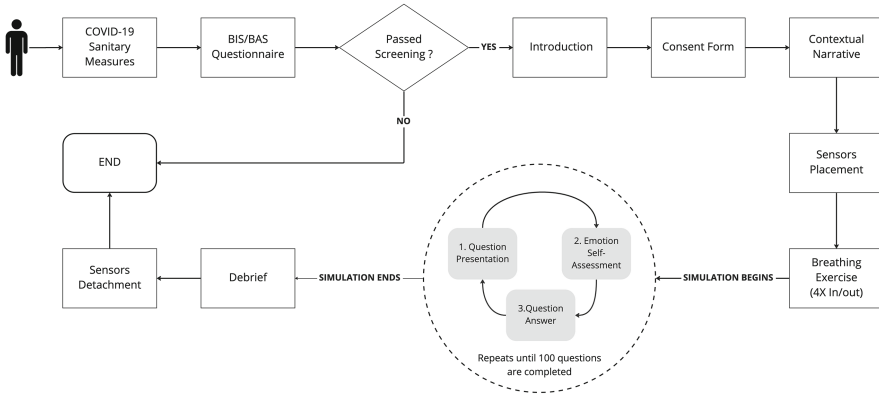


Fig. 3. Stages of the experimental procedure

participants was 18–40 years old. Even though the broader context of the study refers to the vulnerable population (refugees), no actual refugee applicant or any other vulnerable group was included in the study sample. Nevertheless, recruitment embraced participants from foreign countries that would be more likely to have a reference of the border control inquiries.

3.4 Data Collection

The simulation embeds a data collection procedure that synchronizes participants' physiological data with their emotional self-reports. The goal is to enable the participant to experience and disclose emotional states during the physiological data acquisition without disruption or further interference.

Physiological Data. Participants' physiological signals are gathered via BITalino Plugged kit, a low-cost biosignal acquisition device [27]. The sampling frequency of the used ECG is 1,000 Hz, as recommended by [28]. EDA and fEMG measurements are collected at the same sampling rate due to using a single BITalino device to collect the three physiological at the same time. The BITalino device was connected to a high-performance computer via Bluetooth. Participants' physiological information thus is collected from three channels - ECG, EDA, and EMG - and continuously recorded using the compatible software OpenSignals¹.

Sensors were placed on the clean surface of participants' skin, following recommendations of reviewed literature [22, 27]. Figure 4 illustrates the configuration used in the experiments. Two electrodes were used to measure EDA, placed on the thenar and hypothenar eminence of the right hand as illustrated in Fig. 4a.

¹ Available at <https://support.pluxbiosignals.com/knowledge-base/introducing-opensignals-revolution/>

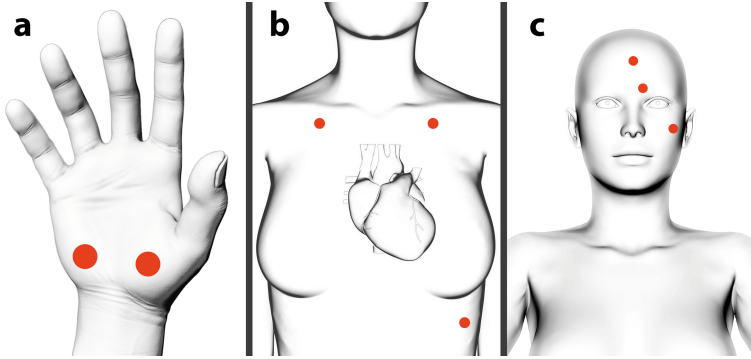


Fig. 4. Electrodes placement for biosignal data collection

For collecting ECG, three electrodes were placed on the participant’s chest, two on both sides of the clavicles and one at the lower left part of the rib cage, as it is illustrated in Fig. 4b. Additionally, facial muscle activity (fEMG) was acquired via three electrodes placed on participants’ face, the reference electrode in the middle of the forehead and the two others on the Corrugator Supercilli and Zygomaticus Major, as in Fig. 4c.

Emotion Self-report. To collect the emotional self-reports of experiment participants on the go, the interface simulation uses a digital tool based on the self-report paradigm proposed by the Geneva Emotion Wheel (GEW) [2]. Self-report paradigms are used to grasp the individual nature of emotional experience, which reflects the integration of mental and bodily changes in the context of particular events. Researchers in emotional labeling point out that even though the results obtained by these paradigms are plausible and interpretable, the statistical analysis is hampered by the abundance of emotion labels collected, making the interpretation complex [2]. This way, embedding the GEW into the simulation allowed the emotional assessment to be done interactively through an adapted version of the tool and provided homogeneity of reports. The choice also benefits the processing and synchronization of physiological data collected through the sensors.

3.5 Preprocessing Data

The multisource dataset required a composite of processing and analysis methods. Constraints faced at this stage caused the removal of the fEMG data from analysis. At first, the ECG and EDA data was downsampled from 1000hz to 100hz to reduce computational overhead. Thereafter, filtering to remove noise from the ECG and EDA data was done automatically as part of the Neurokit2data preprocessing pipeline [18]. Furthermore, we divided the ECG and EDA data into epochs with a duration of 10s, starting one second before the

question presentation and finishing nine seconds after that. Compiling the information into epochs returned a list of numeric identifiers for the physiological reactions regarding each question. The 10s epoch duration was based on the calculation of the participant's question average time (Max: 12 min 32 s, Min: 08 min 15 s). Hence, each participant produced 100 data samples, thus giving us a total of 700 data samples. Two features were extracted from ECG data, ECG Rate, and ECG R Peaks. From EDA data, we obtained EDA Phasic and EDA Tonic features. All these were achieved using the Neurokit2 python library [18].

3.6 Analysis

Following the dataset containing the labeled psychophysiological data of the sample ($N = 7$) was statistically analyzed, focusing on features of the physiological measurements acquired alongside emotional self-reports. Independent t-tests (performed two-sided) were used to draw comparisons between the emotional reports and their specific physiological reactions. A P value of 0.05 was assumed to indicate the statistical significance of the tests. Furthermore, correlation tests were performed to observe the relationship between the physiological data features and the emotion labels indicated by the participants. The statistical analysis was performed using SPSS software (ver. 28.0.1.1).

4 Results

Among the negative valence emotions, the most reported emotion was Disappointment, representing 12.9% ($n = 90$) of the total answers. Fear was reported in 9.4% ($n = 66$), followed by Sadness in 9.1% ($n = 64$) of the answers gathered. Among the positive valence emotions, Relief was the most reported, chosen at 8.0% ($n = 56$) of the time, followed by Interest, reported at 6.4% ($n = 45$), and Pride, 4.0% ($n = 28$). Issues with the self-assessment led the analysis not to include the variations of intensity reported by participants. Manifested differences between the data across participants hampered a separate analysis of the emotional reactions, leading the study to consider the emotions between negative and positive valence.

The analysis has shown that the negative stimuli increased cardiac activity in participants compared to positive stimuli. The mean of the ECG Rate for Positive emotions was $M = .6452$ ($SD = .6760$), and for Negative emotions, it was $M = .6645$ ($SD = .7092$). However, no significant difference was found in the ECG Rate between the emotions of positive and negative valence ($T_{623} = -.330$, $p = .741$). This way, tracing general conclusions about participants' emotional arousal through examination of cardiac activity was unattainable.

A similar issue was found in the features of EDA data. Measurements of the EDA Tonic were scattered and not pronounced. The little difference between the EDA Tonic values could indicate that the quick flow of questions had piled up different stimuli and affected the measurements in the tonic component, which

is distinguished for happening slowly over time. On the other hand, measurements of the EDA Phasic suggested that negative emotion might have increased electrodermal activity response if compared to the positive reports. The mean of EDA Phasic for Positive emotions was $M = -1.5396E-4$ ($SD = 1.243E-3$), while Negative emotions had a mean of $M = -4.9404E-4$ ($SD = 1.859E-3$). Additionally, the parametric test revealed a significant difference in the means of EDA Phasic between groups of positive and negative emotions ($T601.852 = 2.728$, $p = .007$).

Furthermore, Pearson's correlation tests were performed to observe the relationship between the EDA and ECG measurements. Results have indicated that ECG Rate and EDA Phasic have a significant, weak negative correlation ($r = -.316$, $p < .001$). In contrast, ECG Rate and EDA Tonic have a significant moderate, positive correlation ($r = .603$, $p < .001$). Such findings can indicate measures to consider in future studies.

5 Discussion

In an experimental investigation, we showed that our prototype enabled the simultaneous collection of participants' quantitative and qualitative affective information during a simulated interview. This pilot study allowed us to identify what needs to be iterated for the upcoming trials of this study. Moreover, the process of the study illustrated the complexities entangling psycho-physiological studies in HCI. The emotion elicitation method and designed data collection interface had a positive outcome. However, generalized conclusions regarding associations of self-reported emotional experience and the two features of physiological measurements, EDA and ECG, were not attainable.

The analysis of the ECG and EDA measurements across different emotion reports did not allow the distinction of specific emotions experienced by the participants. Still, the results aligned with the previous studies showing that negative emotions are characterized by strong cardiovascular and electrodermal activity [12]. Besides, the results suggested that combinations of measures of ECG Rate could be used alongside EDA Tonic or EDA Phasic to detect positive and negative emotions.

The interview context was appraised negatively among participants, as Disappointment and Fear were the most reported emotions among the responses. For that, the fictitious scenario was considered enough to situate the participants in the research context and allow them to identify with the emotional situation. Besides, the different sorts of appraisal triggered by the simulation suggested that using words from the ANEW list [19] had contributed to the necessary stimuli and evoked different assessments of the given stimuli.

6 Conclusion

In general, emotion recognition is based on automated classifiers trained to identify general emotional states from patterns found in the users' biosignals, whether

explicit (i.e., from facial expressions and gestures) or implicit (i.e., from biosensors acquisition). Converged tasks are required for achieving accuracy and reliability in such systems [11, 14, 22]. HCI practitioners and designers should take advantage of physiological and affective computing methods to develop accurate emotion recognition models. This study adds to the evidence that simulated settings could be used for eliciting emotional reactions, especially if the content of the simulation is framed through validated emotion elicitation approaches, such as the affective words of the ANEW list [19]. The findings of this pilot pave the way for novel methods that enable the understanding of emotional experiences in simulated interviews. We envision applications of the methods for assessing the user experience's emotional aspects with other interactive systems and different contexts.

Limitations and Future Studies. We acknowledge that emotion is a subjective context-dependent experience bound by different factors, whereas the individual appraisal of the situation, culture, age, and gender, among other aspects. Thus, the specificity of each individual appraisal and reaction to the emotional stimuli makes the generalizations about the relationships between the perceived affective information and the physiological reaction measurements very complex. Future research should account for the different physiology of people. Moreover, calibrated individual measurements could help to investigate bodily reaction patterns. Adjustments to experiment settings should cover the definition of a fixed length for the stimuli, the review of the simulation content, and the possibility of a less extensive experiment setting. Also, a larger sample size would be required. Refinements could also approach developing the necessary pipelines and testing different data processing methods, feature extraction, and analysis to achieve generalization. Trends among the data could be analyzed in future research by classification and clustering. Future studies would also benefit from using an alternative experimental design to test the influence of the specific aspects of the simulation.

Ethical Implications. Recognizing the emergent risks connected to the automation of refugee status determination assessment is necessary. The novelty of the IVAs may serve as a tool for facilitating the initial screening at the border control stations. Nonetheless, the emotion monitoring for the adaptability of such systems should be made explicit to the users before the interaction. The access to a person's mental and emotional states may be seen as invasive and place users in a vulnerable position, besides compromising the levels of trust in such interactions. Foremost, the assessment of implicit biosignals should not serve surveillance purposes nor deception detection.

Acknowledgments. The work by Debora C. Firmino De Souza and Pia Tikka has been supported by the EU Mobilias Pluss Top Researcher grant awarded to Pia Tikka by the Estonian Research Council (MOBTT90).

References

1. Scherer, K.R.: Emotions are emergent processes: they require a dynamic computational architecture. *Philos. Trans. R. Soc. B: Biol. Sci.* **364**(1535), 3459–3474 (2009). <https://doi.org/10.1098/rstb.2009.0141>
2. Scherer, K.R., Shuman, V., Fontaine, J.R.J., Soriano, C.: The GRID meets the wheel: assessing emotional feeling via self-report. *Compon. Emotional Meaning* **53**, 281–298 (2013). <https://doi.org/10.1093/acprof:oso/9780199592746.003.0019>
3. Mauss, I.B., Robinson, M.D.: Measures of emotion: a review. *Cogn. Emot.* **23**(2), 209–237 (2009). <https://doi.org/10.1080/02699930802204677>
4. Fontaine, J.J.R., Scherer, K.R., Soriano, C.: *Components of Emotional Meaning: A Sourcebook*. Series in Affective Science, 1st edn. Oxford University Press (2013)
5. Wehrle, T., Kaiser, S.: Emotion and facial expression. *Affect. Interact.* 49–63 (2000). <https://doi.org/10.1007/107202965>
6. Sagar, M., Seymour, M., Henderson, A.: Creating connection with autonomous facial animation. *Commun. ACM* **59**(12), 82–91 (2016). <https://doi.org/10.1145/2950041>
7. Kondinska, A., Kosunen, I., Ajenaghughrure, I.B., Becken, M., Gerry, L.J., Tikka, P.: The Booth: An Exploration of Enculturation Effects of Trust and Empathy towards Refugees. *Worlding the Brain*, Denmark, Aarhus University, 27–29 November 2018 (2018)
8. McNamara, R., Tikka, P.: Well-founded fear of algorithms or algorithms of well-founded fear? Hybrid intelligence in automated asylum seeker interviews (forthcoming)
9. United Nations High Commissioner for Refugees (UNHCR): *Handbook on Procedures and Criteria for Determining Refugee Status and Guidelines on International Protection: Under the 1951 Convention and 1967 Protocol Relating to the Status of Refugees*. (2019th edn.). United Nations High Commissioner for Refugees (UNHCR) (1979)
10. Kosunen, I.: *Exploring the Dynamics of the Biocybernetic Loop in Physiological Computing* (Ph.D. thesis, Series of Publications A. ed.). Unigrafia: Helsinki (2018)
11. Stemmler, G.: Methodological considerations in the psychophysiological study of emotion. In: Davidson, R.J., Goldsmith, H.H., Scherer, K.R. (eds.) *Handbook of Affective Science*, pp. 225–255. Oxford University Press (2002)
12. Stemmler, G.: Physiological processes during emotion. In: Philippot, P., Feldman, R.S. (eds.) *The Regulation of Emotion*, pp. 33–70. Lawrence Erlbaum Associates Publishers (2004)
13. Ajenaghughrure, I.B., Sousa, S.C., Lamas, D.: Measuring trust with psychophysiological signals: a systematic mapping study of approaches used. *Multimodal Technol. Interact.* **4**, 63 (2020)
14. Egger, M., Ley, M., Hanke, S.: Emotion recognition from physiological signal analysis: a review. *Electron. Notes Theor. Comput. Sci.* **343**, 35–55 (2019). <https://doi.org/10.1016/j.entcs.2019.04.009>
15. Tong, E.M.W.: Cognitive appraisals can differentiate positive emotions: the role of social appraisals. In: Fontaine, J.J.R., Scherer, K.R., Soriano, C. (eds.) *Components of Emotional Meaning: A Sourcebook*, 1st edn., pp. 507–511. Oxford University Press (2013)
16. Walker, J.H., Sproull, L., Subramani, R.: Using a human face in an interface. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence, CHI 1994* (1994). <https://doi.org/10.1145/191666.191708>

17. Blythe, M.: Research through design fiction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2014). <https://doi.org/10.1145/2556288.2557098>
18. Makowski, D., et al.: NeuroKit2: a Python toolbox for neurophysiological signal processing. *Behav. Res. Methods* **53**(4), 1689–1696 (2021). <https://doi.org/10.3758/s13428-020-01516-y>
19. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): instruction manual and affective ratings. Technical report C-1, The Center for Research in Psychophysiology, University of Florida (1999)
20. Carver, C.S., White, T.L.: Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *J. Pers. Soc. Psychol.* **67**(2), 319–333 (1994). <https://doi.org/10.1037/0022-3514.67.2.319>
21. Marsella, S., Gratch, J.: Computationally modeling human emotion. *Commun. ACM* **57**(12), 56–67 (2014). <https://doi.org/10.1145/2631912>
22. Cowley, B., et al.: The psychophysiology primer: a guide to methods and a broad review with a focus on human-computer interaction. *Found. Trends Hum.-Comput. Interact.* **9**(3–4), 151–308 (2016). <https://doi.org/10.1561/11000000065>
23. Barocas, S., Hardt, M., Narayanan, A.: *Fairness and Machine Learning: Limitations and Opportunities* (2019). <https://www.fairmlbook.org>
24. Damasio, A.R.: *Descartes' Error: Emotion, Reason and the Human Brain*. Penguin Books, New York (1994)
25. Brave, S., Nass, C.: Emotion in human-computer interaction. In: *The Human-Computer Interaction Handbook*, 2nd edn., pp. 78–94. CRC Press (2007)
26. de Borst, A.W., de Gelder, B.: Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Front. Psychol.* **6** (2015). <https://doi.org/10.3389/fpsyg.2015.00576>
27. Batista, D., Plácido da Silva, H., Fred, A., Moreira, C., Reis, M., Ferreira, H.A.: Benchmarking of the BITalino biomedical toolkit against an established gold standard. *Healthc. Technol. Lett.* **6**, 32–36 (2019). <https://doi.org/10.1049/htl.2018.5037>
28. Němcová, A., Maršánová, L., Smíšek, R.: Recommendations for ECG acquisition using BITalino. In: *Proceedings of the 22nd Conference STUDENT EEICT 2016*, Online, pp. 543–547 (2016)




Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





“NAO Says”: Designing and Evaluating Multimodal Playful Interactions with the Humanoid Robot NAO

Ilona Buchem^(✉) , Lukas Brömming , and Niklas Bäcker 

Berlin University of Applied Sciences, Berlin, Germany
buchem@bht-berlin.de

Abstract. NAO is a small humanoid robot which affords multimodal interaction through speech, non-verbal sounds, visual pattern recognition, gestures and touch. NAO can be animated to move its head, arms, legs in space and to manifest emotional reactions through dialogue, sounds, body movements and light effects. This paper reports on the design, implementation, play-testing and evaluation of a multimodal, playful interaction with NAO in two pre-studies and one pilot study with altogether 209 participants of all age groups. The application “NAO says” was designed based on the popular imitation game “Simon says”, in which three or more players follow the command “Simon says”. In “NAO says” the robot plays the role of Simon and asks the players to play a series of mini-games by imitating body movements and solving simple mathematical riddles. The design of “NAO says” focuses on creating an experience of a less constrained, playful interaction rather than following strict rules of the game. The paper describes the design of the game, the implementation in pilot studies and the results from three evaluations which investigated the perceptions of NAO as a game leader, and perceived psychological stress before and after the playful interaction with the robot. The results indicate that the robot was perceived as a friendly, joyful and pleasant interaction partner and that perceived stress was lower after playing the game.

Keywords: Human Robot Interaction (HRI) · Playful interaction · Interactive games · Humanoid robots · Social robotic game · Stress reduction

1 Introduction

Humanoid robots offer new opportunities for playful interactions with humans. Playful interaction design in Human Computer Interaction (HCI) is rooted in the perspective of humans playful creatures or as “Homo Ludens” engaging in playful, ludic activities, which take place within fixed limits of time and place, and according to freely accepted but binding rules [1]. Playful activities absorb the players and evoke intense feelings of joy and tension or excitement, bringing the players beyond the experience of the “ordinary life” [1]. The distinction between playful interactions, serious games and gamification has been made in relation to the non-utilitarian character of play [2]. Playful

interactions have been associated with such aspects as curiosity, exploration and the experience of wonder [3], as well as with an inherently pleasurable experience [4]. Research studies explored the possibilities and effects of playful interaction in the field on Human Robot Interaction (HRI) including educational robotics. For example, [5] developed an educational robotic system with a driving robot and a programming-board with command-bricks to support tangible, social and playful interactions in context of school education. Other studies with children, focused on the potential of playful human-robot interaction for learning and cognitive development. For example [6] designed playful child-robot interactions for language learning and the results showed promising effects. Studies have shown that playful interactions with technologies can be beneficial not only for children but for human players of all ages. Playful activities develop their potential in physical and social interactions especially through the engagement of the whole-body of all players [7]. These aspects, especially physical and social aspects of playful interactions including whole-body engagement, have played an important role in the design of the application “NAO says” presented in this paper.

The application “NAO says” is based on the popular imitation game “Simon says”, in which three or more players follow the command “Simon says” and act on the follow-up task. If, however, the game leader Simon does not say “Simon says” but a player acts according to the task, this player must quit the game. In “NAO says” the humanoid robot NAO plays the role of Simon and asks the players to engage in playing a series of mini-activities by following the command “NAO says”. NAO is a small humanoid robot which affords multimodal interaction through speech, non-verbal sounds, visual pattern recognition, gestures and touch. NAO can be animated to move its head, arms, legs in space and to manifest emotional reactions through dialogue, sounds, body movements and light effects. The mini-activities include imitating body movements of the robot and solving simple mathematical riddles while playfully interacting through speech and touch. These playful interactions are embedded in a social context and include a group of players following the commands of the robot together at the same time. “NAO says” as a design-led intervention aims to create a pleasurable experience through playful human-robot interaction embedded in less constrained settings, without the pressure to follow strict rules of a game.

This paper reports on the design, programming, implementation, playtesting and evaluation of this multimodal, playful interaction with NAO in two pre-studies and the main pilot study with altogether 209 participants of all age groups. The remainder of this paper is structured as follows. Section 2 outlines the design and the programming of the application “NAO says” with references to the different versions of the game “Simon says”. Section 3 describes the design of the studies themselves, i.e. two pre-studies with university students and the main pilot study, and the research methods applied in these studies. Section 4 presents the results from the two pre-studies and the main pilot study focusing on players’ perceptions of the NAO robot as game leader and the game “NAO says”, and perceived levels of stress before and after playing the game. The paper ends with conclusions and recommendations for further research.

2 Design and Programming

The application “NAO says” was designed to afford playful interactions with the humanoid robot NAO and was inspired by the popular imitation game “Simon says”. The game “Simon says” is structured around playful interactions in a group of players, who follow the command “Simon says”, engaging in a series of playful mini-activities. These mini-activities in the “Simon says” game vary from game to game version, but usually include physical, cognitive and social components which allow to engage gaze, speech, and motion. “Simon says” game has been developed and applied in many different versions. The classic version of the game involves human players playing in the physical room and one of the players taking the role of Simon. Studies on “Simon says” with human players explored its effects for learning outcomes, especially in language learning. The study by [8] showed that playing “Simon says” had a significant effect on listening comprehension of senior high school students. The study by [9] showed that the game improved vocabulary mastery in learning English.

More recently, different versions of the “Simon says” game have been developed using technologies to support the gameplay. For example, [10] developed the “Simon says” game as a mobile application for mobile phones with commands focusing on color identification. A study in the context of long term care (LTC), applied a “Simon says” activity with a robot, in which older adults took turns as leaders [11]. The researchers concluded that robots are promising for social engagement of older adults who suffer from apathy [11]. Another version of the “Simon says” game was developed with a humanoid social robot and included a computational model of turn-taking to support a more natural interaction during the gameplay [12]. Finally, the study by [13] focused on bodily movements and implemented the human pose detection library OpenPose to capture players’ poses [13].

2.1 Design

The design of the “NAO says” game presented in this paper focused on the multimodal and multi-sensory playful interaction of the NAO robot and human players of all ages. The use case scenario for the design of the “NAO says” game was a popular public event “Long Night of Sciences” which takes place every year at research institutions all over Germany on a specific day in June. On this day, scientific and science-related institutions open their doors and invite general public to visit and actively participate in experiments, demonstration, lectures, science shows, and guided tours. Playing “NAO Says” was offered as an interactive game event during “Long Night of Sciences 2022” at Berlin University of Applied Sciences, Germany, on 2nd July 2022. The “NAO says” game was embedded in a social context of this public event with multiple, voluntary participants engaging in playful interactions with the robot. The setting of the scenario was defined to be a university laboratory room, which was open during the event for general public to walk in and participate in the game at defined times. Based on previous experiences from the event “Long Night of Sciences” and the character of the “NAO says” game, the scenario defined families with children and young people as the primary audience and target group.

The “NAO says” game was designed from the human-centered design (HCD) perspective, following the scenario-based design (SBD) approach [14]. Scenarios are task based and descriptive, i.e. events and activities are strung together in purposeful sequences and provide a real-world description of the contents, flow, and dynamics [14]. The design was developed and tested iteratively. The design process included a joint co-creation of a scenario-based script in a project team, joint programming by two authors of this paper, playtesting in two pre-studies with university students, and finally the implementation and evaluation in the main study with 190 participants of all ages during the event “Long Night of Sciences 2022”.

The gameplay in “NAO says” includes a series of playful mini-activities which encompass physical and cognitive tasks. During the gameplay, human players are asked by the NAO robot to follow only when they hear the command “NAO says”. The rule from the classic version of the “Simon says” game, in which a player drops out of the game if he/she follows although there was no command, was not included in the interaction design. The reason not to include this rule was the focus on the playful, more stress-free and less competitive interaction rather than following the strict rules of the game and players having to quit the game.

The “NAO says” gameplay includes a total of ten playful mini-activities. Some of these mini-activities were based on existing animations for NAO, which are available in standard libraries of the Choregraphe software used to program the NAO robot. These included the “Saxophone”, “Elephant”, “Gorilla” and “Take a picture”. These ready-made building blocks for the game were selected as suitable for playful interactions, since they contain both clear body movements and corresponding sounds, which enhance playful engagement. Further existing animations, such as the “Air guitar”, were combined with new sound effects, which were imported as free sound files from the Internet and integrated in Choregraphe. For the purpose of the “NAO says” game, some own animations were programmed in Choregraphe and added to the gameplay. The self-developed animations included: “Stand on one leg”, “Rub tummy, pad head”, “Wave arms above the head”, “Smile” and “Maths”. In total, the following ten mini-activities were used in the “NAO says” gameplay: (1) Gorilla, (2) Elephant, (3) Air guitar, (4) Saxophone, (5) Take a picture, (6) Stand on one leg, (7) Rub tummy, pad head, (8) Wave arms above the head, (9) Smile, and (10) Maths. Figure 1 visualises all mini-activities arranged into categories: animals, music, body and other.

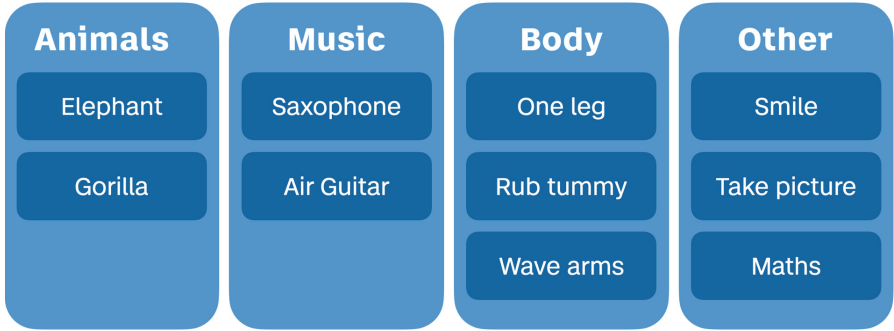


Fig. 1. Playful mini-activities included in the “NAO says” game.

2.2 Programming

The programming of the “NAO says” game was done using the Choregraphe software (Version 2.8.6). The game was designed in the English and German language versions which were tested in two pre-studies with students. Figure 2 visualises the programming of the “NAO says” game in Choregraphe with different elements such as animations, animated say, speech recognition and tactile sensors (bumpers).

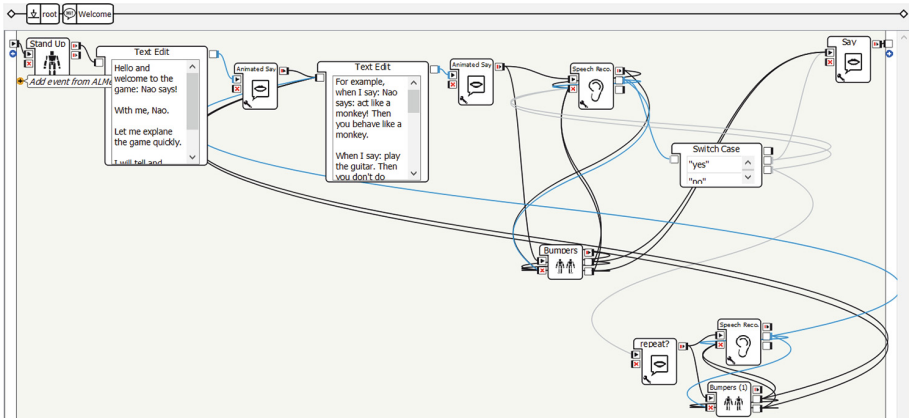


Fig. 2. Programming of “NAO says” in Choregraphe (English version).

When programming the game, four main challenges emerged: (1) How to program new mini-activities and which existing animations can be adapted for the purpose of “NAO says”?; (2) How to make the interaction with NAO possibly seamless with a larger number of humans players and observers present in the room and standing at a relatively large distance from the robot?; (3) How to make the gameplay exciting without repeating the same sequence of mini-activities? These challenges were addressed in the programming phase as described below.

How to program new mini-activities and which existing animations can be adapted to the purpose of the game? In order to create new mini-activities, such as “Stand on one leg” or “Rub tummy, pat head”, the Timeline editor, an integrated tool in Choregraphe, was to store different positions of the robot in a timeline and play one after the other as a fluid sequence of animations. The Choregraphe Timeline editor enables the programmer to adjust every single angle of NAO’s motors using simple sliders. Additionally, this method was combined with the use of the Animation mode. When NAO is in Animation mode, the angles of the joints on the real robot can be adjusted and the joint position saved as a point on the timeline. From there, NAO can be moved to the next position. This procedure results in a fluid animation from the individual positions at the end. When creating animations, special attention must be paid to the length between the individual positions of the animation. If the length is too short, jerky movements can occur. These not only look unnatural, but also cause the motors to heat up more quickly which can even damage the robot. In addition, special care must be taken to ensure that the NAO robot is always in a stable position. The ready-made animation “Gorilla” available in the Choregraphe library demonstrates this problem: In “Gorilla”, NAO drops forward onto its hands and sometimes falls over, either because the animation is executed too quickly or because the ground is not ideally flat. Therefore, creating new animations like “Stand on one leg” was particularly difficult to implement, as NAO had to be kept in a stable position during the entire animation.

How to make the interaction with NAO possibly seamless with a larger number of humans players and observers present in the room and standing at a relatively large distance from the robot? The scenario was designed for participation of multiple players and observers, all present in one room with NAO during the public event at university premises. The presence of many participants enhances the risk of a high volume of background noises which may impede speech recognition of the NAO robot. Therefore, the decision was made to limit the number of mini-activities with human-robot interaction via speech. In fact, the only mini-activity, in which speech input from the participant is necessary, is the maths activity. In the maths activity NAO asks “How much is 3 multiplied by 3?” and expects the answer “nine” from the participants. Also during the game “NAO says” NAO also asks a number of times “Did you understand everything?” and waits until the answer “Yes” is said by a participant. However, these interactions via speech are only possible when there are no background noises in the room and possibly only one participant at a time speaks loudly and clearly. In the scenario at the public event with approx. 20 participants in the room at the same time, a loud and clear response was foreseen to not be feasible. Therefore, it was decided to keep the threshold very low so that everything that sounds similar to “nine” could be recognized by the NAO robot as nine. However, this method had the disadvantage as, for example, “nineteen” or other numbers are also recognised as “nine”. Therefore, the final threshold was set to 30%, i.e. any utterances that sounds 30% like “nine” are recognized as a “nine”.

How to make the gameplay exciting without repeating the same sequence of mini-activities? As described above, a pool of mini-activities was created to provide a variety of non-recurring playful mini-activities, and in this way to enhance the user experience. In order to make the gameplay exciting, the randomization principle was applied in the programming of the game. Randomness of game elements is linked to uncertainty, which

is a frequently overlooked in game design, but an important element for the overall game experience as it holds players' interest and enhances engagement [15]. To incorporate the randomization for the path-finding command “NAO says” before movement, a random variable was included which was then queried with an If-condition. The challenge here was that the positive feedback (praise) given by the NAO robot at the end of the mini-activity was not in the same programming level as the dice rolling of the random number. To address this challenge, a separate random number was chosen for each mini-activity and duplicated this mini-activity. In the final version, different responses of the robot were programmed depending on the use or non-use of the command “NAO says”.

The different versions of the “NAO says” game were tested in the two pre-studies and the final version in the main pilot study at the public event as described below.

3 Methods and Studies

Following the design and the programming phase, the game was play-tested with university students in two pre-studies and the final version was implemented and evaluated during the public event with participants of different ages. Playtesting is a popular method in game research used to test perceptions and preferences of players, allowing designers to modify the game before delivering the final version [16]. The key facts about the pre-studies and the pilot study are summarized below and in Table 1.

Pre-study 1. The first pre-study involved a sample of ten university students, who volunteered to test the English version of the initial version of the “NAO says” game. Participants were asked to fill in an online survey before and after the game. One of the key results from the first pre-study was the wish of students to play the game in the German language version. Therefore the German language version was created in Choregraphe and tested in the second pre-study.

Pre-study 2. The second pre-study involved a sample of nine students, who volunteered to test the German version of the “NAO says” game and did not participate in the first pre-study. This version of the game also included a slower pace of NAO's speaking as the result from the first pilot study clearly indicated the need for slower speed to understand better what to do in each mini-activity. Like in the first pre-study, the participants were asked to fill in an online survey before and after the game.

Pilot Study. The main pilot study took place during the public event “Long Night of Sciences” with participants of different ages. Out of approx. 260–280 participants on that day, 190 persons filled in the evaluation survey which was administered before and after playing the game with NAO. The survey was paper-based to ensure high participation of persons without digital devices and of younger children.

Table 1. Summary of the pre-studies and the main pilot study of the game “NAO says”.

Pre-study 1	Pre-study 2	Pilot study
n = 10	n = 9	n = 190
English version	German version	German version
University students	University students	General public
Classroom setting	Classroom setting	Public event setting
50% female, 50% male	44% female, 56% male	46% female, 51% male, 3% diverse
70% 20–24 years old 30% 25–29 years old	56% 20–24 years old 22% 25–29 years old 11% 30–34 years old 11% 35–39 years old	3% younger than 7 years old 31% 7–18 years old 35% 19–29 years old 5% 30–39 years old 17% 40–49 years old 8% 50–59 years old 1% 60 years old and older

4 Results

The key results from the studies related to: (1) perceptions of the robot and the game “NAO says”, and (2) perceived stress level before and after playing the game are described in the sections below.

4.1 Perceptions of the NAO Robot and the Game “NAO Says”

The data about the perceptions of the participants of the NAO robot as game leader and of the game “NAO says” was collected via online surveys in the pre-studies and via a paper-and-pencil survey in the main study. Both online surveys included additional questions which were not asked during the main study due to the specific conditions of the public event. The online surveys ask the question How did you perceive NAO as a game leader? This question was answered by rating five pairs of semantic items from the Likeability Scale of the Godspeed questionnaires rated on a scale from 1 to 6 [17]. Table 2 summarises the results from the Likeability Scale.

Table 2. Perceptions pre-studies and the main pilot study of the game “NAO says”.

	Pre-study 1	Pre-study 2
unlikely (1) – likeable (6)	M = 5.40 (SD 1.265)	M = 5.89 (SD .333)
unfriendly (1) – friendly (6)	M = 5.30 (SD 1.337)	M = 6.00 (SD .000)
unkind (1) – kind (6)	M = 5.40 (SD 1.265)	M = 5.89 (SD .333)
unpleasant (1) – pleasant (6)	M = 5.20 (SD 1.317)	M = 5.56 (SD 1.014)
awful (1) – nice (6)	M = 5.30 (SD 1.252)	M = 5.56 (SD 1.014)

The results show that in both pre-studies NAO was perceived as a likeable, friendly, kind, pleasant and nice game leader. The data also shows higher values in the pre-study 2 in which the German version of the game was used which may indicate that the use of the local language may have enhanced positive perceptions of the robot.

Next, perceptions of playful interactions with the robot were captured in all three studies using the simple question “How did you like the game?” and asking participants to assess their perception on a scale from 1 = not at all, to 6 = very much. The mean values were as follows: (1) pre-study 1, n = 10: M = 5.10 (SD .738), (2) pre-study 2, n = 9: M = 5.33 (SD .707), (3) pilot-study, n = 190: M = 5.25 (SD = .885). These results indicate that participants in all three studies enjoyed playful interactions with NAO. The foreign language version of the game in English in the first pre-study received the lowest value, which again indicates that the language choice affects user experience. The high average rating of M = 5.25 in the main study with 190 participants show that participants in different age groups liked the game.

4.2 Perceived Level of Stress Before and After the Game

Perceived psychological stress was measured to explore whether there were any changes in how stressed or relaxed participants felt before and after playful interaction with NAO. The data about perceived stress was collected via online surveys in the two pre-studies and via a paper-and-pencil survey in the main study. Psychological stress was reported by the participants before and after playful interactions with NAO using the Perkhofer Stress Scale, which is a validated single item scale [17]. The participants assessed their stress level on the scale from 1 = no stress (“fully relaxed”) to 6 = fully stressed (“anxious”) before and after the game. To explore the differences in perceived stress before and after playing “NAO says”, the dependent samples (paired) t-test was computed at the 95% confidence level and two-tailed p-value using IBM SPSS software. The comparison of means showed that in all three studies the mean values for perceived stress before the game were slightly higher compared to the values after playing the game. In the first pre-study (n = 10) the mean values were M (before) = 2.60 (SD .843) and M (after) = 2.40 (SD 1.174). In the second pre-study (n=9) the mean values were M (before) = 2.56 (SD 1.014) and M (after) = 1.33 (SD 1.000). In the third pre-study the mean values were M (before) = 2.11 (SD .910) and M (after) = 1.74 (SD .917). Table 3 summarises the results for all three t-tests.

Table 3. Paired samples t-test: Perceived psychological stress before and after the “NAO says” game.

	Pairs	Mean	Std. Deviation	Std. Error Mean	t	d	Sig. (2-tailed)
Pre-study 1	10	.200	1.229	.398	.514	9	.619
Pre-study 2	9	1.222	1.302	.434	2.817	8	0.23
Pilot study	190	.374	.949	.068	5.488	189	<.001

The results show that the means of perceived stress before and after the game differed significantly only in the main pilot study. It can be concluded that the perceived level of stressed was statistically lower after the game “NAO says” and changed from 2.11 ± 0.91 to 1.74 ± 0.92 ($p < 0.001$). The results also show that the initial stress level of the 190 participants in the main study during the public event was slightly lower compared to the two pre-studies with students, which may be explained by the leisure character of the event compared to participation in classroom settings. Nevertheless, the participation of volunteers visiting the laboratory during the public event limits the possibilities of generalizing the results of the study. It can be assumed that the participants in the main study differed from the general population and from populations with special needs in regards to their level of initial motivation to participate as well as their interest in and attitudes towards robots. Therefore it is recommended to conduct a broader follow-up study involving a more diverse sample and including variables such as interest, motivation and attitudes towards robots.

5 Conclusions

This paper reported on the design, programming, implementation and evaluation of playful interactions during the game “NAO says” with the humanoid robot NAO in two pre-studies with students and one pilot study with 190 participants of different ages. The exploratory results in all three studies showed that the players perceived NAO as a likeable, friendly, kind, pleasant and nice game leader, and enjoyed playful interactions with NAO. Additionally, there was a significant difference in the perception of own’s psychological stress before and after the game with NAO in the pilot study with 190 participants. The results also indicate possible effects of different language versions of the game on user experience. The results presented in this paper are to be understood as preliminary, exploratory results and as a starting point for further research. Further studies should be conducted with diverse samples and look closer into possible effects of different versions of the game. The paper also pointed out several challenges in the design and programming of the game “NAO says” and how these were addressed. Further studies could explore in more detail which design strategies of playful interactions in games like “NAO says” and which types of feedback from the robot are most effective for specific target audiences. Furthermore, future studies could explore how different types of playful interactions with robots may affect the perception of mood and stress as well as physical stress measures.

References

1. Huizinga, J.: *Homo Ludens: A Study of the Play Element in Culture*. Beacon, Boston (1950)
2. Deterding, S., Khaled, R., Nacke, L.E., Dixon, D.: Gamification: toward a definition. In: CHI 2011, 7–12 May 2011, Vancouver, BC, Canada. ACM (2011)
3. Gaver, W.: Designing for Homo Ludens. *I3 Mag.* **12**, 2–6 (2002)
4. Krause, M.: Designing systems with Homo Ludens in the loop. In: Michelucci, P. (ed.) *Handbook of Human Computation*. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-8806-4_31

5. Pedersen, B.K., Andersen, K.E., Jørgensen, A., Kösslich, S., Sherzai, F., Nielsen, J.: Towards playful learning and computational thinking—Developing the educational robot BRICKO. In: 2018 IEEE Integrated STEM Education Conference (ISEC), pp. 37–44 (2018)
6. Tolksdorf, N.F., Rohlfing, K.J.: Parents’ views on using social robots for language learning. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 634–640 (2020)
7. Rennick Egglestone, S., Walker, B., Marshall, J., Benford, S., McAuley, D.: Analysing the playground: sensitizing concepts to inform systems that promote playful interaction. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011. LNCS, vol. 6946, pp. 452–469. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23774-4_38
8. Azizah, N.L.: The effect of Simon says game towards students’ listening comprehension at the tenth grade of senior high school students (2020)
9. Dalimunthe, M.B.: The Implementation of the Simon Says Game to Improve Students’ Vocabulary Mastery in Learning English at MTS. Laboratorium. UIN-SU MEDAN (2018)
10. Palazzi, C.E., Maggiorini, D., Burattin, A., Ferro, R.: Simon says the color: the digital evolution of an outdoor kids game. In: SimuTools (2010)
11. Lin, Y., et al.: Can robots encourage social engagement among older adults? *Innov. Aging* **4**, 193 (2020)
12. Chao, C., Lee, J., Begum, M., Thomaz, A.L.: Simon plays Simon says: the timing of turn-taking in an imitation game. In: 2011 RO-MAN, pp. 235–240 (2011)
13. Li, C., Imeokparia, E., Ketzner, M., Tshahi, T.: Teaching the NAO robot to play a human-robot interactive game. In: 2019 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 712–715 (2019)
14. Hertzum, M.: Making use of scenarios: a field study of conceptual design. *Int. J. Hum.-Comput. Stud.* **58**, 215–239 (2003)
15. Xu, W., Liang, H., Yu, K., Baghaei, N.: Effect of gameplay uncertainty, display type, and age on virtual reality exergames. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (2021)
16. Amberkar, S., Bennett, K., Delchamps, A., Edwards, R.: Expanding the playtest method for UX research beyond gaming. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 63, pp. 2239–2243 (2019)
17. Bartneck, C., Kulić, D., Croft, E.A., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* **1**, 71–81 (2009)








Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Representation of Air Pollution in Augmented Reality: Tools for Population–Wide Behavioral Change

Grzegorz Pochwatko¹(✉) , Zbigniew Jędrzejewski¹ , Wiesław Kopeć² ,
Kinga Skorupska² , Rafał Masłyk² , Anna Jaskulska³ ,
and Justyna Świdrak⁴ 

¹ Institute of Psychology, Polish Academy of Sciences, Warszawa, Poland
gp@psych.pan.pl

² Polish-Japanese Academy of Information Technology, Warsaw, Poland

³ Kobo Association, Oświęcim, Poland

⁴ August Pi & Sunyer Biomedical Research Institute, Barcelona, Spain

Abstract. Air quality affects health and, unfortunately, has been deteriorating rapidly recently. The problem is significant in smaller towns where the important source of pollution is the heating of households and water from individual sources. Therefore, the inhabitants have an influence on a significant reduction of pollution in their area, but at the same time, they are very often not aware of it. Raising awareness about household-related air pollution plays a vital role as systemic solutions proposed by state and local authorities require the support of local communities. The situation has recently become even more serious as we are facing a crisis caused by the Russian war in Ukraine, which has led to an increase in energy and fuel prices and has postponed restrictions on the use of solid fuels or even incentives to use inferior fuels. Pathologies, such as burning garbage in old-style furnaces, have still not been eliminated. One of the ways of raising citizens' awareness was to be public, easily accessible information about air quality. Many portals, services, and applications currently provide local air quality data, but few people use them. One reason may be that the figures and graphs can be confusing or unattractive to audiences who are not used to reading scientific reports. Visualizing air quality with augmented reality overcomes these obstacles. A mobile application that can use local elements as triggers and a symbolic representation of air quality based on data read in real-time from sensors is simple, attractive for non-experts, and has an additional educational value. We present the experience of creating such an application and prototype tests with the participation of potential users. Unfortunately, the collected results confirm the low awareness of excessive pollution in a given area and its negative impact on health. However, the interest of potential users and positive opinions about the tested prototype fill with optimism.

Keywords: augmented reality · air pollution · user experience · art and science

© The Author(s) 2023

C. Biele et al. (Eds.); MIDI 2022, LNNS 710, pp. 150–158, 2023.

https://doi.org/10.1007/978-3-031-37649-8_15

1 Introduction

Visualizing air pollution in the real world is not a new phenomenon, but it has largely been contained in the domain of arts, with projects aiming to make the local population aware of the pollution's negative effects on their health. A traveling art project, giant lungs made of air filters whose color changed with time when exposed to polluted air, in multiple editions toured the most polluted cities in Poland¹ Such projects have great value, yet they have limitations as they are confined to a specific location and do not show dynamic changes in pollution rates. Still, such campaigns are a crucial step in limiting pollution in the most polluted areas, especially if pollution comes from residential sites. A decrease in emissions can be seen following informative action [4], and such initiatives often pave the way for necessary policy. Still, it is as they garner public support for such initiatives. For these reasons, we have decided to design and verify an Augmented Reality (AR) application representing current pollution levels based on location data. Seeing real-time pollution data on a personal mobile device and having it visualized in a manner that is meaningful to the device-holder at a place where they are currently has the potential to increase awareness of air pollution presence and its negative health effects.

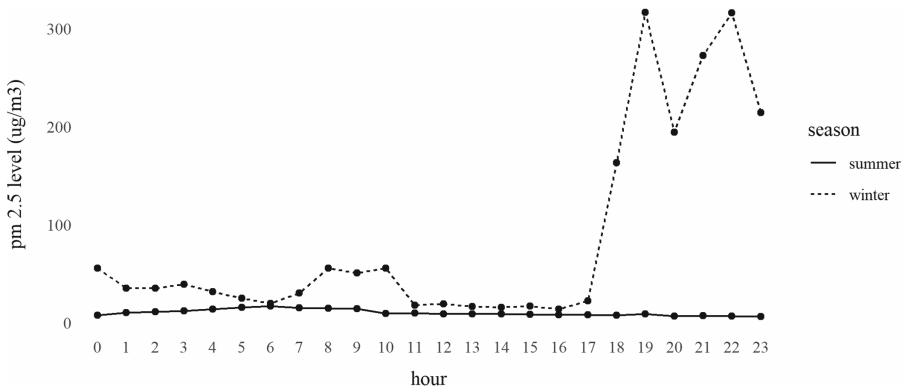


Fig. 1. PM2.5 values measured every hour during one day in winter and summer 2021; a town with up to 25,000 inhabitants, data from the Environmental Protection Inspection [5]

1.1 Household Related Air Pollution

Air quality affects the health of citizens and, unfortunately, recently has been deteriorating rapidly. However, when effects are delayed, people tend to disregard their severity, even though multiple papers closely link negative health outcomes, especially respiratory diseases, and reduced lung capacity, to air pollution

¹ The images from the project, by campaign creators, are available here: <https://www.purpose.com/poland-air-pollution-campaign/>.

[11]. The problem is particularly important in smaller towns where the important source of pollution is the heating of households and water from individual sources. One of the most dangerous pollutants is particulate matter, especially its fine fractions - PM2.5 [1,6,7]. Figure 1 shows how much household heating can contribute to a dangerous increase in PM2.5 levels. The chart shows the pollution levels recorded in hourly intervals in summer and winter in one of the Polish small towns (up to 25,000 inhabitants), which is dominated by single-family housing with its own heating sources. The PM2.5 level in the summer is low. In winter, however, it exceeds the norm many times in the afternoon and in the evening, when citizens usually return to their homes from work and turn on the heating.

1.2 Air Pollution Representation

The choice of the pollution representation is crucial considering the fact that the current methods of presenting information on the air quality turned out to be ineffective or gave inconclusive results (see e.g., [10]). For example, Campaña and Domínguez [3] proposed to present the different types of pollutants as enlarged models of chemical molecules. In another case, Prophet et al. [13] encouraged participants to take care of air quality in a symbolic way - by taking care of a virtual tree in an application using the data of the local pollution measurement station.

In our opinion, the choice of the correct representation of air pollution should not be arbitrary. We do the choice in a multi-stage process, which we started with a participatory workshop with potential users [8] and continued with laboratory research (see e.g., [12]). Initially, four types of visual stimuli were selected and tested: 2 positive vs. negative and 2 concrete vs. abstract. At the time when we were preparing and testing the prototype presented in this paper, the laboratory studies were still in progress, so for the purposes of the AR app tests, we chose a concrete and negative representation - enlarged PM particles.

1.3 Augmented Reality

Augmented reality (AR) is a technology that is gaining in popularity and has a chance to stay ahead of immersive virtual reality due to the availability of interesting solutions for mobile devices[9]. 3D and animated objects can be easily placed in real space thanks to toolkits such as Vuforia² which make developing AR applications easier for various context, varying for workplace instructions, entertainment to product design.

Vuforia is an AR add-on to one of the commercially available development engines for creating games, called Unity3d. The principle is that with the help of an application designed for a mobile device, an additional layer of abstraction

² Vuforia is a comprehensive software development kit (SDK) for Virtual Reality applications available here: <https://developer.vuforia.com/>.

can be added to the area seen by the camera. In this way, the visible area can be expanded with graphic elements, usually in the form of useful information.

For this to happen, it is necessary for the application designer to design and prepare (in appropriate formats) two types of graphic elements. The first is the so-called marker, otherwise known as a trigger - that is, an object to whose presence in the field of view of the camera the application will be “sensitive”, and the second is a target object (static image, animation, 3d model and many others) which appears in the field of view at or near the predefined marker (Fig. 2). Today, AR technology is increasingly used in industry, social media, and everyday products, among others. This is due to several basic needs that the market places on this technology, i.e.: product recognition and training, product visualisation, customer self-service, guided user manuals, and part recognition [14]. The increasing use of AR apps has been noted by several online industry journals. Insider Intelligence [15] notes the increasing share of users using augmented reality functionality, and Forbes [2] notes that this technology could be the future of social media and beyond.



Fig. 2. First draft of the VAPE AR app: scanning QR code to obtain the app (left), scanning the trigger to launch Vuforia AR contents (center), exploring AR contents - pollution representation overlay (right).

2 VAPE Augmented Reality App Design

2.1 Purpose of the AR Application

The aim of the design of the application was to create an engaging experience presenting air pollution with visual cues and to communicate scientific data in a comprehensible way, by doing so, potentially raising awareness of environmental and health-related issues caused by air pollution. The application was designated to be pretested during a scientific event held in Myszków (Poland)³, with the intention of further development in the future. The constraints affected by the out-doors-held event called for an easily accessible, compact and portable setup in addition to minimum device requirements.

³ Myszków is a town with one of the highest air pollution levels in Poland and Europe. It was on the list of 50 cities with the most carcinogenic air in Poland.

2.2 Accompanying Poster

The intention of creating an informative poster complementing the experience with scientific data has affected the decision to employ a trigger-activated AR system as a base design of the application. A 100 cm per 200 cm sized roll-up banner commonly used in advertising at events was chosen as the print medium for the informative poster. The medium allowed easy and self-contained construction, a great level of portability, and re-usability, additionally to the vast canvas surface. The poster was designed and prepared for print in Adobe Photoshop (CS6), composing hand-drawn digital illustrations and text created beforehand in Procreate. The overall design of the poster was intended to be simple with key scientific information, not to overwhelm the participants of the event and to encourage engagement with the application (Fig. 3).



Fig. 3. Testing the prototype - the trigger (left) and a user scanning the trigger (right)

2.3 Implementation of Air Pollution Visual Representation

The VFX representing air pollution was created in Unity's Visual Effect Graph. The node-based Visual Effect Graph allows the creation of procedural animations of a vast number of real-time rendered particles. The particles were set to spawn evenly within a cube shaped-region of base measuring 6 m by 7 m and reaching 5 m high above the ground. The particles were spawned constantly with their basic life-span set to 10 s (not counting the added variation), and the overall capacity of particles of the VFX was set to 100000. The animation of the particles was driven by a turbulence node effecting in an appearance of the particles floating freely in the air. Each of the particles was assigned with a random texture from a set of 16 images using a flipbook particle output method. The images were edited microscopic photographs of actual air pollution particles. The assigned textures were given random sizes ranging from 0.5 cm to 2 cm. A gradient of colours from dark grey to black was assigned to the particles.

In order to raise awareness of health-related issues caused by air pollution, a visualisation of lungs has been chosen additionally to the visualisation of air pollution particles. The 3D model of lungs implemented in the application has been generated by segmentation of trachea and bronchi based on a computed tomography medical scan using 3D Slicer (4.11.20210226) software. The generated 3D model was edited and retopologised using Pixologic Z-Brush (4R7). The model was exported to Unity using an fbx file. The model was enhanced with a VFX animation of the same air pollution particles used in the air floating VFX. The animation shows the particles streaming through the trachea and obscuring the lungs (Fig. 4). The 3D model was visualised and augmented in front of the poster.

2.4 Application Development and Implementation

The application was compiled in Unity (2021.2.0b12) Game Engine with a Vuforia (10.0.12) plug-in installed. The Vuforia plug-in allows easy setup of AR trigger-activated system. The visualisation augmented upon the surroundings of the user was set to be triggered by targeting the smartphone's camera on the illustration of lungs presented in the centre of the informative poster. The scene build in Unity contained a Vuforia image trigger setting, animated Visual Effects (VFX) representing air pollution particles, and a 3D model of lungs with its own VFX animation. The application was built for android devices and preinstalled on a smartphone that was later used as a visualising device during the event. The presented application was a prototype and was not available for download from the internet by the participants, yet such an option of dissemination is possible in the future.

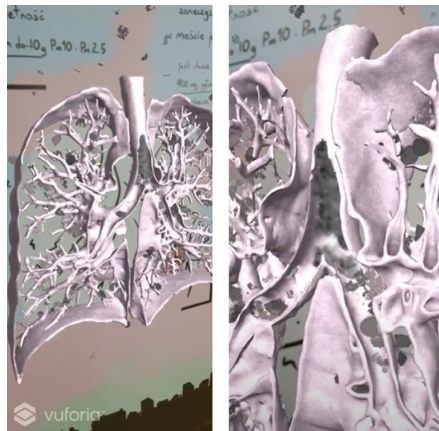


Fig. 4. The animation of the particles entering via the trachea and obscuring the lungs.

3 Pretest Results

The prototype of the AR application was presented to volunteers, participants of an open event promoting pro-ecological behavior in the city of Myszków. The event was co-organized by local authorities and non-governmental organizations. Due to the restrictions caused by the COVID-19 pandemic, all events were held outdoors with appropriate precautions. The presentation of the prototype was one of the activities that the participants of the event (mostly parents with children) could get involved in. In order to better simulate the conditions in which the application will be used in the future, it was not announced in the information inviting to the event (posters hung around the city and spots on local TV). Participants could take part in a pretest, which was organized as one of many stands (e.g., art classes with local artists, Storm Hunters information stand, non-contact sports activities, 3D printing show). Participants who decided to test the AR application were encouraged to provide feedback to the researchers operating the stand. Many participants of the event spontaneously reacted to the information posted on the poster/trigger. As intended, they tried to scan the trigger on the poster with their own phones. In such a situation, the experimenters approached potential users, explained that it was a pre-test of the prototype and, for security reasons, it was not possible to use their own devices yet, but they could see how it works on the device provided for this purpose. They then showed participants how to scan a trigger on a poster, handed over a phone with the app installed, and began observing their interactions. If test subjects watched only the poster and lungs in the AR for a long time, the experimenters encouraged them to turn around and observe the area with the layer of pollution overlaid. Most of the participants spontaneously made such an exploration. Careful observation of the lungs with pollutants entering them was rather avoided. Most users focused on watching pollution particles in the air. Below are representative testimonials from participants in the pretest:

«All this dirt makes a terrible impression. I would definitely like to have such an application to check whether, for example, I can go on a trip with my children.»

«I didn't know that so much dust was flying in the air, especially since it's a nice day today, you can't smell the coal smoke»

«A man wants to do something for his health, he runs, exercises, and here you have something. If you check «on the portal», you can't see it.»

«Maybe it would make people realize that you can't burn just anything»

«Someone would have to want to use it. Some people don't care.»

Thanks to positive and negative comments from users, we have the opportunity to develop the application. There is a need to conduct broader research, which would also indicate ways to popularize the application and ensure its usability, information value and impact on changing user behavior.

4 Conclusions

Visualising key data with mobile AR is a reliable way of grounding the understanding of information within the context of the environment it is displayed over. In this way, people living in a certain area may better realize the invisible presence of air pollution and how it may affect their health - but, as our subjects mentioned, only if they care to stop and check. Still, we hope that our application prototype will serve as a proof-of-concept for larger information campaigns on air pollution so that they are created in a more impactful and understandable way, relevant to their target users.

Acknowledgments. We would like to thank our participants, volunteers and citizens of Myszków, Poland, who participated in our actions; The research leading to these results has received funding from the EEA Financial Mechanism 2014-2021 grant no. 2019/35/HS6/03166.

References

1. Brunekreef, B., et al.: Mortality and morbidity effects of long-term exposure to low-level pm_{2.5}, bc, no₂, and o₃: an analysis of European cohorts in the elapse project (2021)
2. Bullock, L.: AR and social media: Is augmented reality the future of social media? (2018). <https://www.forbes.com/sites/lilachbullock/2018/11/16/ar-and-social-media-is-augmented-reality-the-future-of-social-media/?sh=1eec6c16e141>
3. Campana, F.E., Xavier Dominguez, F.: Proposal of a particulate matter measurement device and an augmented reality visualization app as an educational tool. In: 2020 12th International Conference on Education Technology and Computers, pp. 6–11 (2020)
4. Danek, T., Zareba, M.: The use of public data from low-cost sensors for the geospatial analysis of air pollution from solid fuel heating during the COVID-19 pandemic spring period in Krakow, Poland. *Sensors* 21(15), 5208 (2021). <https://doi.org/10.3390/s21155208>, <https://www.mdpi.com/1424-8220/21/15/5208>
5. Environmental Protection Inspection: Measurement data bank – results from 2021 (2021). <https://powietrze.gios.gov.pl/pjp/archives>
6. Han, M., Yang, F., Sun, H.: A bibliometric and visualized analysis of research progress and frontiers on health effects caused by pm_{2.5}. *Environ. Sci. Pollut. Res.* 28(24), 30595–30612 (2021)
7. Karanasiou, A., et al.: Short-term health effects from outdoor exposure to biomass burning emissions: a review. *Sci. Total Environ.* 781, 146739 (2021)
8. Kopeć, W., et al.: Participatory design landscape for the human-machine collaboration, interaction and automation at the frontiers of HCI (PDL 2021). In: Ardito, C., et al. (eds.) *Human-Computer Interaction - INTERACT 2021*, pp. 564–569. Springer, Cham (2021)
9. Liu11, X., Sohn, Y.H., Park, D.W.: Application development with augmented reality technique using unity 3D and Vuforia. *Int. J. Appl. Eng. Res.* 13(21), 15068–15071 (2018)
10. Mathews, N.S., Chimalakonda, S., Jain, S.: Air: an augmented reality application for visualizing air pollution. In: 2021 IEEE Visualization Conference (VIS), pp. 146–150. IEEE (2021)

11. Nazar, W., Niedozytko, M.: Air pollution in Poland: a 2022 narrative review with focus on respiratory diseases. *Int. J. Environ. Res. Public Health* 19(2) (2022). <https://doi.org/10.3390/ijerph19020895>, <https://www.mdpi.com/1660-4601/19/2/895>
12. Pochwatko, G., et al.: Multisensory representation of air pollution in virtual reality: lessons from visual representation. In: Biele, C., Kacprzyk, J., Kopeć, W., Owsiański, J.W., Romanowski, A., Sikorski, M. (eds.) *MIDI 2021. LNCS*, vol. 440, pp. 239–247. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-11432-8_24
13. Prophet, J., Kow, Y.M., Hurry, M.: Cultivating environmental awareness: modeling air quality data via augmented reality miniature trees. In: Schmorow, D., Fidopiastis, C. (eds.) *AC 2018. LNCS*, vol. 10915, pp. 406–424. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91470-1_33
14. PTC: Top 5 vuforia studio use cases. <https://www.ptc.com/en/blogs/ar/vuforia-studio-use-cases>
15. Williamson, D.A.: Augmented reality in social media (2020). <https://www.insiderintelligence.com/content/augmented-reality-in-social-media>





Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Ukrainian Version of the Copresence Scale

Lyubov Naydonova¹ , Grzegorz Pochwatko² , Mykhaylo Naydonov³ ,
and Justyna Świdrak⁴ 

¹ Institute for Social and Political Psychology, National Academy of Educational Sciences, Kyiv, Ukraine

² Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland
gp@psych.pan.pl

³ Institute of Reflexive Investigation and Specialization, Kyiv, Ukraine
iris_psy@ukr.net

⁴ August Pi & Sunyer Biomedical Research Institute, Barcelona, Spain

Abstract. Millions of Ukrainians are fleeing the war to the EU countries; in particular, Poland has accepted more than three million refugees, among whom are mostly women and children. Some of them continue to work remotely in Ukraine and communicate with friends and relatives at a distance through technology-mediated ways. The aim of our study was to observe the role of perceived copresence during social contacts, both of professional and private character, and validate the Ukrainian version of the copresence scale among Ukrainian migrants. We collected 221 responses in an online and face-to-face study from Ukraine migrants. Analyses revealed that the Ukrainian Perceived Copresence scale has one factor and appropriate internal consistency. Perspectives of future research are proposed.

Keywords: Copresence · Ukrainian · Migrant · Mediated communication

1 Introduction

2 Theoretical Context

2.1 War Migration from Ukraine in Poland

Migration makes it necessary to adapt to a new social environment and at the same time to maintain distance communication with those who remained in the previous place of residence, for this migrants use various technologies to mediate communication. The wave of mass migration from Ukraine as a result of the war and the military aggression of the Russian Federation became a necessary means of preserving life. As of mid-September, according to the UNHCR [10], more than 7 million temporarily displaced persons from Ukraine are in European countries, 1.3 million of them are in Poland. Since the beginning of the full-scale offensive, more than 6 million people have crossed the border from Ukraine to Poland. Distant communication for work and private conversation became an important part of the Ukrainian migrant community in Poland. As communication is needed for mental health, research on the copresence effect in communication is actual, but for these purposes, Ukrainian language methods must be developed and adapted.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 159–166, 2023.

https://doi.org/10.1007/978-3-031-37649-8_16



Fig. 1. Mediated communication (photo by Julia M. Cameron; pixels.com)

2.2 Copresence

Copresence has become a popular research topic in recent years, especially after the COVID-19 outbreak which forced millions of people to stay in isolation for many days or weeks. The term “copresence” was suggested to mark the quality of a communication medium in human-human or human-machine interaction [7] and is defined as the degree of person-to-person awareness which occurs in the computer environment [9]. Our previous study [8] has proved a relationship between the amount of interpersonal communication, housing conditions (a shared or private room, number of children and adults in the household), and copresence with mental well-being in confinement during the COVID-19 lockdowns. Here, we present the validation of the Copresence scale [8] to Ukrainian language, which was used in a study that aimed to replicate these results in another special context of forced disrupted social relations, this time due to migration forced by war. This “war migration situation” has put millions of migrants in new housing conditions, at a distance from their work and relatives, and in emotional positive and negative contacts. Here we describe a study carried out from May to September 2022 in Poland, in which we aimed to validate the Ukrainian version of the copresence measurement instrument.

2.3 Mediated Communication and War Migration

Chen’s study of the role of internet communication in migrant adaptation shows that immigrants who remotely communicate online more frequently with relatives and

friends in their original country are less adaptive in terms of sociocultural adaptation. However, communicating with relatives and friends in the original country is no longer a significant predictor of intercultural adaptation when introducing demographic characteristics into the regression analyses: no significant impact is identified for online communication with fellow immigrants on intercultural adaptation [3]. Copresence may be important to overcome the isolation of people using VR technology as it was approved for anxiety disorder [6]. Migrants experience “co-presence” with their loved ones through social media. Among second-generation Turkish-Dutch migrants who grew up in the Netherlands and migrated to Istanbul in adulthood [1] are in two different models of communication: ambient, fast-paced, background communications and also more direct, immersive, conversational modes. The careful shifting between these modes of social media communication produces migrants’ experiences of transnational emotional intimacy when emotion is mediated through digital media and gives psychological support to migrants. During the war and migration of Ukrainian family members, distant communication is the factor that influences the moral and psychological state of the combatant when family problems impact the psychological state of the service members during deployment in active military zones of anti-terrorist operations [4,5].

3 Method

3.1 Sample and Data Collection

The study was carried out online in the form of a self-report questionnaire. Data were collected from May to September 2022 among Ukrainian adult migrants. The study obtained approval from the ethics committee of the Institute of Psychology at the Polish Academy of Sciences. Participation in the study was completely anonymous and voluntary. 221 Ukrainian migrants participated in the study (195 participants had valid data in both private and work-related conditions; 196 and 219 respectively), and the majority of them were women (72%). 78% of migrant participants entered Poland after the 24th of February 2022, after the Russian military invasion of Ukraine. One person did not want to reveal their gender, and three persons were nonbinary. 56 people were interviewed face-to-face and others via the internet through various platforms and social media groups to reach as heterogeneous groups of potential participants as possible. The average age was 38 (SD = 13, min = 20, max = 72) for Ukrainian women and 31 (SD = 13, min = 18, max = 71) for UA men. Ukrainian participants lived before migration in a city (48%), 34% in a town, and 16% in a village; after migration 85% in a city, 12% in a town, and 3% in a village. Since the study was carried out online mainly among volunteers, it was impossible to avoid the selection bias completely.

3.2 Measures

Copresence. We used the Polish adaptation of the Social Presence Survey from Bailenson, Blascovich, Beall, and Loomis [2] in a variant by Swidrak, Pochwatko, Matejuk [8]

for co-presence measurement. Participants filled the scale twice, separately for work-related (variable copresence-work) and private (copresence-priv) communication. Participants responded on a Likert scale from 1 (I strongly disagree) to 5 (I strongly agree). The score of each variable was calculated by extracting the mean of all items. The Ukrainian version of the Perceived Copresence Scale (PCS-U) was developed by 3-steps translation of instruments procedures. First, the original Polish PCS was translated into Ukrainian by two professionals (a psychologist, and a software developer) from outside the research team, both of whom were bilingual and fluent in Ukrainian and Polish. Next, the two translations were compared, and discrepancies were reconciled until one working draft in Ukrainian was achieved. Second, a bilingual expert panel consisting of two original translators and two science researchers reviewed the draft Ukrainian translation to make cultural adaptations as necessary. Third, the corrected Ukrainian scale was back-translated into Polish by one bilingual translator. The back-translated Polish version was then compared to the original Polish version and reviewed by the original author to ensure that the questions were translated correctly and a cultural and language equivalency was reached.

Housing Conditions. We controlled whether the participant lived alone or with others by asking about the number of adults, children, and small children living in the same household. In analyses we used two variables: the number of adults living with the participant (variable cohab-adult) and the total number of children (variable cohab-kids). We also inquired whether participants had their own room or had to share it (variable ownroom, 0 - no, 1 - yes). Communication conditions were: the number of hours spent at work on mediated communication (h-work), the quantity of video-calls (video-work), and the speed of the internet connection (internet-speed).

Hours of Private Calls and Hours of Work-Related Calls. To measure the quantity of mediated communication, we asked participants to estimate the number of hours they spend daily on private and work-related communication using: a virtual reality head-mounted display, a computer, and a tablet/smartphone. In the next step, we summed the hours separately for work-related and private calls (variables h-work and h-priv).

Internet and Electronic Devices. The Internet connection speed was measured with a single-choice question: My internet connection is (1) slow, (2) average, (3) fast, (4) very fast (variable: internet speed). We also asked about the percentage of calls in both contexts being video calls (variables: video-priv and video-work).

3.3 Data Analysis

Data analysis was carried out in SPSS version 28. We have used descriptive statistics to analyse the demographic data and housing conditions, a principal component analysis (PCA) to test the factorial structure of the Ukrainian Perceived Copresence Scale; and calculated the Cronbach's alpha to measure its reliability. We took an explorative approach, following Gellman and Hill [2]. To test the external validity of the scale, we calculated the correlation between the number of hours spent on work-related and private video calls.

Table 1. Means and standard deviations of co-presence.

Variable	UA migrants (N = 221)	
	M	SD
<i>Co-presence</i>		
Copresence-priv	2.21	1.10
Copresence-work	2.05	0.96

4 Results

4.1 Descriptive Statistics

Housing Conditions and Internet Usage. Housing conditions of Ukrainian migrants differed: 30% of them were working in an office and going to work each day, only 8% worked remotely and 10% worked in a hybrid form (partly remote and partly personally); 22% were students of various levels and forms of learning, 34% were temporarily unemployed, home caregivers or pensioners. Only one in ten respondents lived alone and 28% had own room; 24% lived with a partner, 38% lived with two or three people. Regarding the cohabitants, two thirds of participants lived with children, 21% with small children, 41% with one child. Majority of migrant participants (74%) had mobile/phone internet connection, only 4.5% used a cable internet connection, and 2.3% had the optical fibre connection. Only 1.5% of respondents declared no internet at home. Overall, the vast majority have a fast or rather fast connection (8% very fast, 44% fast, 38% average/medium), with only 9,5% declaring slow connection. Reliability of internet connection was rated as relatively low with half of the sample responding that it was sometimes interrupted, and was slowing down; 38% that it was often interrupted.

4.2 Validation of the Ukrainian Perceived Copresence Scale (PCS-U)

The PCA was used to verify the factor structure of the Ukrainian scale. Bartlett's test of sphericity was significant ($\chi^2 = 1219$ $p < .001$) and the Kaiser-Meyer-Olkin test score was .84. The PCA revealed one factor for the Ukrainian version, explaining 67% of the variance. Reliability analysis revealed very high level of internal consistency of PCS-U (Cronbach's alpha = .928, valid cases N=195). The correlation between copresence and the number of hours spent on work-related and private video calls was .174 ($p < .05$) and .289 ($p < .001$) respectively (Table 2).

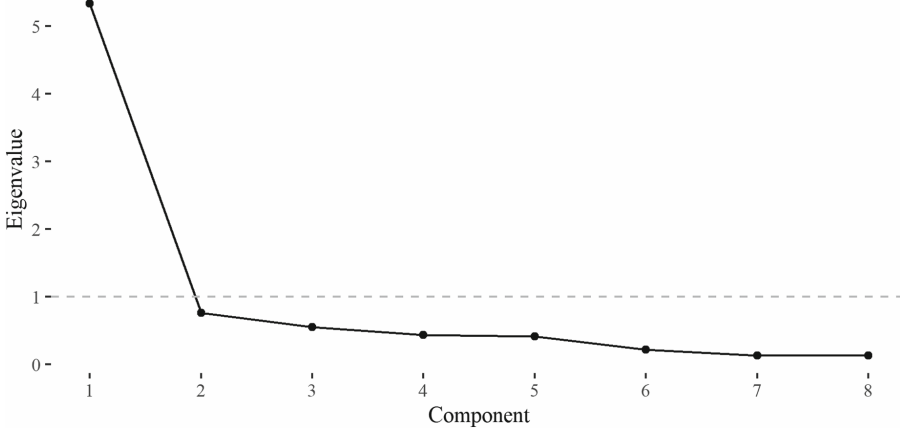


Fig. 2. Perceived Copresence Factor Analysis: scree plot.

Table 2. Perceived Copresence Component Score Coefficient Matrix

UA	PL	Component UA	Component PL
1. В особистому спілкуванні... Було відчуття, ніби ця людина насправді була зі мною в кімнаті	1. W sprawach osobistych... Czułem się tak, jakby ta osoba rzeczywiście była ze mną w pomieszczeniu	.134	.144
2. Я відчув, що ця людина дивиться на мене і усвідомлює мою присутність.	2. Czułem, że ta osoba na mnie patrzy i jest świadoma mojej obecności	.152	.147
3. Було відчуття, ніби екран комп'ютера/телефону не був бар'єром між нами	3. Czułem się tak, jakby ekran komputera/telefon nie stanowiły bariery między nami	.150	.152
4. Я відчував близькість цієї людини	4. Czułem bliskość tej osoby	.150	.158
5. У діловому спілкуванні... Було відчуття, ніби ця людина насправді була зі мною в кімнаті.	1. W sprawach zawodowych... Czułem się tak, jakby ta osoba rzeczywiście była ze mną w pomieszczeniu	.154	.143
6. Я відчув, що ця людина дивиться на мене і усвідомлює мою присутність.	2. Czułem, że ta osoba na mnie patrzy i jest świadoma mojej obecności	.167	.163
7. Було відчуття, ніби екран комп'ютера/телефону не був бар'єром між нами	3. Czułem się tak, jakby ekran komputera/telefon nie stanowiły bariery między nami	.157	.157
8. Я відчував близькість цієї людини	4. Czułem bliskość tej osoby	.159	.158

Extraction Method: Principal Component Analysis.

5 Discussion

The first result of the study is the validation of the Ukrainian Perceived Copresence Scale (PCS-U) as a one-factor measurement instrument with appropriate internal consistency. This gives reason to think about the life situation of migration as one that changes the conditions of work and personal interactions and reduces the feeling of copresence in remote communication, at least during the adaptation period. It means that copresence plays different functions during work and private communications in conditions of war migration; its pattern analysis is the task of the next study. We need to research copresence in private and work communication during the collective trauma healing process. Types of trauma events and the character of conversations about events and emotional stress may be independent variances for future research. Mass migration during the war changed the importance of having a proper physical workspace isolated from other cohabitants to participate in work-related calls efficiently. Copresence in the context of post-war society transformation, peacebuilding, and construction of historical remembering by war migrants, is an important subject for future research for a deeper understanding of copresence. There are several limitations to the study, such as not being representative of the Ukrainian migrant community sample, including gender representation. Future studies of migrants' technology-mediated communication and well-being should focus on research in a more controlled study design.

Acknowledgments. The research described in this publication was funded through an internship awarded to Professor Lyubov Naydonova by the Director of the Institute of Psychology at the Polish Academy of Sciences.

References

1. Alinejad, D.: Careful co-presence: the transnational mediation of emotional intimacy. *Soc. Med.+ Soc.* **5**(2), 2056305119854222 (2019)
2. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Interpersonal distance in immersive virtual environments. *Pers. Soc. Psychol. Bull.* **29**(7), 819–833 (2003)
3. Chen, W.: Internet-usage patterns of immigrants in the process of intercultural adaptation. *Cyberpsychol. Behav. Soc. Netw.* **13**(4), 387–399 (2010)
4. Didyk, N.: The influence of family problem on the psychological state of the servicemen during ATO. *Sci. Stud. Soc. Polit. Psychol.* **41**(44), 68–80 (2018). <https://ssppj.org/index.php/ssj/article/download/134/132>. Accessed 15 Sept 2022
5. Didyk, N.: Socio-psychological support of the servicemen's family members as the factor influencing the moral and psychological state of the combatant during the period of accomplishing tasks in ATO zone. *Ukrainian Psychol. J.* **1**(7), 41–57 (2018)
6. Felhofer, A., Hlavacs, H., Beutl, L., Kryspin-Exner, I., Kothgassner, O.D.: Physical presence, social presence, and anxiety in participants with social anxiety disorder during virtual cue exposure. *Cyberpsychol. Behav. Soc. Netw.* **22**(1), 46–50 (2019)
7. Short, J., Williams, E., Christie, B.: *The Social Psychology of Telecommunications*. Wiley, Toronto; London; New York (1976)
8. Świdrak, J., Pochwatko, G., Matejuk, P.: Copresence and well-being in the time of COVID-19: is a video call enough to be and work together? In: Biele, C., Kacprzyk, J., Owsiniński, J.W., Romanowski, A., Sikorski, M. (eds.) *MIDI 2020. AISC*, vol. 1376, pp. 169–178. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-74728-2_16

9. Tu, C.H.: The measurement of social presence in an online learning environment. *Int. J. E-learn.* **1**(2), 34–45 (2002)
10. UNHCR: UNHCR operational data portal: Ukraine refugee situation. <https://data.unhcr.org/en/situations/ukraine>. Accessed 15 Sept 2022






Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Modular Platform for Teaching Robotics

Damian Nagajek¹, Michał Rapala¹, Kamil Wołoszyn²,
Krzysztof Turchan², and Krzysztof Piotrowski²

¹ Space Research Centre of the Polish Academy of Sciences, Zielona Góra, Poland
{dnagajek,mrapala}@cbk.waw.pl

² IHP - Leibniz-Institut für Innovative Mikroelektronik, Frankfurt (Oder), Germany
{woloszyn,turchan,piotrowski}@ihp-microelectronics.com
<https://cbkpan.pl>, <https://www.ihp-microelectronics.com>

Abstract. The Fourth Industrial Revolution causes changes in the economy and shifts the job market demand towards workforce with high technical skills. To keep an undisturbed economic growth we have to encourage more young people to develop competences in STEAM (science, technology, engineering, arts and mathematics). This has met with a response from some education systems which have prepared special programmes focusing on developing technical skills among students. One of the desired field is robotics, which involves constructing and programming. We have already conducted some workshops for high school students in this subjects and we would like to find the correct teaching tools to attract primary school students. Our idea is to create a modular platform, which elements could be used as black boxes, to teach robotics to young children. We have noticed that Arduino based kits are a bit too complicated and we decided to test a LEGO Technic set equipped with an external microcontroller. We have verified the interest level of children and the difficulty and time needed for a teacher to master the whole teaching platform. According to our study, LEGO attracts students much more than Arduino and is easier in operation and less time consuming during classes for teachers.

Keywords: Autonomous robots · Recognition algorithms · LEGO · Robotics · Neural networks · OpenVINO · Microcontrollers · Python

1 Motivation

The 21st century brought us the Fourth Industrial Revolution (4IR). This industrial development is fueled by rapid changes in the digital technologies and industry, especially in the areas of artificial intelligence and advanced robotics. 4IR causes incredible modification of the global production and supply network by introducing smart technologies, wider communication between machines, internet of things (IoT), autonomous process monitoring, self-diagnostic etc.

The transformation from traditional manufacturing to full automatization of industrial processes totally shifts the role of human in the whole system; stronger

dependence on robotics moves us from the main workforce in a factory to auxiliary function of maintenance and emergency service. Undoubtedly, this situation has got many advantages like: less strains put on workers body, faster and cheaper production, better repeatability of a production process etc. Inevitably, the fast pace of drastic changes during the 4IR brings many challenges and dangers; especially socio-economic risks are raised in [10, 11] by some authors. One of the problems that arises is lack of qualified workforce that can operate at the newly created intelligent workplaces. According to Spoetti and Windelband [14] vocational education and training of the workforce will be highly relevant to successful implementation of the 4IR. This burdens the sector of education with an expectation to follow the needs of the industrial development and employment, what has a strategic importance for a stable and undisturbed economic growth.

The mentioned challenges affecting schools have been already spotted in some countries. In Poland, the Ministry of Education and Science has commissioned a study which defines an actual demand for particular professions and which of them the economy lacks the most. The published documents from 2021 [12] and 2022 [13] confirm the statements from the study by Spoetti and Windelband [14]. Three professions occurred on the list each year: mechatronics engineer, automation technician and robotics technician. In response, the Ministry of Education and Science in cooperation with the GovTech Center of the Poland's Prime Minister's office have created an education programme destined for primary schools called "The Laboratory of the Future". The idea is to financially support entities that develop students' competences in STEAM (science, technology, engineering, arts and mathematics). New workshops will be created, which should encourage pupils to learn these subjects. A huge focus has been put on robotics, more precisely on a controller programming.

2 Observations

Our laboratory together with a student research group called "KNIK" from the University of Zielona Góra and our partner in the EU's SpaceRegion project - IHP institute from Frankfurt Oder have organized in 2022 some controller programming workshops for high and vocational schools' students.

We have used Arduino starter kits composed of an Arduino Uno microcontroller board, a breadboard, jumper wires, a power unit, transistors, sensors, LEDs, motors, resistors, buttons, an LCD display etc. All in all, it is a quite versatile but simple to use set to introduce pupils to electronics and programming. During some classes also custom made microcontroller boards designed by IHP have been used.

In the previously mentioned "Laboratory of the Future" programme, schools are free to select appliances' models, since only a mandatory equipment type is specified. A microcontroller board is one of them. We have assumed that many will choose Arduino based sets due to relatively low price, good availability, comprehensive documentation and huge community support.

Not a lot of students have attended the organised workshops. From the destined group, only about 10% in the high school and about 20% in the vocational

school. The Arduino sets have seemed to be a bit too laborious and complicated to operate, since for many of them it has been a first contact with a microcontroller programming. Also handling of the accessories with universal connectors is a bit different than what students are used to - connecting equipment with dedicated plugs. Additionally, they have been a bit older and more skilled than primary school students to whom the new ministerial programme is devoted.

Another problem that can occur in many schools is lack of specialized teachers. A computer science tutor is skilled in programming but can have inadequate knowledge in electronics. Additionally, the time needed to put together a working unit based on the Arduino is relatively long, and taking into account very complex primary school syllabus, effective utilization of lesson's time is crucial.

3 Idea

In cooperation with our partner IHP we have looked for a solution which would appeal more to younger people. One of the most widespread toy types are bricks especially LEGO sets. Almost all children are familiar with them and they are generally appreciated.

Another advantage of a LEGO set with a microcontroller is simplicity of construction process. All parts can be easily and tightly connected, and the amount of possible configurations is vast. It has to be underlined that for primary school students, the playing aspect is very important, and good entertainment can help to stay focused and encourage them to learn. From our observations, quite raw Arduino sets, very utilitarian in their nature, are not really rewarding for students. On the contrary, LEGO gives a possibility to make your own construction move by adding motors and sensors in the right spots. It gives more freedom in model creation and saves lesson time due to relatively simple way of connecting parts. Students can be very creative and engaged in building new constructions and simultaneously, without a lot of effort, learn how to design control systems and program them.

For our tests, we have chosen a LEGO set featuring electric motors and a wheeled platform, see Sect. 4. Additionally, some external electronic parts and a control software written in Python have been implemented. We believe that these elements are relatively easy to handle, and the learnt skills would be useful at the job market. We had decided to give the construction and programming tasks to a PhD student inexperienced in such subjects to mimic a teacher's situation and to verify how challenging and time consuming is gaining a new knowledge. Young students reaction and interest level have been checked during a public science exhibition at IHP institute in Frankfurt.

4 Construction

A robot that has been created by the PhD student has been a wheeled platform with a camera. He has used some out of the box parts to simplify the development process and to make it more reproducible. The construction components are

shown in Fig. 1. The entire based structure has been assembled with elements from the LEGO Technic 42114 Volvo 6 × 6 set (Fig. 1a [4]). Motion system has consisted of four Lego Large Motors 88013 (Fig. 1f [3]) installed on the platform. They have been connected to a Raspberry Pi Build HAT (Fig. 1h [7]) which is an extension (shield) for the main computer - Raspberry Pi B4 (Fig. 1d [6]). An Intel Neural Compute Stick 2 (Fig. 1c [1]) and a Raspberry Pi Camera HD v2 (Fig. 1g [8]) have been linked to the main computer. The power has been supplied by a LiFePO4 battery (Fig. 1e [2]) connected to a step-down Voltage Inverter LM2596 (Fig. 1b [9]).

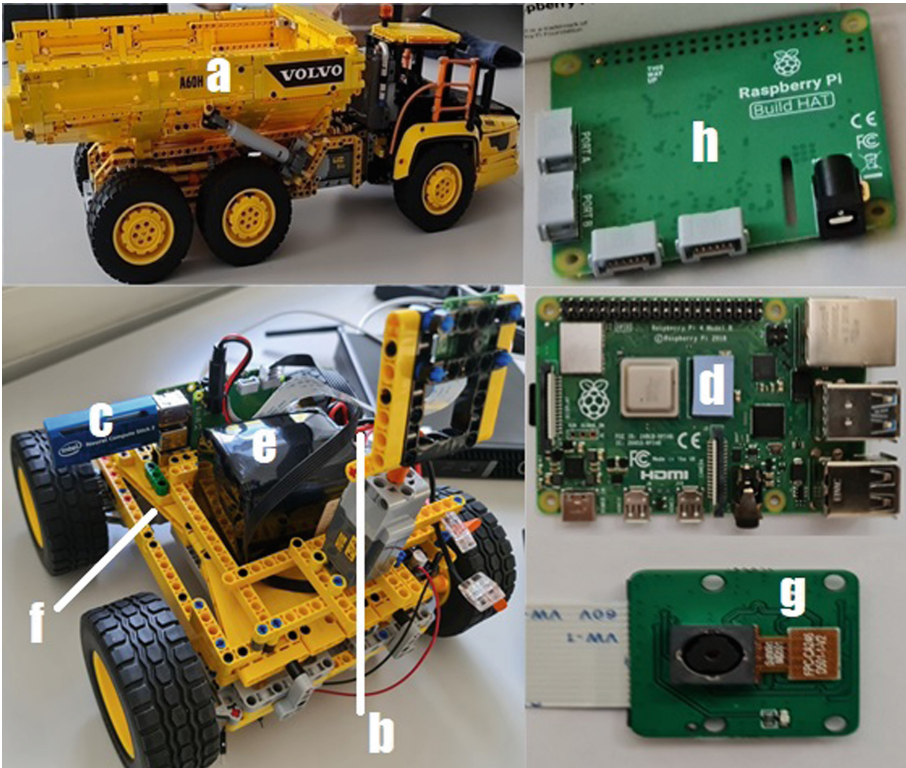


Fig. 1. Used components.

Only the compatible components have been selected, thus assembling of the entire platform has been relatively straightforward. A first three-wheeled prototype had been constructed within one week. It had been rear-wheel drive, and its drivetrain had included a front-mounted self-aligning wheel. Unfortunately, such a solution had encountered some mobility problems, so it has been decided to transform the entire structure to a four-wheeled vehicle. Still, a simple attachment of the wheels via LEGO cross axle has caused them to fall out of their

mounts during sudden turns. Moreover, a rather low torque and fast movement of the motors have made a precise control of the whole structure relatively cumbersome. It has oscillated or on the grippy surface has not moved at all. To solve these problems, we have applied a gear multiplier equipped with a dedicated hub for each wheel. In addition, we have had to upgrade the entire structure to make it more stable and to withstand the forces exerted by relatively heavy battery and electronic parts. A final construction is shown in Fig. 2.

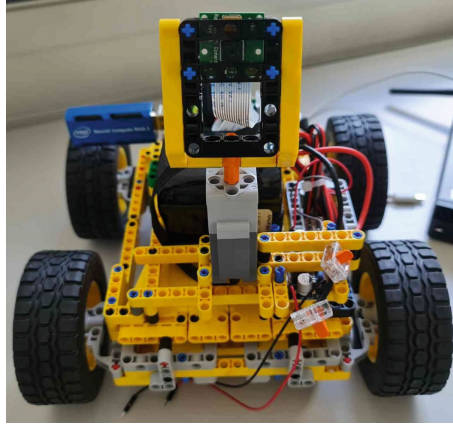


Fig. 2. Final construction.

The Raspberry Pi 4B has been the main computer board for this setup. It has been extended with the Raspberry Pi Build Hat which is a dedicated add-on board (shield) for the Raspberry Pi computer. Generally, this shield is compatible with various LEGO motors and can control up to four different units or sensors of distance, color, pressure etc. Moreover, it allows users to precisely control motors and read data from encoders i.e. an absolute motor position, its current speed or rotation direction. Additionally, a motors control via PWM signal is possible. A minimal possible rotation angle of used LEGO Large Motors 88013 has been 6° .

The Intel Neural Compute Stick 2 (Intel NCS2) is a plug-and-play USB hardware deep neural-network inference accelerator for computer vision and deep learning. The acceleration is working by assisting the main computer's processing unit (CPU) by taking over mathematical calculations required for running deep learning models. Moreover, the Intel NCS2 is supported with an OpenVINO (Open Visual Inference and Neural Network) library, which includes multiple pre-trained models e.g. face or text detection, formula recognition etc. A wide variety of them allows users to choose the most suitable for desired tasks. The biggest advantage of OpenVINO is its application simplicity. This library can be used as a ready-made module. A person without specialized knowledge is able to apply it without modifications and understanding of its working principles.

Inexperienced students can treat it as a black box and quickly create a fully functioning device.

For people and facial recognition task we have used the compatible Raspberry Pi v2 camera with an 8 MPx resolution. This device works with drivers included in Raspbian - Raspberry Pi's operating system, moreover it has hardware support what totally limits a consumption of CPU's computing power.

The whole system has been powered by LiFePO4 battery connected to the step-down Voltage Inverter LM2596. A connection diagram is shown in Fig. 3.

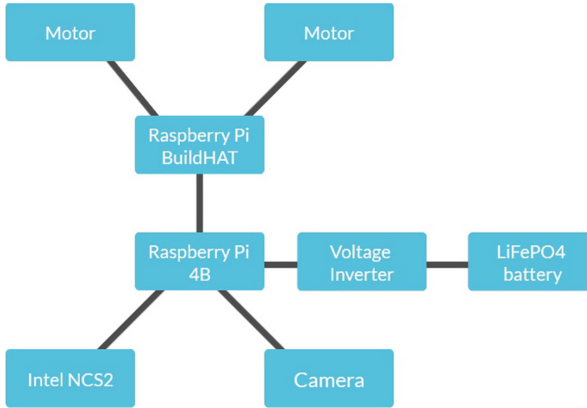


Fig. 3. Connection diagram.

5 Algorithm's Working Principles

The main task of the used algorithm has been recognizing a human posture and robot's movement control. It has had to follow and hold a desired target (a human) in the middle of the observed area.

We used an OpenVINO pretrained model from the Object Detection Models library called "person-detection-retail-0013" [5]. The control algorithm has been designed to center the detection frame in the middle of the observed area and then try to follow the target and hold it in that position (Fig. 4). The size of the detected target has had to be kept within defined limits what has forced the robot to zoom in or out by moving forward or backward.

Software code changes implemented by users can be kept to a minimum. It is only necessary to declare the camera resolution, the desired data detection frame size and the motors' speed. The whole algorithm works in real time. Users have to bare in mind that a wide accepted model size limit gives better error tolerance but can cause some control problems like lack of a trigger to move. On the other side, a too small limit can cause permanent movement due to defined, finite minimal possible motor's rotation angle. The software gets caught in an infinite loop while trying to center the detection frame because the minimal

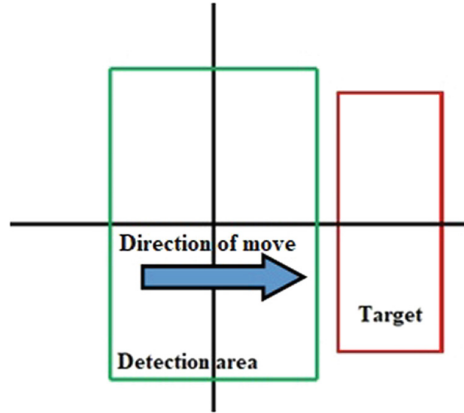


Fig. 4. Visualization of the algorithm.

movement value is bigger than the accepted margin. A similar situation will occur if the engines' speed is too high; the control unit is not able to proceed collected data fast enough and to send control commands on time. Also a too high video resolution can cause delays between target recognition and motors movement due to limited CPU's calculation power.

Generally, the camera constantly sends data to the computation unit. When a target occurs in the detection area and is recognized, the main computer transmits a control signal to motors via the Raspberry Pi Build HAT. The motors are running until the target's frame is centered in the middle of the detection area. The platform rotation is caused by movement of the wheels placed on different sides in the opposite direction. Afterwards motors shut down and wait until target leaves the center of the detection area. Moreover, the robot tries to keep the correct target size by moving forward and backward by rotating all wheels in the same direction. The robot follows a detected person to keep him within a certain distance. This has been very encouraging for children, who have thought that the robot has been scared by them and has run away to keep a safe distance every time they have come closer.

All pretrained models are implemented in .xml and .bin files and can be used like black-boxes in the modular fashion. Their content do not need to be changed by users. E.g. changing our software to recognize particular objects instead of people comes down to replacement of the complete .xml and .bin files to ones containing the correct detection model. This totally simplifies the software development and can be quickly executed even by inexperienced primary school students guided by a teacher who has completed only a short training.

6 Tests

Real life tests have been conducted during the Doors Open Day 2022 at IHP institute. The robot has been placed at the floor next to a stand. Other experi-

ments have been also presented in the same space. The forward-backward operation feature has been disabled to prevent the robot from moving away from the stand and colliding with someone.

The reaction of the targeted age group - children aged 7–15 years has been very enthusiastic. The bright yellow colour of the robot attracted their attention. Smaller children have been mostly interested in playing with the robot - moving from side to side and observing how it has followed them. Older primary school students generally have wanted to know how the whole thing works and what more it could do. We have received questions about used components and software. All in all, the presented robot has drawn much more attention and vigorous reaction than any of the Arduino sets used during the organized by us workshops mentioned before.

7 Conclusions

Our observations show that teaching primary school students using LEGO sets with additional electronic extensions seems to be a much better idea than introducing them to the Arduino. LEGO is well recognised among children, they know how to use it and what can be constructed with it. Younger children associate it with good fun and creativity freedom, older ones additionally with complex mechanics and possibility to control the components movements of LEGO Technic sets. Complementing LEGO with external microcontrollers and simple artificial intelligence creates a perfect educational set for primary school students. They can learn programming through playing and improve their creativity and technical skills by building diverse interactive constructions fulfilling various tasks. These can encourage them to develop competences in STEAM what is the aim of the “Laboratory of the Future” educational programme and will be desired by the economy changed under the influence of the Fourth Industrial Revolution.

Acknowledgements. This work was supported by the European Regional Development Fund within the INTERREG V A BB PL2014-2020 Programme, “Reducing barriers using the common strengths”, project “SpaceRegion: Cross-border integration of the space sector”, grant number 85038043.

References

1. Intel Neural Compute Stick 2 (2022). https://assets.hardwarezone.com/img/2018/11/Intel_Compute_Stick_2.jpg. Accessed 23 Nov 2022
2. JuBaTec - LiFePO4 12V 6Ah battery (2022). <https://bilder.jubatec.net/item/images/4652/middle/LiFePo4-12Volt-6Ah-C-1-min-4652.png>. Accessed 23 Nov 2022
3. LEGO Large Motor 88013 (2022). <https://www.lego.com/cdn/cs/set/assets/blt5f34c8d5e611bbe8/88013.jpg?fit=bounds&format=jpg&quality=80&width=1500&height=1500&dpr=1>. Accessed 23 Nov 2022

4. LEGO Technic - 6x6 Volvo Articulated Hauler (2022). <https://image.ceneostatic.pl/data/products/95856255/i-lego-technic-42114-zdalnie-sterowane-wozidlo-przegubowe-volvo-6x6.jpg>. Accessed 23 Nov 2022
5. OpenVINO Documentation - Object Detection Models (2022). https://docs.openvino.ai/2021.2/omz_models_intel_person_detection_retail_0013_description_person_detection_retail_0013.html#example. Accessed 23 Nov 2022
6. Raspberry Pi B4 (2022). <https://asset.conrad.com/media10/isa/160267/c1/-/pl/002138864PI00/image.jpg?x=400&y=400>. Accessed 23 Nov 2022
7. Raspberry Pi Build HAT (2022). <https://cdn.shopify.com/s/files/1/1217/2104/products/build-hat-2.png?v=1652768694&width=600>. Accessed 23 Nov 2022
8. Raspberry Pi Camera HD v2 (2022). https://cdn1.botland.com.pl/106921-pdt_540/raspberry-pi-camera-hd-v2-8mpx-oryginalna-kamera-do-raspberry-pi.jpg. Accessed 23 Nov 2022
9. Step-Down Voltage Inverter LM2596 (2022). https://lispol.com/uploaded/products_images/1459345578_09016600.jpg. Accessed 23 Nov 2022
10. Kuzmenko, O., Roienko, V.: Nowcasting income inequality in the context of the Fourth Industrial Revolution. *SocioEconomic Challenges* 1(1), 5–12 (2017). <https://doi.org/10.21272/sec.2017.1-01> <https://doi.org/10.21272/sec.2017.1-01> <https://doi.org/10.21272/sec.2017.1-01> <https://doi.org/10.21272/sec.2017.1-01>
11. Mertl, J., Valenčík, R.: The socioeconomic consequences of industrial revolution. *Cent. Eur. J. Manag.* 3 (2016). <https://doi.org/10.5817/CEJM2016-1-4>
12. Polish Minister of Education and Science: Obwieszczenie Ministra Edukacji i Nauki z dnia 27 stycznia 2021 r. w sprawie prognozy zapotrzebowania na pracowników w zawodach szkolnictwa branżowego na krajowym i wojewódzkim rynku pracy (2021)
13. Polish Minister of Education and Science: Obwieszczenie Ministra Edukacji i Nauki z dnia 28 stycznia 2022 r. w sprawie prognozy zapotrzebowania na pracowników w zawodach szkolnictwa branżowego na krajowym i wojewódzkim rynku pracy (2022)
14. Spöttl, G., Windelband, L.: The 4th industrial revolution - its impact on vocational skills. *J. Educ. Work* 34(1), 29–52 (2021). <https://doi.org/10.1080/13639080.2020.1858230>







Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





A Method for Co-designing Immersive VR Environments with Users Excluded from the Main Technological Discourse

Wiesław Kopec^{1,2,4} , Anna Jaskulska⁴ , Barbara Karpowicz¹,
Grzegorz Pochwatko^{2,4} , Monika Kornacka^{3,4} , and Kinga Skorupska^{1,4}  

¹ Polish-Japanese Academy of Information Technology, Warsaw, Poland
kinga.skorupska@pja.edu.pl

² SWPS University of Social Sciences and Humanities, Warsaw, Poland

³ Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

⁴ Kobo Association, Warsaw, Poland

Abstract. With the advent of new interfaces and modes of interaction related to virtual, augmented and mixed reality (VR, AR, MR) or voice user interfaces (VUI) a need to explore new approaches to foster rapid software prototyping and development emerges. Drawing from our experiences in human, cooperative, and collaborative aspects of software engineering, based on IT-empowerment and participatory design with older and younger adults, we propose and examine some methods, practices and tools for co-designing software for the immersive extended reality continuum (XR), including VR, AR or VUI. In a series of empirical and field studies we examined various experimental setups for stakeholders' participation within and across different steps and phases of the software development process for immersive extended reality environments (IERE). In this article we provide an overview of selected methods, practices and tools, that we found best support communication, collaboration, and cooperation among stakeholders, especially the members of vulnerable groups who are often excluded from the main technological discourse and need more empowerment.

Keywords: Virtual Reality · Participatory Design · Software Development

1 Introduction

Intensive growth of virtual, augmented and mixed reality solutions opens up new possibilities for better immersive experience of software solutions for end users. However, the rapid growth of those areas that comprise the immersive extended reality continuum (XR), brings new challenges to the software development process and teams. Despite the wide range of existing methods, tools and approaches for developing diverse mobile and desktop environments, including user-centered design coupled with end-users participation, i.e. co-design and

participatory design approaches, there is still a need to select and tailor the most promising solutions to new interfaces and modes of interaction of the XR continuum. Recently increasing pace of both hardware capabilities and its proliferation puts pressure on software developers to deliver more content as more diverse groups of users enter the market for Immersive Extended Reality Environments (IERE)¹. Therefore, the objective of this paper is to present our findings regarding the methods, practices and tools aimed to facilitate direct involvement of users of varying ICT-proficiency levels, particularly from vulnerable groups, in the participatory design process for the development of new interfaces and modes of interaction related to virtual, augmented and mixed reality (VR, AR, MR). In a series of empirical and field studies [1, 2, 6, 9–11, 18, 23] we examined various experimental setups for stakeholders’ participation within and across different steps and phases of the software development processes. Therefore, we provide an overview of selected methods, practices and tools, that we found best support communication, collaboration, and cooperation among stakeholders, especially members of vulnerable groups who are often excluded from the main technological discourse.

2 Related Work

Participatory design, or co-design, engages users in the development processes [20, 21], allowing them to directly shape the designed solutions at their core. Users can provide insights at different stages of the process [5, 15] - or, according to Ladner [14], become designers themselves, in Design for User Empowerment. Without this, even solutions meant to work towards Social Good can be tinted by stereotypes [24], which can become evident in unexpected places due to unconscious biases held by the project team, putting into question the ethical aspects of such solutions, as some some stereotypes may trickle down into the immersive environments built. [4] To prevent this, user engagement during the design and development is crucial. One way of approximating it is by forming Living Labs. The term *Living Lab* itself was coined by William Mitchell from MIT [16] and indicated a space where routines and everyday life interactions of users and new technology could be observed and recorded [3, 22], to examine the users’ needs and avoid business risks [19]. However, besides problems with maintaining long-term and sustainable user communities [17] a realistic Living Lab, requires significant investments. Some alternatives, like lightweight living labs [9], were proposed where users were animated with interesting activities which keeps them engaged. This approach is also useful for new and emerging interfaces such as

¹ In this paper we define **Immersive Extended Reality Environments (IERE)** as environments which use new technological solutions running software to enhance or replace the experience of reality by appealing to the users’ various senses. What these environments have in common is the use of the real life metaphor, thus making participatory design rely more on ethnographic studies which evaluate and elaborate on the nature of users’ interactions with technologies which have a chance to become truly embedded in their daily lives.

extended or mixed reality (XR), like VR, AR or voice interaction. There is also a lot of room for further research to establish a comprehensive and ubiquitous interface that combines all of the extended reality technologies. In this vein, it is possible to engage users in a distributed Living Lab environment [11] aimed at testing and developing these solutions, and eventually even engage them in start-up development teams, as in the SPIRAL method described in one of the previous CHASE workshops [9]. Our RAPID approach draws from all of these methods, while opening the discussion for further exploration of new and emerging interfaces on the extended reality continuum in the context rapid software development.

3 RAPID Approach Outline

Based on our previous experience with empowering older adults and other vulnerable groups for participatory design, we developed an instant environment called **RAPID (Rapid and Agile Participatory Interactive Development)** devoted to lightweight Living Lab conditions. The RAPID approach consists of three phases divided into steps and for each of these we list related methods and expected outcomes.

3.1 I. Preliminary Phase: Team Formation

This preliminary stage consists of two steps, that can be realized in a few days or partially omitted, in particular step 1. in experienced teams or even step 2. based on communities such as Living Labs or local activity groups.

Step 1. Core XR Team Formation and Empowerment. Internal workshops with methods presentation, discussion and roles assignment. This is an important step when we prepare the team for direct cooperation with users as they may hold some stereotypes about them as well. Internalization of methods is important, especially to ensure unbiased and open cooperation with vulnerable users and among each other in the in-team role assignment, i.e. product owner, team leader, technology advisor, developer, designer etc.

Methods and Tools that Can be Used:

- evidence from previous successful projects with direct user participation in different contexts, i.e. videos, artifacts (actual applications, products) which may come from outside of the ICT-area, for example, from participatory Social Design projects,
- interviews with designers successfully engaging in such projects, sharing the benefits and challenges from their points of view
- ideas about what the team knows about the users, and what is unknown, which can later be verified

- brainstorming for ideas and insights about the solution with core team members, further facing their stereotypes of the other group (users),
- affinity diagramming for collection, categorization and analysis of ideas,
- mind mapping for methods, ideas, team roles and responsibilities.

Expected Result: Team members understand the importance and goal of the process and are open to direct cooperation with end-users, while remaining aware of their unconscious biases.

Step 2. XR Team Extension: User Engagement. Preliminary trials with representatives of the group of end users, user engagement, recruitment and selection in order to extend the core team (from local communities, groups, Living Labs). User engagement session. Preliminary trials with end-users as potential team members. Controlled audio-visual immersion into the IERE world. Direct interaction (i.e. controllers) optional, not required.

Methods and Tools that Can be Used:

- showcasing interaction with various IERE interfaces (Brain-Computer Interface (BCI), Voice User Interface (VUI) and various types of XR solutions)
- brainstorming possible everyday uses of such interfaces to invoke creativity
- engaging users in a simple, largely passive, VR environment (i.e. cardboard or video pass-through)
- advanced XR environment (i.e. full headset or an immersive AR game)
- fully fledged demos and trials (i.e. fluent and attractive)
- semi-structured individual and group interviews with users, screening forms.

Expected Result: Recruitment of two to six end-user representatives with high motivation and a creative flair for the main co-design phase session and team participation.

3.2 II. Main Phase: RAPID IERE Development

Main phase consists of two stages: user empowerment stage (steps 3 and 4) and co-design development stage (steps 5, 6 and 7). Each stage can be realized in one day, or extended as needed.

Step 3. Introduction: Discussion of the Goal of the Workshops and Current Pre-workshop Expectations of Participants. An informative and motivating introduction is key, as both the potential users and the development team need to realize the purpose and importance of their work. Additionally, the participants ought to introduce themselves, to provide their background on technology use, and their experience with the specific topic and their own expectations of these workshops. When engaging vulnerable groups it is crucial

to not strain their resources and to ensure that the benefit is mutual for the parties involved. Next, in order to set baseline expectations, we introduce our participants to the idea of the project without putting it in the context of specific technologies nor equipment.

Methods and Tools that Can be Used:

- ice-breaking games to learn about everyone’s motivations and aspirations
- division into sub-teams, each with at least one representative of a different group (target user, programmer or designer), to facilitate debates and allow everyone enough speaking time to voice their opinions
- semi-structured interviews
- brainstorming, affinity diagramming, mind mapping
- co-creating mood and context boards exploring the potential of the project
- sharing insights accross different sub-teams

Expected Result: The participants understand the importance and goal of the workshops and share their experiences with the subject matter of the workshop (e.g. in our cases: banking technology and ATMs, potential uses of Smart Home technology with VUIs and BCI, workplace stress, intrusive thoughts and relaxation techniques).

Step 4. Immersion with Interaction: Hands-On Presentation of the IERE Technology. This step is crucial for the target group to get a feel of what is possible in the technology of their choice, so that they are not limited by their presuppositions or lack of experience in the development stages that follow. It may also be beneficial for the target group to witness the development team also discovering something new and engaging within the demos. Moreover, experiencing immersion may get the users excited about this technology, awakening their creativity. In this step the target users try IERE gear and engage in interactive activities, which to some extent may be similar to the ones being developed later, either in scale, means of interaction or goal (Fig. 1).

Methods and Tools that Can be Used:

- Showcasing passive commercially available low-cost products that are within their reach and their key functions: Cardboard 360 movies and experiences, XR experiences, VUI in their own smartphones
- Engaging the users in active experience creation: capturing 360 photos with their own phones, setting functions of their own voice assistants
- Explaining the idea of the Internet of Things and how experiences can be combined and co-dependent
- Letting all of the participants engage with commercial higher quality games and applications available on higher-end devices, which display different and more-advanced aspects of IERE and the range of possible interactions with the real world and each of them through the Internet of Things



Fig. 1. In one of our workshops older adults played the NVIDIA® VR Funhouse game, to become aware of the wide range of what is possible in VR

Expected Result: The participants understand the range of possibilities that exist in IERE and get excited about the technology.

Step 5. Development - Environment in IERE. While working on any type of project that involves software engineering many environmental factors come into consideration, such as type of devices, software choices or general theme. While in our cases most of those factors were predetermined by the concepts of the project, hardware solutions available to us at the time and team's skill, we were able to involve all of our participants in the process of designing XR playing space and brainstorming on the potential future uses of other IERE solutions, while also gaining valuable insights in their real life preferences.

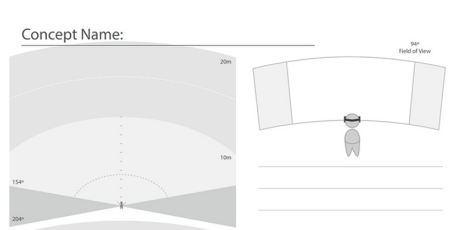


Fig. 2. We printed out VR Sketch Sheets, to be used for web-based XR environment design from <https://blog.prototypr.io/>, however, using them directly with members of our target group proved to be too reliant on an unknown metaphor and too detached from the familiar experience.

Methods and Tools that Can be Used:

- Semi-structured interviews
- Showing existing large scale projects and their interactions, eg: Google Earth VR Street View in different locations, engaging in unscripted interactions with audio assistants

- Evaluating existing solutions as design case studies: eg: Rating different aspects of the presentation of locations, talking with virtual assistants
- Presenting the specific case and evaluating it in the same format
- Using paper prototyping tools to jot down insights: XR Paper Prototyping with the use of Sketch Sheets 2)²
- Creating flow diagrams to depict the discussed interaction methods and aspects

Expected Result: The participants are involved in project defining decisions, can visualize and evaluate the suitability of different environments, weigh the dangers and benefits associated with each choice, and come up with their own ideas regarding environmental variables.

Step 6. Development of the Concept. Working in IERE environments opens a whole new range of possibilities regarding UI and UX, as such interfaces rely on approximations of real interactions to a greater extent, therefore it is of utmost importance to have as much feedback as possible. By using quick photoshop mockups, 3D modeling software and interaction and flow mock-ups in web-based tools like Proto.io, we were able to gather important insights on what our participants expect from our projects, while allowing them to quickly see mock-ups of their ideas (Fig. 3).



Fig. 3. Older adults were shown web-based VR mock-ups from Proto.io on smartphones

Methods and Tools that Can be Used:

- Paper prototyping of interface elements and their functionalities, which are later turned into quick clickable mock-ups on e.g. Adobe Experience Design
- Proto.io for UX tests on the initial UI
- Adobe Photoshop rapid UI prototyping

² The Sketch Sheets are available from: <https://blog.prototypr.io/vr-paper-prototyping>.

- Unreal Engine as a quick means of creating simple VR project
- Autodesk Maya 3D for 3D modeling purposes

Expected Result: The participants conceptualize and create UI and UX elements while having the ability to give immediate feedback to our rough interpretations.

Step 7. Development - Functionalities. This step, as one too technical for easy and quick implementation, focuses solely on the participants' ideas and insight without prototyping the functionalities there and then. Participants explore ideas and conceptualize their implementations and restrictions, to go along with the project concept and future and existing assets. It is important to give freedom to participants' ideas and leave out any implementation restrictions. While applicable pre-made sets of interactions can be presented during discussion as a starting point, it is not necessary.

Methods and Tools that Can be Used:

- Passive interactions with environments in Proto.io, Unreal or Unity
- Wizard-of-Oz method
- Brainstorming with mind-mapping the functions in each environment
- Engaging with other fully functional IERE interfaces to remind of the possible range of interactions
- Affinity diagramming of most commonly suggested concepts

Expected Result: The participants conceptualize project functionalities, without software/engine restrictions.

3.3 III. Closing Phase: XR Product Delivery and Testing

This phase is from the classic user-centered design approach, so we do not elaborate on it.

Step 8. Testing with End Users. We would like to point out three important groups of methods, i.e. qualitative interviews and observations and quantitative sensor-based methods and tools.

Methods and Tools that Can be Used:

- Semi-structured qualitative interviews,
- Hands-on usability tests,
- Eye-tracking and other sensor-based quantitative methods.

Expected Result: Verifying the results with extended team members and new participants, as well as the extensive network of contacts they all have to gather fresh perspectives. Testing is important and can be done outside of the work setting, by presenting the demo at a conference or attending fairs, events where members of the target group may be present.

4 Discussion

Our RAPID approach was developed and verified in design cases, for example a series of design workshops of a virtual simulator of ATM and well-being XR environments [12], while its user-empowerment elements were tested in research related to Voice User Interfaces (VUI) [13] and Brain Computer Interfaces (BCI) [8] for IoT, as well as in a case study of an all round XR-hackathon with several dozens of stakeholders from various groups, including software engineers and content developers [6,7]. Overall, the RAPID approach is well-suited for creating XR solutions with users who need to be empowered, in our case older adults³, and we expect this to be true for other vulnerable groups at risk of being viewed through the lens of stereotypes or just feeling vulnerable in new circumstances.

4.1 Phase I: Team Formation

In the case of participatory design high motivation and engagement are crucial, which is why RAPID **Phase I** is essential for the success of the approach. Here are some takeaways from our practice:

1. One good practice is to **have the team discuss the most promising target user candidates**, in terms of their involvement and the ease of collaboration. As this approach is fast, it is important to gather people who enjoy working together, are interested in the project and want to express themselves. On one hand this may seem counter-intuitive, as we give voice to the target group members who are outspoken already, but on the other hand, they do belong to the target group, and they are more likely to be early-adapters, so we in a way, we do design for them when we work on the early version of the solution.
2. We found that granting the potential target groups **access to technology**, which may otherwise be out of their reach (either financially, or because of the setup) is a very good way to guarantee engagement, and usually it is enough to convince them to participate in the development process.
3. Vulnerable users, who are shown technologies they had no experience with before, need assistance from dedicated staff who encourage them and ensure the first impressions are smooth.

4.2 Phase II Development

After the users excluded from the main technological discourse join the development team, empowerment is even more important. They are now among people who are tech-saavy, driven and may dominate the discussion. Some takeaways from our studies indicate that:

³ The groups needing empowerment may be not as obvious, as in different contexts we enter different relationships where knowledge or power are not equally distributed. This is the case with students vs teachers, employers vs employees but also clients and contractors or even scientists in interdisciplinary teams.

1. One strong need of our participants was to **clearly understand the technology and the nature of the development process**, as well as the end goal of the project and each of the involved parties' stake in it. For example, in XR workshops it was reflected in the preferred use of familiar terminology, such as “simulated” instead of “virtual”.
2. **Taming the technology:** surprisingly every participant had almost no issues with getting used to VR headset and its controllers. The participants treated both stationary and room scale VR experiences as a novelty. They unanimously agreed that room scale VR felt more impressive; however, some participants concluded that cardboards are better suited for beginner users. The same was true of using VUIs, but in this case some participants decided this solution could work for other members of the target group, but not for them - **showing some prevailing intergroup stereotypes** to be aware of, even when engaging in participatory design.
3. **Giving a place to start:** The discussion inspired by Google Earth VR Street View in 3D was valuable, as it gave the participants space to imagine other interactive aspects of the possible solution, without the need to prototype it first. The users explored potential scenarios for the ATM use, as well as locations, pointing to specific places and recalling different situations, as well as mentioning their own insecurities about them.
4. **Drawing and prototyping:** It is not necessary for the target group to draw and conduct prototyping by themselves, as designers may do this for them following instructions and descriptions they provide. Overall, Design for Empowerment, where users become designers, does not have to mean that they have to become craftspeople as well. For XR setting, paper prototyping did not seem to work for us, as the metaphor of changing 2D depictions into 3D ones was too detached from the target project for users lacking experience, therefore we found low-fidelity prototyping to be better suited for generating insights.
5. **Not being shy about half-finished products:** In our ATM case, thanks to the participants' interaction with 3D mock-ups we gathered multiple insights concerning both the build in terms of product design of the ATM as well as its UI, discussed in the context of existing ATMs. We have also gathered ideas regarding preferred relaxation environments, including all immersive aspects such as sounds, colors, movement and avatar placement and their realism, down to the wind movement and animal species.

4.3 Phase III Closing

The UX tests were conducted both by the users and the development team in order to maintain the fresh approach and view on the different aspects of the ATM and its usability, controls, environment and the final purpose as a training simulation and the well-being environment (Fig. 4).



Fig. 4. Example of immersion verification with eye-tracking and psychophysiology

4.4 Other Considerations

Scheduling. RAPID is designed to be concise but scalable. This means, that it can be extended or compressed to just five days, as follows:

- Day 1 I. Preliminary Phase - team assembly and empowerment;
- Day 2 I. Preliminary Phase - user recruitment and engagement;
- Day 3 II. Main Phase - user empowerment;
- Day 4 II. Main Phase - RAPID prototyping;
- Day 5 III. Closing Phase - preliminary product delivery and testing.

Moreover taking into account the typical iterative nature of agile approach and rapid prototyping, some steps can be omitted i.e. core team empowerment or user recruitment, based on previous cycles of software development process or previous projects. Below we present a possible schedule of implementation of the RAPID approach as a sprint.

It is vital to schedule the meetings with enough time for questions and digressions since the whole team and the users alike need to feel their presence and insights are of importance - to an extent even if they relate to the issues outside of the main goal of the workshop. While they may not be crucial for the end product, these questions are essential for the process that relies on open sharing. Often such workshops run long, so ordering catering is another good practice.

Venue. The RAPID approach will work best if the venue changes, depending on the focus of each phase.

- **I. Preliminary Phase:** while Stage 1 with the team can be conducted anywhere, it is advisable to get out of the usual place of work to allow the team to think out of the box in a more creative setting, which does not resemble the regular work setup. Stage 2 on the other hand should be done in an open social environment, such as a fair, a community center or any other open venue in which potential participants may feel empowered to take the novel gear on a trial run.
- **II. Main Phase:** This phase can happen in any flexible workshop space with the possibility to create a round-table to facilitate open discussion
- **III. Closing Phase:** Ideally the venues in this case would be environments similar to where the end product is expected to be used.

4.5 General Discussion

Despite the fact that many methods and tools were developed in recent years, there is still a room for discussion to propose a better, more comprehensive approach for the software development teams to address the challenges of constantly changing hardware landscape of IERE. For instance, many of use cases described at premier software engineering and HCI conferences were based on the rapid prototyping using cardboard platforms, thus became more or less obsolete or hard to implement, as support for these weakened or ceased. On the other hand, there are also some shortcomings of paper prototyping that are not easy to overcome, because new interfaces and modalities are not just a simple extension of the WIMP paradigm (windows, icons, menus, pointers). In contrast, methods for prototyping 2D interfaces are well-established in both software engineering and human-computer interaction with plethora of effective tools, including low fidelity mock-ups that can be prepared even by the end-users themselves.

Thus, drawing from our research experience, we concluded that if even users who are generally excluded from the main technological discourse can creatively and competently engage in such 2D prototyping activities with a proper empowerment. On the other hand, during our research on software engineering for advanced and multi modal IERE interfaces we observed that unlike prototyping mobile or web-based flat environments and applications there is a gap between proficient software designers, developers and end-users. This gap is evident in the field of paper prototyping of immersive interactive environments, which constitutes an important barrier to direct cooperation with end-users, content designers and software engineers. Even empowered users have significant difficulties in bridging the gap between low-fidelity paper prototypes and hi-fidelity immersive multidimensional interaction. This situation resembles the case of architectural design, where drawings and schemes are not enough for many end users to fully understand and imagine the final solutions, hence the industry practice of 3D visualisation of models and spaces, more recently with the use of VR-based solutions. In our case, the same metaphor applies, where prototyping for IERE has to “hit closer to home” to enable the users to fully understand the functionalities of end products and to be able to contribute feedback and relevant insights. These low-fidelity functioning prototypes are generated based on insights from the earlier steps in the development cycle, which rely heavily on the empowerment of users, consisting of demonstrations, free use of technology and discussions with the team, unhindered by fears and false preconceptions, and open to constructive criticism.

5 Conclusions

In this paper we outlined the RAPID approach with example methods and tools based on a series of case studies. It is an attempt to discuss and explore new approaches to rapid software prototyping and development relying on user-empowerment of users typically excluded from the main technological discourse and facing unconscious biases in the context of immersive extended reality environments (IERE). The proposed approach allowed us to gain valuable insights on

the development process of Immersive Extended Reality Environments, including such aspects as UX, UI, 3D models, web-based mock-ups and current and future functionalities, but it remains to be seen whether the software produced with this approach perform reliably better and are more likely to be used by these groups. At the same time we are excited about the prospects of the discussions and follow-up studies, further testing the limits of such sprints, and refining such IERE development methods to facilitate the creation of great experiences directly with users excluded from the main technological discourse, empowered to share their opinions, needs and aspirations.

Acknowledgments. This study constitutes an example of a bottom-up participatory research initiative done in the spirit of transdisciplinary collaboration between scientists, practitioners and volunteers. It was conducted without a dedicated grant to further the understanding of key concepts in HCI in the context of immersive virtual environments (IVR). Therefore, we would like to thank the many people and institutions gathered together by the Living Lab Kobo and HASE Research Group. First, we would like to thank all the members of HASE research group (Human Aspects in Science and Engineering) and Living Lab Kobo for their support of this research. In particular, the members of XR Lab and XR Center of Polish-Japanese Academy of Information Technology (PJAiT), Emotion-Cognition Lab SWPS University (EC Lab), Psychophysiology Lab of the Institute of Psychology Polish Academy of Sciences, Kobo Association, as well as Living Lab Kobo community, especially older adults, for supporting recruitment and their participation in the lab studies.

References

1. Balcerzak, B., et al.: Press F1 for help: participatory design for dealing with online and real life security of older adults. In: 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), vol. 1, pp. 240–243. IEEE (2017)
2. Balcerzak, B., Kopeć, W., Nielek, R., Warpechowski, K., Czajka, A.: From close the door to do not click and back. security by design for older adults. In: Shakhovska, N., Stepashko, V. (eds.) CSIT 2017. AISC, vol. 689, pp. 40–53. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-70581-1_3
3. Ballon, P., Pierson, J., Delaere, S.: Test and experimentation platforms for broadband innovation: examining European practice (2005)
4. Brey, P.: The ethics of representation and action in virtual reality. *Ethics Inf. Technol.* **1**(1), 5–14 (1999). <https://doi.org/10.1023/A:1010069907461>
5. Davidson, J.L., Jensen, C.: Participatory design with older adults: an analysis of creativity in the design of mobile healthcare applications. In: Proceedings of the 9th ACM Conference on Creativity & Cognition, pp. 114–123. ACM (2013)
6. Kopeć, W., Balcerzak, B., Nielek, R., Kowalik, G., Wierzbicki, A., Casati, F.: Older Adults and hackathons: a qualitative study. In: Proceedings of the 40th International Conference on Software Engineering, ICSE 2018, ACM, New York (2018)
7. Kopeć, W., et al.: VR hackathon with goethe institute: lessons learned from organizing a transdisciplinary VR hackathon. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York (2021)

8. Kopeć, W., et al.: Older adults and brain-computer interface: an exploratory study. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York (2021)
9. Kopeć, W., Nielek, R., Wierzbicki, A.: Guidelines towards better participation of older adults in software development processes using a new spiral method and participatory approach. In: Proceedings of the CHASE 2018: International Workshop on Cooperative and Human Aspects of Software, ICSE 2018. ACM, New York (2018). <https://doi.org/10.1145/3195836.3195840>
10. Kopeć, W., et al.: Hybrid approach to automation, RPA and machine learning: a method for the human-centered design of software robots. arXiv preprint [arXiv:1811.02213](https://arxiv.org/abs/1811.02213) (2018)
11. Kopeć, W., Skorupska, K., Jaskulska, A., Abramczuk, K., Nielek, R., Wierzbicki, A.: Livinglab pjait: towards better urban participation of seniors. In: Proceedings of the International Conference on Web Intelligence, WI 2017, pp. 1085–1092. ACM, New York (2017). <https://doi.org/10.1145/3106426.3109040>
12. Kopeć, W., et al.: VR with older adults: participatory design of a virtual ATM training simulation. *IFAC-PapersOnLine* **52**(19), 277–281 (2019)
13. Kowalski, J., et al.: Older adults and voice interaction: a pilot study with google home. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA 2019, pp. 1–6. Association for Computing Machinery, New York (2019)
14. Ladner, R.E.: Design for user empowerment. *Interactions* **22**(2), 24–29 (2015)
15. Lindsay, S., Jackson, D., Schofield, G., Olivier, P.: Engaging older people using participatory design. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 1199–1208. ACM (2012)
16. Niitamo, V.P., Kulkki, S., Eriksson, M., Hribernik, K.A.: State-of-the-art and good practice in the field of living labs. In: 2006 IEEE International Technology Management Conference (ICE), pp. 1–8. IEEE (2006)
17. Ogonowski, C., Ley, B., Hess, J., Wan, L., Wulf, V.: Designing for the living room: long-term user involvement in a living lab. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1539–1548. ACM (2013)
18. Orzeszek, D., et al.: Beyond participatory design: towards a model for teaching seniors application design. In: CEUR Workshop Proceed (2017)
19. Pallot, M., Trousse, B., Senach, B., Scapin, D.: Living lab research landscape: from user centred design and user experience towards user cocreation. In: First European Summer School “Living Labs” (2010)
20. Sanders, E.B.N.: From user-centered to participatory design approaches. In: Design and the Social Sciences: Making Connections, pp. 1–8 (2002)
21. Sanders, E.B.N., Stappers, P.J.: Co-creation and the new landscapes of design. *Co-design* **4**(1), 5–18 (2008)
22. Schumacher, J., Feurstein, K.: Living labs-the user as co-creator. In: 2007 IEEE International Technology Management Conference (ICE), pp. 1–6. IEEE (2007)
23. Skorupska, K., Núñez, M., Kopeć, W., Nielek, R.: Older adults and crowdsourcing: Android tv app for evaluating tedx subtitle quality. In: Proceedings of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 286–296. ACM (2018)
24. Vines, J., Pritchard, G., Wright, P., Olivier, P., Brittain, K.: An age-old problem: examining the discourses of ageing in HCI and strategies for future research. *ACM Trans. Comput.-Hum. Interact.* **22**(1), 2:1–2:27 (2015). <https://doi.org/10.1145/2696867>


Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Improving the Usability of Requests for Consent to Use Cookies

Kristina Lapin^(✉)  and Laima Volungevičiūtė

Vilnius University, Vilnius, Lithuania
kristina.lapin@mif.vu.lt

Abstract. An HTTP cookie (hereinafter cookie) is a piece of information that maintains a state in a stateless HTTP protocol. Recently established privacy and security regulation imposes an obligation on the service provider to obtain the user's consent to use cookies. This paper is aimed at studying usability guidelines of requests for consent to use cookies (hereinafter consent requests). The consent requests have usability issues that make it difficult for the users to choose the right privacy and security options. A study of privacy and security regulation is aimed at extracting design requirements. Manipulative designs also known as dark patterns are explored and applied to assess consent requests of two of the most popular Lithuanian news portals. An evaluation revealed the presence of dark patterns in consent requests as well as violation of privacy and security requirements. As a result, usability guidelines on the design of cookie consent requests are developed.

Keywords: HTTP cookies · cookie consent requests · dark patterns · security · privacy

1 Introduction

An HTTP cookie is a small piece of information passed between a web server and a browser [1]. This is a collection of information the server creates when a user visits a website [2]. The information stored in the cookies helps the website to identify the user or restore the options set by the user – it can be login, the pages visited by the user on the website, or other usability options, such as language and font size.

Cookies are used to check whether two received requests were sent from the same web browser and to remember the state of the website, for example, stored information about what goods are placed in an e-shop cart. Although the purpose of cookies is to track the state of the web page which would be lost when the user leaves the domain, more detailed purposes are often distinguished [3]: (a) *strictly necessary cookies* for the smooth functioning of the website; (b) *preferences cookies*, such as language, font size, login name, and password; (c) *statistics cookies* collect information about the user's behavior on the website, such as links the user visited; (d) *marketing cookies* track user activity on the Internet to help deliver personalized advertising.

Depending on the collected content, cookies may endanger users' privacy [4]. Users' devices and their information are recognized as personal space in the European Union

(hereinafter EU) law. According to EU law, to process users' personal data, service providers must inform them about the methods of data processing and obtain their consent [5]. Implementing the law, cookie consent requests are introduced on the websites. However, the consent requests require additional mental effort that occurs unexpectedly when opening the page, thus usability issues arise.

The design of consent requests usually does not help the users understand what they are agreeing to. Even more, designers use knowledge of user behavior (e.g., psychology, A/B testing) to generate as many consents as possible when designing consent requests. Such misleading designs are called dark patterns [6]. They are also widely used in social blogs, online stores, mobile apps, and even computer games [7]. Consent requests include various dark patterns, such as information hiding, manipulation of element positions, formatting, blocking of the website content, the abundance of choices, hard-to-see text or links to additional settings, and pre-marked choices [8]. This manipulation degrades the user experience of using the website.

This paper aims to formulate usability guidelines facilitating the design of consent requests that meet privacy and security requirements, are easier to understand, and are less annoying to users. For this purpose, the next chapter explores essential privacy and security requirements. The third section examines dark patterns found in cookie consent requests. Further, the evaluation of consent requests of the most popular Lithuanian news portals based on revealed requirements and dark patterns is conducted. Finally, the usability guidelines facilitating the design of more usable cookie consent requests are formulated.

2 Privacy and Security Regulation

The laws regulating the processing of personal data in the European Union are defined in the General Data Protection Regulation (hereinafter referred to as GDPR). In this regulation, personal data is defined as any information relating to a person whose identity can be or has already been identified [9]. According to the EU directive on privacy and electronic communications supplementing it (hereinafter referred to as E-Privacy Directive) [3], service providers must obtain the user's consent to use cookies that are not necessary for the provision of the service [10]. To obtain consent, service providers use requests on their websites that are accepted when the user first visits the website. To process personal data, service providers have to receive the consent of the person concerned. The GDPR sets out the rules for consent-based data processing, some of which relate to the usability of the consent requests [11]:

- design must ensure that the users understand what they are consenting to;
- consent must be given freely and provided in clear and understandable language;
- consent is not considered freely given if the person does not have a free choice or cannot refuse consent;
- the individual must be able to refuse consent as easily as it is to give it [5, 8].

According to the E-Privacy Directive, cookies may be used provided that users are clearly and accurately informed about the purposes for which they are used [11]. If cookies are not necessary for the functionality provided, such as tracking cookies for

market research, it may be necessary to obtain user consent [12]. Users must be informed about what information is provided to the device they use. The user must also be able to opt-out of cookies when other users have access to personal information stored on the device. Information about the purposes and subsequent uses of cookies must be provided together with information about the user’s right to refuse cookies before they are used. If the user decides to refuse cookies, the service provider must still provide the minimum services – for example, the website with restricted content.

To process user data, at least one of the several conditions for the processing of personal data provided for by EU laws must be met (e.g. obtaining the consent of the individual). If service providers seek to obtain a person’s consent, it should meet certain requirements and recommendations. Table 1 summarizes the requirements and recommendations that consent should be sought.

Table 1. Privacy and security requirements for cookie consent requests

Identifier	Condition	Description
R1	Given free will	Consent must be freely given. The GDPR recommends that consent should not be considered freely given if the individual has no other choice or option to opt-out
R2	Concrete	Consent must be expressed by clear, affirmative action
R3	Informed	The user must be provided with accurate information about the purposes of data processing and future uses in clear and understandable language
R4	Unambiguous	Consent must be expressed for each purpose for which user data will be processed
R5	Clear	Consent requests must be separated from other matters
R6	Continuous	If consent is requested electronically, the request should be clear, and concise and not unnecessarily interrupt the use of the service
R7	Easy disagreeing	Opting out of cookies should be as easy as accepting them
R8	The right to disagree	Information about the user’s right to refuse the use of cookies must be provided together with information about the purposes of using cookies

3 Dark Patterns in Cookies Consent Requests

Harry Brignull defined dark patterns as a carefully crafted user interface to trick users into doing things they might not otherwise do [13]. Dark patterns are increasingly found in different digital platforms such as social blogs and online stores [7] as well as on the official websites of major companies such as Facebook, Amazon, LinkedIn, and Uber [14–16]. One of the dark patterns – “Privacy Zuckering” – was named after the Facebook

founder due to the difficulties of managing privacy on Facebook. Other dark patterns include trick questions, sneak into the basket, Roach motel, price comparison prevention, hidden costs, bait and switch, confirmshaming, disguised ads, forced continuity, and friend spam [15]. Gray et al. categorized Brignull's identified dark patterns into five types: nagging, obstruction, sneaking, interface interference, and forced actions [16]. A study by Luguri and Strahilevitz revealed that the more aggressively dark patterns are applied, the more likely users are to succumb to manipulation and make choices that are against their interests [17]. Even seemingly small design decisions, such as highlighting the consent option or moving the decline button to another window, can significantly increase the likelihood of obtaining user consent [5].

Dark patterns are often observed in cookie consent requests. A study on the top 1,000 EU websites revealed that 57.4% of requests use at least one dark pattern in 2019 [18]. The following design solutions containing dark patterns are met in consent requests [8]:

- *Consent walls* are pop-up windows that contain part of the service provider's privacy policy or informative text about the use of cookies on the website. They clearly separate consent request from other matters, thus satisfying the GDPR requirement (R5). However, consent walls block access to the website until users express their consent. This can be seen as not freely given consent (R1 unsatisfied). Thus, the dark pattern of forced action can be assigned. The dark pattern of nagging is also observed as a consent request is displayed immediately after visiting the website.
- *Tracking walls* are a type of consent walls, in which the user can only accept the use of cookies or leave the site. This design uses the dark pattern of forced action which is more aggressively applied than on consent walls. The tracking wall is a sort of barrier that prevents the user from taking the desired action. Therefore, this design is also a case of the obstruction dark pattern.
- *Reduced service* is provided when the users do not accept the privacy settings. If no alternative is provided to access the full content (for example, a paid service option), the user is forced to agree with the use of cookies to reach the full functionality of the website. This situation can be described by the dark pattern of forced action. Also, redirecting a user to a version of a site with restricted content or functionality could be regarded as nagging.
- *Manipulation of configurations* encourages users to accept the use of cookies. Visual manipulations such as different sizes, position, formatting (use of different colors and fonts), an abundance of options, and pre-marked options aim to make one option more noticeable, and more attractive than the other and secretly encourage the user to choose it. Such manipulation is characterized by the dark pattern of interface interference.

Dark patterns exploit the users' limited attention because they often multitask while browsing the Internet. Designers use attention diversion techniques in request configurations, which draw the user's attention to one part of the website to divert it from other parts that would be more useful to the user [19]. A study conducted by Utz and others shows that such design decisions as moving the position of the request from the top to the bottom of the screen or highlighting the consent button influence the choice of users to express consent or disagreement with the use of cookies. The study observed that

the probability of consent to the use of cookies increased from 0.16% to 83.55% when pre-tagged options were used in the request [18].

4 Evaluation of Consent Requests in Lithuanian News Portals

Evaluation of the design of cookie consent requests was conducted on the two most visited news portals in Lithuania: Delfi.lt and Lrytas.lt [20]. In the assessed requests, consent was expressed using the “I agree” button in the main request window (R2 is met) (see Figs. 1 and 2). However, an option to disagree was not presented in the first window of any request (R7 is violated), therefore disagreement requires more clicks than expressing consent. Each request allows the user to change specific goal settings separately (R4 is met) and the consent request was clearly separated from other questions (R5 is met).

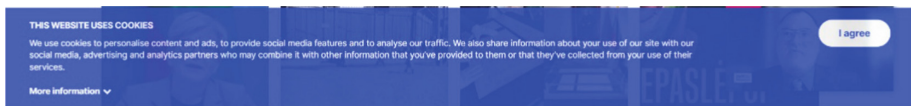


Fig. 1. The first window of cookie consent request on Delfi.lt (<https://www.delfi.lt/en/> – the English version of Delfi.lt)

Delfi.lt consent request is designed in notification bar style and does not interrupt the use of the service (R7 is met). Lrytas.lt uses a consent wall-type request, it is displayed immediately after opening the website which is a case of nagging. However, its sole purpose is to inform the users about the use of cookies and to obtain their consent, so it can be said that it is not an unnecessary interruption (R6 is met).



Fig. 2. The first window of cookie consent request on Lrytas.lt (<https://www.lrytas.lt/english> – the English version of Lrytas.lt). Although the English version is activated, the request is presented in Lithuanian. Below two options are presented: “More options” – on the white button and “I agree” – on the red one.

Although information about the user’s right to disagree was provided only by Lrytas.lt in the first request window, this right was indicated in both examined sites (R8 is met). Lrytas.lt provided an option in the main window that leads to the cookie setting window, so the user can understand that consent is not the only option. In the Delfi.lt

consent request, the cookie settings are hidden under the less visible “More information” dropdown. The user may not even find the hidden settings without carefully examining the request and think that the only option is to accept the setting of cookies (Fig. 3). This raises doubts about whether the consent received is freely given (R1 is violated). The same doubts are raised by Lrytas.lt developers decision to block the site’s content and secretly encourage users’ consent by highlighting the “I agree” button.

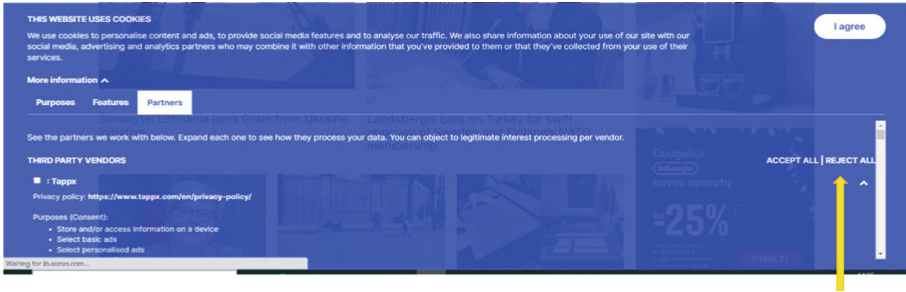


Fig. 3. Expanding “More Information” on Delfi.lt. It is hard to notice that the rejection of all options is available, because of the small text on the right (indicated by a yellow arrow). The user cannot decline individual options.

Lrytas.lt provides the users with detailed information about the data collected and the purposes of processing – for each partner it is indicated what information is collected, the expiration date of the cookie is presented and the goals are clearly stated (R3 is met) (Fig. 4).

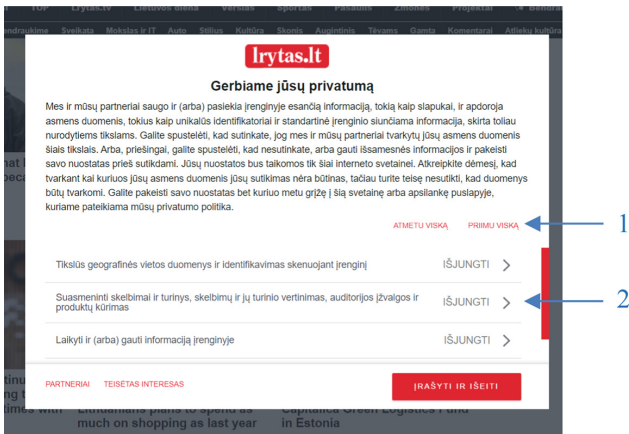


Fig. 4. Presentation of the categories of collected data. Label 1 indicates “Accept all” and “Reject all” options. Label 2 shows an option to decline individual categories. This implicitly means that all options by default are enabled.

Part of Delfi.lt information related to partners is not translated from English, which may not be understandable to all Lithuanian users (R3 is violated).

The four dark pattern categories were found on the assessed sites. Dark patterns of nagging, obstruction, forced action, and interface interference are observed in all sites examined. The obstruction pattern makes it difficult to reject cookies when such an option is not provided in the first request window. The dark pattern of interface interference is observed on Delfi.lt where users are not informed about the user right to disagree with the use of cookies in the first request window. Moreover, Lrytas.lt tries to hide from users that some options are enabled by default. The dark pattern of interface interference manifests itself in the consent option that is presented in all requests more attractively than the option that directs to cookie settings. Delfi.lt used a less noticeable drop-down element for cookie settings. The dark pattern of forced action is observed only at Lrytas.lt in which the content is blocked until the users express their choice.

All in all, both examined consent requests have not complied with at least two privacy and security requirements. Four dark patterns were detected on Lrytas.lt, three – on Delfi.lt.

5 Design Guidelines for Cookie Consent Requests

Revealed violations of privacy and security requirements highlight the need for easily applicable guidelines that would support developers in ensuring the privacy and security of consent requests as well as following the GDPR and avoiding dark patterns. Based on the revealed defects, the guidelines for ensuring privacy and security requirements have been formulated (Table 2).

As an example of applying guidelines, the main page of Delfi.lt cookie consent request is redesigned (Fig. 5) by adding the disagree option (G3). Both options require clicking the button which is a clear affirmative action (G1). The information that the user can refuse to use cookies is shown on the first page (G7).

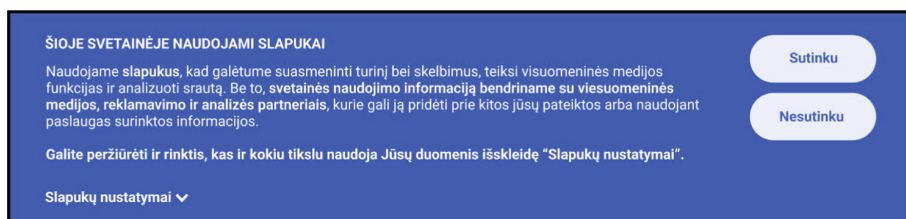


Fig. 5. The redesigned main page of Delfi.lt consent request (in Lithuanian): both options (agree and disagree) are presented as equivalent choices to meet the guidelines G1, G3, and G7.

The example of the application of other guidelines is based on the redesign of the Lrytas.lt site (Fig. 6). Guideline G13 requires facilitating the reading of the lengthy texts. In user interface design this is usually achieved by designing a proper visual hierarchy. Another solution is introducing associative icons. Both solutions are involved in the examined page: icons facilitate quick scanning; the summary of each purpose is

provided in boldface. Further details of each purpose can be obtained by expanding these options. The options to accept or reject all options facilitate the user's decision to refuse more options (G8). Each changeable option indicates its status, whether it is enabled or disabled (G5). More sophisticated design decisions were made to support the quicker perception of texts. A relation between the collected data and their processing purpose (G9, G10) is visualized using graphs and color codes. This solution was popularized by the pribot.org tool for a showing of a privacy policy on Twitter.

Table 2. Usability guidelines for consent requests

Identifier	Description
G1	Consent must be expressed by clear affirmative action, such as clicking a button or checking a checkbox
G2	To ensure that consent is given freely, consent and tracking walls should not be used to obtain consent
G3	Disagreeing with the use of cookies should be as simple as accepting their use. The "Disagree" option has the same importance as the "Agree" option, thus both options must require the same number of mouse clicks
G4	If cookies are used for several purposes, consent should be requested for each purpose separately
G5	If the request uses elements whose state can be changed (e.g., disable/enable, agree/disagree), the setting must have a clear indication of the state (i.e. whether it is enabled/disabled or agreed/disagreed)
G6	The default settings can only be used to activate strictly necessary cookies
G7	The users must be informed about their right to refuse the use of cookies in the first request window
G8	If the consent request requires the user to express an opt-out or objection to more than one legitimate interest, data processing, or partner purpose, an option to facilitate changing the settings of many elements must additionally be provided, such as to reject all listed objectives
G9	The user must be provided with accurate information about the purposes of data processing, data collected, cookies used, and their possible uses in the future, including how cookies can be used and how long they will be valid
G10	If a query contains a lot of information in one place, it should be presented a in visual structure highlighting important points
G11	The user must be informed if the data is shared with third parties
G12	The request must include a link to the website's privacy policy, where the user can find more information about data processing and storage
G13	Getting to know the purposes of data processing should take acceptable time
G14	The design has to ensure that conditions are created for making an informed decision. The elements used in the request must not distract the users' attention or encourage them to make specific decisions
G15	User-relevant information should not be hidden under hard-to-see elements or on additional request pages. If due to the design of the request, it is not possible to present all the information on one page, the request should provide a link where users can find the full text of the information

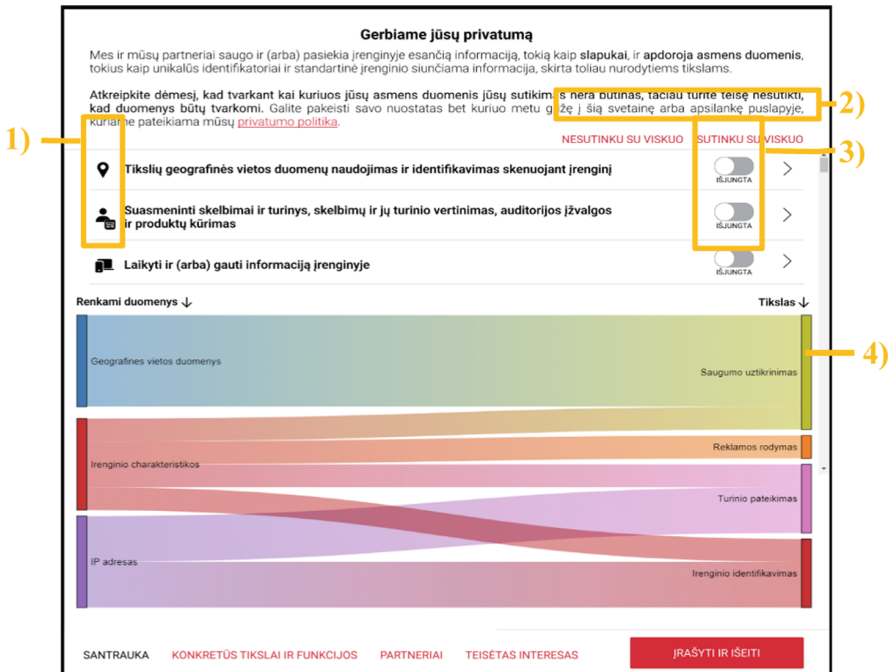


Fig. 6. Redesign of data processing purposes (in Lithuanian): 1) the text of processing purpose is augmented by icons to facilitate quicker content perception (G13); 2) an option to accept or reject all options is provided (G8); 3) all elements whose state can be changed have a clear indication of the state, in the example – disabled (G5); 4) the chart relates the purposes (right edge) with a processed data (left edge), mouseover highlights selected options (G9, G10).

6 Conclusions

This paper aims to establish usability guidelines that would support the usability of consent requests. Recently established regulation for the processing of personal data requires service providers to obtain the user's informed consent to use cookies that are needed for the provision of the service. Implementing this law, each website offers a consent form that contains all the required information. These forms interrupt the current activity and require users to focus on reading a complex legal text. The very fact of the interruption is annoying. The unusable design adds frustration. Therefore, many users decide to minimize the disturbance by choosing the quickest solution that, unsurprisingly, consciously or not, is beneficial for service suppliers. However, unethical behavior will not go unnoticed, thus damaging the users' perception of the provider's brand value. As all sites are required to obtain consent, the sites with more usable solutions will benefit.

The study of GDPR revealed the requirements for ensuring the privacy and security of processing personal user data in web services. These requirements relate usability of consent requests design. Since providers are interested in a specific user choice, the existence of dark patterns in consent requests was checked.

Investigation into dark patterns revealed the presence of four types of dark patterns in the examined consent requests. It was found that the first screen in both examples hides the possibility to reject or choose appropriate cookies on a website. The texts are provided in the form of text walls that hinder their readability. The consent wall prevents the usage of the service until the user will accept with consent request. So, the user's free will is questionable.

Although service providers have an interest to push desirable behavior by accepting all cookies, care for brand value requires considering the users' interests, too. Developed usability guidelines are aimed at facilitating the design of consent requests in future designs. The redesign examples illustrate the usefulness of the developed guidelines. Most of the developed guidelines simply suggest the design decision, such as providing a button that could be clicked or including both options on the first screen. Others are formulated more abstractly. For example, facilitating readability or avoiding distraction. The latter guidelines are subject to further investigation to provide more specific design solutions that would be ready to apply.

References

1. Kristol, D.M.: HTTP Cookies: standards, privacy, and politics. *ACM Trans. Internet Technol.* **1**, 151–198 (2001). <https://doi.org/10.1145/502152.502153>
2. Using HTTP cookies – HTTP—MDN. <https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies>. Accessed 28 Sept 2022
3. Koch, R.: Cookies, the GDPR, and the ePrivacy Directive. <https://gdpr.eu/cookies/>. Accessed 28 Sept 2022
4. Hamed, A., Kaffel-Ben Ayed, H., Kaafar, M.A., Kharraz, A.: Evaluation of third party tracking on the web. In: 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013), pp. 471–477 (2013). <https://doi.org/10.1109/ICITST.2013.6750244>
5. Nouwens, M., Liccardi, I., Veale, M., Karger, D., Kagal, L.: Dark patterns after the GDPR: scraping consent pop-ups and demonstrating their influence. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3313831.3376321>
6. Brignull, H.: Bringing Dark Patterns to Light. <https://harrybr.medium.com/bringing-dark-patterns-to-light-d86f24224ebf>. Accessed 28 Sept 2022
7. Mathur, A., et al.: Dark patterns at scale: findings from a crawl of 11K shopping websites. *Proc. ACM Hum.-Comput. Interact.* **3**, 81:1–81:32 (2019). <https://doi.org/10.1145/3359183>
8. Gray, C.M., Santos, C., Bielova, N., Toth, M., Clifford, D.: Dark patterns and the legal requirements of consent banners: an interaction criticism perspective. *arXiv:2009.10194* [cs] (2021). <https://doi.org/10.1145/3411764.3445779>
9. Data protection under GDPR. https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm. Accessed 28 Sept 2022
10. Santos, C., Bielova, N., Matte, C.: Are cookie banners indeed compliant with the law? Deciphering EU legal requirements on consent and technical means to verify compliance of cookie banners. <http://arxiv.org/abs/1912.07144> (2020). <https://doi.org/10.48550/arXiv.1912.07144>
11. General Data Protection Regulation (2016). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=LT>
12. Buthlezi, M.P., Looock, M.: User online privacy and identity management behaviors: a comparative study. In: 2014 Annual Global Online Conference on Information and Computer Technology, pp. 53–57 (2014). <https://doi.org/10.1109/GOCICT.2014.14>

13. Brignull, H.: Dark Patterns: inside the interfaces designed to trick you. <https://www.theverge.com/2013/8/29/4640308/dark-patterns-inside-the-interfaces-designed-to-trick-you>. Accessed 29 Sept 2022
14. Schlosser, D.: LinkedIn Dark Patterns. <https://medium.com/@danrschlosser/linkedin-dark-patterns-3ae726fe1462>. Accessed 29 Sept 2022
15. Brignull, H.: Dark Patterns. <https://www.darkpatterns.org/index.html>. Accessed 11 Feb 2021
16. Gray, C.M., Kou, Y., Battles, B., Hoggatt, J., Toombs, A.L.: The dark (patterns) side of UX design. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3173574.3174108>
17. Luguri, J., Strahilevitz, L.: Shining a Light on Dark Patterns. Social Science Research Network, Rochester, NY (2019). <https://doi.org/10.2139/ssrn.3431205>
18. Utz, C., Degeling, M., Fahl, S., Schaub, F., Holz, T.: (Un)informed consent: studying GDPR consent notices in the field. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 973–990. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3319535.3354212>
19. Chatellier, R., Delcroix, G., Hary, E., Girard-Chanudet, C.: Shaping choices in the digital world. From dark patterns to data protection: the influence of UX/UI design on user empowerment. Technical report, CNIL (2019)
20. gemiusAudience: September overview of the most popular Lithuanian websites (in Lithuanian). <http://www.gemius.lt/interneto-ziniasklaidos-naujienos/gemiusaudience-rugsejo-mensio-apzvalga-6411.html>. Accessed 12 Oct 2022








Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Transdisciplinary Approach to Virtual Narratives - Towards Reliable Measurement Methods

Grzegorz Pochwatko¹ , Daniel Cnotkowski² , Paweł Kobylński² ,
Paulina Borkiewicz³ , Michał Pabiś-Orzeszyna⁴ , Mariusz Wierzbowski² ,
and Laura Osęka¹ 

¹ Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland
gp@psych.pan.pl

² Laboratory of Interactive Technologies, National Information Processing Institute,
Warsaw, Poland
daniel.cnotkowski@opi.org.pl

³ Visual Narratives Laboratory vnLab, Lodz Film School, Łódź, Poland

⁴ Film and Audiovisual Media Department, University of Lodz, Łódź, Poland

Abstract. We have recently observed intense growth in the film industry's interest in VR creations. Cinematic VR artists encounter challenges that result from discrepancies between established techniques of storytelling, stylistic conventions, and organizational culture indicative of traditional modes of film practice and the requirements of the new medium and new audience. We propose a transdisciplinary approach to cinematic VR research. Thanks to the cooperation of art & science - a collaboration between psychologists, information technology specialists, film scholars, and filmmakers will contribute to the emergence of a new VR narrative paradigm. We use a number of quantitative and qualitative methods to study the perception of cinematic VR works, an illusion of spatial presence and copresence, attention, emotions, and arousal of its users, narrative understanding, and character engagement. We measure participants' reactions in many independent ways: in addition to subjective assessments and declarative methods, we use more objective data: eye tracking, multi-point position skeleton tracking, and psychophysiological responses. We show the effectiveness of the adopted approach by studying three artistic cinematic VR works: narrative and non-narrative, live-action, and animated. We compare the user experience and present the possibilities of interpretation and feedback benefits for art.

Keywords: cinematic VR · research app · usability · presence

1 Introduction: Motivation and Related Work

Any VR experience allows the subjects to choose their paths of visual attention [4, 5] and spatial behavior [11, 12, 14, 15], which leads to different emotional reactions [7, 9, 16]. VR designers that employ any system of attention cues and plan to manipulate emotional arousal in immersive storytelling, may be interested in measuring the effectiveness of their actions. Tools are emerging, but so far, they are limited (e.g., based on the eye, and head movements only [13]). Our transdisciplinary multi-dimensional

approach addresses the problem of reliable measurements of the audience response to cinematic VR experiences.

The number of virtual training and research environments in which participants are free to explore limited or (nearly) unlimited space increases [1, 8, 18, 19]. Analysis of participants' movement, physiological responses and declarations allows inferences on the impact of the experience, approaching or avoiding objects, being relaxed or stressed, the response to virtual agents (passive or active characters).

Cinematic VR experiences (CVR) are limited compared to computer-generated 3D environments because they do not allow users to walk freely or interact with objects and characters. CVR is typically experienced in a linear story, but the use of spherical projection and central placement of the participant can make the experience different for each viewer, depending on the creator's intention. CVR can be more immersive than computer-generated environments due to its higher level of photorealism, which can lead to **greater spatial presence and co-presence** illusions. Some CVR experiences can elicit intense arousal in viewers due to the inclusion of primary distress signals, such as sudden loud noises or changes in lighting [3, 10]. Increased arousal may not be related to the content of the narrative. However, high tonic activity (TA) can make it easier to elicit a response to important parts of the story. It can also make it difficult if the creator puts too many primal stimuli and TA reaches the ceiling level. Situations in which the story evokes arousal are much more interesting. These can be e.g., the appearance of characters, interaction between them, and music. Reliable measurement enables us to register, e.g., increased heart rate or skin conductance level as a response to objects and characters.

Successful storytelling in CVR depends on participants being placed in the right direction at critical moments, giving them the physical ability to follow what is visible and audible and respond to it in a predicted way.

Art creators and filmmakers are looking for a **new VR narrative paradigm**, and want to do it systematically. VR technology that made it possible to shoot stereoscopic 360° videos and create interactive experiences also allows us to analyze participants' reactions to the virtual content. The use of this information depends on close, transdisciplinary collaboration between art & science.

2 Overview of the Cinematic VR Research Method

2.1 The Research Application

The research application was created using the Unity engine, which is the core of the method and enables the collection of behavioral data and communication with external devices. However, the transdisciplinary approach to CVR research also involves the use of a variety of other methods to provide additional insights into viewer reactions to CVR and ensure the reliability of the research.

2.2 Screening

Before inviting participants to the lab, we ensure that there are no health issues or medications that could interfere with psychophysiological measurements (PF), and that participants have not consumed any psychoactive substances, alcohol, or coffee. We also verify that participants' vision is corrected if necessary.

2.3 Measures Used Before VR Experience

Before watching the experience, participants complete questionnaires to assess psychological variables, individual differences, attitudes, and traits that may impact their perception of the experience or be affected by it. The specific methods used in these questionnaires may be tailored to the content of the experience or used consistently across all experiences, such as the mood scale and the Self-Assessment Manikin (SAM). [2, 6]. We measure temperamental traits, which are relatively stable human characteristics, to better understand participants' motor behavior and emotional reactions. We also assess participants' mood and emotional state to determine the positive or negative impact of the experience on them.

2.4 Baseline Measures

Research with the use of PF and behavioral measures require a baseline measurement in neutral conditions.

Calibration: 1) PF measures (e.g., EDA, ECG, EEG¹ if applicable) - require external equipment, and trained personnel; 2) eye-tracker - standard procedure from the manufacturer; 3) body motion and position (head, arms, legs, torso and waist). Participants perform a series of standard movements (head and whole body rotations) - providing data for the correction of motor artifacts in the PF signal.

Baseline Sequence: 1) 3 min in a quiet, minimalist environment, eyes open²; 2) 360 neutral and affective videos - reference to PF response. Videos by Li et al. [7]: Abandoned City, valence 3.33/9, arousal 3.33/9, duration 50 s; Spangler Lawn, v. 5.9/9, a. 3.27/9, d. 58 s; Seagulls, v. 6/9, a. 1.6/9, d. 60 s; Walk Among Giants, v. 5.79/9, a. 2/9, d. 60 s; Tahiti surf, v. 7.1/9, a. 4.8/9, d. 60 s.

2.5 Cinematic VR Experience Test

After completing the baseline, the target CVR experience begins. Participants start watching from a fixed position and are free to move around (3DoF) and explore the experience (6DoF). Both body and eye movements are recorded. Thanks to markers, it is possible to time synchronize the PF signal recorded with external equipment.

2.6 Measures Applied After VR Experience

Participants complete a series of questionnaires. Presence and co-presence, repeated measure of mood and SAM, user experience and declarative evaluation of the baseline and target CVR is measured. Questionnaires specific for the tested VR experiences are introduced. An in-depth interview regarding the content and impressions related to the experience is conducted.

¹ EDA - electrodermal activity, ECG - electrocardiogram, EEG - electroencephalogram.

² provides comparative data for PF analyzes of reactions to stimuli; (for EEG recording add a 3-minute measurement with eyes closed).

2.7 Digital Markers Calibration

Synchronization with external devices (e.g. BIOPAC, NEUROSCAN) requires the ability to send digital triggers via LPT or USB TTL adapter. Unity allows us to send time/event related and conditional markers with millisecond precision, which is important when analyzing subtle changes in PF signal. We use 8-bit markers (1–255). Recording frequency of 1000 Hz or 2000 Hz enables precise reading of the marker. A test is performed before each series of studies. A series of markers in 1–1000 ms intervals is sent and registered on an external device with 20000 Hz resolution. The marker times are compared and corrected if necessary.

3 Current Research - Method

3.1 Participants

247 people participated in the study (157 females, 88 males, 2 persons did not indicate gender or chose a different answer), mean age 31.65 SD = 9.32. Half of the participants were involved in creative activities, while the other half were recipients of art (interested in, e.g., cinema or art exhibitions - screened to rule out the distracting influence of low motivation) (Fig. 1).



Fig. 1. One of the participants of the cinematic VR experience study.

The study involved volunteers recruited through online ads. Screening was performed to exclude people who had contraindications to participate in VR experiments (e.g. problems with stereoscopic vision) or used substances that could influence PF measurements (e.g. anti-arrhythmic drugs).

3.2 Materials and Apparatus

Measures. Due to length restrictions we limit the scope of this paper and analyze only a part of the data. The full list of measurements includes: 1. questionnaires related to

the CVR productions and individual differences that may affect reception of them, 2. mood and well-being repeated measure, 3. PF data (eye movements and fixations, pulse, EDA), 4. post-VR questionnaire that included:

- User experience experimental setting and equipment scale (from 1 definitely not to 5 definitely yes)
 - instructions: *The instruction given by the experimenter was clear and understandable*
 - lab.comfort, .hot, .tooSmall, e.g., *The room where the simulation took place made me feel comfortable,*
 - appar.comfort *The apparatus used during the study was comfortable,*
 - HMD.heavy *The HMD was too heavy,*
 - HMD.hot *It was too hot with the HMD on,*
 - HMD.unhyg *The HMD looked unhygienic,*
 - Bothering.wires *Wires connected to HMD bothered me,*
 - Bothering.noVisual (...) *after putting on HMD, I couldn't see the experimenter,*
 - Bothering.observed *It bothered me that other people were watching me,*
 - afraid.fall *After putting on the HMD, I was afraid that I might bump into something or fall over,*
 - afraid.damage *The equipment used for the training was too delicate, I was afraid that I might damage something.*
- User state - Self-Assessment Manikin (SAM) [2,6]
 - valence (positive - negative),
 - arousal (high - low),
 - dominance/control (low - high);
- Short Presence Scale (SPS) created by the authors, based on the MEC-SPQ [17] (e.g., *I felt as if I were taking part in events, not just watching them*): Spatial Presence - Self Location (SPSL), four items; Spatial Presence - Possible Actions (SPPA), four items; Suspension of Disbelief (SoD), two items; Attention Engagement (AE), two items.

Hardware. 64bit PC Intel Xeon I7, RAM: 32 GB, GTX 1080TI (Win10), a Neuroscan device, 6 HTC ViveTrackers 2.0, Vive Pro Eye HMD, accessories: elastic wrist, ankle, belt, chest straps with tracker mounts and replacement facial interface from VRCover with single use hygienic covers.

Experimental Application. Experiment application was prepared in Unity Engine. It consisted of a single scene environment, with a simple cubical room as a neutral starting point for the stimulus. Important features were: playback of hi-res 360 videos with ambisonic audio, sending event markers using parallel port and gathering user positional and eye tracking data with high resolution. Video playback was solved by usage of Unity Store asset designed for uninterrupted 4K video playback with stereoscopic 360 video support. To solve ambisonic audio playback we have used a Facebook 360 spatial decoder, synchronising its playback internally with video. This approach combined with splitting audio and video streams into two separate files, proved itself sufficient to meet aforementioned assumptions. For parallel port communication an external

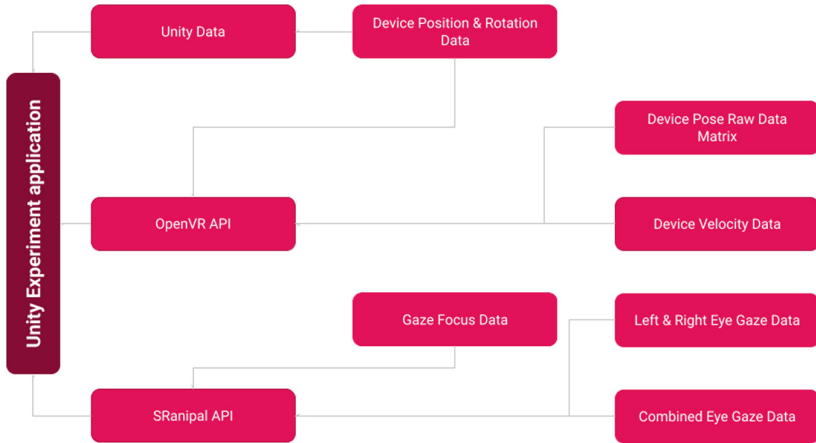


Fig. 2. Diagram of sources of gathered data.

library was used. To achieve best precision of event markers delivery we have developed solution allowing us to prepare event markers in advance and separate task of timing and sending those to parallel port output into a separate thread, with help of C# built in Stopwatch class millisecond precision of the markers was achieved. Highest resolution of data gathering possible was limited by Unity to 90 Hz as the core of the application is constrained by the application framerate (consistent with HMD's framerate), which proved to be sufficient. Data we gathered was obtained as follows: from within scene - positional and rotational data of each connected VR device and accessory, using OpenVR API we also gathered positional data, while additionally collecting velocity and pose matrix data of those devices, by using SRanipal API we gathered eye tracking data consisting of combined eye, left and right eye gaze data and current eye focus point. Outgoing event markers were recorded allowing us to sync with PF data (Fig. 2).

3.3 Procedure

Six HTC Vive Trackers were attached at ankles, wrists, belt and chest areas with special anti-slip straps. Participant was seated in a chair where electrodes and sensors were applied. Then, the HMD was put on, electrode readouts were checked, and per-user eye tracker calibration was performed. After setup all lights in the experiment room were turned off, baseline and target registration followed under supervision.

Table 1. Global scores and Personal characteristics

Variables	range	Global		Gender M		Gender F		Art recipient		Art creator	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SAM:sad/happy	1-5	3.83	0.91	3.80	0.94	3.84	0.90	3.79	0.87	3.88	0.96
SAM:calm/aroused	1-5	1.98	1.04	1.93	1.01	2.01	1.06	2.03	1.08	1.92	1.00
SAM:controlled/in control	1-5	3.66	1.25	3.71	1.11	3.64	1.32	3.58	1.35	3.75	1.14
Instructions	1-5	4.93	0.32	4.92	0.35	4.93	0.30	4.92	0.33	4.93	0.31
Lab.comfort	1-5	4.76	0.56	4.72	0.56	4.78	0.56	4.78	0.60	4.74	0.53
Lab.hot	1-5	1.46	0.99	1.53	1.03	1.43	0.98	1.47	1.01	1.46	0.99
Lab.tooSmall	1-5	1.25	0.74	1.30	0.75	1.22	0.74	1.25	0.73	1.25	0.75
Appar.comfort	1-5	3.96	1.14	4.03	1.10	3.91	1.16	3.96	1.20	3.94	1.08
HMD.heavy	1-5	2.76	1.36	***2.34	1.32	***3.00	1.33	2.71	1.42	2.82	1.31
HMD.hot	1-5	1.81	1.08	1.70	1.07	1.88	1.08	1.80	1.09	1.83	1.07
HMD.unhyg	1-5	1.20	0.56	1.21	0.59	1.20	0.55	1.19	0.59	1.22	0.54
Bothering.wires	1-5	1.75	1.07	1.84	1.12	1.69	1.02	1.67	0.97	1.84	1.15
Bothering.noVisual	1-5	1.33	0.81	1.30	0.75	1.34	0.85	1.40	0.91	1.25	0.71
Bothering.observed	1-5	1.27	0.67	1.20	0.61	1.32	0.70	+1.35	0.76	+1.2	0.55
Afraid.fall	1-5	1.58	1.01	1.56	0.97	1.58	1.02	1.62	1.12	1.54	0.90
Afraid.damage	1-5	1.50	0.90	1.57	0.88	1.46	0.91	1.52	0.96	1.49	0.84
SPSL	1-7	4.21	1.40	4.20	1.33	4.22	1.45	***4.43	1.42	***3.99	1.37
SPPA	1-7	3.97	1.58	**3.62	1.54	**4.16	1.59	***4.24	1.61	***3.69	1.53
SPAI	1-7	5.98	1.17	6.02	1.05	5.94	1.24	6.00	1.12	5.95	1.23
SPSD	1-7	3.70	1.73	3.58	1.76	3.75	1.72	***4.07	1.72	***3.33	1.68

Significance level: *.05, **.01, ***.001, +marginally significant.

4 Results

4.1 User Experience - Experimental Setting and Equipment Evaluation

The overall user experience rating was very good. The pattern of the results of individual aspects was as predicted - see Table 1 “Global”: 1) The **instructions** provided by the experimenter and in the app were clear and understandable; 2) **Lab comfort** was rated very high, which is consistent with temperature and lab size ratings; 3) **Apparatus** was comfortable, although rated lower than lab in general. HMD - appropriate weight and temperature, hygienic (important in times of epidemic threat); 4) Neither **equipment** nor **experimental situation** bothered the participants. Participants were not bothered by the fact that they did not have visual contact with the experimenter after wearing HMD, or that someone could observe them during the VR experience; 5) Participants were not afraid to fall or bump into something. They also did not see the equipment as fragile and easy to damage. 6) **Differences**; Women considered HMD significantly heavier than men. Art recipients were more concerned about being watched than the art creators. Surprisingly we noticed that some aspects of comfort differed between narrative vs. non-narrative and live-action vs. animated conditions.

Table 2. Contents characteristics

Variables	range	Narrative		Non-narrative		Live-action		Animated	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
SAM:sad/happy	1-5	3.88	0.91	3.74	0.91	***3.62	0.94	***4.24	0.72
SAM:calm/aroused	1-5	1.95	1.03	2.05	1.07	***2.15	1.10	***1.66	0.84
SAM:controlled/in control	1-5	***3.86	1.17	***3.25	1.31	***3.45	1.32	***4.06	0.98
Instructions	1-5	4.92	0.36	4.95	0.22	4.94	0.31	4.91	0.33
Lab.comfort	1-5	4.80	0.55	4.67	0.57	4.76	0.55	4.76	0.59
Lab.hot	1-5	**1.33	0.87	**1.74	1.17	1.46	1.01	1.47	0.96
Lab.tooSmall	1-5	1.30	0.81	1.14	0.55	**1.12	0.52	**1.48	1.00
Appar.comfort	1-5	3.99	1.12	3.88	1.20	3.86	1.15	4.14	1.10
HMD.heavy	1-5	2.78	1.37	2.73	1.36	***3.01	1.36	***2.31	1.25
HMD.hot	1-5	***1.64	0.99	***2.16	1.17	1.89	1.09	1.65	1.03
HMD.unhyg	1-5	1.19	0.57	1.24	0.56	1.17	0.48	1.26	0.69
Bothering.wires	1-5	1.69	0.98	1.89	1.22	***1.91	1.15	***1.46	0.83
Bothering.noVisual	1-5	1.30	0.80	1.39	0.85	1.32	0.79	1.33	0.86
Bothering.observed	1-5	1.28	0.65	1.26	0.71	1.26	0.68	1.29	0.65
Afraid.fall	1-5	*1.67	1.03	*1.39	0.95	***1.42	0.94	***1.88	1.07
Afraid.damage	1-5	1.45	0.83	1.61	1.03	**1.60	0.98	**1.31	0.67
SPSL	1-7	4.12	1.45	4.41	1.27	**4.03	1.52	**4.57	1.06
SPPA	1-7	3.99	1.65	3.93	1.45	**3.76	1.59	**4.37	1.51
SPAI	1-7	5.93	1.16	6.08	1.20	***5.78	1.30	***6.34	0.76
SPSD	1-7	3.73	1.77	3.64	1.66	**3.48	1.63	**4.13	1.83

Significance level: *.05, **.01, ***.001, +marginally significant.

4.2 User State - Emotion, Arousal and Control

A key ethical issue was to ensure that participants were in a good emotional state upon completion of the study. This goal was achieved: participants felt rather happy, calm and in control. As expected, there were no differences in SAM scale (Table 1). Differences appeared between types of content (Table 2). As expected, participants felt more in control in narrative compared to non-narrative condition. They also felt more happy, aroused and in control in the animated compared to live-action condition.

4.3 Presence

Global level of presence was moderate (SPSL and SPPA, see Table 1 “Global”). SPPA were rated lower than SPSL. There are **individual differences**; men rated SPPA below the midpoint of the scale, whereas women rated it higher, above the midpoint of the scale. More differences were found between art creators and recipients: Art creators rated SPSL&PA significantly lower than recipients. They scored also lower in SPSP. Ratings in all presence sub-scales were lower in **live-action** compared to **animated** condition (Table 2).

5 Discussion

We presented scientific tools for researching virtual narratives - 360 3D CVR. Our methods are comfortable and safe for participants. They rate high conditions in the lab and the equipment used (in terms of e.g. comfort and hygiene). There are also no significant distractions from equipment or procedure (comp. SAM and presence scales). After the study, participants feel happy, calm and in control. The level of self-location is slightly above average, indicating illusion of presence. The level of possible actions is lower, because CVR is not interactive. This may lower the overall presence level, as CVR photorealism may trigger an expectation of interaction. Another reason could be experimental situation: participation in the study may trigger participants' urge to carefully analyze the presented materials. Attention involvement is very high and suspension of disbelief low. This may cause the participants to notice all the flaws of the cinematic experience while being distracted by the internal and external stimuli. This is in line with the differences observed between art creators and recipients. Creators (being more critical) have lower scores on the presence subscales except for attention involvement. They analyse the experience more, and therefore they have lower suspension of disbelief and presence. Differences between animated and live-action experiences may be due to the contents of the experiences. Live-action VR is very realistic, whereas animated VR looks artificial - expectations differ.

The unexpected differences between narrative vs non-narrative can be explained by artifacts that possibly appeared in between-subjects design. The lab and HMD temperature ratings differed, which coincided with the execution of one of the conditions (experiences were tested right after they were produced), they could cause systematic differences in ratings. It needs further investigation, as there is no theoretical reason for them to differ this way (especially as there were no differences in other factors, e.g. gender).

There are differences in self assessment of emotional state between contents conditions, but not in art-related and gender groups. This further supports the explanation coming from the differences in the tested narratives which match the more global classification of contents groups. A further investigation is needed with a more comparable material - for example an experience with the same content but a different form: live-action vs animated. It is worth noting that despite differences, the participants in all groups are satisfied, calm and in control after completing the study.

6 Conclusions

In terms of user experience and predictability of results, we have created and presented an effective method of CVR research. It can provide an insight into the perception of CVR experiences that can be used by art creators to better understand their audience and improve their means of expression. The obtained data is also attractive for various fields of science. This gives hope for closer cooperation between art&science and the improvement of both CVR productions and tools for studying them.

Acknowledgements. The publication is based on research carried out as part of the project “New Forms and Technologies of Narration”. Project financed under the program of the Minister of Education and Science under the name “Regional Excellence Initiative” in 2019–2023, project number 023/RID/2018/19, financing amount: PLN 11865100.

The methods used in these studies were partly developed as part of the National Science Center OPUS no. 2012/05/B/HS6/03630.

References

1. Barteit, S., Lanfermann, L., Bärnighausen, T., Neuhaus, F., Beiersmann, C., et al.: Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. *JMIR Ser. Games* **9**(3), e29080 (2021)
2. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **25**(1), 49–59 (1994)
3. Graja, S., Lopes, P., Chanel, G.: Impact of visual and sound orchestration on physiological arousal and tension in a horror game. *IEEE Trans. Games* **13**(3), 287–299 (2020)
4. Kobylinski, P., Pochwatko, G.: Visual attention convergence index for virtual reality experiences. In: Ahram, T., Taiar, R., Colson, S., Choplin, A. (eds.) *IHIET 2019. AISC*, vol. 1018, pp. 310–316. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-25629-6_48
5. Kobylinski, P., Pochwatko, G., Biele, C.: VR experience from data science point of view: how to measure inter-subject dependence in visual attention and spatial behavior. In: Karwowski, W., Ahram, T. (eds.) *IHSI 2019. AISC*, vol. 903, pp. 393–399. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11051-2_60
6. Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O.: Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* **30**(3), 261–273 (1993)
7. Li, B.J., Bailenson, J.N., Pines, A., Greenleaf, W.J., Williams, L.M.: A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Front. Psychol.* **8**, 2116 (2017)
8. Mallam, S.C., Nazir, S., et al.: Effectiveness of VR head mounted displays in professional training: a systematic review. *Technol. Knowl. Learn.* **26**(4), 999–1041 (2021)
9. Ménard, M., Richard, P., Hamdi, H., Daucé, B., Yamaguchi, T.: Emotion recognition based on heart rate and skin conductance. In: *PhyCS*, pp. 26–32 (2015)
10. Öhman, A., Soares, J.J.: On the automatic nature of phobic fear: conditioned electrodermal responses to masked fear-relevant stimuli. *J. Abnorm. Psychol.* **102**(1), 121 (1993)
11. Rothe, S., Althammer, F., Khamis, M.: GazeRecall: using gaze direction to increase recall of details in cinematic virtual reality. In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 115–119 (2018)
12. Rothe, S., Buschek, D., Hußmann, H.: Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technol. Interact.* **3**(1), 19 (2019)
13. Rothe, S., Höllerer, T., Hußmann, H.: CVR-analyzer: a tool for analyzing cinematic virtual reality viewing patterns. In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 127–137 (2018)
14. Rothe, S., Hußmann, H.: Guiding the viewer in cinematic virtual reality by diegetic cues. In: De Paolis, L.T., Bourdot, P. (eds.) *AVR 2018. LNCS*, vol. 10850, pp. 101–117. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95270-3_7
15. Rothe, S., Hußmann, H., Allary, M.: Diegetic cues for guiding the viewer in cinematic virtual reality. In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2 (2017)
16. Slater, M., Khanna, P., Mortensen, J., Yu, I.: Visual realism enhances realistic response in an immersive virtual environment. *IEEE Comput. Graphics Appl.* **29**(3), 76–84 (2009)

17. Vorderer, P., et al.: MEC spatial presence questionnaire. Retrieved Sept. 18, 2015 (2004)
18. Xie, B., et al.: A review on virtual reality skill training applications. *Front. Virtual Reality* **2**, 645153 (2021)
19. Zhu, Y., Li, N.: Virtual and augmented reality technologies for emergency management in the built environments: a state-of-the-art review. *J. Saf. Sci. Resilience* **2**(1), 1–10 (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Towards Gestural Interaction with 3D Industrial Measurement Data Using HMD AR

Natalia Walczak¹(✉), Franciszek Sobiech¹, Aleksandra Buczek¹, Mathias Jeanty²,
Kamil Kupiński¹, Zbigniew Chaniecki¹, Andrzej Romanowski¹,
and Krzysztof Grudzień¹

¹ Lodz University of Technology, 116 Zeromskiego Street, 90-924 Lodz, Poland
natalia.walczak@dokt.p.lodz.pl

² Arts et Métiers Institute of Technology, 151, Boulevard de l'Hôpital, 75013 Paris, France

Abstract. Despite the spread of augmented reality (AR) systems and its applications onto a number of various areas, the adoption of AR in industrial context is relatively limited. We decided to conduct an exploratory user study to define the eventual singularities that might be associated with the barriers for HMD AR technology adoption in the industrial settings, as recent works presented potential benefits of its applications with regard to specific 3D measurement data interpretation. The task-based study was designed to engage users with interaction of volumetric data of static and time series nature. We compared actions of users performed in lab vs. in situ conditions simulating real, process tomography measurement data visualisations for granular bulk solids flow in large containers. Study results revealed concrete directions for further work that might eventually enable wider adoption of HMD AR systems in the industrial context in terms of specific gestural interaction and visualisation techniques development.

Keywords: augmented reality · gestural interaction · measurement data visualization · industrial data analytics

1 Introduction

Technological progress significantly influences the development of systems supporting access to information and data visualization. Increasing the access paths to the sense of sight, and changing perception of the real environment, by providing additional information generated by computer systems against the background of the natural world, is crucial for applying augmented reality (AR) in different domains of life [1–3]. The study of the ways of communicating and interacting with augmented reality elements is an essential subject of research undertaken on many levels and fields of application of AR. The use of this technology is changing the way we work. AR fulfills the function of supporting activities performed physically by providing the user with a visual perception beyond the real world [4, 5]. AR allows you to enrich the real world with additional content, such as video or image, which allows the user to perform activities simultaneously in the real and digital world. Special attention is given to the use of AR in industrial applications [6].

There is a potential to further enrich the experience and usability of the AR Head-mounted devices (HMD AR) in the industrial context with gestural interaction and input. This would be especially helpful for monitoring and maintaining the reach of ongoing processes in measurement data, where HMD AR brings the extra value of freeing hands from the need to constantly hold the devices, a requirement for smaller AR-enabled equipment such as tablets. Gesture interaction with hands-free devices, like touch screen surfaces, is widely explored in literature [7]. However, as the understanding how the context of use influences gesture interaction with 3D data is still challenging, we designed an experiment conducted in both lab conditions as well as in situ with real industrial flow rig settings. This exploratory study investigates how users would interact with the measurement data of a 3D nature with the aid of HoloLens HMD in different settings.

2 Related Work

The interaction issue with AR systems has been well investigated. From this point of view, we can analyse users in terms of interaction with objects/data visible in augmented reality. Most researchers focus only on visualizing objects/data and developing better methods for gesture recognition. However, interaction with the augmented reality world also includes interaction with AR objects or datasets and directly influencing their form to analyse information hidden in objects or datasets.

2.1 Gestural Interaction and Data Visualization in AR Systems

The comparison of interaction in AR systems, between hand gesture-based interaction and multi-touch interaction, in terms of visual contexts, shows the advantage of hand gestures. [8]. Hand gesture interaction is faster than multi-touch interaction in regard to task completion time. There have been studies conducted to determine which gestures are the most intuitive for users [9]. Performing movements such as scaling, moving, deleting and approving are often used during user studies. However, the common approach is to show directly what the user should achieve and then they must make a move by which they want to achieve a specific goal. Additionally, gestural interaction is investigated on a general, universal object without special purposes; what causes tasks to be dedicated to object manipulation without a wider, determined aim. Such gesture interaction can cause a lack of understanding of the full interaction of users with the problem posed to solve and limit natural user interactions with AR data. Another study worth noting is the research on stock exchange data visualization and its use in AR [10]. The 3D representation of financial data with hand gesture interaction was only evaluated in the possibility of data analysis regarding limited time and fulfilling tasks.

An interesting gesture study involved the manipulation of different scale objects, rotating a house, and rearranging its rooms [11]. Authors explored how the scale of AR affects the gestures people expect to use to interact with 3D holograms. It was shown that one or two hands gestures were applied depending on the manipulated object size. In the case of large objects, the participants used both hands and, in the case of small objects, they did it with two fingers. The tasks were not complex and consisted of a sequence of separate gestures. The objects and work with objects were not analysed, only gestures.

2.2 Augmented Reality in an Industrial Setting

The direct application of the AR system in industry is widely researched, yet seldom implemented. The possibility of using popular modern AR systems based on mobile devices such as smartphones/tablets and smart glasses (Apple ARKit, Google ARCore, and Microsoft HoloLens) in an industrial context was investigated in terms of localisation quality in a large industry area [12]. The impact of using AR systems during device assembly instead of using a paper manual is also widely examined [13–15]. The most important limitation to be noted is the effort involved in creating a manual in AR compared to a paper one and considering the possibility of a serious mistake. Augmented Reality has also been tested in monitoring industrial flow processes. It has found application in drug diagnosis and simulations [16] or in-situ analysis and monitoring of measurement data analysis and monitoring [17]. As shown [18], most AR applications in the industry involve assembly processes by providing instructions to users on how to perform scheduled activities. These include remote assistance, improved user safety in industry space, or industrial process inspection & monitoring on site making. Some of the technologies involved relate to different sizes of displays (primarily tablets), projected AR views, and HMD use.

3 Experimental Study Description

The main goal of the study is to reveal what kind of gestures participants will use when conducting 3D data analysis in augmented reality and if there will be any differences depending on the context by recruited ($n = 20$) participants. The prototype AR app was based on 3D data model visualisation supporting baseline performance and enabling basic manipulations helpful for fundamental tasks performed with the data in normal conditions [19]. The chosen datasets were the electrical capacitance tomography (ECT) type [20, 21] for the gravitational flow in the silo-discharging process. Figure 1 presents both in-lab, in situ experimental space as well as types of projected AR visuals that can be treated as two types of interface alignments [22]. The observations and interviews were conducted during 20 experiment sessions. 15 sessions were conducted within lab conditions -- an empty classroom space. The remaining 5 took place in situ, at the semi-industrial tomography flow measurement lab. Each session started with a brief introduction to the tomography system and image interpretation in the context of the process.

The participants were required to perform 4 tasks (as described in Table 1.), with no time constraints. They were encouraged to speak out loud about what they wanted to do and how, which allowed researchers to gather more data by making notes on their comments. Afterwards, semi-structured interviews were recorded, and the observations were archived.

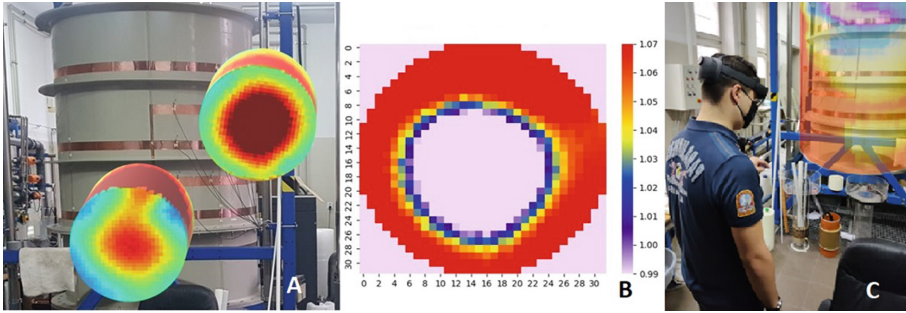


Fig. 1. Visualisations of application used in laboratory (A & C) experimental conditions. A: in situ silo flow 3D AR projection; B: Horizontal slice of time series flow topogram; C: in lab tomography visualisations projected as free AR holograms. (Source: Personal collection)

Table 1. Short descriptions of the four tasks prepared for each participant to complete.

Task name	Description
T0	Involved verifying the user's understanding of what is visible on the 3D data model. The participant was to determine the state of the process and show understanding of visuals' colour spatial distribution in terms of low or high value of the flowing material concentration
T1	The participants were given an opportunity to propose their own gesture for the interaction with the 3D model (moving and rotating the cylinder) and then to use the implemented gesture
T2	The 3D image data had to be divided into separate portions, which allowed participants to look inside the 3D image of the silo space
T3	The users were to use all the gestures they had learned earlier and add new ones to show the researcher how they would present and explain the silo flow process based on the 3D data

4 Results Overview

The results of the 4 conducted tasks lead us to indicate 4 main aspects of interactions: (i) location and position of user relative to the projected object, (ii) projected object displacements, (iii) object rotation and (iv) slicing & extracting sub-elements to get deeper insight into the projected visuals (as illustrated on Fig. 2).

Each of the identified gesture and interaction groups was analysed further to look for patterns throughout the study session and form conclusions and recommendations:

- (i) localisation and positioning: obtained results have shown that participants were less eager to move around the cylinder rather than just looking at the cylinder. Occasionally, they tried to move closer to the visualization, yet no significant differences between lab and in situ industrial conditions were observed.
- (ii) displacement: Most of visual objects displacement moves were one-hand driven, except when in real in situ conditions where users tend to use both hands for

object grabbing and manipulation. Notably, we identified individual cases where participants utilized index finger pointing gestures for moving the object to the desired destination.

- (iii) rotation: Most of the participants performed this task with the implemented gesture of rotation by grabbing the object with two hands, yet a considerable portion of users tried to rotate the object with one hand as well.
- (iv) slicing & extracting: extracting single images from the stack of images (3D data) was observed in more complex tasks (T2 and T3). Again, for in situ conditions users tend to use both hands while generally one-hand interaction was preferred by the participants. Notably, slicing was the most diverse action in regard to preferred gestures proposed by the users. While single finger gestures were noticed for empty lab condition, no single finger gestures were observed in situ.

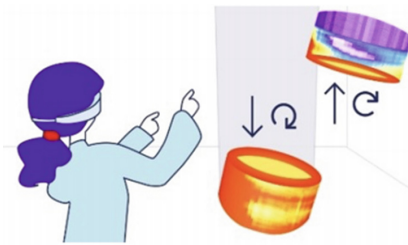


Fig. 2. Visual representation of the actions of the user - moving and rotating the part of the 3D object during gestural interaction with 3D AR data. (Source: Personal collection)

Figure 3 shows the mean rating of each category evaluated by users and the mean weight each category has, with the width of the bar indicating the weight [23]. Based on the graph we note that on average the physical and temporal demand were not the highest rated, we also can note that performance is the most highly rated and weighted category, which indicates that most users were preoccupied by their performance during the tasks. The frustration is also very highly weighted, which indicates that for a lot of users the frustration was important. Overall, mental demand is not rated highly, which is promising for implementation of the technology, as it shows that the complexity of the tasks did not increase because of HoloLens, although effort is one of the two highest rated categories.

5 Discussion and Conclusions

This exploratory study demonstrated patterns of possible use of HMD AR technology for the specific, industrial flow inspection application. Initial results revealed that behavior of a group of users while interacting gesturally with a virtual 3D data visual might be different for safe, lab conditions than for real industrial settings. In open space situations, where there are little to no potential risks while using the application, users tend to focus more on optimal solutions like operating with one hand. Analysis of data connected with object movement gestures shows that when creating an AR application for an industrial environment, it is important to implement grabbing functionality for both 1 and 2 hands.

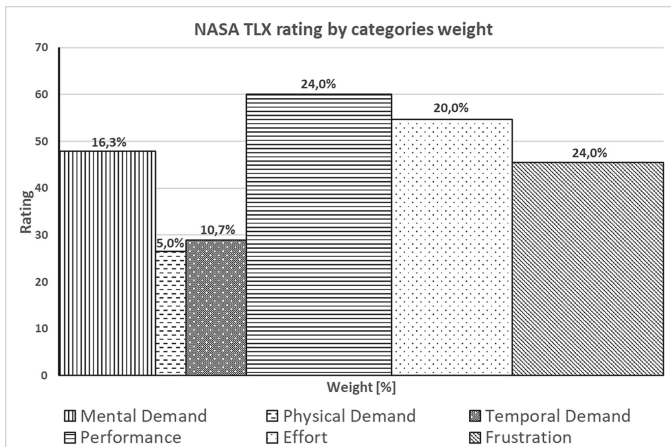


Fig. 3. Bar chart showing the mean rating of each category measured in the NASA TLX and their mean weight. (Source: Personal collection)

Some minor problems revealed in the study might have been associated with the following factors such as: big-scale silo and its projections size coupled with limited HoloLens display area; habits of using everyday touchscreen gestures designed for flat, tangible surfaces that are not fully transformable to in-the-air space; etc. In the context of designing for HoloLens 2, the focus should be on improving the comfort of work. Some users have reported that wearing goggles is uncomfortable. They had them on their heads for about 10–15 min. Long work in the goggles may be problematic for users due to the uncomfortable mounting on the head.

The most frustrating problem for the user was the absence of recognition of the gestures they wanted to use. It made subsequent attempts more nervous. Despite a few shortcomings, users were enthusiastic about AR and how to use it. This points to the fact that along with the popularization of Augmented Reality in everyday life, they will show a desire to immerse themselves in it. Main conclusions derived from this study are as follows:

- Significant difference between the gestural interactions performed by the participants within a neutral and an industrial environment was spotted in.
- There are physical and technical limitations of HoloLens HMDs. Hence the context of working within the industrial settings must be considered when designing a particular solution.
- There might be some benefits of simultaneous implementation of several different types of gestures for the same operation/command/task to accommodate the needs of different users.

Results suggest further exploratory research on this topic is recommended. Revealed patterns show we can highlight a need to mix single- and two-hand- gestures while building applications for industrial use. Furthermore, it was noticed that some users treat working with a 3D dataset as working with a physical object, while others treat it as a flat touchscreen-alike visual. Some differences were observed when comparing the users'

behavior in a semi-industrial settings and in an empty room to be further investigated in the future. All the examined cases suggest the need to consider both the surroundings and the context, when designing augmented reality applications for industrial settings. Furthermore, it would be interesting to explore possible combinations of the gestural interaction with some other sensing technologies, such as EMG [24] or ultrasound-based [25], to involve some machine learning algorithms [26, 27] for optimising the mixture of gestural, voice and traditional input [28] as well as further explore eye-tracking modality to track attention and performance of the users [29–31].

References

1. Zhang, Y., Nowak, A., Romanowski, A., Fjeld, M.: On-site or remote working?: an initial solution on how COVID-19 pandemic may impact augmented reality users. In: 2022 International Conference on Advanced Visual Interfaces (AVI 2022), Article no. 65, p. 3. ACM, New York (2022). <https://doi.org/10.1145/3531073.3534490>
2. Gerup, J., Soerensen, C.B., Dieckmann, P.: Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. *Int. J. Med. Educ.* **18**(11), 1–18 (2020). <https://doi.org/10.5116/ijme.5e01.eb1a>
3. Juanes, J.A., Hernández, D., Ruisoto, P., García, E., Villarrubia, G., Prats, A.: Augmented reality techniques, using mobile devices, for learning human anatomy. In: Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2014), pp. 7–11. ACM, New York (2014). <https://doi.org/10.1145/2669711.2669870>
4. Dubois, E., Nigay, L.: Augmented reality: which augmentation for which reality?. In: Designing Augmented Reality Environments (DARE 2000), pp. 165–166. ACM, New York (2000). <https://doi.org/10.1145/354666.354695>
5. Aromaa, S., Aaltonen, I., Kaasinen, E., Elo, J., Parkkinen, I.: Use of wearable and augmented reality technologies in industrial maintenance work. In: Proceedings of the 20th International Academic Mindtrek Conference (AcademicMindtrek 2016), pp. 235–242. ACM, New York (2016). <https://doi.org/10.1145/2994310.2994321>
6. Gattullo, M., Evangelista, A., Uva, A.E., Fiorentino, M., Boccaccio, A., Manghisi, V.M.: Exploiting augmented reality to enhance piping and instrumentation diagrams for information retrieval tasks in industry 4.0 maintenance. In: Bourdot, P., Interrante, V., Nedel, L., Magnenat-Thalmann, N., Zachmann, G. (eds.) EuroVR 2019. LNCS, vol. 11883, pp. 170–180. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-31908-3_11
7. Wobbrock, J.O., Ringel Morris, M., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, pp. 1083–1092. ACM, New York (2009). <https://doi.org/10.1145/1518701.1518866>
8. Minseok, K., Sung, H., Kyeong-Beom, P., Jae Yeol, L.: User interactions for augmented reality smart glasses: a comparative evaluation of visual contexts and interaction gestures (2019)
9. Piumsomboon, T., Clark, A., Billingham, M., Cockburn, A.: User-defined gestures for augmented reality. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013. LNCS, vol. 8118, pp. 282–299. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40480-1_18
10. Rumiński, D., Maik, M., Walczak, K.: Visualizing financial stock data within an augmented reality trading environment. *Acta Polytechnica Hungarica* **16**(6), 223–239 (2019)
11. Pham, T., Vermeulen, J., Tang, A., MacDonald Vermeulen, L.: Scale impacts elicited gestures for manipulating holograms: implications for AR gesture design. In: Proceedings of the 2018

- Designing Interactive Systems Conference (DIS 2018), pp. 227–240. ACM, New York (2018). <https://doi.org/10.1145/3196709.3196719>
12. Feigl, T., Porada, A., Steiner, S., Löffler, C., Mutschler, C., Philippsen, M.: Localization limitations of ARCore, ARKit, and HoloLens in dynamic large-scale industry environments. In: Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - GRAPP, pp. 307–318 (2020). <https://doi.org/10.5220/0008989903070318>
 13. Redžepagić, A., Löffler, C., Feigl, T., Mutschler, C.: A sense of quality for augmented reality assisted process guidance. In: 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR), pp. 129–134 (2020). <https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00046>
 14. Büttner, S., Prilla, M., Röcker, C.: Augmented reality training for industrial assembly work - are projection-based AR assistive systems an appropriate tool for assembly training? In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI 2020), pp. 1–12. ACM, New York (2020). <https://doi.org/10.1145/3313831.3376720>
 15. Hebenstreit, M., Spitzer, M., Eder, M., Ramsauer, C.: An industry 4.0 production workplace enhanced by using mixed reality assembly instructions with Microsoft HoloLens. In: Hansen, C., Nürnberger, A., Preim, B. (Hrsg.) Mensch und Computer 2020 - Workshopband., Gesellschaft für Informatik e.V. (2020). <https://doi.org/10.18420/muc2020-ws116-005>
 16. Meyer, S.: Augmented reality in the pharmaceutical industry a case study on HoloLens for fully automated dissolution guidance (2021)
 17. Nowak, A., Zhang, Y., Romanowski, A., Fjeld, M.: Augmented reality with industrial process tomography: to support complex data analysis in 3D space. In: 2021 ACM International Symposium on Wearable Computers (UbiComp 2021), pp. 56–58. ACM, New York (2021). <https://doi.org/10.1145/3460418.3479288>
 18. Fernando de Souza Cardoso, L., Martins Queiroz Mariano, F.C., Zorzal, E.R.: A survey of industrial augmented reality. *Comput. Ind. Eng.* **139**, 106159 (2020). <https://doi.org/10.1016/j.cie.2019.106159>. ISSN 0360-8352
 19. Sulikowski, P., Zdziebko, T.: Deep learning-enhanced framework for performance evaluation of a recommending interface with varied recommendation position and intensity based on eye-tracking equipment data processing. *Electronics* **9**(2), 266 (2020). <https://doi.org/10.3390/electronics9020266>
 20. Hampel, U., et al.: A review on fast tomographic imaging techniques and their potential application in industrial process control. *Sensors* **22**(6) (2022). <https://doi.org/10.3390/s22062309>
 21. Rymarczyk, T., Kłosowski, G., Kozłowski, E., Tchórzewski, P.: Comparison of selected machine learning algorithms for industrial electrical tomography. *Sensors* **19**(7) (2019). <https://doi.org/10.3390/s19071521>
 22. Sulikowski, P., Zdziebko, T.: Horizontal vs. vertical recommendation zones evaluation using behavior tracking. *Appl. Sci.* **11**(1), 56 (2021). <https://doi.org/10.3390/app11010056>
 23. Hertzum, M.: Reference values and subscale patterns for the task load index (TLX): a meta-analytic review. *Ergonomics* **64**, 869–878 (2021). <https://doi.org/10.1080/00140139.2021.1876927>
 24. Woźniak, M., Pomykalski, P., Sielski, D., Grudzień, K., Paluch, N., Chaniecki, Z.: Exploring EMG gesture recognition-interactive armband for audio playback control. In: 2018 Federated Conference on Computer Science and Information Systems, pp. 919–923 (2018)
 25. Soleimani, M., Rymarczyk, T.: A tactile skin system for touch sensing with ultrasound tomography. *TechRxiv*. Preprint (2022). <https://doi.org/10.36227/techrxiv.21332655.v1>
 26. Rymarczyk, T., Król, K., Kozłowski, E., Wołowicz, T., Cholewa-Wiktor, M., Bednarczuk, P.: Application of electrical tomography imaging using machine learning methods for the

- monitoring of flood embankments leaks. *Energies* **14**, 8081 (2021). <https://doi.org/10.3390/en14238081>
27. Romanowski, A., et al.: Interactive timeline approach for contextual spatio-temporal ECT data investigation. *Sensors* **20**, 4793 (2020). <https://doi.org/10.3390/s20174793>
 28. Pomykalski, P., Woźniak, M.P., Woźniak, P.W., Grudzień, K., Zhao, S., Romanowski, A.: Considering wake gestures for smart assistant use. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA 2020)*, pp. 1–8. ACM, New York (2020). <https://doi.org/10.1145/3334480.3383089>
 29. Sulikowski, P., Ryczko, K., Bąk, I., Yoo, S., Zdziebko, T.: Attempts to attract eyesight in e-commerce may have negative effects. *Sensors* **22**, 8597 (2022). <https://doi.org/10.3390/s2228597>
 30. Sulikowski, P., Kucznerowicz, M., Bąk, I., Romanowski, A., Zdziebko, T.: Online store aesthetics impact efficacy of product recommendations and highlighting. *Sensors* **22**, 9186 (2022). <https://doi.org/10.3390/s22239186>
 31. Schrader, A., et al.: Toward eye-tracked sideline concussion assessment in eXtended reality. In: *ACM Symposium on Eye Tracking Research and Applications (ETRA 2021)*, pp. 1–11, Article no. 7. ACM, New York (2021). <https://doi.org/10.1145/3448017.3457378>





Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Polish Adaptation of the Cybersickness Susceptibility Questionnaire (CSSQ-PL)

Laura Osęka¹ , Grzegorz Pochwatko¹  , and Justyna Świdrak^{1,2} 

¹ Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland
{gp, jswidrak}@psych.pan.pl

² August Pi and Sunyer Biomedical Research Institute, Barcelona, Spain

Abstract. We present the Polish adaptation of the Cybersickness Susceptibility Questionnaire (CSSQ) by Freiwald et al. [4] and its short version. To validate the scale, we analyzed the results of 85 participants of the VR artistic experience “Nightsss” in versions with 3 and 6 degrees of freedom. Both long and short versions are characterized by high internal consistency. The validity was analyzed by evaluating the difference between the scores obtained in the CSSQ-PL and the level of experienced cybersickness measured with the Simulator Sickness Questionnaire. The scale is suitable for screening participants in research conducted with the use of immersive virtual reality (VR). Further studies are necessary to confirm the predictive power of the scale.

Keywords: virtual reality · cybersickness · immersion · questionnaire

1 Introduction

Cybersickness may be described generally as a bodily discomfort induced by exposure to VR content. It may be one of the enduring hindrances in order to use the full potential of VR technology. The symptoms of cybersickness are similar to the ones which people suffer from when they get motion sick. There is a number of different definitions of cybersickness. In general, cybersickness is a set of symptoms produced by VR exposure that temporarily affects the well-being of the VR user and impedes the use of VR [4]. The most common symptoms are nausea, vertigo, disorientation and pallor [10]. It is most often categorized as a form of simulator sickness [16], although there is a distinction between the experience of cybersickness and simulator sickness. The dominant symptoms are different, i.e., in cybersickness, the disorientation symptoms tend to occur more often.

1.1 Measuring Cybersickness

There are two types of cybersickness measures, objective and subjective. Objective measures include physiological markers, such as bradycardic activity, respiration rate [3, 8], heart rate [15], galvanic skin response [5] and behavioral

measures, such as early termination of VR experience [9], delivery of incomplete task [15], impeding the further VR use [2]. Subjective measurements, on the other hand, are different multi-item questionnaires, such as Simulator Sickness Questionnaire (SSQ; [6]). The subscale measuring the experienced symptoms level consists of 16 items. It gives a possibility to generate a total sickness score and the levels of oculomotor discomfort, disorientation, and nausea [18]. It is commonly appreciated as a measurement of cybersickness in VR.

Other measurements that are offered for evaluating cybersickness are Cybersickness Questionnaire (CSQ; [17]) and Virtual Reality Sickness Questionnaire (VRSQ; [7]). CSQ was developed to capture the occurrence of cybersickness in VR context, and because it is based on SSQ, uses the SSQ answers to estimate dizziness and troubles with focusing. The VRSQ is a questionnaire which is aimed to measure motion sickness in a virtual reality environment [7]. It is a modified version of SSQ, and consists of 9 items grouped into two components, namely oculomotor and disorientation.

1.2 Predicting Cybersickness

Cybersickness Susceptibility Questionnaire predicts the tolerance to VR exposure. The questionnaire consists of three parts and it is aimed to be used before the VR environment exposure. It was created to demonstrate the effect of biological, chemical and psychological occurrence of cybersickness [4]. The first part of CSSQ collects demographic questions, the second part evaluates the physical health and fitness levels. There are questions aimed at investigating the headaches, stomach complaints, alcohol and substance consumption within 24 h prior to the VR exposure as well as acute complaints, as all of these are the indicators of a higher probability of cybersickness occurrence [4]. The last part of the questionnaire evaluates the sensibility to motion sickness caused by various types of motion, such as driving as a passenger, traveling by train, or sitting on a roller coaster ride.

Most of the questionnaires which measure cybersickness focus mainly on the outcomes, not on what triggers the problem to occur. It is important to combine both the physiological and psychological factors of cybersickness to better understand this phenomenon, as the occurrence of cybersickness is one of the reasons impeding people from the use of VR [14]. Moreover, there is some evidence that cybersickness may be negatively related to presence in VR. The sense of presence presumably suppresses cybersickness because the attention is directed to the unpleasant motions from the body [18]. Thus, in some studies, screening potential participants for their susceptibility to cybersickness may help reduce the risk of unpleasant or poor VR experiences for participants and the dropout rate for researchers simultaneously.

The main objective of the presented study is to prepare the Polish version of CSSQ and test its psychometric characteristics to evaluate its usefulness in research and other VR applications, e.g., physiotherapy or education.

2 Method

The adaptation of the scale was a part of a larger study on emotional reactions to immersive art. The study design was similar to that used by the Authors of the original scale [4]. The Nightsss experience has been chosen as a material to be shown to participants, as it contains two parallel versions, i.e., an interactive and a cinematic one [11, 12]. The content of both was identical, and the difference was in the interactivity level. In the interactive version, the participants were allowed to walk freely in the 3D environment and have some interactions with objects, while in the Cinematic VR (CVR) version, they were only able to watch the experience in the form of a 360 3D movie. The procedure was conducted on the HTC Vive Pro Eye HMD with the wireless addon and the Intel I9 2.8 GHz PC with Gigabyte Geforce RTX 3070. The study obtained ethics approval from the Ethics Committee of the Institute of Psychology Polish Academy of Sciences.

2.1 Participants and Apparatus

Eighty five Caucasian (Polish) participants (26 male, 58 female, 1 nonbinary or other) took part in the experiment. The mean age was respectively 31, 31 (SD = 9 in both groups), and 22. Levene's test for homoscedasticity of variance detected no differences ($p = .296$).

2.2 Measures

Polish adaptation of SSQ [1] was used. The original scale is divided into two parts: one with the questions about the current health state, alcohol, and medicine consumption and the other measuring the symptoms of simulator sickness, on 4 points scale, from none to severe. The results range was 0–179.5. We omitted non-diagnostic questions as the whole procedure was aimed to be not very bothersome for participants.

CSSQ [4] consists of three parts. The first collects demographic data, second consists of questions considering the health and fitness of participants, including two questions on a 5-point Likert scale and 11 binary ones. The last part measures vulnerability to motion sickness with thirteen questions on a 5-point Likert scale from 0 - very rarely to 4 - very often (Table 1).

2.3 Stimuli and Procedure

After entering the vrLab, participants were asked to fill the first part of the Polish adaptation of the SSQ (demographic data, questions about health condition, past illness, alcohol, and drug intake, and sleep quality) [6] together with CSSQ-PL. Next, they were informed about the equipment, procedure, the possible effects of using the VR headset, and the collected data. Then the GSR and HRV electrodes were placed. In the first part of the VR experiment, the participants stood in an area limited to 1,5 sq m and viewed short video clips of neutral, positive, and

negative valence from the Public Database Of 360 Videos With Corresponding Ratings Of Arousal And Valence [13]. Next, the participants saw either CVR standing in one place or interactive “Nightsss” where they were able to move around an area limited to about 7sq m. The experience lasted around 8min. After removing the VR goggles, the participants were asked to fill out the SSQ and evaluate their feelings and opinion with open-ended questions.

2.4 Validation

The language adaptation to Polish was performed as follows: (1) the scale was translated by three independent researchers from English to Polish, (2) all three versions were compared to select the final version, (3) the subscale’s reliability and validity were tested. The questions with a binary answer scale in the “Health and fitness” subscale were coded as yes = 1 and no = 0. Reliability of the “Health & Fitness” subscale was tested by point biserial correlation between the item score and the total subscale score. The “Motion Sickness” subscale’s reliability was evaluated by analyzing the alpha Cronbach’s scores if an item would be removed (Table 1). Next, we calculated the sum of the “Health and fitness” and “Motion sickness” subscales as a sum score of CSSQ items excluding the demographic variables. We also calculated the total score of the SSQ-post and its subscales (Nausea, Oculomotor, Disorientation), following Biernacki et al. [1] to test the validity of the adapted scale. Next, we visually analyzed the distributions of the variables and normalized the variables to a 0–1 scale, and calculated the correlation matrix.

3 Results

3.1 Language Adaptation

The final version of the Polish scale and the original scale published by Freiwald et al. [4] can be found in Table 1. Individual translations used to build the final scale can be obtained on request from the Authors.

3.2 Reliability

Due to the content of the “Health and Fitness” subscale, which refers mainly to the medical records and is constructed mainly on binary variables, the reliability analysis was carried out only for the “Motion sickness” subscale. The overall Cronbach’s $\alpha = .87$, which indicated very high reliability. Cronbach’s α and items correlations per item can be found in Table 1. Based on the analysis, we also prepared a short version of the “Motion sickness” subscale, which included six items (selected on the basis of itemwise analysis to maintain high internal consistency - Cronbach’s α above .80). The subscale is characterized by high reliability ($\alpha = 0.82$) and can be used for a quick screening. Items selected for the short scale are marked with an asterisk in Table 1.

Table 1. Cybersickness Susceptibility Questionnaire - PL.

English [4]	Polish			
<i>Demographic data</i>	<i>Dane demograficzne</i>	Scale		
Age, Height, Gender, Ethnicity	Wiek, wzrost, płeć, pochodzenie	Freetext		
How often do you use VR?	Jak często korzystasz z rzeczywistości wirtualnej?	Likert scale		
<i>Health and Fitness</i>	<i>Zdrowie i stan fizyczny</i>		biserial corr	p-value
How often do you suffer from a migraine headache?	Jak często cierpisz na migrenowe bóle głowy?	Likert scale	.762	< .001
How often do you suffer from stomach discomfort?	Jak często cierpisz na problemy żołądkowe?	Likert scale	.788	< .001
Did you consume alcohol in the last 24 h?	Czy spożywał_ś alkohol w ciągu ostatniej doby?	tak/nie yes/no	.309	.004
Did you consume drugs in the last 24 h?	Czy zażywał_ś narkotyki w ciągu ostatnich 24 godzin?	tak/nie yes/no	n/a	n/a
Did you consume medication in the last 24 h?	Czy brał_ś jakieś leki w ciągu ostatnich 24 godzin?	tak/nie yes/no	.325	.002
Do you suffer from a cold or flu at the moment?	Czy chorujesz obecnie na przeziębienie lub grypę?	tak/nie yes/no	n/a	n/a
Do you suffer from an ear infection at the moment?	Czy masz obecnie infekcję ucha?	tak/nie yes/no	n/a	n/a
Do you suffer from a respiratory disease at the moment?	Czy chorujesz obecnie na jakąś chorobę układu oddechowego?	tak/nie yes/no	.084	.447
Are you suffering from a lack of sleep at the moment?	Czy cierpisz obecnie na niedobór snu?	tak/nie yes/no	.215	.048
Are you suffering from an eye disease?	Czy cierpisz obecnie na jakąś chorobę oczu?	tak/nie yes/no	.257	.017
Have you been prescribed new glasses recently?	Czy przepisano Ci ostatnio nowe okulary?	tak/nie yes/no	.257	.021
Do you suffer from a limitation of your vestibular system?	Czy masz zaburzenia układu przedsionkowego?	tak/nie yes/no	n/a	n/a
Do you suffer from a limitation of your oculomotor system?	Czy masz zaburzenia układu okoruchowego?	tak/nie yes/no	n/a	n/a
<i>Motion Sickness</i>	<i>Choroba lokomocyjna</i>		α if Item Deleted	Item-Total Corr.
When I drive a car, I feel sick as a passenger	Jest mi niedobrze gdy jadę samochodem jako pasażer/pasażerka.*	Likert scale	.850	.712
Reading as a passenger makes me sick.	Czytanie w czasie podróży przyprawia mnie o mdłości	Likert scale	.873	.515
I feel sick when I am sailing. Likert scale	Jest mi niedobrze podczas żeglowania.	Likert scale	.858	.588
I feel sick on small boats.	Jest mi niedobrze kiedy płynam łódką.*	Likert scale	.856	.637
I feel sick when I go by train.	Jest mi niedobrze kiedy jadę pociągiem.*	Likert scale	.869	.406
I feel sick when I go backwards by train.	Jest mi niedobrze kiedy jadę tyłem do kierunku jazdy w pociągu.	Likert scale	.859	.605
I feel sick when I ride the bus or am a co-driver in a car.	Jest mi niedobrze kiedy jadę autobusem lub siedzę obok kierowcy w samochodzie.*	Likert scale	.855	.673
I feel sick when I ride the bus sitting backwards.	Jest mi niedobrze kiedy jadę autobusem tyłem do kierunku jazdy.*	Likert scale	.851	.709
I feel sick when I am in an airplane.	Jest mi niedobrze podczas lotu samolotem.	Likert scale	.868	.408
I feel sick on rotating swivel chairs.	Robi mi się niedobrze na krześle obrotowym.*	Likert scale	.863	.523
I feel sick when I ride a roller coaster.	Jest mi niedobrze podczas jazdy kolejką górską.	Likert scale	.866	.475
I feel sick when I ride a carousel.	Jest mi niedobrze na karuzeli.	Likert scale	.855	.655
I feel sick when swinging on a swing.	Jest mi niedobrze gdy huśtam się na huśtawce.	Likert scale	.866	.450

3.3 Distributions

Overall, most participants reported very low to low susceptibility to cybersickness ($M = 7.36, SD = 6.86$) and very low to low cybersickness after the VR exposure ($M = 16.37, SD = 15.93$). The distributions were strongly left-skewed for all subscales (see Fig. 1).

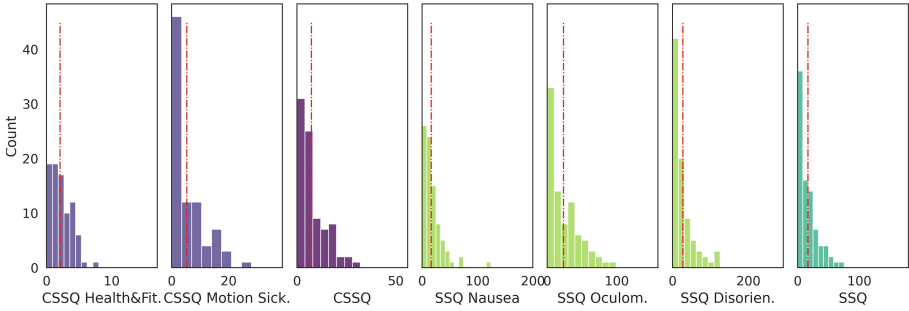


Fig. 1. Distribution of subscales

3.4 Validity

Due to the skewness, we decided to test the predictive power of the CSSQ scale by applying the authors’ original approach. First, we normalized both CSSQ and SSQ total scores to 0–1 scale, and extracted the difference between the estimated susceptibility (CSSQ) and the obtained cybersickness levels (SSQ) to evaluate whether these scores matched. The variable had close-to-normal distribution with a mean of $-.011$ and standard deviation of $.212$. The mean was very close to zero and 80% of the participants had scores within one standard deviation (see Fig. 2).

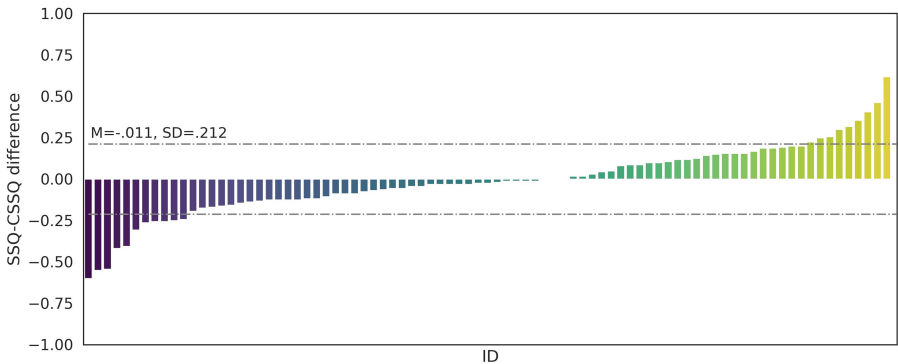


Fig. 2. Difference between the prediction (CSSQ) and the obtained cybersickness levels (SSQ)

Moreover, we tested the strength of the linear association between the subscales and demographic variables (Fig. 3). We found a strong positive correlation of SSQ with CSSQ ($r = .58$), a moderate correlation of gender with both CSSQ and SSQ. Analyzing the correlations between the CSSQ and SSQ subscales, we can observe that “Health & Fitness” subscale was weakly correlated only with SSQ total score ($r = .32$), while the “Motion Sickness” subscale correlated weakly to moderately with all SSQ measures. All subscales within the same construct correlated in the expected direction. Age did not correlate with any of the variables. Moreover, female gender and lower height were related to higher levels of motion sickness sensitivity and overall cybersickness, although the association was relatively weak ($r = .35 \times .03$). Since the sample consisted of more women than men, and there was a strong negative correlation between gender and height, it is possible that the correlation with height was an artifact.

4 Discussion and Future Directions

We have demonstrated the reliability and validity of the Polish version of the CSSQ. Both long and short versions are characterized by high internal consistency. At the same time, there are no items to be removed to significantly improve consistency. Unfortunately, we also observed a strong skewness of the results. This may be due to the fact that potential participants were previously informed about possible contraindications for VR studies and resigned from participation at the recruitment stage. Unfortunately, no satisfactory solution to this limitation is available since a different approach would be against the rules of ethics. CSSQ-PL scores match the levels of actual symptoms of cybersickness. This makes the questionnaire a good tool for screening participants in VR research and applications, including gaming, physiotherapy, or education, and protects potential participants from unpleasant experiences. In future research, it is necessary to 1) recruit a much larger sample in order to at least partially reduce the impact of skewness of the distribution on the results and 2) select a longer and more diverse VR material in order to better verify the accuracy of the predictions.

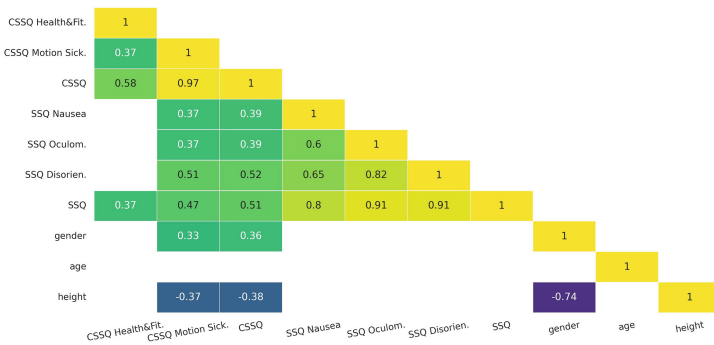


Fig. 3. Pearson’s correlation matrix. Scales were normalized to 0–1. Visible correlations are significant at $p < .005$

Acknowledgments. The publication is partly based on research carried out as part of the project “New Forms and Technologies of Narration”. Project financed under the program of the Minister of Education and Science under the name “Regional Excellence Initiative” in 2019–2023, project number 023/RID/2018/19, financing amount: PLN 11865100.

References

1. Biernacki, M.P., Kennedy, R.S., Dziuda, L.: Simulator sickness and its measurement with simulator sickness questionnaire (SSQ)/zjawisko choroby symulatorowej oraz jej pomiar na przykładzie kwestionariusza do badania choroby symulatorowej-SSQ. *Med. Pr.* **67**(4), 545–556 (2016)
2. Davis, S., Nesbitt, K., Nalivaiko, E.: A systematic review of cybersickness. In: *Proceedings of the 2014 Conference on Interactive Entertainment*, pp. 1–9 (2014)
3. Dennison, M.S., Wisti, A.Z., D’Zmura, M.: Use of physiological signals to predict cybersickness. *Displays* **44**, 42–52 (2016)
4. Freiwald, J.P., Göbel, Y., Mostajeran, F., Steinicke, F.: The cybersickness susceptibility questionnaire: predicting virtual reality tolerance. In: *Proceedings of the Conference on Mensch and Computer*, pp. 115–118 (2020)
5. Gavgani, A.M., Nesbitt, K.V., Blackmore, K.L., Nalivaiko, E.: Profiling subjective symptoms and autonomic changes associated with cybersickness. *Auton. Neurosci.* **203**, 41–50 (2017)
6. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* **3**(3), 203–220 (1993)
7. Kim, H.K., Park, J., Choi, Y., Choe, M.: Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* **69**, 66–73 (2018)
8. Kim, Y.Y., Kim, H.J., Kim, E.N., Ko, H.D., Kim, H.T.: Characteristic changes in the physiological components of cybersickness. *Psychophysiology* **42**(5), 616–625 (2005)
9. Kinsella, A.J.: The effect of 0.2 Hz and 1.0 Hz frequency and 100 ms and 20–100 ms amplitude of latency on simulator sickness in a head mounted display. Ph.D. thesis, Clemson University (2014)
10. LaViola, J.J., Jr.: A discussion of cybersickness in virtual environments. *ACM Sigchi Bull.* **32**(1), 47–56 (2000)
11. Lewandowska, W.: Nightsss (cinematic VR) (2021). <http://vnlab.film school.lodz.pl/pracownia-vr-ar/wizualny-poemat-erotyczny/>
12. Lewandowska, W., Frydrysiak, S.: Nightsss (VR) (2021). <http://vnlab.film school.lodz.pl/pracownia-vr-ar/wizualny-poemat-erotyczny/>
13. Li, B.J., Bailenson, J.N., Pines, A., Greenleaf, W.J., Williams, L.M.: A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Front. Psychol.* **8**, 2116 (2017)
14. Mousavi, M., Jen, Y.H., Musa, S.N.B.: A review on cybersickness and usability in virtual environments. In: *Advanced Engineering Forum*, vol. 10, pp. 34–39. *Trans. Tech. Publ.* (2013)
15. Nalivaiko, E., Davis, S.L., Blackmore, K.L., Vakulin, A., Nesbitt, K.V.: Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* **151**, 583–590 (2015)

16. Stanney, K.M., Kennedy, R.S., Drexler, J.M.: Cybersickness is not simulator sickness. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 41, pp. 1138–1142. SAGE Publications, Sage, CA (1997)
17. Stone III, W.B.: Psychometric evaluation of the simulator sickness questionnaire as a measure of cybersickness. Ph.D. thesis, Iowa State University (2017)
18. Weech, S., Kenny, S., Barnett-Cowan, M.: Presence and cybersickness in virtual reality are negatively related: a review. *Front. Psychol.* **10**, 158 (2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



**Special Session: Advances
in Collaborative Robotics**



NARX Recurrent Neural Network Model of the Graphene-Based Electronic Skin Sensors with Hysteretic Behaviour

Jakub Możaryn^(✉) 

Department of Mechatronics, Institute of Automatic Control and Robotics,
Warsaw University of Technology, ul. Sw. A. Boboli 8, 02-525 Warsaw, Poland
jakub.mozaryn@pw.edu.pl

Abstract. The electronic skin described in the article comprises screen-printed graphene-based sensors, intended to be used for robotic applications. The precise mathematical model allowing the touch pressure estimation is required during its calibration. The article describes the recurrent neural network model for graphene-based electronic skin calibration, in which parameters are not homogeneous, and the touch force characteristics have visible hysteretic behaviour. The presented method provides a simple alternative to the models known in the literature.

Keywords: Electronic skin · Graphene nanoplatelets · Force-sensitive resistors · Recurrent Neural Networks · Hysteresis

1 Introduction

Skin is the thin layer of tissue forming the natural outer covering of the body. It is a protective barrier against mechanical, thermal and physical injury and hazardous substances. It also acts as a sensory organ (touch, temperature, strain, moisture). The electronic skin (e-skin) and its applications in many domains have attracted researchers worldwide for over three decades [8]. E-skin as an interface that mimics a biological tissue is being developed in medicine (smart health care [18], prosthesis [20]), and robotics [3]). In robotics, it allows for increasing safety in human-machine cooperation, the agility of robotic manipulation [19], and soft robotics [2, 4].

To calibrate and establish the measurement characteristics of e-skin sensors, a standard approach based on the measurement of the force exerted by a reference device. In the known research, linear, quadratic, and cubic polynomial models were fitted to data to calibrate the example sensor [5, 9, 21]. The best results were obtained using a Huber regression with a quadratic polynomial model. In [15] e-skin semi-automatised calibration procedure using industrial robot FANUC LR Mate 200iC equipped with a reference force sensor was presented where second-order exponential function and logistic function were used. To assess fitting results, two parameters were analysed: adjusted coefficient of determination (ARS) and root-mean-square error (RMSE). However, in all the

abovementioned research [15] the e-skin graphene sensor's characteristics manifest hysteretic behaviour, and the visible effect of hysteresis on the modelling accuracy has not yet been discussed.

Hysteresis is the dependence of the state of a system on the state's history, and we can find this phenomenon in many science fields as physics, chemistry, and engineering. Besides the subject-specific models, some hysteretic models capture general features of many systems with hysteresis that can be classified as algebraic models, transcendental models, differential models and integral models [16]. One of the promising solutions is to use artificial neural networks to model the system's data with hysteretic behaviour. They have distinct advantages over linear identification methods, i.e., the approximation of multivariable nonlinear functions, the simple gradient-based adaptation of model parameters and a rapid calculation of neural network equations. In contrast to analytical models, the design procedure of the ANN does not demand an exact knowledge of model physical equations and physical parameters that describe the model, but only values of model variables in the causal form. Several neural network architectures with recurrent layers and memory capacity were proposed recently, e.g. recurrent neural networks for ultra-capacitors [1], physics-informed deep neural networks for mechanical dampers [10] or extended Preisach neural network [6] among others.

The presented research aims to propose a novel method of e-skin graphene sensor modelling using the NARX recurrent neural network that can describe the hysteretic behaviour of the sensor.

2 Graphene-Based Electronic Skin

The e-skin used for the research consisted of two layers [13, 14]. The first is a conductive layer of comb electrodes printed on plastic foil and connected along columns and rows. In contrast, the second one comprises FSR graphene sensors arranged in a rectangular pattern placed on a plastic foil. In the research, a matrix with the size of a single sensor (approx. 5×5 mm) was used. The e-skin controller measures the pressure for each sensor and transmits the position and touch force exerted on the active surface. The data is sent to a computer, processed and saved. The computer software enables the visualization of the touch results as a colour-coded image. Figure 1 presents the hand touch measurement for the 16×32 FSR matrix.

The measurement acquisition setup comprised the FANUC LR Mate 200iC manipulator, the R-30iA Mate manipulator controller, the e-skin with a driver, the OnRobot Hex-e 6-axis force and torque measuring device with a controller, and a general-purpose PC. Data from e-skin sensors and the reference Hex-e sensor were acquired during the calibration procedure. Data acquisition was subdivided into the 'loading phase', when the force exerted on the particular sensor of the e-skin by the robotic arm increased, and the 'unloading phase', when the force decreased (Fig. 2).

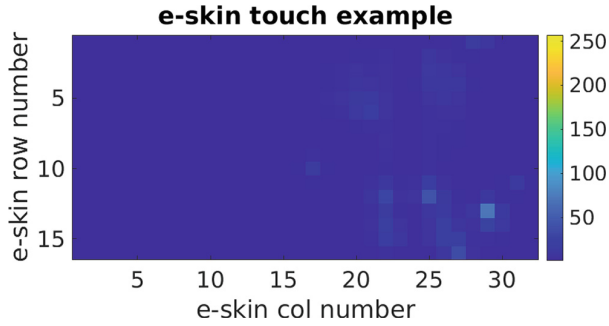


Fig. 1. Example hand touch visualization for the size of 16×32 cells; Gaussian blur used for post-processing.

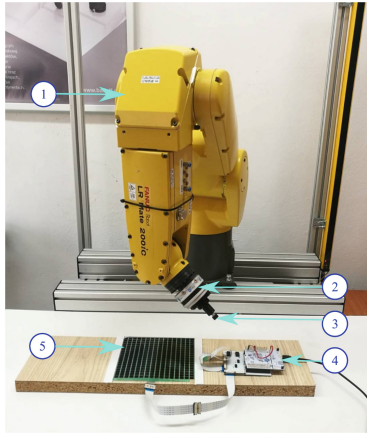


Fig. 2. Illustrative representation of the calibration station: 1. LRMate 200iC manipulator, 2. reference e-Hex sensor, 3. robot tool, 4. e-skin electronic system, 5. e-skin.

3 Neural-Network Modelling

Nonlinear AutoRegressive eXogenous model (NARX) in the form of the recurrent artificial neural network (R-ANN) was used to model the e-skin graphene sensor to estimate the pressure (force exerted on sensor) based on the sensor readouts. NARX-RNN nonlinear model extends the autoregressive linear model with the exogenous input (ARX), popular in time-series nonlinear modelling. NARX-RNN model is the input-output model, where the output in each step is described by input signal with noise. NARX-RNN has memory capabilities. It memorizes previous data and can be used to model hysteretic behaviour. While making a decision, it considers the current input and what it has learned from the inputs it received previously. Output from the previous step is fed as input to the current step creating a feedback loop.

In such a case, the nonlinear part of the NARX-RNN model was described as

$$y_{NN}(k) = F_{NN}(k) = f(V(k), V(k-1), F_{NN}(k-1)) \tag{1}$$

where $y_{NN}(k)$ - output of the NARX-RNN, $F_{NN}(k)$ - estimated output (touch pressure) at step k , $V(k)$ - sensor readouts at the step k , k - sample number, $t = kT_p$ - time, T_p - sampling time.

Exemplary scheme of used in the research NARX-RNN model is presented in Fig. 3.

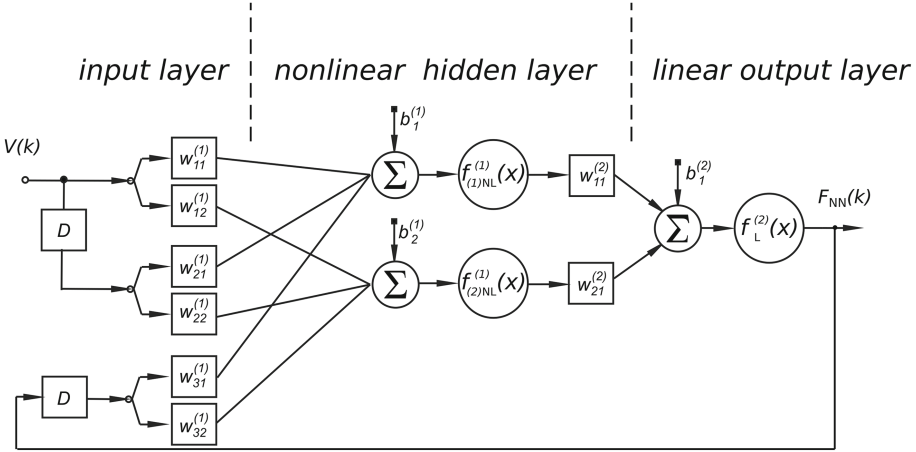


Fig. 3. Used in the research NARX-RNN architecture (2 neurons in the nonlinear hidden layer, D - delay ($k - 1$)).

The NARX-RNN comprises the nonlinear hidden layer that processes the input data as

$$\begin{cases} y^{(1)}_{(j)NL}(k) = f^{(1)}_{(j)NL}(x^{(1)}_{(j)NL}(k)); f^{(1)}_{(j)NL} \in [-1, 1] \\ x^{(1)}_{(n)NL}(k) = b^{(1)}_{(n)} + w_{p+1,n}^{(1)}y_L^{(2)}(k-1) + \sum_{i=1}^p w^{(1)}_{i,n}x_i(k) \end{cases} \tag{2}$$

where $y^{(1)}_{(j)NL}(k)$ - output signal of the j -th neuron in the nonlinear layer, $x^{(1)}_{(j)NL}(k)$ - input signal to the j -th neuron in the nonlinear layer, n - the number of neurons in the nonlinear layer, p - the number of neuron inputs in the nonlinear layer, $w^{(1)}_{i,j}$ - the weight of the i -th input to j th neuron in the nonlinear layer, $x_i(k)$ - i -th input to the network (with tapped delay line, D inputs in Fig. 3), $b^{(1)}_{(n)}$ - the threshold offset of the n -th neuron in the nonlinear layer. The second term of (2) describes the recurrent feedback loop with delay (D in Fig. 3). In the considered case, the output of the RNN network $y^{(1)}_L(k)$ is

the output signal from a linear output layer, with 1 linear neuron described as follows

$$\begin{cases} y_{\text{NN}}(k) = y_{\text{L}}^{(2)}(k) = f_{\text{L}}^{(2)}\left(x_{\text{L}}^{(2)}, k\right) = x^{(2)}, f_{\text{L}}^{(2)} \in R \\ x_{\text{L}}^{(2)}(k) = b_{\text{L}}^{(2)} + \sum_{i=1}^n w_{(\text{L})i,1}^{(2)} h_i(k) \end{cases} \quad (3)$$

where $y_{\text{NN}}(k)$ - the network output, equal to the linear layer output, $x_{\text{L}}^{(2)}(k)$ - the input signal to the neuron in the linear output layer, $w_{(\text{L})i,1}^{(2)}$ - the weight of the i th input to neuron in the linear output layer, and $b_{\text{L}}^{(2)}$ - the neuron's bias in the linear output layer.

For the training and evaluation of the NARX-RNN model, the estimation error was calculated as follows

$$e_{\text{NARX-RNN}}(k) = y_{\text{NN}}(k) - y(k) \quad (4)$$

4 Results

4.1 Research Method

The Mean Squared Error (MSE) has been used to describe the performance function $E_{\text{NARX-RNN}}$ of the NARX-RNN model during training and testing. The changes in the weights in the i -th iteration were used in the NARX-RNN models according to the Levenberg-Marquardt algorithm, and variable metrics method [12]. The training algorithm stop conditions were defined because of the possible large step of each iteration. The mentioned conditions are usually estimated as the assumed minimum value of the performance function and the maximum number of training iterations. In the case described in this article, the stopping conditions were: epochs = 1000, $E_{\text{NARX-RNN}} \leq 10^{-6}$. The early stop method was used during training of the NARX-RNN model in a controlled way by segmentation of the dataset into three subsets namely: **Training subset** used during NARX-RNN training (70% of data), **Validation subset** used for NARX-RNN validation during training and to prevent the data overfitting (15% of data), **Testing subset** not used in the training phase, only used for comparison of the models during final evaluation (15% of data). The NARX-RNN models were trained per iteration in batch mode [7], while weights and biases were initialized using the Nguyen- Widrow initialization procedure [17]. The values of the sensor readouts and the touch pressure measured by the HEX device were normalized to the range [0, 1] to avoid the early stop due to neuron saturation. The quality of the NARX-RNN model of the e-skin graphene sensor was evaluated based on the MSE performance function, and the goodness of fit between the estimated data and the reference data was calculated as Root Mean Square Error (RMSE) [11].

4.2 Modelling

The proposed NARX-RNN model was evaluated on the exemplary sensor (row 6 and column 16 of the sensor matrix). The sensor-measured characteristic is

presented in Fig. 4. The two types of nonlinear hidden layer neuron transfer functions were used: hyperbolic tangent sigmoid transfer function (*tansig*) and symmetric saturating linear transfer function (*satlins*). The RMSE values for the NARX-RNN models with the different numbers of neurons in the hidden layer before and after training are presented in Table 1.

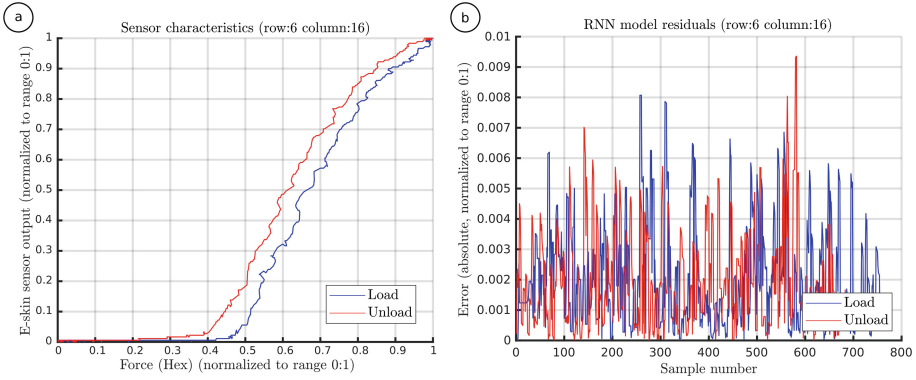


Fig. 4. (a) E-skin sensor measurement characteristic. (b) Residuals for e-skin sensor RNN model (1 neuron described with *satlins* function).

Table 1. Quality indices of the RNN models for ‘loading’ and ‘unloading’ phases.

NL function	N_{NL}	epochs	$RMSE_{load}$	$RMSE_{unload}$
<i>tansig</i>	1	0	0.4322	0.3981
	1	187	0.0028	0.0030
	2	0	0.9629	0.9559
	2	68	0.0028	0.0027
	5	0	0.5648	0.5697
	5	76	0.0028	0.0026
<i>satlins</i>	1	0	0.3582	0.3554
	1	14	0.0028	0.0030
	2	0	0.6814	0.6734
	2	17	0.0028	0.0030
	5	0	0.6354	0.6282
	5	21	0.0028	0.0026

4.3 Discussion

Obtained results, presented in Table 1, indicate that touch pressure estimation improves after training around 200–300 times. The estimation errors do not depend significantly on the number of neurons in the hidden layer, even if only one neuron is used. Moreover, the number of iteration epochs is a few times lower in the case of *satlins* function describing neurons in the hidden layer (14–21 epochs), compared with *tansig* function (68–187 epochs). The obtained quality indices are similar for the ‘loading’ and ‘unloading’ phases, thus properly modelling the hysteretic behaviour of the sensor.

5 Summary

The article presents the possibility of touch pressure estimation of the e-skin graphene pressure sensors with hysteretic behaviour using NARX recurrent neural networks. Increasing the number of neurons in the nonlinear hidden layer did not improve the generalization properties of the model. Moreover, the neurons described by the simple *satlins* function give similar results with fewer epochs than those described with *tansig*.

The presented method provides a simple alternative to the e-skin graphene pressure sensors models known in the literature. Further research should be done to extend the developed model to the sensor matrix and the layered deep neural networks for e-skin calibration and data compression.

Acknowledgements. This research was funded by Warsaw University of Technology, Faculty of Mechatronics Dean, grant number 504/04701/1141/44.000000.

References

1. Alimi, A., Assaker, I.B., Mozaryn, J., Ávila Brande, D., Castillo-Martínez, E., Chtourou, R.: Electrochemical synthesis of mno₂/nio/zno trijunction coated stainless steel substrate as a supercapacitor electrode and cyclic voltammetry behavior modeling using artificial neural network. *Int. J. Energy Res.* **46**(12), 17163–17179 (2022). <https://doi.org/10.1002/er.8380>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/er.8380>
2. Cai, M., Jiao, Z., Nie, S., Wang, C., Zou, J., Song, J.: A multifunctional electronic skin based on patterned metal films for tactile sensing with a broad linear response range. *Sci. Adv.* **7**(52) (2021). <https://doi.org/10.1126/sciadv.abl8313>
3. Dahiya, R.: E-Skin: From Humanoids to Humans [Point of View]. *Proc. IEEE* **107**(2), 247–252 (2019). <https://doi.org/10.1109/jproc.2018.2890729>
4. Dai, Y., Gao, S.: A flexible multi-functional smart skin for force, touch position, proximity, and humidity sensing for humanoid robots. *IEEE Sens. J.* **21**(23), 26355–26363 (2021). <https://doi.org/10.1109/jsen.2021.3055035>
5. Dawood, A.B., Godaba, H., Ataka, A., Althoefer, K.: Silicone-based capacitive e-skin for exteroception and proprioception. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8951–8956 (2020). <https://doi.org/10.1109/IROS45743.2020.9340945>

6. Farrokh, M., Dizaji, F., Dizaji, M.: Hysteresis identification using extended preisach neural network. *Neural Process. Lett.* 1–25 (2022)
7. Hagan, M., Demuth, H., Beale, M., Orlando, D.: *Neural Network Design*. Martin Hagan, Stillwater, OK (2014)
8. Hammock, M.L., Chortos, A., Tee, B.C.K., Tok, J.B.H., Bao, Z.: 25th anniversary article: the evolution of electronic skin (E-Skin): a brief history, design considerations, and recent progress. *Adv. Mater.* **25**(42), 5997–6038 (2013). <https://doi.org/10.1002/adma.201302240>
9. Holgado, A.C., Tomo, T.P., Somlor, S., Sugano, S.: A multimodal, adjustable sensitivity, digital 3-axis skin sensor module. *Sensors* **20**(11), 3128 (2020). <https://doi.org/10.3390/s20113128>, <http://dx.doi.org/10.3390/s20113128>
10. Hu, Y., Guo, W., Long, Y., Li, S., Xu, Z.: Physics-informed deep neural networks for simulating s-shaped steel dampers. *Comput. Struct.* **267**, 106798 (2022). <https://doi.org/10.1016/j.compstruc.2022.106798>, <https://www.sciencedirect.com/science/article/pii/S004579492200058X>
11. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**(4), 679–688 (2006)
12. Kelley, C.: Iterative methods for optimization, *SIAM front. Appl. Math.* **18** (1999)
13. Klimaszewski, J., Janczak, D., Piorun, P.: Tactile robotic skin with pressure direction detection. *Sensors* **19**(21) (2019). <https://doi.org/10.3390/s19214697>, <https://www.mdpi.com/1424-8220/19/21/4697>
14. Klimaszewski, J., Władziński, M.: Human body parts proximity measurement using distributed tactile robotic skin. *Sensors* **21**(6) (2021). <https://doi.org/10.3390/s21062138>, <https://www.mdpi.com/1424-8220/21/6/2138>
15. Klimaszewski, J., Wildner, K., Ostaszewska-Liżewska, A., Władziński, M., Możaryn, J.: Robot-based calibration procedure for graphene electronic skin. *Sensors* **22**(16) (2022). <https://doi.org/10.3390/s22166122>, <https://www.mdpi.com/1424-8220/22/16/6122>
16. Mayergoyz, I.D.: *Mathematical Models of Hysteresis and Their Applications*. Academic Press, Cambridge (2003)
17. Nguyen, D., Widrow, B.: Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In: 1990 IJCNN International Joint Conference on Neural Networks, pp. 21–26. IEEE (1990)
18. Oh, J.Y., Bao, Z.: Second skin enabled by advanced electronics. *Adv. Sci.* **6**(11), 1900186 (2019). <https://doi.org/10.1002/advs.201900186>
19. Wang, F.X., et al.: Multifunctional self-powered e-skin with tactile sensing and visual warning for detecting robot safety. *Adv. Mater. Interfaces* **7**(19), 2000536 (2020). <https://doi.org/10.1002/admi.202000536>
20. Yang, J.C., Mun, J., Kwon, S.Y., Park, S., Bao, Z., Park, S.: Electronic skin: recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics. *Adv. Mater.* **31**(48), 1904765 (2019). <https://doi.org/10.1002/adma.201904765>
21. Zhu, L., et al.: Large-area hand-covering elastomeric electronic skin sensor with distributed multifunctional sensing capability. *Adv. Intell. Syst.* **4**(1), 2100118 (2022). <https://doi.org/10.1002/aisy.202100118>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202100118>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Proximity Estimation for Electronic Skin Placed on Collaborative Robot Conductive Case

Jan Klimaszewski^(✉)  and Przemysław Białorucki

Department of Mechatronics, Institute of Automatic Control and Robotics,
Warsaw Univeristy of Technology, ul. Sw. A. Boboli 8, 02-525 Warsaw, Poland
{jan.klimaszewski, przemyslaw.bialorucki.stud}@pw.edu.pl

Abstract. One of the key issues in human-machine collaboration is human safety. Safe human-robot interaction can be implemented using an electronic skin (e-skin) that detects the human's proximity to the collaborative robot (cobot) casing even before the collision with the machine. The detection delay of such a situation should be as small as possible to be able to stop the machine safely. This paper presents an analysis of the results of estimating the proximity of a human to a robot with an electronic skin placed on its surface. The proximity estimation system works by measuring the capacitance of an open capacitor, and the placement of the capacitor on the conductive robot case significantly affects system performance. This paper outlines which parameters have the most important influence on this performance.

Keywords: Collaborative robotics · Electronic skin · Human proximity estimation · Open capacitor

1 Introduction

In recent years, an increasing number of areas can be observed that enable human-robot collaboration in the same workspace. The demand for such robotic applications is emerging in industry, education, agriculture, medical services, security and space exploration [1]. Cobots (collaborative robots) have become an essential component of Industry 4.0 [2]. Equipping cobots with additional sensors allows them to comply with special safety measures, avoid collisions with humans and manipulate objects more agilely.

In order to ensure that the effects of collisions with humans are minimised, the construction used in cobots can be lightweight and compliant, and there is a reduction in the power and strength of such machines [3]. Another approach is to equip cobots with an electronic skin (e-skin), which will allow the detection of a human approaching the robot's casing [4].

Department of Mechatronics, WUT.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 242–249, 2023.

https://doi.org/10.1007/978-3-031-37649-8_24

Detection of human proximity can be implemented in a number of ways. For this purpose, the standard approach is to measure physical properties such as the reflection of light [5], the reflection of a sound wave [6], the reflection of an electromagnetic wave in radar [7,8], as well as the detection of changes in a selected parameter in the area around the sensor, such as the electrical permeability of the sensor [9,10], the magnetic permeability of the sensor [9], or the temperature of the sensor [11].

The aim of the manuscript is to present selected aspects of the operation of a capacitive proximity estimation mechanism for cobot enclosure. Capacitive proximity estimation is implemented by measuring the operating parameters of an open capacitor, one plate of which is an e-skin [4]. An important aspect addressed in the paper is the placement of the e-skin on a conductive machine case, which significantly affects the proximity estimation.

The paper is organised as follows. In Sect. 2, the developed test stand consisting of e-skin placed on industrial robots metal case is presented. In Sect. 3, the tests results are described and analysed. Finally, Sect. 4 provides the summary and further investigation proposal.

2 Developed Test Stand

The main components of the hardware setup are the e-skin together with the measuring controller, a Fanuc M10iA industrial robot and a PC. The e-skin together with its electronic measurement system is described in detail in [4]. In brief, the e-skin is implemented as a rectangular array of graphene force sensitive resistors (FSR). The touch pressure value is measured via a conductive comb electrode layer. The electronic circuit for measuring the proximity of objects to the e-skin consists of a rectangular signal generator with a frequency dependent on the capacitance of an open capacitor formed by the conductive e-skin layer and a reference plate. The NUCLEO-F44RE board with an STM32 microcontroller clocked at 90 MHz performs the counting of the number of periods of the rectangular signal in fixed time interval T . The implementation consists of counting the occurrences of a rising edge via an external interrupt at a specified fixed time interval T . The length of this interval depends on the relevant settings of the microcontroller's internal timer such as PSC (Prescaler), CKD (Internal Clock Division) and ARR (AutoReload Register). The default length of the time interval is set to 728,178 μs . The PC performs the measurement data logging transmitted serially via the USB port. The principle of measuring the proximity of an object to the e-skin is based on the properties of an open capacitor. The capacitance of an open capacitor changes when an object, with an electrical permeability different from that of a vacuum, comes into proximity. Changes in this capacitance cause a change in the frequency of the signal generated by the electronic circuit, which is measured indirectly by the microcontroller. Placing a conductive robot enclosure in the vicinity of a system operating in this way causes significant problems with proximity estimation. A previous study [4] confirmed the effectiveness of estimating the proximity distance of a human body part to

an e-skin placed on a dielectric material. As part of the research described in the current manuscript, the e-skin was placed on an industrial robot arm (Fig. 1). Bringing the hand close to the robotic arm mimics the working conditions of a cobot in which the robotic arm approaches a human.

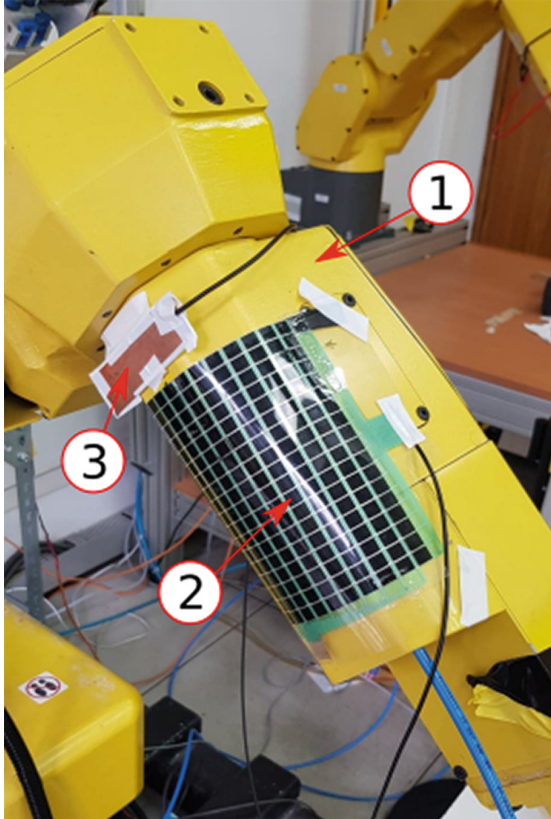


Fig. 1. Hardware setup consisting of: the Fanuc M10iA robot (1), the e-skin (2) and open capacitor reference plate (3).

3 Research Results

The tests conducted on the test stand were divided into two stages. In the first stage, the estimation of the proximity of the human hand to the e-skin at a distance of approximately 20 mm and smaller was analysed without modifications relative to the tests conducted in [4]. In the next stage, modifications were made to the measurement system and the experiment was repeated. The modifications mainly concerned the operating parameters of the microprocessor chip used.

3.1 Initial Results

The experiment consisted in approaching an open human hand oriented parallel to the surface of the e-skin, which was placed on the robot as described in Sect. 2. The size of the hand was about 10 cm × 20 cm. The approach velocity had a vector parallel to the normal surface of the e-skin, but its modulus was not recorded. The result of the measurement is the number of periods n of the signal generated by the electronic measurement system measured over a fixed time interval T . Changes in this number indicate a change in the relative electrical permeability in the area of the sensor which makes it possible to infer the detection of an object in its vicinity. During this measurement, the constant time interval during which the periods of the generated signal are counted was set to the default value and is 728.178 μs. The results obtained are shown in the Fig. 2. The graph shows an experiment involving bringing a human hand close to the e-skin device. The e-skin proximity response with the measurement system configured by default is small and relatively difficult to distinguish from measurement noise. The determination of the signal-to-noise ratio (SNR) was calculated on the basis of the formula (1).

$$SNR = \left(\frac{A_{signal}}{A_{noise}} \right)^2, \tag{1}$$

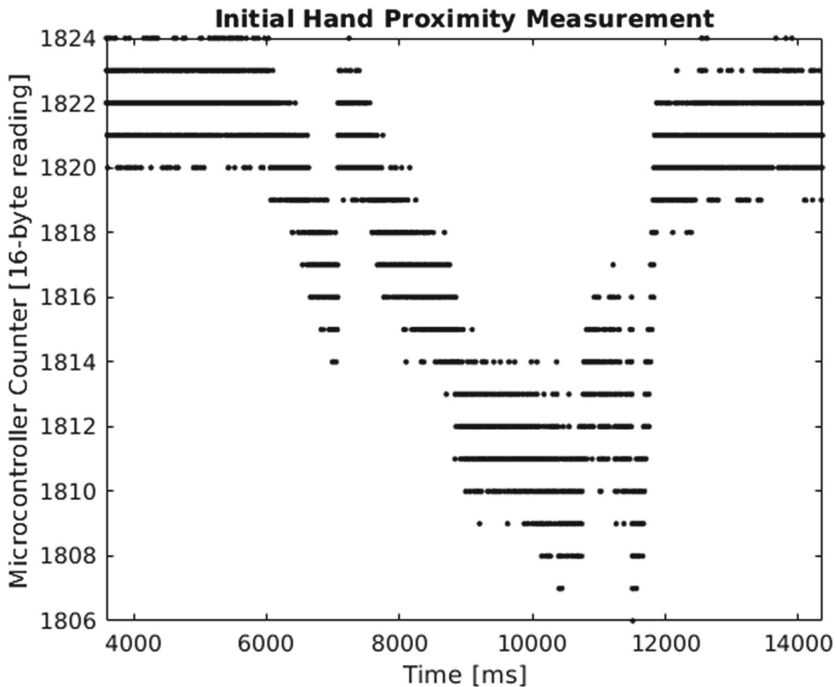


Fig. 2. Initial proximity estimation results.

where:

A_{signal} - root mean square amplitude of hand proximity signal,

A_{noise} - root mean square amplitude of noise signal.

In regards to initial results, the proximity of the hand to the e-skin A_{signal} was determined to be 10,282 while A_{noise} was determined to be 1,090. As a result, SNR was calculated as 89,047. Such a result does not allow the estimation of the proximity distance, but is sufficient to detect the presence of a hand in the vicinity of the e-skin. In order to improve the quality of the detection and allow the estimation of the proximity distance of the object, it was necessary to enhance the response of the system in relation to the noise present.

3.2 Developed Improvement

In order to improve the quality of the circuit's response, the fixed time interval T in which the n periods of the signal generated by the electronic circuit are counted was increased. This was implemented by modifying the microcontroller's internal timer settings. The value of the PSC prescaler was increased from 0 to 3 which had the effect of increasing the value of T from 718 μs to 2,872 ms. After the modifications were made, the experiment was repeated by bringing the hand close to the e-skin placed on the robot housing. The Fig. 3 shows an example of the resulting measurement system. As a result of the modifications, a more evident response of the e-skin placed on the robot housing to the proximity of a human hand was obtained. The determination of the SNR was calculated based on the formula (1). In terms of the example shown in Fig. 3 the hand proximity to e-skin A_{signal} was determined to be 66,338 while A_{noise} was determined to be 4,759. As a result, SNR was calculated as 194,335. This represents a much better result than the initial performance described in the Sect. 3.1.

3.3 Discussion

A summary of the measurement results obtained is shown in the Table 1.

Table 1. Time reactions for e-skin pressure detection.

Event	n_{noise}	n_{max}	SNR	T	PSC
Initial setup	6 ± 1	12 ± 2	89,047	718 μs	0
Improvement	6 ± 1	70 ± 2	194,335	2,872 ms	3

The relatively small response of the system to an approaching hand is due to the increased capacitance of the open capacitor resulting from the presence of the conductor under the e-skin layer. This capacitance has been measured and is approximately 400 pF. In comparison, the change in capacitance due to the approach of an object with the properties of a human hand is of the order of a few pF. The change in capacitance caused by the object being targeted

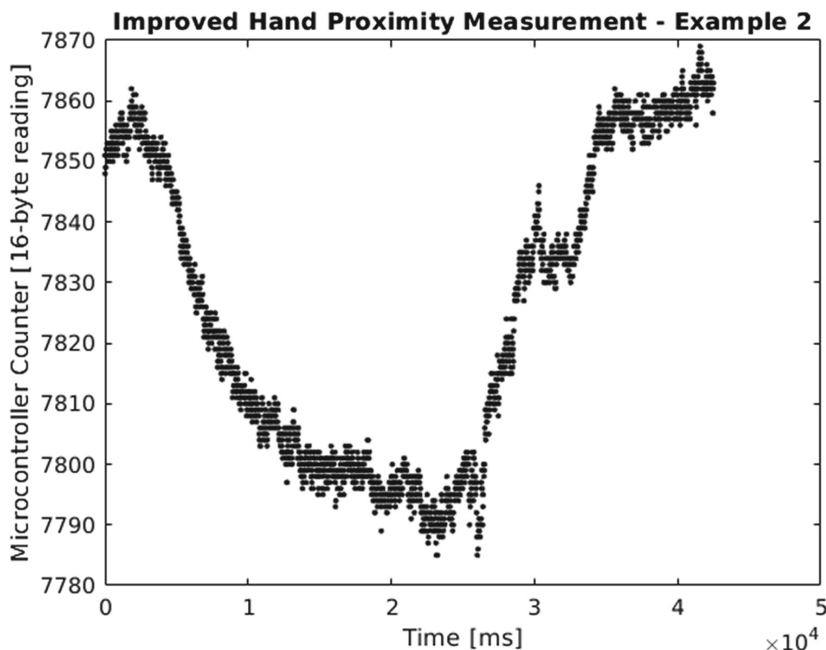


Fig. 3. Improved proximity estimation results.

for detection is two orders of magnitude smaller which results in little change in the frequency generated by the electronic circuit. In spite of these problems, an efficient detection of the proximity of the human hand to the robot was successfully achieved.

The negative effect of improving the quality of the response of the e-skin placed on the metal housing of the robot to the approach of a human hand is an increase in the response time of the system to an approaching object. In the target application for safe human-robot collaboration, the measurement delay is a key determinant of whether the collaborative robot will have time to stop and avoid a dangerous collision. At the example TCP speed of an industrial robot of 4,000 mm/s, increasing the time T from 718 μs to 2,872 ms results, in the extreme case, in a robot response delayed by 2,154 ms which, at the above tool speed, translates into an increase in the potential braking distance of 8,616 mm. This extension of the braking distance is significant given the reaction distance of the e-skin to the human hand of approximately 20 mm. The increase in response time also results in a significant reduction in the potential for de-noising filtering of the measurement results. When applied to, for example, eliminate outliers, this will result in a further increase in braking distance.

4 Summary

Using e-skin for capacitance-based proximity estimation when placed near conductors poses problems. They are related to the occurrence of a much higher

capacitance of the open capacitor resulting from the presence of a conductor in the sensor's working area than from changes in capacitance resulting from the proximity of objects. This problem is very important, taking into account the target application of e-skin as a device supporting safe robot-human interaction - the housing of robots is often made of conductive materials. In the manuscript, in order to minimise the problem at hand, the time interval during which the signal periods generated by the electronic system are counted was increased. This improved the response of the system and increased the SNR of the measurement system. The negative cost of obtaining such results is the increased response time of the robotic system. Further work on this topic may allow a more precise and robust estimation of the proximity distance of the human to the e-skin.

References

1. Bloss, R.E.: Collaborative robots are rapidly providing major improvements in productivity, safety, programming ease, portability and cost while addressing many new applications. *Ind. Robot.* **43**, 463–468 (2016)
2. Sherwani, F., Asad, M.M., Ibrahim, B.: Collaborative robots and industrial revolution 4.0 (IR 4.0). In: 2020 International Conference on Emerging Trends in Smart Technologies (ICETST), pp. 1–5. IEEE (2020)
3. Svarny, P., Tesar, M., Behrens, J.K., Hoffmann, M.: Safe physical HRI: toward a unified treatment of speed and separation monitoring together with power and force limiting. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7580–7587. IEEE (2019)
4. Klimaszewski, J., Władziński, M.: Human body parts proximity measurement using distributed tactile robotic skin. *Sensors* **21**(6) (2021). <https://www.mdpi.com/1424-8220/21/6/2138>
5. Cheng, G., Dean-Leon, E., Bergner, F., Rogelio Guadarrama Olvera, J., Leboutet, Q., Mittendorfer, P.: A comprehensive realization of robot skin: Sensors, sensing, control, and applications. *Proc. IEEE* **107**(10), 2034–2051 (2019)
6. Paulet, M.V., Salceanu, A., Neacsu, O.M.: Ultrasonic radar. In: 2016 International Conference and Exposition on Electrical and Power Engineering (EPE), pp. 551–554 (2016)
7. Geiger, M., Waldschmidt, C.: 160-GHz radar proximity sensor with distributed and flexible antennas for collaborative robots. *IEEE Access* **7**, 14977–14984 (2019)
8. Matheoud, A.V., Sahin Solmaz, N., Frehner, L., Boero, G.: Microwave inductive proximity sensors with sub-pm/Hz^{1/2} resolution. *Sens. Actuat. A: Phys.* **295**, 259–265 (2019). <http://www.sciencedirect.com/science/article/pii/S092442471831567X>
9. Purcaru, D., Gordan, I.M., Purcaru, A.: Study, testing and application of proximity sensors for experimental training on measurement systems. In: 2017 18th International Carpathian Control Conference (ICCC), pp. 263–266 (2017)
10. Tsuji, S., Kohama, T.: Tactile and proximity sensor using self-capacitance measurement on curved surface. In: 2017 IEEE International Conference on Industrial Technology (ICIT), pp. 934–937 (2017)
11. Mikita, H., Azuma, H., Kakiuchi, Y., Okada, K., Inaba, M.: Interactive symbol generation of task planning for daily assistive robot. In: 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012), pp. 698–703 (2012)



Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Finite Element Method Based Toolchain for Simulation of Proximity Estimation Using Electronic Skin

Anna Ostaszewska-Lizewska¹  and Jan Klimaszewski² 

¹ Department of Mechatronics, Institute of Metrology and Biomedical Engineering,
Warsaw Univeristy of Technology, ul. Sw. A. Boboli 8, 02-525 Warsaw, Poland
anna.ostaszewska@pw.edu.pl

² Department of Mechatronics, Institute of Automatic Control and Robotics,
Warsaw Univeristy of Technology, ul. Sw. A. Boboli 8, 02-525 Warsaw, Poland
jan.klimaszewski@pw.edu.pl

Abstract. The emergence of new areas of human-robot cooperation creates the need to ensure human safety in this regard. Therefore, there is a need to develop new sensors to detect the presence of a human in the vicinity of a robot. One such sensor is an electronic skin (e-skin). Manufacturing and testing new e-skin prototypes is labor-intensive. This paper presents a software toolchain developed to simulate the operation of an e-skin used to detect human proximity. The toolchain is based on the finite element method and has been developed exclusively with free and open-source software. The presented toolchain makes it possible to test e-skin modifications without the need for a physical prototype and significantly reduces implementation costs. The developed solution is multi-platform and allows parallel and multi-threaded calculations conducted on multiple machines simultaneously. This paper presents modeling results obtained for a simplified e-skin sensor, which are consistent with experimental results on the actual model.

Keywords: Finite element modelling · Sensor prototyping · Sensor optimization · Electronic skin

1 Introduction

We are witnessing a revolution - robots, formerly reserved for industry and the professionals who work in it, are occupying more and more extensive areas of public space in, for example, medical services or education. The work of robots cooperating in spaces shared with humans should be supervised by various types of safety mechanisms and sensors. [1]. There is a number of concepts how to avoid collision. One of them is electronic skin (e-skin) [2], which was originally intended to mimic the sense of touch by a biological organ and was used in grippers and

Department of Mechatronics, WUT.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 250–259, 2023.

https://doi.org/10.1007/978-3-031-37649-8_25

end-effector mechanisms. Current research on e-skin can be divided into those aimed at providing reception of the same stimuli that are received by humans (e.g., temperature [3]), or extending the range of measurements to include those that are not received by human skin (e.g., magnetic fields [4]). At the Faculty of Mechatronics of the Warsaw University of Technology, research is being carried out on the production of e-skin, which, in addition to measuring the pressure force [5], also detects objects at a certain distance [6]. The time, cost and amount of work required to produce and test successive e-skin prototypes introducing improvements and new ideas is significant.

The aim of this paper is to describe an open-source software toolchain developed to simulate the operation of an e-skin electronic printed circuit using the finite element method (FEM). This will enable, among others, the testing of subsequent prototypes of the device without the need for a physical prototype. The main objective of this task is to improve the quality and development of estimating the proximity of human body parts based on e-skin. To this end, the mathematical modeling results developed in the paper are verified by laboratory test results.

The paper is organised as follows. In Sect. 2, the prototype of e-skin is presented. In Sect. 3, finite element method and software toolchain is presented as well as the model of e-skin and the results obtained. Section 4 displays the effect of modelling results comparison with measurements. Finally, Sect. 5 provides the summary and further investigation proposal.

2 Prototype of an E-Skin

The prototype e-skin model used to measure the proximity of human body parts is described in [6]. E-skin consists of two layers of flexible foil 1. One has comb electrodes made of conductive material. On the second layer, resistive touch fields are made based on graphene. Touch estimation is based on measuring the resistance of the variable touch fields in relation to the touch pressure exerted on them. In [6] the principle of operation of the approximation estimation was implemented only with the conductive layer acting as one of the plates of an open capacitor. The second cover is a copper reference plate. The capacity of this setup changes when there is an object nearby with a different permittivity than the surrounding air. This change is measured using a custom electronic driver based on a rectangular signal generator.

In order to confirm the quality of the developed toolchain for modelling the functionality of estimating the object's proximity to the e-skin, a simplified model was proposed. A simplified e-skin model was built of two $110 \times 140 \times 1.6$ mm plates placed on a plane. The plates, together with the surrounding air as a dielectric, constituted an open capacitor, similar to that described in [6].

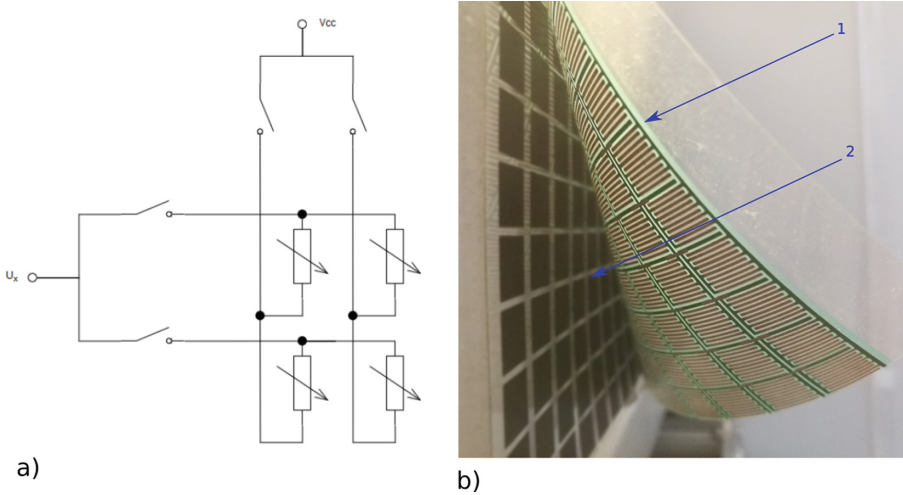


Fig. 1. a) Design principle for FSR matrix of size 2×2 [5]; b) the e-skin [6]: 1 - a conductive layer of comb electrodes printed on plastic foil, 2 - FSR sensors arranged in a rectangular pattern placed on a plastic foil.

3 Finite Element Method for Simplified E-Skin Modelling

Finite element method (FEM) [7–9], along with finite difference method (FDM) and finite-volume method (FVM), is one of numerical methods for solving differential equations numerically, which were originally defined on meshes of data points. The idea of mesh is to split the computational domain into a number of simple geometric elements, usually triangles (for 2D) or tetrahedral (for 3D), and to approximate the solution function over each element with the weighted residual concept. Finally the global solution is obtained by combining the individual solutions. The origins of FEM date back to the early 1960s, and its first application was the analysis of aircraft structures. FEM has been rapidly developed over the years and nowadays it is considered to be the most flexible method [10]. It can be used for a wide range of numerical problems in hydraulic, mechanical, aeronautical [11] and electrical engineering [12] for solving problems involving irregular geometries and steep gradients.

3.1 Toolchain

Due to the huge popularity of FEM, the market offers many packages for FEM analysis. They differ in terms of their features, terms of operation, functionality and areas of application. In case of this project, the most important was the division of tools into commercial and open-source. Commercial packages enable completing the entire project within one environment, but the cost of the license is usually high. When deciding to use free solutions, it is often necessary to select separate software for the implementation of each of the individual modelling

stages: defining the geometry of the system, creating a mesh, performing the FEM analysis and graphical visualisation of the results. Figure 2a) shows the developed toolchain which consists of OCTAVE¹, NETGEN², ELMERFEM³ and PARAVIEW⁴.

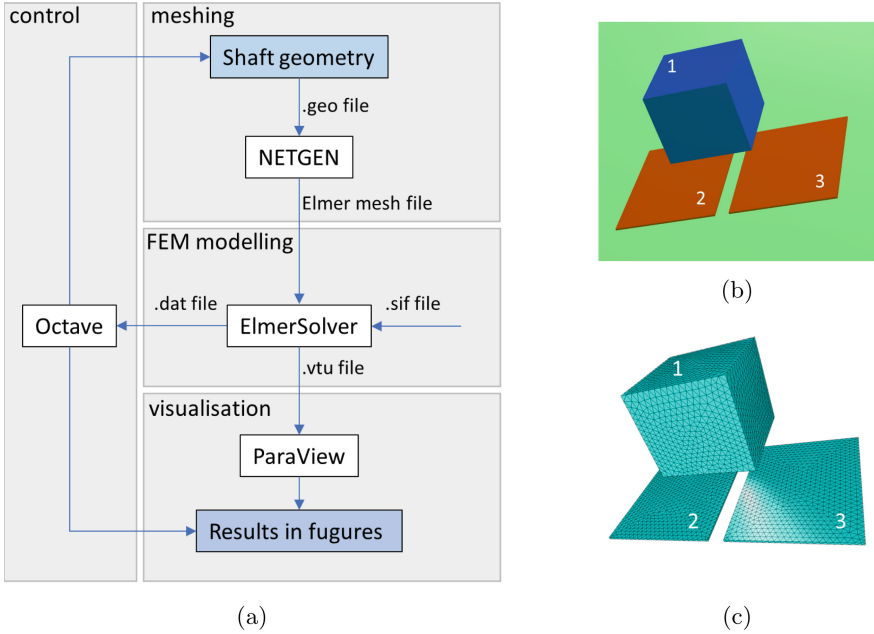


Fig. 2. Toolchain (a) for sensor FEM modelling with open-source software; e-skin mathematical model, 1-water-box, 2,3-capacitor plates: (b) geometry (NETGEN), (c) mesh (ELMER).

NETGEN is an automatic 3D tetrahedral mesh generator that contains modules for mesh optimisation and hierarchical mesh refinement [13].

ELMER FEM is a finite element software for the numerical solution of partial differential equations and multiphysical problems [14], enabling the modeling of mechanical, thermal, electrical, and magnetic systems.

ParaView is a data visualization application that enables qualitative and quantitative analysis techniques [15]. The data obtained were explored interactively in 3D, but it is also possible to use batch processing.

OCTAVE is a high-level programming language developed under the GNU Project for scientific computing and numerical computation. It may be also used

¹ <https://octave.org/>.
² <https://ngsolve.org/>.
³ <http://www.elmerfem.org/blog/>.
⁴ paraview.org.

for batch processing and in presented toolchain it was used to automate a series of calculations.

All programs were run in the LINUX MINT 19.1 operating system. Although each of them can also be used in Windows, they have been optimised for LINUX.

Data flow between applications was carried out using text files. This solution allowed for the validation of the results at each stage of modelling and batch processing. The automation of the calculations was carried out by OCTAVE, which generated batch .geo files for consecutive variants of the system's geometry, ran meshing in NETGEN, and then FEM computation in ELMER FEM. Finally, the modelling results were processed in the OCTAVE environment to create graphs and in PARAVIEW, which was used to visualise the distribution of the magnetic field.

3.2 Modelling

The effect of mesh generated from GEO geometry import is displayed in Fig. 2 b). The model is set up of two capacitor plates of dimensions $110 \times 140 \times 1.6$ mm placed on a plane, a $80 \times 80 \times 80$ mm cuboid of water as an object approaching an open capacitor. Those three elements were placed inside a sphere of atmospheric air with a diameter of 500 mm. The next variants of the modelled system differed in the distance between the water box and the plane of the capacitor plates. This distance varied from 1 to 50 mm with an increment of 1 mm.

Each of the 50 .geo models was meshed in the NETGEN batch processing environment. For the project described, Delaunay tetrahedral meshing modules [16] were used. An example of the meshing result is presented in 2c). The maximum height of a tetrahedral element was set up to 0.005 m, hence each final mesh consisted of 25553 points, 138243 elements, which established 21648 surface elements. Then each of the meshes was saved in a format dedicated for ELMER FEM. The ELMER FEM environment provides an opportunity to solve Maxwell's Eqs. (1-4) that govern the macroscopic electromagnetic theory.

$$\nabla \cdot \vec{D} = \rho \quad (1)$$

$$\nabla \cdot \vec{B} = 0 \quad (2)$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (3)$$

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t} \quad (4)$$

where: ∇ denotes the three-dimensional gradient operator, \vec{D} is the electric flux density, \vec{B} is the magnetic flux density and \vec{E} is the electric field strength, and \vec{H} is the magnetic field strength.

For linear materials the fields and fluxes are simply related with Eqs. (5) and (6).

$$\vec{B} = \mu \vec{H} \quad (5)$$

$$\vec{D} = \epsilon \vec{E} \quad (6)$$

where the permittivity $\epsilon = \epsilon_0 \epsilon_r$ is defined through the permittivity of vacuum ϵ_0 and the relative permittivity of the material ϵ_r . In the modelled stationary case the electric field may be expressed with a help of an electric scalar potential ϕ according to Eq. (7).

$$\vec{E} = -\nabla \phi \quad (7)$$

Assuming linear material law and using the Eq. (1) gives Eq. (8)

$$-\nabla \cdot \epsilon \nabla \phi = \rho \quad (8)$$

where ρ is the charge density. The energy density of the field is expressed by Eq. (9).

$$e = \frac{1}{2} \vec{E} \cdot \vec{D} = \frac{1}{2} \epsilon (\nabla \phi)^2 \quad (9)$$

Thus the total energy of the field may be computed from Eq. (10).

$$E = \frac{1}{2} \int_{\Omega} \epsilon (\nabla \phi)^2 d\Omega \quad (10)$$

where Ω is any volume with closed boundary surface. As there is only one potential difference ϕ present then the capacitance C may be computed from (11).

$$C = \frac{2E}{\phi^2} \quad (11)$$

The StatElecSolve module by Leila Puska, Antti Pursula, Peter Råback was used for this purpose. This module enables for calculating the electrostatic potential in a linear dielectric material and a conductive medium. Depending on the potential, different domain variables can be calculated as well as physical parameters such as capacitance. The mesh definition together with the .sif file defining the used constants, solvers and simulation boundary conditions constituted batch files for individual models calculated in ElmerSolver which is the module of ELMER FEM. In this project solvers using conjugate gradient-oriented algorithms were used. For electric potential Dirichlet boundary conditions were used to define the value of the potential on boundaries of capacitor plates (0 V and 3 V). The simulation results were saved both in .dat files (containing individual values of electric capacity of the system) and in .vtu files dedicated to PARAVIEW 3.

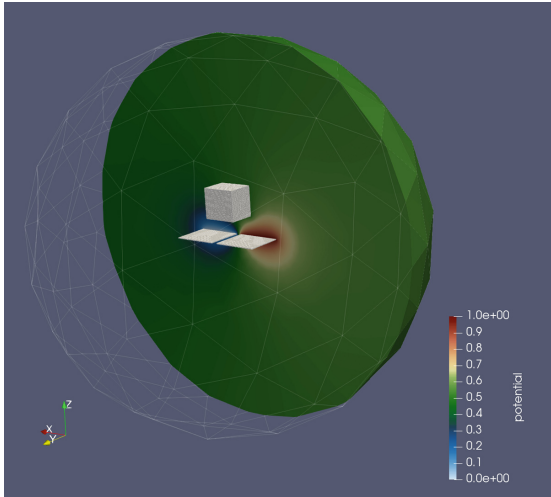


Fig. 3. E-skin modelling visualisation (PARAVIEW) - the electric potential distribution.

4 Tests

The section contains a description of the tests performed to verify the correctness of the simulations' results. For this purpose, a robotic workstation was developed using the Fanuc M10iA industrial robot. It goes on to describe the comparison of the results obtained by simulation and measurement.

4.1 Experimental Setup

For initial tests, a plastic $80 \times 80 \times 80$ mm thin-walled box filled with water was proposed as an object approaching an open capacitor. The test object was suspended over a simplified e-skin model by attaching it with thin nylon lines to the robot's handle. This allowed precise positioning of the object's distance from the e-skin and eliminated the influence of robotic components on the measurement. Preliminary measurements were used to determine the number n of periods of the signal generated by the e-skin electronic driver depending on the distance of the test object from the e-skin. A dedicated electronic circuit described in [6] was used to determine this value. The circuit indirectly measures the frequency f depending on the capacitance of the open capacitor based on a number n counted at a fixed time T of $720 \mu\text{s}$. As a result, each measurement point of the Fig. 4a) plot was determined from about 10,000 of recorded measurements of n values.

4.2 Comparison of the Measurement Results with the Simulation Model

The simplified e-skin model was modelled and simulated using the toolchain described in Sect. 3. As a result of the simulation, the model presented the capac-

itance of an open capacitor as a function of the distance of a container of water near it. In order to compare the measurement results with the simulation results, it was necessary to calculate the capacitance of the open capacitor based on the measured value of n . The relationship is described by the formula (12).

$$C(d) = \frac{k}{R} \cdot \frac{1}{f(d)}, \tag{12}$$

where: C - capacity [pF] of the open capacitor, d - distance [mm] to the object being approached, $f = n/T$ - frequency [1/s] of the signal generated by the electronic system, k - experimental constant coefficient characterising the system, R - resistance characterising the system, $k/R = 0,00001$ [pF/s].

The measurement results had the same character of changes as those obtained in simulations, but the dynamics of their changes differed significantly. Therefore, a linear function was developed to transform the measurements of the physical system to match the simulation results. The transformation was made according to formula (13).

$$C_t(d) = a \cdot (C(d) + b) + c, \tag{13}$$

where: $a = 50.0, b = -3.0, c = -7.5$ - constant transformation coefficients; a is unitless; b and c in [pF].

Finally, a comparison of the simulation results obtained and the transformed measurement data is shown in Fig. 4b).

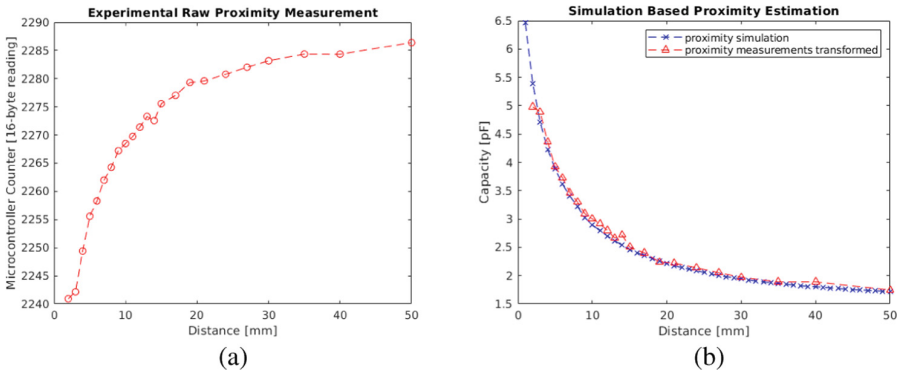


Fig. 4. (a) Test stand proximity measurements; (b) proximity simulation verification.

5 Summary

The mathematical modelling proposal presented in the article reduces the costs of creating prototypes and testing them in the laboratory. This enables faster acquisition of knowledge and reduced material consumption for producing prototypes for research. The developed toolchain allows for the optimisation of sensors

used in robotics (and not only) without the need to incur the costs of software licenses for FEM. Although all the described software is free, there is no technical limitation to using it for more complicated geometry and for many more solvers for modelling other physical phenomena for the same project. ELMER FEM enables parallel computation - it historically used message passing but now is developed towards multithreading using OpenMP. ELMER FEM can be used on multiple cores, and it also supports several ways of geometry partitioning. The only inconvenience of using the described environment is the need to master several tools and transfer files between them. However, using four different applications makes it possible to optimise the workflow through parallel data processing on hardware with adequate computing power and an operating system that works better with a particular application. As the developed toolchain is free, it significantly reduces the optimisation and implementation costs of the developed solutions. The software used to build the toolchain is also open source, which allows for code validation and modification required by the newly developed materials. Commercial solutions do not offer such possibilities.

References

1. Li, G., Liu, S., Mao, Q., Zhu, R.: Multifunctional electronic skins enable robots to safely and dexterously interact with human. *Adv. Sci.* **9**(11), 2104969 (2022)
2. Dahiya, R.: E-Skin: from humanoids to humans [Point of View]. *Proc. IEEE* **107**(2), 247–252 (2019)
3. Fastier-Wooler, J.W., Dau, V.T., Dinh, T., Tran, C.D., Dao, D.V.: Pressure and temperature sensitive e-skin for in situ robotic applications. *Mater. Des.* **208**, 109886 (2021). <https://www.sciencedirect.com/science/article/pii/S0264127521004391>
4. Liu, S., et al.: Highly flexible multilayered e-skins for thermal-magnetic-mechanical triple sensors and intelligent grippers. *ACS Appl. Mater. Interfaces* **12**(13), 5675–15685 (2020). <https://doi.org/10.1021/acsami.9b23547>, PMID: 32134626
5. Klimaszewski, J., Janczak, D., Piorun, P.: Tactile robotic skin with pressure direction detection. *Sensors* **19**(21), 4697 (2019). <https://doi.org/10.3390/s19214697>
6. Klimaszewski, J., Władziński, M.: Human body parts proximity measurement using distributed tactile robotic skin. *Sensors* **21**(6) (2021). <https://www.mdpi.com/1424-8220/21/6/2138>
7. Finlayson, B.: *The Method of Weighted Residuals and Variational Principles: With Application in Fluid Mechanics, Heat and Mass Transfer*. Educational Psychology, Academic Press, Cambridge (1972). <https://books.google.pl/books?id=KHHVNESp5UoC>
8. Reddy, J.: *An Introduction To The Finite Element Method*. McGraw-Hill series in mechanical engineering, Tmh, New York (1984). <https://books.google.pl/books?id=DrsXnwEACAAJ>
9. Zienkiewicz, O.C., Taylor, R.L., Zhu, J.Z.: *The Finite Element Method: Its Basis and Fundamentals*. Elsevier, Amsterdam (2005)
10. Rapp, B.E.: Engineering mathematics. In: *Microfluidics: Modelling, Mechanics and Mathematics*, pp. 21–50. Elsevier (2017)
11. Zhu: The finite element method. Standards Information Network, March 2018

12. Silvester, P.P., Ferrari, R.L.: Finite elements in one dimension. In: Finite Elements for Electrical Engineers, pp. 1–27. Cambridge University Press, Cambridge, September 1996
13. Schöberl, J.: NETGEN an advancing front 2D/3D-mesh generator based on abstract rules. *Comput. Vis. Sci.* **1**(1), 41–52 (1997)
14. Råback, P., Forsström, P.L., Lyly, M., Gröhn, M.: Elmer-finite element package for the solution of partial differential equations. In: EGEE User Forum. CSC Espoo, Finland (2007)
15. Ahrens, J., Geveci, B., Law, C.: Paraview: an end-user tool for large data visualization. *Vis. Handb.* **717**(8) (2005)
16. Cignoni, P., Montani, C., Scopigno, R.: Dewart: a fast divide and conquer Delaunay triangulation algorithm in ed. *Comput.-Aided Des.* **30**(5), 333–341 (1998)


Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Collaborative Robotics. Safety and Ethical Considerations

Monika Róžańska-Walczuk^(✉) 

Department of Mechatronics, Institute of Automatic Control and Robotics,
Warsaw University of Technology, ul. Św. A. Boboli 8, 02-525 Warsaw, Poland
monika.rozanska@pw.edu.pl

Abstract. Every year, collaborative robots get closer to humans and cooperation with them takes place not only in industrial spaces, where specialized employees work with them, but also people who do not have knowledge in the field of engineering and robotics. Therefore, great attention is paid to safety in the cooperation of robots and humans. In addition, the aspect of ethics and their ethical behavior towards a human co-worker, companion or petitioner is more and more often taken into account. Knowledge of potential safety hazards is important to secure safety early in robots' design and development process. Therefore security is one of main issues raised in the article. The most important safety standards from the point of view of collaborative robotics are presented. In the article described example of cobots acting increasingly role as members of our society. Access to them is becoming more and more common - they are household members, waiters or airport staff. Presented in the paper issue of ethics in reference to robots and AI are becoming increasingly significant impact on human. It deals with topics of physical and ethical safety in cooperation between humans and robots. Reference has been made to the safety standards. Due to proximity of technology in humans lives, access to them, and even dependence on them, that issue was particularly emphasized by the author. The paper is source of references to considerations of human safety in robotized environments and ethics in robotics applications.

Keywords: Collaborative robotics · Safety · Ethics · Robot Safety Standards

1 Introduction

Robotics has been playing an important role in industry for over 40 years. As statistics show, the development in the robotics industry is still very dynamic [14]. Information from [15] reports on the market situation after the pandemic indicate that the pandemic period, although initially associated with the suspension or postponement of some investments, was overall equally beneficial for

Department of Mechatronics, WUT.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 260–269, 2023.

https://doi.org/10.1007/978-3-031-37649-8_26

the industrial, logistics and other sectors. The interest in production automation technologies, in particular digital solutions, has increased, and openness to innovation has increased.

Increasingly, robots are used in space exploration, education, healthcare, intralogistics, and agriculture [3]. In recent years, the distance between humans and robots has narrowed. Positive changes in the perception of robots are noticeable in the research carried out on the acceptance of human-robot cooperation [12]. There is a noticeable change in attitude compared to the results of the [13, 18] study conducted a few years ago on behaviour, emotions and attitudes towards robots. The research, described in the article [17], showed that assigning human characteristics to robots causes negative feelings due to a strong belief in the uniqueness of human nature. The negative attitude towards interaction with robots indicated the lack of acceptance and readiness for technological changes in society.

Collaborative robots (so-called *Cobots*) (Fig. 1) are a relatively new type of robots, more and more commonly used in production plants, laboratories, warehouses, etc. Robots also play the role of waiters, couriers and guides. The cooperating robots differ in terms of construction, weight, and tools with which they are equipped. This is due to the increased emphasis on safety issues [12] that must be ensured when the human and robot are in the same workspace. For this reason, it is so important to change people's approach to collaborating with a robot, as the goal is to share a workspace. A robot that works safely between (or with) people can improve product flow and increase production by automating new spaces and processes. The combination is the most reliable and effective combination of a robot and a human.



Fig. 1. Cobots

A great deal is currently being said about collaborative robots that work hand in hand with humans without any additional protective measures. The robots were designed to perform heavy tasks and it was definitely dangerous to enter the working area of a robot while it was manipulating a heavy element. Trends in robotization have been changing for several years (Fig. 2). The significant difference in the way of operation between cobots and traditional robots is speed they work with. Whereas traditional robots are designed to operate at a high level of speed and accuracy, cobots are intended to be safe. Moreover they are easy programmable and use. Due to high-speed work of the industrial robot arm, they are isolated from the workspace when operating to protect workers from dangerous, fast moving parts (Fig. 2 on the right). Collaborative robots are designed to work alongside people without barriers or fences, what is presented on the Fig. 2 (on the left). The cobots are able to immobilize themselves with the slightest touch preventing injury or any danger to nearby people, thanks to built-in sophisticated sensors. They assist human co-workers, accelerate tasks or assume monotonous and tedious work, leaving more complicated tasks to the human workers. Simultaneously, they are much easier to set up. Industrial robots are proper for large companies that production is standardized and repeatable. Smaller companies can benefit from the flexibility and cost-effectiveness of cobots.

Robot manufacturers set new safety standards that allow for unprotected collaboration. There is a very low level of tolerable risk which is considered acceptable when the robot is in the same environment as the worker.

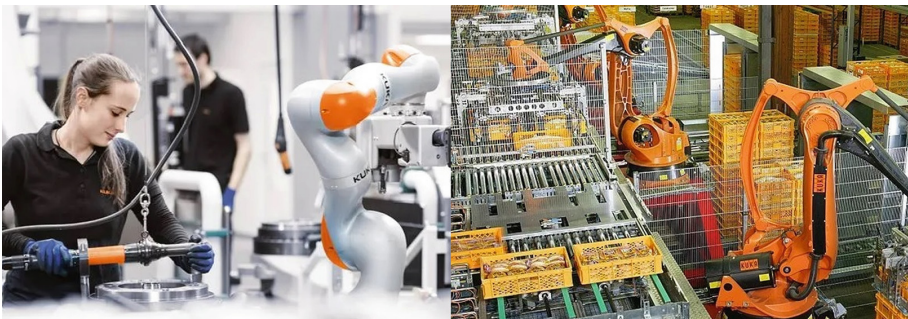


Fig. 2. Cobots vs. traditional Industrial Robots [source: <https://www.kuka.com>]

The paper aims to discuss the topics of physical and ethical safety in cooperation between humans and robots. In the dynamically developing field of robotics, the issue of safety is of particular importance. The observed trend is to construct robots that will be safe for humans. Due to the fact that cobots are intended to cooperate in an environment where accidental contact with it may occur. The essential is exploration of the potential hazards of co-operation while robots interact with humans closely. On the other hand, looking at the potential of AI and robotics, new technologies should be implemented in a sustainable and socially ethical way.

The paper is organised as follows. In Sect. 2, safety collaborative robots and human, collected information about safety standards in this area, is presented. In Sect. 3, ethical issues are described in reference to the first British Ethical Standard. Finally, Sect. 4 provides the summary and further development consideration.

2 Safety

Due to the rapid development of collaborative robots on the market, their increased market share (in production spaces and at trade fairs), their presence is no longer unusual. The features that distinguish them from traditional robots are:

- a light structure,
- compact design,
- rounded edges,
- hidden wiring,
- different tooling than in case of traditional robots,
- different functionality.

Collaborative robots are equipped with a number of solutions that allow for the detection and response to a collision with operators and elements of the environment [3]. They move slower than traditional machines, which is also due to safety reasons. The possibility of easier programming is also a big difference. This opens up new possibilities for creating completely new scenarios of robotization applications, as well as increasing production efficiency. In the era of the employee market or due to other unexpected situations, such as a pandemic, cobots are more conducive to maintaining stability and continuing work in workplaces. This is a strong argument for introducing robotic solutions. The goal is to develop companies and change the tasks performed by the production staff, rather than replacing them.

2.1 Safety Standards

Safety standards are developed to protect personnel from the risks associated with the nature of their work and their place of employment. They are formulated in such a way as to impose minimal restrictions or interference with the level of services provided.

The selection of protective measures is related to all elements of the robotic system, i.e. the type and specificity of the robot itself, connection with other machines, equipment with which the robot and machines are provided [1]. Proper selection of the system components is essential for the work to be performed correctly, effectively and for the necessary changes to be made.

The risk assessment of a robot and robotic application is based on the assumptions set out in the PN-EN ISO 12100 [1] standard. Due to the different nature of the hazards associated with various applications of industrial robots, the ISO 10218 standard has been divided into two parts [5,6].

2.2 Collaborative Robot

It is not efficient for the production process when a robot inside a cell fenced off from a human being during its active work, has to be stopped in the event that the operation requires human participation. This is the case with standard robotic applications. Cobots enable the use in industry of repetitive and efficient work of machines with individual skills and experience of the operator. This is important because people are able to approach the solution of the problem and task in a non-standard way, while cobots show high endurance, precision and strength at work. In other, non-industrial solutions, robots can even approach humans directly, being a waiter or a nurse. The common feature of all cobots is their safety in contact with humans.

The challenge for cobot applications is to work without protective fences. The boundaries between the workspace of people and cooperating machines are completely or significantly limited. An additional disadvantage that must be taken into account is the unpredictable human movements. It is extremely difficult in the context of calculations related to speed, reflexes or unexpected entry into the work area of additional people. Standards for industrial robots ISO 10218-1 [5] and ISO 10218-2 [6] have been in force since 2011. The technical specification ISO/TS 15066 [8] was created to supplement and regulate the safety requirements of cooperating systems robots and their work environment. It is the world's first technical specification that focuses on the safety aspects of human cooperation with collaborative robots. It provides detailed guidelines and tips for conducting a risk analysis for collaborative robot applications. The idea of allowing a cobot to come into direct contact with a human when no pain or injury is caused, prompted the creation of a new technical standard. The ISO/TS 15066 [8] standard defines for the first time the limits of speed and power allowed during robot-human cooperation and the risk assessment of the application.

Research on the e-skin prototype [9, 10] could open up new possibilities at the human machine interface (*HMI*) level. The use of e-skin will increase the security of cooperation and improve the interface by collecting data, which will be processed into information about the type of contact. Studying the topic may help to read the emotions of people working with the robot [19].

2.3 Intralogistic Robotics

Cobots, or collaborative robots, are a large group of AGV (*Automated Guided Vehicles*) products. Analysing the development of robotics and safety in human-robot collaboration, we should also mention ARM-class mobile robots (*Autonomous mobile robots*). They are another type of AGV robots. Their main task is to transport products using advanced technology and are equipped with artificial intelligence [20].

Workers' concerns about AGV are similar to those about cobots. Even more visible and emphasised are the users' concerns about security in this case. AGVs

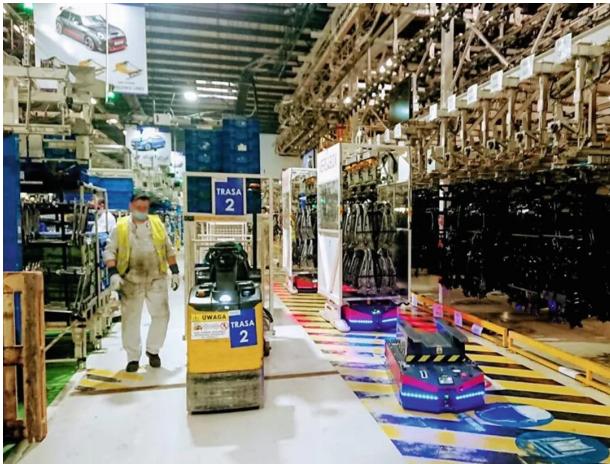


Fig. 3. AMR, VersaBox, [source: <https://versabox.eu>]

are equipped with sensors and other security devices. The security system characteristic of AMR class robots is best defined in the harmonised standards PN-EN 1525 [16], ISO 3691-4 [7]. The regulations mainly concern the speed of movement of the AMR robot (Fig. 3): and the required security systems with which it should be equipped. The specific requirements are related to the two spaces in which the robot can work.

3 Ethics

In industry, robots commonly build, arrange, rearrange, transport, pack, and inspect things. For a long time, they have also been a visible support in medicine, e.g. they perform surgical procedures and dispense prescription drugs in pharmacies. Both medical robots and social robots establishing contacts with people evoke emotions and build relationships with the user. Referring to the [17,18] experiment on attitudes towards robots using the NARS scale, it can be concluded that a positive or negative attitude towards robots can be manifested through emotions, evaluations and reactions, and these have an impact on human well-being. These potential ethical threats are recognised as having a stronger and deeper impact than physical threats, so it is important to consider various ethical harms and countermeasures.

Like most robots, social robots, cobots, AMR use artificial intelligence to decide what movement to make in response to information received through cameras and other sensors. The ability to react in a way that seems close to human behavior has been trained through research. It forms a perception that is social and emotional intelligence. It studies how people can read thoughts, feelings and even touch [9,10].

3.1 First Ethic Standard

The British Standards Institution (BSI) has published the world's first ethical standard for the design, production, sale and use of social robots.

Robots and robotic devices are increasingly used in industrial and non-industrial environments. The psychological factor is also taken into account as it takes into account how they affect the people with whom they share space and tasks. In addition to speech recognition, the use of vision systems to recognise emotions, new methods are also used. The distance between human and robot functioning is shortened (Fig. 4).



Fig. 4. RCH patient, Miles, working with NAO [source: Alvin Aquino/RCH]

BS 8611 [2] provides guidance on identifying the potential ethical harm of the growing number of robots and autonomous systems used in everyday life. The standard also provides additional guidance to eliminate or reduce the risks associated with these ethical risks to an acceptable level. The British Standard addresses ethical issues in social, application, commercial/financial and environmental terms. It includes four pages of examples of ethical risks, related ethical risks and mitigation measures (e.g. Table 1). Helpful comments are also included, with examples of how mitigation and reduction of negative impacts can be verified. The standard collects the requirements and guidelines for the design of the robot, the protective measures used for protection and the method of developing clear information for the user.

In [2] section of ethical hazard identification reference is made to groups of humans or animals that are likely to be affected by a new robot or application, despite the fact the definitions in clause do not refer specifically to animals. In subsequent subsection other standards for risk assessment are also referred to namely BS EN ISO 12100:2010 for machines, and BS EN ISO 14971 for medical devices. The conclusion of the risk assessment, BS 8611 stated that *'As a general principal, the ethical risk of a robot should not be higher than the risk of a human operator performing the same action.'*

Issues related to the idea of risk assessment are nothing new in safety standards. However, the method of assessing robots in terms of ethical risks is an

Table 1. Example of ethical risk associated with societal and the robot application [2]

Ethical issue	Ethical hazard	Ethical risk	Mitigation	Comment	Verification/Validation
Societal	Loss of trust (human robot)	Robot no longer used or is misused, abused	Design to ensure reliability in behavior	In unexpected behaviour occurs, ensure traceability to help explain what happened	User validation
Application	Inappropriate 'trust' of a human by a robot	Malign or inadequate human control	Model of appropriate human control	Design robot to reach safe state with respect to any other tasks the robot is executing	Software verification; expert guidance

interesting addition to the set of tools that a designer, a robotics should use. The IEEE Standard Association, in 2016, launched a global initiative on ethics in autonomous systems and artificial intelligence. The result of a global initiative by the IEEE Standards Association was the document Ethically Aligned Design (EAD), which was based on information gathered from the wider public. In subsequent editions, the EAD documents expand on ethical issues and related recommendations.

Faced the ethical dilemmas in robotics, likewise, the AIs matter should be concerned [4]. The matter refers to role of AIs behave in Society, e.g. computer programs capable of making decisions for approving home loans. Including AIs in daily decisions process for numerous profound and important questions it is an increasingly common practice. Robots are becoming increasingly involved in our daily lives. In [11, 21] is extensively discussed AI and robotics of certain roles that ethics plays in the prosperity of humanity. The author [11] suggest that trust and cooperation play key role in this process.

4 Summary

The article presents the most important issues related to the development of robotics, because more and more challenges are faced by collaborative, social robots, AMR. They must be more reliable due to their responsible tasks, as well as a closer or even direct contact with people. The most important common feature of all these robots is their safe contact with humans. The presence of a robot forces the improvement of order, work organisation and changes the approach to the use of common space. The result is greater efficiency and greater work safety.

The basic role of safety systems in machine control systems is to protect human health and life. In the face of the dissemination of the Industry 4.0 strategy, many machines receive new functionalities. Due to the dynamic development of robotics, the changing nature of cooperation, new ISO/FDIS 10218-1 and -2 standards are already being developed, which are to eventually replace ISO 10218-1: 2011 and 10218-2:2011.

Another issue that is ethics in robotics is a major challenge to create a new generation of robotics standards. Ethical standards are a big step forward for people to trust new technologies. Without ethical standards, it will be difficult to gain universal trust and acceptance among the general public. Furthermore, ethical concerns can be incorporated into learning, planning and control algorithms. The issue of ethics, comparably to the safety, are currently considered in a very wide range, on various levels and in various fields of the new technologies.

The last important point in the development of technology in the coming years will be the mutual communication between the robot and safety devices, which will enable information about the exact position of the robot, as well as the human. Currently, the robot knows where it is, but does not know who or what is approaching it until it collides. The safety systems also do not recognize the exact position of the robot, nor do they know exactly what obstacle they are dealing with. The situation would be completely different if the robot could communicate with security systems and both knew their position and what obstacle was approaching it. This is the next step in the development of collaborative robots.

References

1. Bezpieczeństwo maszyn - Ogólne zasady projektowania - Ocena ryzyka i zmniejszanie ryzyka [Safety of machinery - general principles for design - risk assessment and risk reduction]. Standard PN-EN ISO 12100:2012, PKN, PL (2012)
2. Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems. Standard BS 8611:2016, UK (2016). <https://shop.bsigroup.com/ProductDetail/?pid=00000000030320089>
3. Bloss, R.: Collaborative robots are rapidly providing major improvements in productivity, safety, programing ease, portability and cost while addressing many new applications. *Ind. Robot Int. J.* **43**(5), 463–468 (2016). <https://doi.org/10.1108/IR-05-2016-0148>
4. Burton, E., Goldsmith, J., Koenig, S., Kuipers, B., Mattei, N., Walsh, T.: Ethical considerations in artificial intelligence courses. *AI Mag.* **38**, 22–34 (2017)
5. Robots and robotic devices - Safety requirements for industrial robots - Part 1: Robots. Standard, ISO, Switzerland (2011)
6. Robots and robotic devices - Safety requirements for industrial robots - Part 2: Robot systems and integration. Standard, ISO, Switzerland (2011)
7. Industrial trucks - Safety requirements and verification - Part 4: Driverless industrial trucks and their systems. Standard, ISO, Switzerland (2020)
8. Robots and robotic devices - Collaborative robots. Standard, ISO, Switzerland (2016)
9. Klimaszewski, J., Janczak, D., Piorun, P.: Tactile robotic skin with pressure direction detection. *Sensors* **19**(21), 4697 (2019). <https://doi.org/10.3390/s19214697>
<https://www.mdpi.com/1424-8220/19/21/4697>
10. Klimaszewski, J., Władziński, M.: Human body parts proximity measurement using distributed tactile robotic skin. *Sensors* **21**(6), 2138 (2021). <https://doi.org/10.3390/s21062138>. <https://www.mdpi.com/1424-8220/21/6/2138>
11. Kuipers, B.V.: Trust and cooperation. *Front. Robot. AI* **9**, 65 (2022). <https://doi.org/10.3389/frobt.2022.676767>

12. Paliga, M.: Human-cobot interaction fluency and cobot operators' job performance. The mediating role of work engagement: a survey. *Robot. Auton. Syst.* **155**(104191) (2022). <https://doi.org/10.1016/j.robot.2022.104191>
13. Piçarra, N., Giger, J.C., Pochwatko, G., Gonçalves, G.: Validation of the Portuguese version of the negative attitudes towards robots scale. In: *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, vol. 65, pp. 1–10 (2014). <https://doi.org/10.1016/j.erap.2014.11.002>
14. Piątka, Z., (ed.): Raport rynkowy: Roboty współpracujące i mobilne roboty agv, automatyka podzespoły aplikacje [Robotics market report: collaborative robots and mobile AGV robots, automation components applications]. *Automatyka Podzespoły Aplikacje* **5**(151), 54–81 (2019). <https://ulubionykiosk.pl/wydawnictwo/apa/3264>
15. Piątka, Z., (ed.): Raport: Roboty współpracujące i mobilne- rynek po pandemii [Robotics market report: collaborative robots and mobile robots-the post-pandemic market]. *Automatyka Podzespoły Aplikacje* **10**(151), 48–67 (2021). <https://ulubionykiosk.pl/wydawnictwo/apa/4316>
16. Wózki jezdniowe - Bezpieczeństwo - Wózki bez operatora i ich układy [Safety of industrial trucks - driverless trucks and their systems]. Standard, PKN, Poland (1999)
17. Pochwatko, G., et al.: Polish version of the negative attitude toward robots scale. In: *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, vol. 65, pp. 1–10 (2014). https://doi.org/10.14313/JAMRIS_2-2015/25
18. Różańska-Walczuk, M., Pochwatko, G., Świdrak, J., Kukielka, K., Możaryn, J.: Wybrane predyktory postawy wobec robotów społecznych [Selected predictors of attitude towards social robots]. In: *Prace Naukowe Politechniki Warszawskiej*, vol. 1, pp. 15–24. Elektonika PW (2016)
19. Soleimani, M., Friedrich, M.: E-skin using fringing field electrical impedance tomography with an ionic liquid domain. *Sensors* **22**(13), 5040 (2022). <https://doi.org/10.3390/s2213504>. <https://www.mdpi.com/1424-8220/22/13/5040/htm>
20. Walcki, M.: Normy bezpieczeństwa w kontekście robotyzacji z użyciem AMR [Safety standards in the context of robotization with AMR] (2020). <https://versabox.eu/pl/normy-bezpieczenstwa-a-robotyzacja/>. VersaBox
21. Winfield, A.: Ethical standards in robotics and AI. *Nat. Electron.* **2**, 46–48 (2019). <https://doi.org/10.1038/s41928-019-0213-6>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Versatile Robotic Workstation for Electronic Skin - Problems and Solutions

Jan Klimaszewski^(✉) 

Department of Mechatronics, Institute of Automatic Control and Robotics,
Warsaw University of Technology, ul. Sw. A. Boboli 8, 02-525 Warsaw, Poland
jan.klimaszewski@pw.edu.pl

Abstract. In the face of the Fourth Industrial Revolution (Industry 4.0), collaborative robots have become one of the key pillars of development. Thanks to their sensors, they allow increased flexibility and safety while working in a shared space with humans. One such sensor is the electronic skin (e-skin), which enhances human-robot collaboration through physical contact. This paper presents the developed versatile robotic workstation that allows, among other things, the calibration of e-skin touch measurements. In particular, the problems encountered with the use of a standard industrial robot are presented and ways to solve them are discussed. The presented approach allows the automatic acquisition of calibration measurements of e-skin sensors within the reach of the robot.

Keywords: Collaborative robotics · Electronic skin · Test stand · Force-sensitive resistors · Robot Operating System

1 Introduction

For more than three decades [1] researchers around the world have supported the development of electronic skin (e-skin) technology. In robotics, the most important potential applications of e-skin are to improve the safety of human-machine collaboration and the agility of the robot [2]. E-skin devices currently being developed worldwide often allow the estimation of the value and location of tactile pressure [3]. There are also an increasing number of applications that allow the measurement of the proximity of human body parts to the e-skin surface [4]. Accurate testing and calibration is particularly important in light of the development in recent years of low-cost e-skin manufacturing techniques [5]. As the number of e-skin sensors increases, this problem becomes more relevant. As an example, previous study [6] performed an effective semi-automated calibration of e-skin, where the matrix studied had 16 rows and 18 columns, giving a total of 288 sensors. In order to calibrate and accurately test the e-skin, researchers

Department of Mechatronics, WUT.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 270–278, 2023.

https://doi.org/10.1007/978-3-031-37649-8_27

typically construct a stand to apply a precise pressure to one selected e-skin sensor. The most naive approach is presented by [7], where reference touch pressure is implemented in the form of weights of known mass manually placed at selected locations on the surface of the e-skin. One can find more sophisticated approaches using XY-axis tables [8] or a Z-axis platform [9] with a load cell. In [10], an interesting test rig was described to implement e-skin stretching measurements using a motor and linear gearbox. In the case of the heterogeneous parameters of the e-skin sensors, the calibration procedures carried out on the presented benches would be labour-intensive and difficult to repeat in practice. In the literature reviewed, no publication was found presenting an e-skin test bench allowing universal and automated measurements. This manuscript describes a robotic workstation designed for the calibration and testing of e-skin fabricated as an array of multiple sensors with heterogeneous measurement parameters. The developed stand consists of an industrial standard robot, a reference force sensor and a customised tactile tool tip. Thanks to the use of an industrial robot, the repeatability of the experiments performed and the ease of automation of the measurement process are ensured. Experiments with e-skin performed at the stand may include, for example, testing several hundred regularly arranged e-skin sensors for their calibration, measuring the reaction to touch in terms of normal and lateral forces exerted on the e-skin, testing the estimation of the proximity of objects to e-skin skin, and many other scenarios possible with the use of an industrial robot. To carry out these tests, it is not necessary to change the elements of the robotic workstation significantly, but only to change the control programs. The paper describes the solutions used to integrate the robot control system, the reference sensor measurement system and the e-skin. Problems that may arise in this type of robotic application and the proposed way to solve them are discussed.

The paper is organised as follows. In Sect. 2, the developed robotic workstation and its software integration is presented. In Sect. 3, the example measurements results are given. Finally, Sect. 4 provides the summary and further investigation proposal.

2 Developed Robotic Workstation

The first part of this section describes the developed robotic workstation in terms of hardware. The use of an industrial robot as one of the components has a number of significant advantages. First and foremost, it allows universal use of the developed stand - it enables automation of the measurement acquisition process, ensures repeatability, precise setting of the pressure direction and, for example, testing the proximity of various types of objects to the e-skin due to the possibility of precise positioning of the robot. The next part of the section presents the workstation in terms of software integration. This task was the main challenge in developing the workstation. In order to ensure the required functionality of the developed robotic workstation and the possibility of using it, for example, for calibration, proximity testing, tactile touch testing, and other e-skin tests, it is necessary to ensure simultaneous robot movement and recording:

tactile pressure measurements from the e-skin, measurements of the force exerted from reference device and TCP position measurements of the industrial robot. Measurements from different components should be synchronized in time. This chapter presents an exemplary e-skin test performed on the proposed workstation that meets all the above requirements.

2.1 Hardware Setup

The main components of the robotic workstation are a Fanuc LR Mate 200iC manipulator with a Mate R-30iA control cabinet (CC), an OnRobot Hex-e 6-axis reference device with a controller for measuring contact force, and the e-skin with a custom controller [3]. The central device integrating the measurements and communication with the devices is a general-purpose PC. Figure 1 shows pictures of the robotic workstation prepared for data acquisition.

As shown in Fig. 1, the Hex-e device (2) was attached to the end of the Fanuc robot kinematic chain (1). A custom tool tip with a compliant element (3) was then attached to the Hex-e. The use of the compliant element allows for a slower build-up of contact force with the robot's standard positional control, and is an important part of the functionality of testing the tactile parameters of the e-skin. This is particularly important for highly sensitive tactile sensors (e.g. graphene-based). The e-skin (5) and its measurement controller (4) were placed within reach of the robot.

For control and measurement acquisition, all devices (1, 2 and 5) were connected to a PC using the appropriate drivers. A USB port was used for the e-skin, while the Fanuc robot and Hex-e device were connected via Ethernet to two separate network cards. The general connection diagram of the robotic workstation devices is shown in Fig. 2.

2.2 Software Integration

In order to meet the measurement objectives associated with the station, time-synchronised acquisition of measurements from the e-skin, the Hex-e device and the position recording of the Fanuc robot is necessary. A major problem in automating measurements on a robotic workstation was the software integration of communication with component devices. The following part of the manuscript shows how to calibrate the coordinate systems of an industrial robot and how to record the robot's TCP position while integrated with the measurement systems of the e-skin and Hex-e reference device.

Robot Calibration. Before using the robotic workstation, it is necessary to define the tool and user frames for the manipulator. The tool coordinate frame is best defined so that the TCP (Tool Center Point) is unambiguous with the tactile tool tip being used. The user frame was then defined so that the origin of the coordinate frame was defined at one corner of the e-skin, while the direction of the X and Y axes are consistent with the columns and rows of the e-skin

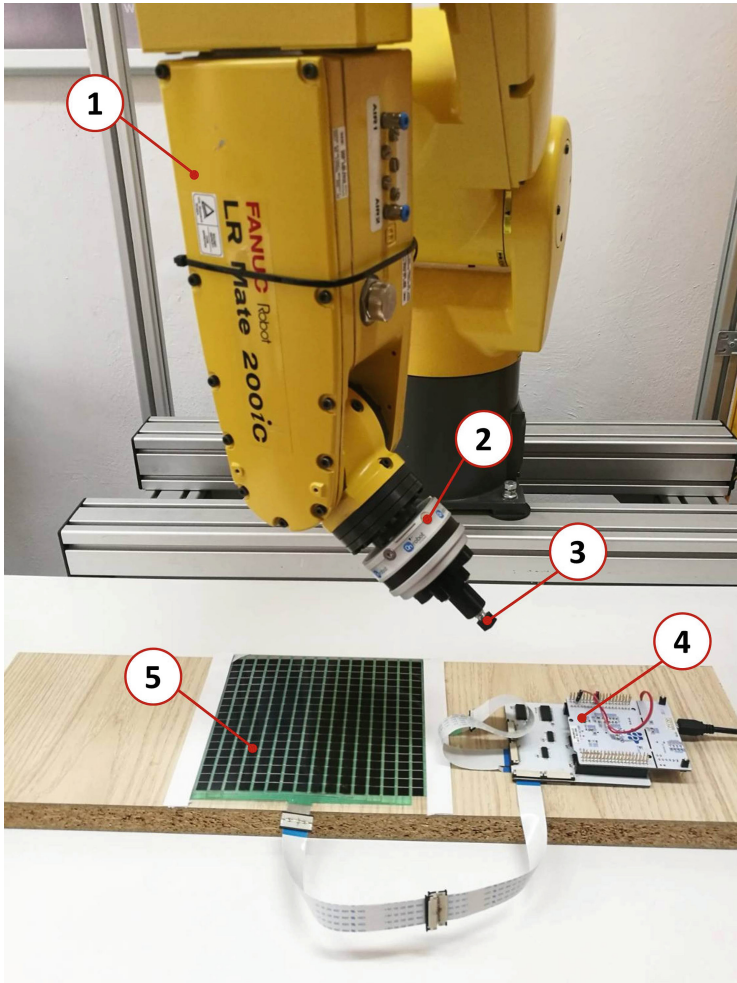


Fig. 1. Illustrative representation of the developed robotic workstation [6]: 1. LRMate 200iC manipulator, 2. reference e-Hex sensor, 3. robot tool, 4. e-skin driver, 5. e-skin.

sensors. The tool and user coordinate frames were defined using three-point methods. With regard to the tool frame, it is important to easily physically identify the physical touch point of the tool tip. With regard to the user frame, it is necessary to identify physically with TCP the three points associated with the e-skin: the point defining the center of the corner sensor of the e-skin, the center of any sensor located in the x-axis direction, and any point located in the x-y plane of the e-skin. It is worth emphasizing that the TCP and the centers of the selected sensors should be easy to visually identify and available for physical touch.

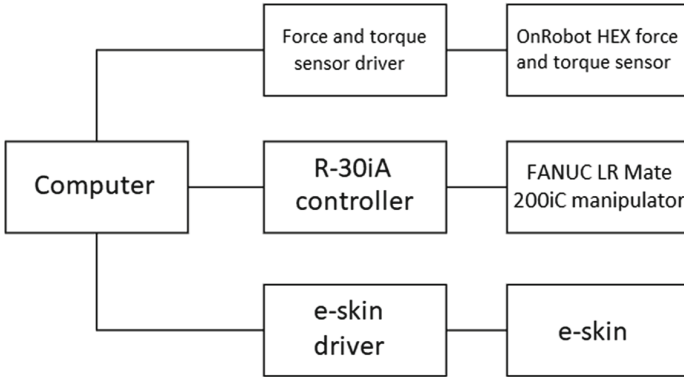


Fig. 2. Connection diagram of developed robotic workstation. [6].

TCP Position Acquisition. For the purpose of communicating with the Fanuc robot and recording its position, packages of the ROS¹ environment in the Indigo version running on the Linux distribution Kubuntu were used. For integration with the Fanuc robot, ROS-Industrial libraries were used. The process of running ROS for the Fanuc robot is complicated. In addition to the standard installation of ROS on a PC, it is necessary to install packages designed to run on the Fanuc CC. The installation process is based on compiling and loading the files into the CC with the Fanuc RoboGuide v7.70 (v7.70P/53 7DA7/53) software.

After network parameters configuration default ROS-Industrial package nodes allow to acquire only part of the status data from the Fanuc CC. This data includes information about joints angular positions. In order to capture the current position of the robot's TCP, it is necessary to calculate it based on the kinematic parameters of the robot. In addition it is necessary to manually publish the custom tool frame tf transformation.

Hex-e and E-skin Acquisition. To implement measurements from the Hex-e device, a dedicated device controller was used along with a provided sample program in C++ implemented as a ROS node. In terms of measurements from the e-skin, the exact details of the measurement controller are described, for example, in [3]. From the programming perspective, multi-threaded integration of measurements with the ROS node responsible for measurements from Hex-e was implemented.

¹ <https://www.ros.org/>.

Example Application. The example developed application is designed to collect the data necessary for e-skin calibration. To do this, it is necessary to exert pressure on successive e-skin sensors and simultaneously record measurements from the e-skin and the reference Hex-e sensor. A recording of the robot's position is not crucial for determining calibration, but it can be helpful in distinguishing between sensors, the loading and unloading phases of the exerted pressure, and thus in analyzing the hysteresis of e-skin sensors. The developed application generally consists of a PC-based ROS server program written in Python (1), a robot motion program (2) and ROS client for the robot both running on the CC (3), and a PC ROS node responsible for data collection (4). It is worth noting that there is a limitation on the number of simultaneously running programs in the robot CC. In order to simultaneously run the robot movement using (2) and the ROS client (3), it is necessary to set the "User Tasks" parameter in the CC configuration high enough. The robot's motion program (2) is responsible for executing a repetitive pressure trajectory on successive e-skin sensors. A ROS client (3) is additionally run on the robot's CC, responsible for sending messages about the robot's current state to a server (1) running on a PC. The robot's state, depending on the needs, can describe the robot's position or, for example, operating states such as: downward motion, upward motion or travel between sensors. The ROS server on the PC (1) receives the robot status messages sent by the robot's client program (3) and publishes a message on the corresponding ROS topic to record or not the measurement data. This is especially important to avoid large volumes of unnecessary measurement data being recorded. The node responsible for data recording (4) listens for messages on the ROS topic and starts recording measurement data of the e-skin, the Hex-e reference device and the robot's position into files at appropriate times.

3 Example Results

Figure 3 shows selected measurement results from the operation of an example application on the developed robotic workstation. It is worth noting that, for example, data from the e-skin and the Hex-e device were acquired with different sampling frequencies: approximately 50 samples/s for the e-skin sensor and approximately 10 samples/s for the Hex-e sensor. Before using these measurements, it is necessary to solve the problem of different acquisition frequencies, e.g. by resampling. Figure 4 shows data from the selected sensor pressed during the test.

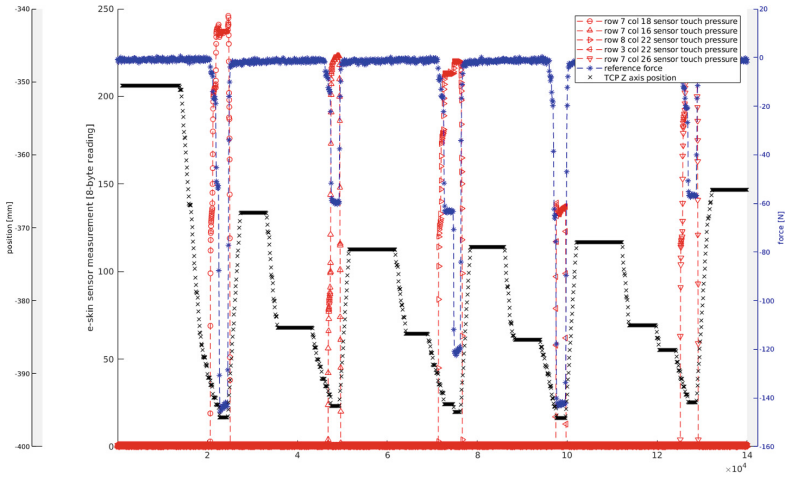


Fig. 3. Simultaneous acquisition of measurements from e-skin, Hex-e and Fanuc robot for five e-skin sensors.

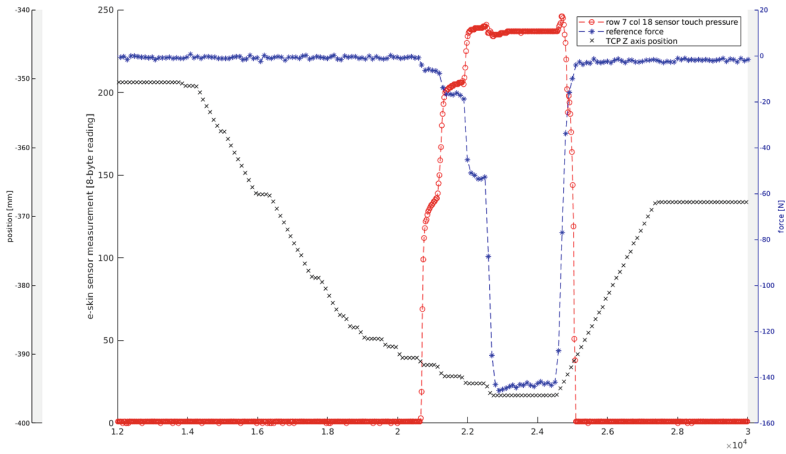


Fig. 4. Simultaneous acquisition of measurements from e-skin, Hex-e and Fanuc robot for one e-skin sensor.

4 Summary

The manuscript describes the hardware and software framework for a e-skin versatile robotic workstation based on a standard Fanuc industrial robot. Using the example of the e-skin test described in Sect. 2.2, it was shown that the proposed workstation meets all the requirements to provide the required e-skin testing functionality in various scenarios. All the devices included in the station were integrated and work with the use of the PC and ROS programming system. Functional tests on the workstation showed that it could be used, for example, to

calibrate the normal pressure of e-skin sensors characteristic. The application of the workstation, however, is much more versatile and can be used, for example, for testing sheer touch force e-skin reaction [3] or proximity estimation [4].

As outlined in the manuscript, the use of a standard industrial robot, e.g. Fanuc, is possible; however, its integration with the standard ROS programming system used increasingly in robotics development is a problem on its own. In the future, it is planned to replace Fanuc robot with the ES5 from EasyRobots company², which has built-in integration with the ROS system.

The application of robotization in the field of e-skin testing leads to significant time savings (the acquisition of measurements from several hundred e-skin sensors can take several hours [6]) and improves the quality of measurements due to the repeatability achieved with an industrial robot. In order to further develop the workstation, it is worth solving the problem of manual calibration of the robot against the e-skin to further deepen the automation of the workstation and the associated benefits.

References

1. Hammock, M.L., Chortos, A., Tee, B.C.K., Tok, J.B.H., Bao, Z.: 25th anniversary article: the evolution of electronic skin (E-skin): a brief history, design considerations, and recent progress. *Adv. Mater.* **25**(42), 5997–6038 (2013)
2. Li, G., Liu, S., Mao, Q., Zhu, R.: Multifunctional electronic skins enable robots to safely and dexterously interact with human. *Adv. Sci.* **9**(11), 2104969 (2022)
3. Klimaszewski, J., Janczak, D., Piorun, P.: Tactile robotic skin with pressure direction detection. *Sensors* **19**(21), 4697 (2019). <https://doi.org/10.3390/s19214697>
4. Klimaszewski, J., Władziński, M.: Human body parts proximity measurement using distributed tactile robotic skin. *Sensors* **21**(6), 2138 (2021). <https://www.mdpi.com/1424-8220/21/6/2138>
5. Pagoli, A., Chapelle, F., Corrales-Ramon, J.A., Mezouar, Y., Lapusta, Y.: Large-area and low-cost force/tactile capacitive sensor for soft robotic applications. *Sensors* **22**(11), 4083 (2022)
6. Klimaszewski, J., Wildner, K., Ostaszewska-Lizewska, A., Władziński, M., Możaryn, J.: Robot-based calibration procedure for graphene electronic skin. *Sensors* **22**(16), 6122 (2022). <https://www.mdpi.com/1424-8220/22/16/6122>
7. Soleimani, M., Friedrich, M.: E-skin using fringing field electrical impedance tomography with an ionic liquid domain. *Sensors* **22**(13), 5040 (2022). <https://doi.org/10.3390/s22135040>
8. Holgado, A.C., Tomo, T.P., Somlor, S., Sugano, S.: A multimodal, adjustable sensitivity, digital 3-axis skin sensor module. *Sensors* **20**(11), 3128 (2020). <https://doi.org/10.3390/s20113128>
9. Zhu, L., et al.: Large-area hand-covering elastomeric electronic skin sensor with distributed multifunctional sensing capability. *Adv. Intell. Syst.* **4**(1), 2100118 (2022). <https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202100118>
10. Dawood, A.B., Godaba, H., Ataka, A., Althoefer, K.: Silicone-based capacitive e-skin for exteroception and proprioception. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8951–8956 (2020)

² <https://easyrobots.pl/>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



**Special Session: Interacting with Virtual
Reality Applications**



VR Game for Powerlifting Training

Krzysztof Popielski^(✉) , Katarzyna Matys-Popielska ,
and Anna Sibilska-Mroziewicz 

Department of Mechatronics, Warsaw University of Technology, Warsaw, Poland
krzysztof.popielski.dokt@pw.edu.pl

Abstract. Virtual reality applications are becoming more and more popular. In addition to apparent uses like providing entertainment, VR applications are finding use in fields such as education, engineering, and architecture. The market for VR games used in medicine and sports is also thriving. The applications allow monitoring of an athlete's progress, training advanced movements specific to a given sport, and are difficult to reproduce during traditional training. A significant advantage of this type of solution is the increased safety of the user and, thus, a lower risk of injury. The article presents a VR application designed for the training of a powerlifting triathlon. This sport consists of three exercises. They are performed by both strength athletes and those training in other sports to prepare for the season properly. Due to the fact that they are simple multi-joint exercises, they fall into the collection of exercises often performed by amateur trainers. Despite their significant popularity and the undoubted advantages of performing them, it is often observed that they are performed incorrectly, which significantly increases the risk of injury. The purpose of the application is to enable safe training and learning of correct movement patterns of powerlifting exercises, regardless of the user's level of proficiency.

Keywords: Powerlifting · VR application · VR in sport

1 Introduction

Powerlifting is a discipline that belongs to the strength sports group. It consists of three exercises: a squat with a barbell on one's back, a bench pressing in a supine position, and deadlifting [1]. Significantly, exercises specific to this sport are performed not only by powerlifters but also by bodybuilders, strongmen, CrossFit practitioners, and during rehabilitation [2] (e.g., deadlifting is used to stabilize the spine in cases of discopathy [3]). These movements are also treated as a complement during preparatory training for all kinds of sports (such as snowboarding, skiing, or cycling) [4]. Due to the large number of applications of these exercises and their sophistication, they are associated with a significant number of injuries. For triathletes alone, the statistics are 1.0–4.4 injuries per 1000 h of training [5]. For CrossFit practitioners, the injury statistics are 0.2 to 18.9 injuries per 1000 h of training, with as many as 8.7% requiring surgery [6]. The reasons for this phenomenon are found in too much load, too little rest time between

sets, and incorrect technique of performing these exercises. The last reason is important because it can be reduced by improving exercisers' technique and movement patterns. This paper proposes a solution based on learning movement patterns using a virtual reality application.

VR applications supporting training and rehabilitation are increasingly common in the daily functioning of athletes [7], physiotherapists [8], and ordinary non-specialists. They allow both to conduct enjoyable workouts in a diverse environment (which is extremely useful, for example, in the physiotherapy of children [9]) and provide immediate feedback and the ability to monitor their progress [10–12]. In addition, for professional athletes, are used to analyze and optimize the movement patterns used [13–16]. Furthermore, the argument for the research is the variety of sports and exercises and other health and movement fields (such as applications for training surgeons) for which applications using virtual reality have already been created to teach [7, 17]. The literature data mentions the use of VR technology to assist strength athletes, however, mainly in the motivational sphere. Unknown to the authors are solutions for learning the movement patterns of powerlifting using VR applications [18].

2 VR Application Development

The application's main purpose is to allow one to safely learn the exercises included in the powerlifting or correct previously acquired, incorrect movement patterns – performing exercises without load will not cause injury. The critical element in these exercises is the correct handling of the barbell, which was taken into account when creating the application. In a significant number of cases, according to specialists, this is a sufficient condition to reduce the frequency of injuries resulting from incorrect exercise significantly. This statement is supported by the fact that the guidance of the barbell forces the body position of the exerciser. After consulting with a powerlifting coach, this was accepted as a sufficient condition for the VR application.

2.1 VR Application - Selecting the Type of Training

A VR game was created that met the outlined objectives. The player can choose a specific exercise. In the first stage, a video showing the correct performance of the exercise is displayed.

Then the player can choose one of four levels of difficulty. Levels of difficulty differ in the number of visible elements that determine the correct trajectory of the barbell and additionally allow the player to monitor his progress.

2.2 VR Application - Proper Training

Before performing an exercise, the player is informed of the correct starting point. For standing exercises (deadlift and squat), it is the virtual barbell's position (Fig. 1).

The user's position on the bench is additionally controlled by the horizontal bench press. During the execution of the exercise, the path of the barbell movement is examined, and the transition along the correct trajectory is checked, as well as the maintenance of

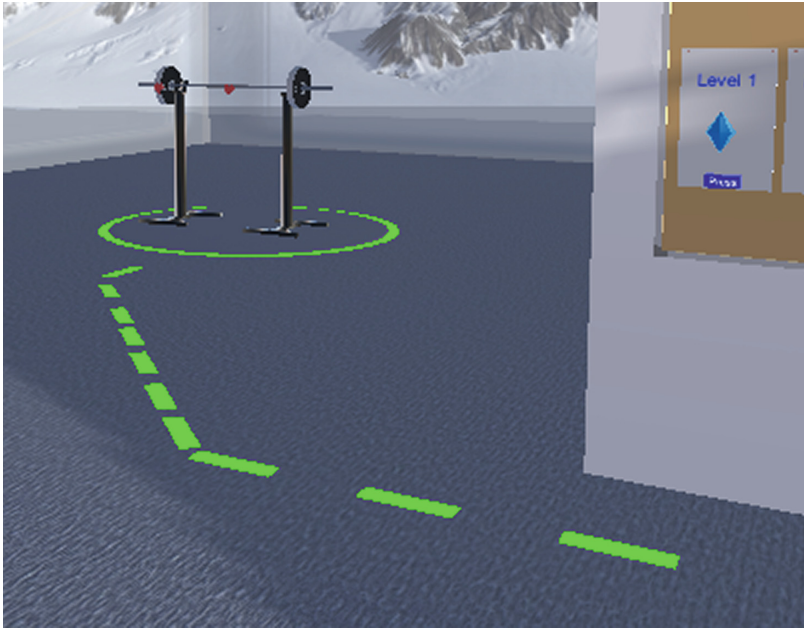


Fig. 1. Squat training station with a barbell, with visible starting points

the horizontal position of the barbell (required for each trained exercise). In addition, as the user performs repetitions, advice from a “virtual coach” appears, including critical elements of each exercise. The user performs five repetitions of an exercise. A special counter helps him control the number of repetitions. After completing the appropriate number of repetitions, the player proceeds to check the results.

2.3 VR Application - Results and Progress Monitoring

The results are then assigned to a 5-grade scale and presented to the player in an iconic, currently popular, and easy-to-understand manner. The results are displayed in three tables. The first gives the user feedback on each of the repetitions for a given exercise, with easy-to-understand graphs. The number two board describes how to improve exercise performance. The last board graphically depicts the overall averaged training score in a humorous way.

2.4 VR Application - Environment

The entire training room is located in a mountainous environment (Fig. 2) which is meant to make the learning process even more enjoyable. In addition, the app includes motivating voice comments and elements to increase mental immersion. All these elements aim to extend training time by making it more attractive.

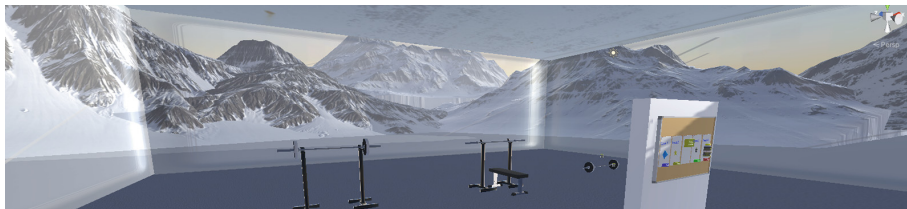


Fig. 2. Environment of application

3 Conclusion and Further Steps

A game was created that allows players to improve their technique in powerlifting exercises. At the time of writing, the prototype application has been tested by a strength triathlete coach. The testing results are auspicious, and necessary corrections have been made, so the application is ready for further research on a larger group of people. In addition, the use of VR games for learning or simulating sports [19, 20] and rehabilitation [8] allows us to assume that the application has a significant case to meet its goal, which is to reduce the number of injuries during training. The prototype application has also received positive feedback from physiotherapists.

The next step of the work will be to conduct tests on a larger number of users with different proficiency levels. The tests will be conducted under the supervision of a powerlifting coach. The test subjects will be divided into two groups. Each participant will get to watch a demonstration video at the beginning. Subjects in group one will then perform each of the exercises (at an adequate, safe load for the user's strength) with no prior in-game training. Participants in group two will do the in-game training first and then the gym training second. At the very end, the progress will be evaluated by the trainer. Thus, the real impact of performed exercises in virtual reality on the correctness of authentic movement patterns will be analyzed. The impact of the simplifications adopted, including analysis of the barbell movement alone without reference to characteristic body parts, on the correctness of the training performed will also be examined. The results of the study will be described in subsequent articles.

References

1. Radenković, L., Nešić, L.: The physics of powerlifting. *Eur. J. Phys.* **39**(3), 034002 (2018). <https://doi.org/10.1088/1361-6404/aaa90e>
2. Del Vecchio, L., Daewoud, H., Green, S.: The health and performance benefits of the squat, deadlift and Bench Press. *MOJ Yoga Phys. Ther.* **3**, 40–47 (2018). <https://doi.org/10.15406/mojypt.2018.03.00042>
3. Fischer, S.C., Calley, D.Q., Hollman, J.H.: Effect of an exercise program that includes deadlifts on low back pain. *J. Sport Rehabil.* **30**, 672–675 (2021). <https://doi.org/10.1123/JSR.2020-0324>
4. Bengtsson, V., Berglund, L., Aasa, U.: Narrative review of injuries in powerlifting with special reference to their association to the squat, bench press and deadlift. *BMJ Open Sport Exerc. Med.* **4**, 382 (2018). <https://doi.org/10.1136/bmjsem-2018-000382>

5. Ferland, P.M., Comtois, A.S.: Classic powerlifting performance: a systematic review. *J. Strength Cond. Res.* **33**, S194–S201 (2019). <https://doi.org/10.1519/JSC.0000000000003099>
6. Ángel Rodríguez, M., García-Calleja, P., Terrados, N., Crespo, I., Del Valle, M., Olmedillas, H.: Injury in CrossFit®: a systematic review of epidemiology and risk factors. *Phys. Sportsmed.* **50**, 3–10 (2022). <https://doi.org/10.1080/00913847.2020.1864675>
7. Farley, O.R.L., Spencer, K., Baudinet, L.: Virtual reality in sports coaching, skill acquisition and application to surfing: a review. *J. Hum. Sport Exerc.* **15**, 535–548 (2020). <https://doi.org/10.14198/JHSE.2020.153.06>
8. Rose, T., Nam, C.S., Chen, K.B.: Immersion of virtual reality for rehabilitation - review. *Appl. Ergon.* **69**, 153–161 (2018). <https://doi.org/10.1016/j.apergo.2018.01.009>
9. Lino, F., Arcangeli, V., Pia, D., Chieffo, R., Antonietti, A., Mandelbaum, D.E.: The virtual challenge: virtual reality tools for intervention in children with developmental coordination disorder. *Children* **8**, 270 (2021). <https://doi.org/10.3390/CHILDREN8040270>
10. Rego, P.A., Moreira, P.M., Reis, L.P.: Proposal of an extended taxonomy of serious games for health rehabilitation. *Games Health J.* **7**, 302–309 (2018). <https://doi.org/10.1089/g4h.2017.0138>
11. Wood, G., Wright, D.J., Harris, D., Pal, A., Franklin, Z.C., Vine, S.J.: Testing the construct validity of a soccer-specific virtual reality simulator using novice, academy, and professional soccer players. *Virtual Real.* **25**, 43–51 (2021). <https://doi.org/10.1007/S10055-020-00441-X/FIGURES/3>
12. Cannavò, A., Praticcò, F.G., Ministeri, G., Lamberti, F.: A movement analysis system based on immersive virtual reality and wearable technology for sport training. In: Proceedings of the 4th International Conference on Virtual Reality (ICVR 2018), pp. 26–31. ACM Press, New York (2018). <https://doi.org/10.1145/3198910.3198917>
13. Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., Craig, C.: Using virtual reality to analyze sports performance. *IEEE Comput. Graph. Appl.* **30**, 14–21 (2010). <https://doi.org/10.1109/MCG.2009.134>
14. Levac, D.E., Huber, M.E., Sternad, D.: Learning and transfer of complex motor skills in virtual reality: a perspective review. *J. Neuroeng. Rehabil.* **16**, 121 (2019). <https://doi.org/10.1186/s12984-019-0587-8>
15. Arndt, S., Perkis, A., Voigt-Antons, J.N.: Using virtual reality and head-mounted displays to increase performance in rowing workouts. In: Proc. 1st Int. Work. Multimed. Content Anal. Sport. Co-located with MM 2018 (MMSports 2018), pp. 45–50. (2018). <https://doi.org/10.1145/3265845.3265848>
16. Le Noury, P., Buszard, T., Reid, M., Farrow, D.: Examining the representativeness of a virtual reality environment for simulation of tennis performance. *J. Sports Sci.* **39**, 412–420 (2021). <https://doi.org/10.1080/02640414.2020.1823618>
17. Ahir, K., Govani, K., Gajera, R., Shah, M.: Application on Virtual Reality for Enhanced Education Learning, Military Training and Sports. *Augment. Hum. Res.* **5**(1), 1–9 (2019). <https://doi.org/10.1007/s41133-019-0025-2>
18. Gulec, U., Isler, I.S., Doganay, M.H., Gokcen, M., Gozcu, M.A., Nazligul, M.D.: PowerVR: interactive 3D virtual environment to increase motivation levels of powerlifters during training sessions. *Comput. Animat. Virtual Worlds* **34**(2), e2045 (2022). <https://doi.org/10.1002/CAV.2045>
19. Popovic, M.B.: *Biomechanics*. Academic Press, Massachusetts (2019)
20. Stone, J.A., Strafford, B.W., North, J.S., Toner, C., Davids, K.: Effectiveness and efficiency of virtual reality designs to enhance athlete development: an ecological dynamics perspective *Science Motricité Movement Sport Sciences. Mov. Sport Sci. Mot.* **102**, 51–60 (2018). <https://doi.org/10.1051/sm/2018031>







Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





A Case for VR Briefings: Comparing Communication in Daily Audio and VR Mission Control in a Simulated Lunar Mission

Kinga Skorupska^{1,2(✉)}, Maciej Grzeszczuk¹, Anna Jaskulska^{1,3},
Monika Kornacka^{1,2}, Grzegorz Pochwatko^{3,4}, and Wiesław Kopec^{1,2,3}

¹ Polish-Japanese Academy of Information Technology, Warsaw, Poland

² SWPS University of Social Sciences and Humanities, Warsaw, Poland
kinga.skorupska@pja.edu.pl

³ Kobo Association, Warsaw, Poland

⁴ Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

Abstract. Alpha-XR Mission conducted by XR Lab PJAIT focused on research related to individual and crew well-being and participatory team collaboration in ICE (isolated, confined and extreme) conditions. In this two-week mission within an analog space habitat, collaboration, objective execution and leisure was facilitated and studied by virtual reality (VR) tools. The mission commander and first officer, both experienced with virtual reality, took part in daily briefings with mission control. In the first week the briefings were voice-only conducted via a channel on Discord. During the following week last briefings were conducted in VR, using Horizon Workrooms. This qualitative pilot study employing participatory observation revealed that VR facilitates communication, especially on complex problems and experiences, providing the sense of emotional connection and shared understanding, that may be lacking in audio calls. The study points to the need to further explore VR-facilitated communication in high-stake environments as it may improve relationships, well-being, and communication outcomes.

Keywords: remote collaboration · virtual reality · communication

1 Rationale

In high-stake environments, where status needs to be reported carefully and often the mode of communication plays a crucial role. The choice between different communication spaces: audio, video or VR, may affect the quality of communication, relationships and well-being of the participants. Especially in the context of manned space flight communication between the crew and Mission Control

K. Skorupska and M. Grzeszczuk—The first two authors contributed equally to this study.

© The Author(s) 2023

C. Biele et al. (Eds.): MIDI 2022, LNNS 710, pp. 287–297, 2023.

https://doi.org/10.1007/978-3-031-37649-8_29

(MC) this choice may greatly impact goal completion, as problems encountered in space are analysed, simulated and solved by specialized teams on Earth.

To explore this topic we devised a pilot study to observe how the quality of briefings may change if they are moved from the audio space to virtual reality. In a two-week simulated Lunar mission we had the commander and the first officer take part in daily briefings with MC. In the first week the briefings were voice-only conducted via a locked channel on Discord. During the following week last briefings were conducted in VR, using Horizon Workrooms. This qualitative preliminary study employing participatory observation gathers insights, considerations and future research directions by exploring such research questions as:

1. To what extent does the technology, audio transmission and VR constitute a barrier to communication? (explored in Sect. 3.1)
2. What are the differences between the impressions of participants of remote audio and VR briefings? (explored in Sect. 3.2)
3. What is the potential of using VR-communication to improve the well-being of people experiencing ICE (isolated, confined and extreme) conditions? (explored in Sect. 3.3)

1.1 The Role and Forms of Communication

Communication is a form of exchanging information, thoughts and feelings. It can be job related, inter-team coordination, or used by mission control to manage the performance of crew members [4]. Communication can be an important factor in isolation missions, stabilizing the well-being as a carrier of relations with friends, family or a wider group of associates [22]. The level of copresence during mediated communication (usually higher in rich media like VR) influences well-being [21]. Natural communication usually takes place in a multimodal setting, integrating many senses such as sight, hearing and touch. The sender naturally uses those to encode the message not only in the words or the way they are uttered, but also in their posture, gestures or emotions that are conveyed through facial expressions [1]. Traditional media - like email, telephone or videoconference - negatively affect the process by filtering out some information carriers [23], which matters, as we sometimes send contradictory signals (e.g. when hurt, we say “it’s fine”). In Mehrabian’s study, only 7% of the actual message is attributed to words, 38% is vocal and 55% is the body language [13]. Moreover, communication between diverse stakeholders who experience different conditions, such as being under stress in a quick response scenario, is also challenging. While there exist communication protocols and best practices, a shared mental model based on prior training and relevant experience in the personal, interpersonal, and institutional dimensions is necessary to foster effective communication and collaboration, facilitating relationship-building and shared understanding [8].

1.2 ICE Conditions in Human Space Flight

The subject of human exploration of outer space raises an increased interest in the safety of astronauts staying in the hostile environment of a space mis-

sion. Unique factors associated with it, such as microgravity, altered day-night cycle, sense of danger, isolation and confinement, affect human behavior, performance and well-being, but their long-term effects are still not well understood. ICE (Isolated, Confined and Extreme) environments are therefore used for their exploration, combining presence of various type of stressors to uncover compound effects they might have on the human being [24]. Experience shows that the impact of ICE conditions is often underestimated by both the crew and the support staff [3]. The planned return of man to the Moon, and in the longer term, plans to place a man on the surface of Mars, prompt scientists to look for ways to obtain the data, knowledge and experience necessary to prepare the crews for the mission. Building analog habitats allows to simulate selected mission conditions in an artificially created environment. Depending on the project, the emphasis is on study of life support, food self-sufficiency or isolation issues [19].

1.3 VR for Collaboration and Work

The use of Virtual Reality (VR) as an environment to meet together, whether for entertainment such as computer games or for professional purposes such as working on a joint project is called “Social VR” [14]. Trials showed that, if the process is properly prepared, e.g. the proper division of roles and competences is ensured so that each team member knows what tasks are ahead and how to use the tools provided, the use of VR is well received and positively influences the sense of community in the team [9]. A virtual multi-user experience (VMUE) can also be used as relatively low cost means of simulating the concurrent activities of crews interacting with a simulated work environment. VR applications include training, as well as research on team behavior and interaction during simulated task loads. It can be used to simulate unfamiliar environments, low-gravity conditions or complex equipment maintenance, enabling true collaboration, not just dry procedural training [12, 20]. ESA is exploring the potential of VR and Augmented Reality (AR) also to increase the safety of teams working during missions - the EdcAR project could reduce the number of mistakes during procedures, such as the need for medical assistance when an on-board medic needs help [7]. The recent COVID-19 pandemic gave also an opportunity to explore mental health issues exposed by isolation [5] and the potential use of VR as a tool to reduce anxiety, depression, perceived stress, and hopelessness while increasing social connectedness [17].

2 Method

2.1 Simulated Lunar Mission and Its Crew

The research team engaged in a simulated Alpha-XR Mission conducted by XR Lab PJAIT between the 27th of April 2022 and 13th of May 2022 at an analog space habitat. It was initiated by a 3-day training, followed by two weeks

of confinement in ICE (isolated, confined and extreme) conditions. The participatory research conducted during the mission related to individual and crew well-being and team collaboration. It was facilitated and studied with the use of virtual reality tools, questionnaires and ethnographic methods. There were five study participants, who, as crew members consented to the research conducted as part of the Alpha-XR mission, and other research ongoing in the analog space habitat. All of the crew members were quite comfortable with information and communication technologies. The two commanders who took part in this particular study were very familiar with VR, having worked with it in a professional setting before and owning an Oculus headset.

2.2 Briefings with Mission Control

Every day of the mission, from the flight, lunar landing and the following EVA missions to return to Earth started with a remote briefing with the Mission Control team on Earth. The briefings would start at 9:00 AM and last about an hour, depending on how many things had to be discussed and who took charge of the meeting (either mission control members or the two commanders). Mission Control briefings started with the commanders reporting on the status of the habitat, including the water and energy consumption, graywater levels, food supply. Information about the crew was also included, such as calorie intake, health concerns, any conflicts or changes noticed as well as morale and performance evaluations. Next, the status of the tasks planned for the previous day was discussed, as well as the tasks for the day. Most tasks were related to the planned EVAs¹, habitat upgrades and related engineering tasks as well as the XR research that was being conducted as the main aim of the mission (Fig. 1).

name of the objective	execution (achieved / partially / issues / neglected)	comment
Saliva Collection	achieved	
Outreach Session	achieved	
Participatory prototyping session: VR, AR, XR	achieved	
BORP	neglected	no longer required
SWAMP	achieved	some minor visual improvements may be done before EVA on T+8
Tablet holder	achieved	3 straps, 3 tablet holders, 6 adapter plates
MC - Briefing Questionnaire 1	achieved	
MC - Check the dust sensor with IT team	partially	to be worked on by excursion crew during EVA
hydroponic setup test, observations gathered	partially	
aeroponic setup test, observations gathered	delayed	no roots, not enough pre-shower water
Grafana review	neglected	ran out of time

Fig. 1. Overview of task statuses from Day 6, as discussed on Day 7 of the mission

¹ EVA, that is Extravehicular Activities, were planned for every second day, and they constituted a major part of the simulation, with the exploration of the lunar surface, building up the SWAMP temporary moon base environment (adding sensors and the Internet connection) or testing different analog astronaut suits, such as the BORP suit as well as documenting the lunar surface.

Audio Briefings. Audio briefings were held on a dedicated channel on the habitat's Discord server, and the Commanders participated in them using either their personal laptops, or the personal mission tablets given to them at the start of the mission as well as personal headphones. During the audio briefings all of the participants looked at a shared Google Spreadsheets file, which had a sheet dedicated to the activities and habitat and crew statuses of each day. Additionally, there was an EVA file, with all the objectives and tasks related to the next EVA (Fig. 2).



Fig. 2. Example activities reported during daily briefings. On the left engineering of the BORP suit collar, on the right an EVA mission in progress viewed from a lunar rover camera by HabCom, that is the two crewmates staying behind as support.

VR Briefings. Two commanders and one member of MC participated in the VR briefings using Oculus 2 VR standalone headsets, on the first day in the same room, and the following days in separate rooms. Additionally, one member of MC connected to these briefings using a desktop application without using their webcam. The VR briefings were held using the Meta Horizon Workrooms and after the first meeting the whiteboard functionality was utilized to discuss tasks and their progress: an image of a relevant snippet of the Google Spreadsheets file was captured and uploaded to the Horizon Workrooms and pulled up to the whiteboard.

2.3 Briefing Evaluation Tools

All data was gathered using a timed survey built with Movisens: at 10:00 AM, after each briefing the commanders had an alarm on their tablets, which took them to the installed Movisens app where they jumped right into filling out a brief questionnaire. If this alarm was missed, then they received another prompt at a later time. Most questionnaires were submitted a few minutes after 10:00.

The questionnaires consisted of a query on the mode of participating in the meeting (Audio or VR), technical difficulties, discomfort and three longer questions asking to rate the experience on sliding visual analog bipolar scales, expressed underneath by ratings from 0 to 100.

- Meeting satisfaction was measured with a scale developed by Rogelberg et al. [18] after some adjustments. We have adjusted the labels of the scales to instead of “stimulating/boring” read “engaging/non engaging”, instead of “enjoyable/unpleasant” to “pleasant/unpleasant” and instead of “satisfying/annoying” to “meets expectations/doesn’t meet expectations”.
- Social presence questions are inspired by Nowak and Biocca [16] and Bailenson, Blascovich, Beall and Loomis [2] (but see also Swidrak, Pochwatko and Matejuk [21]); social interaction by Nezlek, Imbrie and Shean [15].
- The measures related to the mental state recorded after each meeting were based on Kashdan et al. [11] and Ciarocco, Twenge, Muraven, and Tice [6].

Additionally, meeting notes and impressions after the briefings were recorded in a journal by the executive commander and the crew kept written journals of their experience, as part of an autoethnographic project of another crewmate (Fig. 3).

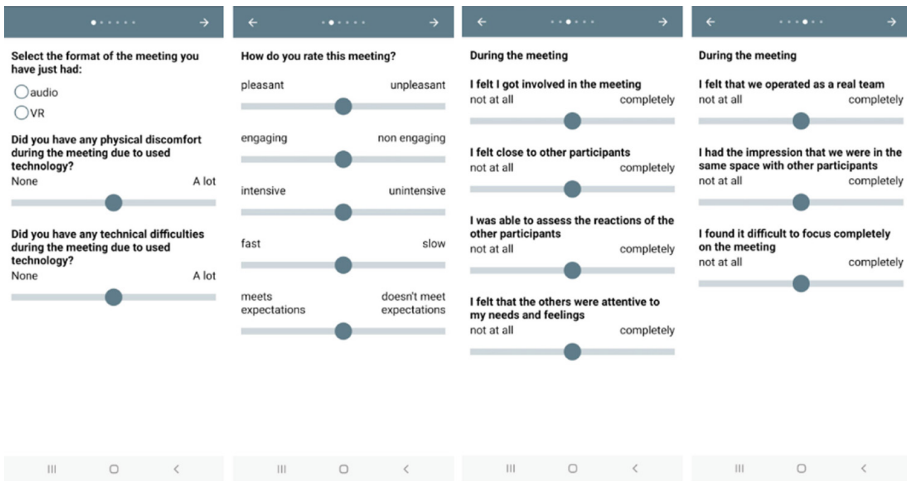


Fig. 3. Movisens application, showing the first four screens (out of six) of the post-briefing questionnaire employing visual analog bipolar scales. The visible questions are derived from the Meeting Satisfaction scale [18]

3 Results and Discussion

The quantitative data consists of ratings from two participants of 3 audio meetings (6–8.05.2022), followed by three VR meetings (9–12.05.2022). The data for

5.05 and 10.05 was excluded as it was incomplete. The data was gathered after a week of regular audio meetings, so that participants could get used to holding daily briefings in their default mode before introducing first the questionnaire, after each audio briefing, to establish a baseline, and then the VR meeting mode.

3.1 Technology and Overall Rating

Both audio and VR were rated as comfortable and free of difficulties with VR causing slightly more physical discomfort. Journal notes on the first VR meeting read: *There were two short freezes, but other than that it was enough to conduct the meeting. (...) We need to do it in two different rooms, because it is confusing to hear each other twice.* The overall rating of the experience in answer to the question “How do you rate this meeting?” is strongly in favour of VR, as it was rated as nearly double as pleasant, engaging, intensive and fast when compared with audio meetings (see Fig. 4).

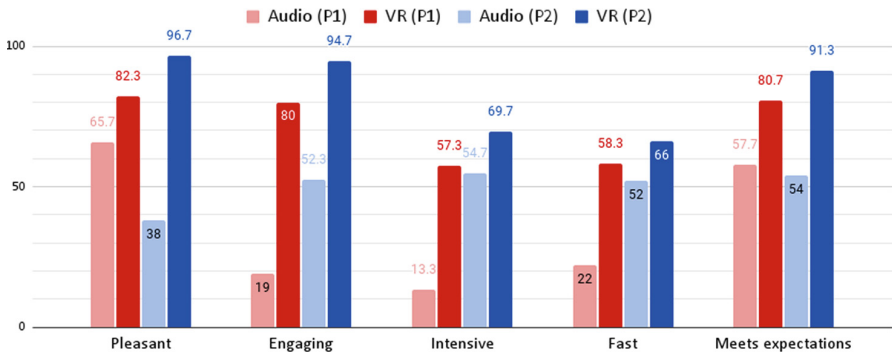


Fig. 4. Rating averages of the audio and VR meetings

Overall, the briefings held in VR were rated at 86/100 vs 55.9/100 for Audio, on whether they have met expectations.

3.2 Impressions During the Meeting

The impressions of the audio meetings recorded in the journal were that the meetings felt slow and not engaging, as if the Mission Control Staff was experiencing time slower, it felt as if they even talked slower. This is recorded as the cause that the commanders were likely to lose focus during audio briefings and drift off, or start doing other tasks meanwhile. The impressions of the perception of time by MC and their subjective reported effect on the commanders may in part be attributed to the issues studied by Kansas et al., who found that with increased quantity of communication, there is an increasing chance of astronauts displacing the problems they experience onto ground control and experiencing

more negative effects of ICE [10]. Meanwhile, journal notes from the first VR meeting read: *It was awesome! Much better than the audio meetings we've had. The pacing was better (it was same time but more filled with information) and it was nice to share the same space, because we could use gestures, especially to explain some engineering that we're doing (picture + gestures for VKK helmet) and gestures for the BORP on how it has broken at the collar attachments and the lights and the gloves. We also weren't as frustrated and shared more, as we felt we can be understood and gauge the others' impressions of what we are saying. The commander said he never wants to have audio briefings again:)* Especially the use of gestures, enabled by VR communication, seems to be highly beneficial for communication quality while explaining engineering issues or expressing emotions. Similar impressions are visible in survey results in Fig. 5, as the participants felt as if they participated in face-to-face meetings, sharing the same space with others, who were attentive to their needs.

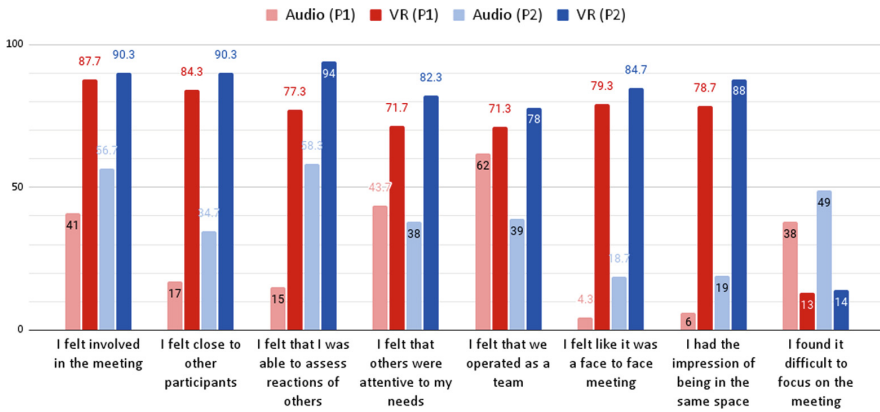


Fig. 5. Average impressions of the Audio and VR Meetings from P1 and P2

3.3 Mental State After Each Meeting

Mental states were mostly more positive after the use of VR, as shown in Fig. 6. The only exception here was being “sad” and “anxious” for P2, however that may be attributed either to the slowly approaching ending of the mission, which also coincided with VR meetings or other, unrelated, circumstances. However, on other occasions the whole crew mentioned that as the mission was drawing to an end they felt progressively sadder. Overall, after having used VR for briefings the participants were a bit more content, enthusiastic, joyful, and relaxed, while after audio briefings they were slightly more mentally exhausted. An interesting research direction is the effect of VR on mood and feelings about one’s cognitive performance. In this study, the crew mentioned increasing problems as more time in confinement passed, as indicated by this journal entry: *We have cognitive*

decline that we don't recognize, because it sneaks up on us. Astronauts have it too - why does that happen though? (...) This is actually a bad problem because it frustrates people. We have trouble learning new stuff. Forget e.g. chords, have trouble learning and explaining new games, have trouble remembering where stuff is either physically or on the drive.

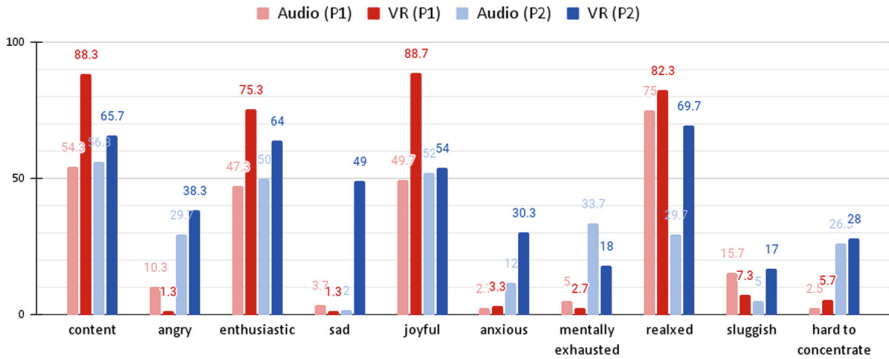


Fig. 6. Average mental states of P1 and P2 after each type of the meetings

4 Conclusions and Future Work

This pilot study points to possible positive effects of enriching audio-only communication with VR, especially in stressful settings, requiring frequent reports and demanding high understanding between participants. VR facilitates communication, especially when explaining complex problems and experiences and it provides the sense of connection and empathy, that may be lacking when it comes to audio calls. VR-meetings forced participants to fully focus on them and to be in the moment, despite their perceived cognitive decline and the stress of meeting daily goals. This resulted in feeling of being in a shared space, not only physically, but also in terms of common understanding and emotional connection. Future work ought to explore the effect of substituting other modes of communication, either audio or video, with VR in longer, larger studies.

Acknowledgments. We would like to thank the many people and institutions gathered together by the distributed Living Lab Kobo and HASE Research Group (Human Aspects in Science and Engineering) for their support of this research. This research is part of ALPHA-XR – a new long-term proposal for a transdisciplinary research framework programme that fits in the Strategic Innovation Area of Human and Robotic Exploration also known as Terrae Novae programme.

References

1. Acarturk, C., Coskun, M., Emil, S.: Multimodal communication in instructional settings: an investigation of the functional roles of gestures and arrows. *Rev. Signos* **54**, 867–892 (2021)

2. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Interpersonal distance in immersive virtual environments. *Pers. Soc. Psychol. Bull.* **29**(7), 819–833 (2003)
3. Bishop, S.: Evaluating teams in extreme environments: from issues to answers. *Aviat. Space Environ. Med.* **75**, C14–21 (2004)
4. Bondaruk, A., Kozuba, J.: Selected aspects of aviation communication. *Sci. Res. Educ. Air Force* **19**, 79–88 (2017)
5. Choukér, A., Stahn, A.: COVID-19-the largest isolation study in history: the value of shared learnings from spaceflight analogs. *NPJ Microgravity* **6**, 32 (2020). <https://doi.org/10.1038/s41526-020-00122-8>
6. Ciarocco, N., Twenge, J., Muraven, M., Tice, D.: Measuring state self-control: reliability, validity, and correlations with physical and psychological stress. Unpublished manuscript (2007)
7. ESA: Reality to the rescue (2017). https://www.esa.int/Enabling_Support/Preparing_for_the_Future/Discovery_and_Preparation/Reality_to_the_rescue. Accessed 30 Nov 2022
8. Doyle, E.E.H., Paton, D.: Decision-making: preventing miscommunication and creating shared meaning between stakeholders (2017). <https://doi.org/10.1007/11157201631>
9. Heinonen, H., Burova, A., Siltanen, S., Lähteenmäki, J., Hakulinen, J., Turunen, M.: Evaluating the benefits of collaborative VR review for maintenance documentation and risk assessment. *Appl. Sci.* **12**(14), 7155 (2022). <https://www.mdpi.com/2076-3417/12/14/7155>
10. Kanas, N., et al.: Psychosocial issues in space: results from shuttle/Mir. *Gravitational Space Biol.* **14**, 35–45 (2001). <https://link.gale.com/apps/doc/A176373056/AONE>
11. Kashdan, T.B., Farmer, A.S., Adams, L.M., Ferrisizidis, P., McKnight, P.E., Nezlek, J.B.: Distinguishing healthy adults from people with social anxiety disorder: evidence for the value of experiential avoidance and positive emotions in everyday social interactions. *J. Abnorm. Psychol.* **122**(3), 645–55 (2013)
12. Mastro, A., Federico, M., Benyoucef, Y.: A multi-user virtual reality experience for space missions. *J. Space Saf. Eng.* **8**, 134–137 (2021). <https://doi.org/10.1016/j.jsse.2021.03.002>
13. Mehrabian, A.: *Nonverbal Communication*. Taylor & Francis, Milton Park (2017)
14. Mütterlein, J., Jelsch, S., Hess, T.: Specifics of collaboration in virtual reality: how immersion drives the intention to collaborate (2018)
15. Nezlek, J., Imbrie, M.: Depression and everyday social interaction. *J. Pers. Soc. Psychol.* **67**, 1101–1111 (1995). <https://doi.org/10.1037/0022-3514.67.6.1101>
16. Nowak, K.L., Biocca, F.: The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators Virtual Environ.* **12**(5), 481–494 (2003). <https://doi.org/10.1162/105474603322761289>
17. Riva, G., et al.: COVID feel good-an easy self-help virtual reality protocol to overcome the psychological burden of coronavirus. *Front. Psychiatry* **11**, 563319 (2020). <https://doi.org/10.3389/fpsy.2020.563319>
18. Rogelberg, S., Allen, J., Shanock, L., Scott, C., Shuffler, M.: Employee satisfaction with meetings: a contemporary facet of job satisfaction. *Hum. Resour. Manage.* **49**, 149–172 (2010). <https://doi.org/10.1002/hrm.20339>
19. Schlacht, I., et al.: Existing and new proposals of space analog, off-grid and sustainable habitats with space applications (2016)

20. Silva, G., Morgado, L., Cruz, A.: Impact of non-verbal communication on collaboration in 3D virtual worlds: case study research in learning of aircraft maintenance practices. In: Beck, D., et al. (eds.) iLRN 2017. CCIS, vol. 725, pp. 25–34. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60633-0_3
21. Świdrak, J., Pochwatko, G., Matejuk, P.: Copresence and well-being in the time of COVID-19: is a video call enough to be and work together? In: Biele, C., Kacprzyk, J., Owsiński, J.W., Romanowski, A., Sikorski, M. (eds.) MIDI 2020. AISC, vol. 1376, pp. 169–178. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-74728-2_16
22. Vanhove, A., Herian, M., Harms, P., Luthans, F., Desimone, J.: Examining psychosocial well-being and performance in isolated, confined, and extreme environments (2015)
23. Wickman, L., Tsai, A., Walters, R.: Isolation and confinement issues in long duration spaceflight, pp. 1–9 (2008)
24. Zivi, P., De Gennaro, L., Ferlazzo, F.: Sleep in isolated, confined, and extreme (ICE): a review on the different factors affecting human sleep in ice. *Front. Neurosci.* **14**, 85 (2020). <https://doi.org/10.3389/fnins.2020.00851>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Construction and Evaluation of a Laboratory Stand for Testing MV Switchgears Using VR Technology in the Power Industry

Tadeusz Daszczyński¹(✉), Kacper Berdek¹, and Dariusz Naruszewicz²

¹ Warsaw University of Technology, Politechnika Square 1, 00-661 Warsaw, Poland
tadeusz.daszczyński@pw.edu.pl

² Pradma Dariusz Naruszewicz, Lubelska Street 33, 10-408 Olsztyn, Poland

Abstract. The purpose of this study was to build and evaluate a laboratory stand for testing MV switchgears using VR technology in the power industry. In the course of work, an exemplary program of operation of the application created by PRADMA was developed. The developed program was designed to effectively introduce the application user to the basic information about moving the virtual avatar, interacting with the prepared elements, and to provide information on the construction of the MV switchgear and its components. In order to confirm the advantages of VR technology and the quality of the developed VR application program, tests of the VR application were carried out with the help of students of the Warsaw University of Technology. The results of the evaluation questionnaire created for the purposes of the tests were used to develop conclusions regarding the use of VR in the power industry. The evaluation questionnaire questions related to key issues such as the quality of the mapped virtual elements, the level of realism preserved in the virtual environment, the effectiveness of the developed program or the substantive values of the entire exercise. The summary contains the answer to the theses put forward in the paper regarding the profitability and usefulness of using VR applications in the power industry.

Keywords: evaluation · switchgear · power engineering · virtual reality · application · user · avatar · interaction

1 Introduction

At the turn of the last twenty years, significant progress has been made in the field of computer sciences and information technology. New solutions and applications of modern technologies have been intensively searched for in both the science and entertainment sectors in order to facilitate and optimize everyday life. This translated directly into the development of technology which is virtual reality, which with each successive year manifests its application in subsequent areas of life.

Virtual reality (VR) is used in those industries where making a mistake may result in damage to health or exposure to costs related to the destruction of equipment. Due to the affordability and wide range of possibilities of this technology, it is gaining more and

more popularity all over the world. Among others, Schneider Electric used virtual reality as a training medium in the field of electrical power devices offered by the company [1]. Another example of the use of virtual reality in practice is a project implemented by Enea Operator, consisting in mapping a training means for learning work under voltage. This project involves the development of training scenarios for selected Main Power Supply Points and MV stations [2].

Computer-generated artificial reality is not limited to a computer that generates virtual space. Necessary to use VR technology are devices that are the medium between humans and the virtual world. The combination of such devices enabling real-time interaction of the real world and the computer-created space, creating an extension of the real world, is called AR (Augmented Reality) [3, 4]. This interaction applies to both adding virtual objects in the real world and placing real objects in the virtual space, creating a mixture of both these environments.

The main purpose of this work was to create and evaluate a laboratory stand for testing MV switchgear with the use of VR technology. As part of the work, an exemplary program of operation of a VR application developed by PRADMA and Institute of Power Engineering in Warsaw University of Technology (WUT) was developed in order to create a comfortable and effective environment for teaching purposes in the field of power engineering, with the specification of issues related to medium voltage switchgear. Then, on the basis of the developed scheme, application tests were performed and an evaluation questionnaire was conducted to develop the results of the tests.

During the work, the profitability of the VR application, the impact of its use on the learning outcomes and other aspects of the use of virtual reality, such as its impact on the well-being of users or the level of complexity of operating devices in virtual space, were analyzed.

2 The MV Switchgear

The main aspect of the VR application that was used for the work is the virtual model of the MV switchgear. A switchgear is a set of electrical power equipment operating at the same rated voltage, used to distribute electricity. It consists of a structure equipped with busbars and insulating elements, as well as electric power equipment serving as distribution, protection or measurement [5]. The medium voltage switchgear implemented in the application is based on the real model of the switchgear produced by Elektrometal Energetyka SA presented in Fig. 1.

MV switchgears with power equipment are characterized by characteristic electrical quantities, the values of which depend on the method of execution, electrical solutions and materials used. These values are crucial when selecting the equipment, and the designers make every effort to achieve the highest possible values of current, voltage or temperature to which the switchgear will be adapted, while keeping its dimensions and production costs as small as possible. The standardization of the e2ALPHA switchgear in terms of limit values and design applications is described in standards, i.e. [6]. Due to the nature of the work, the focus was mainly on the design properties of the discussed switchgear.

MV switchgear as an electrical power device should be as reliable and safe as possible for its users. For the benefit of its users, switchgears are equipped with a number of

interlocks related to the user's inaccessibility to conductive elements during its operation under voltage. One of the most important interlocks in the e2ALPHA switchgear are interlocks for the withdrawable unit compartment door while the withdrawable unit is in the operating state, the interlock of closing the earthing switch during the operating state of the movable unit or the interlocking of opening the cable compartment while the earthing switch is in the open position.



Fig. 1. MV switchgear from Elektrometal Energetyka SA (picture taken in WUT).

3 Development of a VR Application Operation Program

The aim of the program development was to adapt the user of the application to the virtual reality environment, to familiarize with the rules of moving inside the virtual space and to perform two pre-prepared didactic exercises in the field of testing the MV switchgear and its components.

3.1 Description of Equipment and Installations

For the proper functioning of the VR application, a set of necessary devices was used, which included:

- VR glasses, two touch controllers included in the Oculus Quest 2 set [7],

- A computer equipped with a modern processor and graphics card, with sufficient computing power,
- Monitor displaying a view from the perspective of using VR glasses, used to analyze the respondent's movements in real time,
- USB type cable - USB Type C, connecting the glasses with a computer unit in order to transmit the image to the monitor and increase the computing power of the glasses.

The created laboratory stand is shown in Fig. 2. In the further description of the application, for easier identification, the user is the person who is currently using VR glasses, and the avatar is the character that the user moves in the virtual world.



Fig. 2. Laboratory setup in the Laboratory of Power Apparatus and Switching Process (picture taken in WUT).

3.2 Description of the Application Interface Using the Tutorial Example

After launching the VR application, the user will see a menu on the glasses screen containing the selection of a specific part of the exercise. The virtual application menu is presented on the Fig. 3. The selection is made using the main button, targeting the selected option with the indicator generated by the controller. During the first contact with the program, the target option will be to select the tutorial first.

After selecting the tutorial, we will be greeted with a visual message along with the tutor reading its content, in which we will be informed about the purpose of the tutorial. After accepting the message with the main button, you will find yourself in a virtual room. The teacher will then inform the user about the rules of navigating in the virtual world. The avatar is rotated by physically turning the user's head or by using the knob on the top panel of the controller. Moving around the virtual room takes place through physical movement of the user or teleportation with an avatar. After pointing the controller pointer at any place on the virtual room floor, the user will see a circle

appearing in the selected place. After pressing the circle, the avatar will be teleported to the indicated place. Such an application is intended to facilitate the user's movement and to limit the user's movement within the security zone.

Then, the user will be informed about the possibility of interacting with objects using the main controller button and about the possibility of catching virtual items. Using the middle finger button, the user can grab the switchgear element and rotate it freely by moving the controller. After executing the instructions, the user will be asked to check the acquired skills by pressing the green button on the wall and dragging the switch in the trolley into the vicinity of the green zone. After completing the tutorial, the user will be informed about the possibility of leaving it.



Fig. 3. Basic menu of the VR application.

3.3 VR Exercise 1

After completing the tutorial, the user from the application menu should enter the main part of the first exercise, entitled "Switchgear construction". In this part of the exercise, the user was presented with the construction of MV switchgear on the example of the e2ALPHA switchgear described in chapter two. The user of the application will get to know the basic structural elements of the switchgear, view its devices, read the information prepared about them and view construction diagrams displayed on the virtual board.

After entering the exercise part concerning the construction of the switchgear, the user will be transported again to the virtual room and will be informed about the purpose of the exercise. Then the user will be instructed to approach the switchgear located in the room and to click it with the main button. After these steps are completed, a short animation will be shown showing the elements separating from the switchgear and moving towards the center of the room.

The element to which the controller will be approached will be highlighted in yellow, which means that one can interact. After pressing the button, the avatar will be transported

to the room where the user can freely move the isolated element, read the information on the board and view a photo or diagram of the actual device. Example of isolated element of MV switchgear (circuit breaker) is shown in Fig. 4.

After examining a given element, the return button on the wall can be pressed, which results in returning to the training field and marking the examined element on the list of elements on the wall, presented in Fig. 5.



Fig. 4. Separated elements of the virtual switchgear (left) and the virtual model of the e2BRAVO circuit breaker (right).

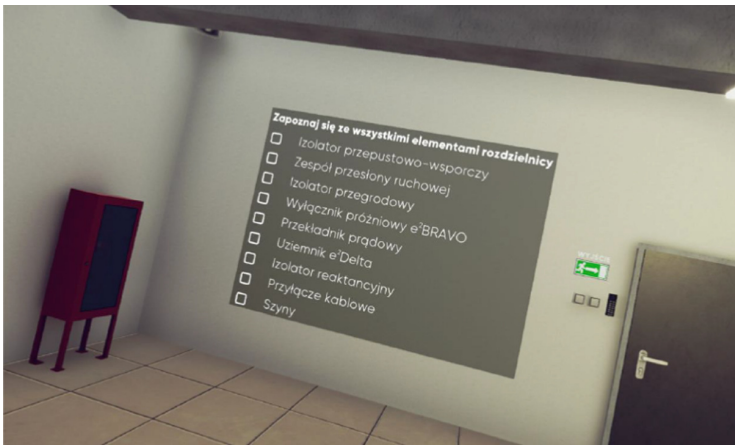


Fig. 5. The list of possible to view elements in MV switchgear.

Additionally, on one of the walls of the virtual room there is a table with diagrams of devices and elements used in the exercise. After pressing an arrow associated with a given element, its detailed diagram will be displayed.

3.4 VR Exercise 2

In the second exercise, entitled “Switchgear diagnostics”, the user was supposed to use a trolley to pull out the MV circuit breaker located in the withdrawable compartment of the switchgear. Then he was subjected to the task of finding five damages to the MV switchgear and its elements caused by the action of an electric arc. In order to open the withdrawable compartment of the switchgear, it was first necessary to release the lock preventing the opening of the compartment door in the absence of safe conditions. This is done by then opening the switch contacts by pressing the red button responsible for controlling the switch and closing the earthing switch via the green control button of the earthing switch. If any wrong steps are taken, the user will be notified that they cannot be performed. On Fig. 6 the MV circuit breaker used in the exercise is shown. The maneuvering of devices and electrical apparatus in the model is carried out in accordance with the principles of operating works at MV switchgears.



Fig. 6. The MV circuit breaker in VR application

4 Testing of VR Application

For the purpose of testing VR applications, a group of forty-three students of electrical and related faculties was assembled. Before performing the VR application test, each student was acquainted with the construction of a real MV switchgear located at the test site, shown in Fig. 1. In order to verify the correctness of the switchgear implementation into the virtual environment.

Then, each student started the application program discussed earlier, and then conveyed his feelings from the virtual exercise by performing a prepared evaluation questionnaire. The basis of the survey questions were questions developed by the author of

this work and, inter alia, in close cooperation with Dr. M. Marzec, Dr. M. Kotyśko and Dr. E. Waszkiewicz, employees of the Faculty of Social Sciences of the Department of Clinical Development and Education at the University of Warmia and Mazury. The questions were arranged for the purposes of the study in order to develop conclusions from the use of VR applications. The survey was made with the use of the Microsoft Forms survey editor and consisted of 45 questions concerning the feelings associated with the use of VR, the accessibility of this technology, the developed exercises and the level of mapping virtual elements. The results of the survey were then used to develop the results of the tests and the quality of the prepared application scheme.

5 Summary and Conclusions

The proposed VR application operation program is a proven example for substantive classes in the field of power engineering, detailing issues related to the MV switchgear. It allowed for quick and effective adaptation of the user to the virtual world, while leaving him as much knowledge gained during the prepared exercises as possible. The test participants showed a high level of focus and interest during the execution of the developed VR application program.

The laboratory stand created for the purpose of testing the application allowed for the effective conduct of the research participants through the developed diagram of the VR application. The process of creating a laboratory stand allowed to learn about the affordability and ease of installation and operation of equipment adapted to virtual reality operations.

The prepared evaluation questionnaire allowed to collect the users' feelings about the prepared exercises and the overall use of VR. The test participants' answers confirmed the belief that VR technology is attractive in the academic environment. According to the respondents, the exercises prepared in VR made it possible to effectively remember the presented content about the switchgear and its elements, which confirmed the sense of using VR as a way to supplement standard teaching techniques. The models implemented in the VR application were recreated realistically enough to ensure a positive reception of the application among the respondents. The entire VR exercise was received positively by the test participants and they firmly supported the attractiveness of the presented exercises and the opportunities created by virtual reality.

Current attempts to use VR technology indicate the trend of the increasingly common use of VR technology in subsequent sectors of life. In the future, the discussed VR application will be developed in a project from Polish Ministry of Education and Science "Development and implementation of a cognitive training platform using VR and AR technologies".

"Publication co-financed from the state budget under the program of the Minister of Education and Science under the name "Science for Society" "project number NdS/532684/2021/2022 amount of funding 13 500 zł total value of the project 1 355 390 zł".

References

1. SCHNEIDER ELECTRICS Virtual reality haptics training (2022). <https://www.gotouchvr.com/schneider-electrics>. Accessed 22 Aug 2022
2. ENEA OPERATOR (2019). <https://media.enea.pl/pr/430400/enea-operator-wykorzysta-wirtualna-rzeczywistosc-do-szkolenia-kadry-te>. Accessed 22 Aug 2022
3. Wu, H.-K., Lee, S.W.-Y., Chang, H.-Y., Liang, J.-C.: Current status, opportunities and challenges of augmented reality in education. *Comput. Educ.* **62**, 41–49 (2013). ISSN 0360–1315
4. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., Macintyre, B.: Recent advances in augmented reality. *IEEE Comput. Graph. Appl.* **21**(6), 34–47 (2001)
5. Elektrometal Energetyka SA, Rozdzielnicza średniego napięcia e2ALPHA – karta katalogowa (2022). <https://elektrometal-energetyka.pl/wp-content/uploads/2021/09/Karta-katalogowa-e2ALPHA-K-1.2.5-1.pdf>. Accessed 22 Aug 2022
6. IEC 62271:2022 High-voltage switchgear and controlgear
7. META QUEST 2. Oculus.com/quest-2 (2022). Accessed 22 Aug 2022





Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Virtual Reality Simulations of the Snake Robot

Anna Sibilska-Mroziewicz , Ayesha Hameed , Jakub Możaryn  ,
and Andrzej Ordys

Faculty of Mechatronics, Warsaw University of Technology,
ul. św. Andrzeja Boboli 8, 02-525 Warsaw, Poland
jakub.mozaryn@pw.edu.pl

Abstract. The following paper introduces a new way of presenting the results of engineering simulations. The object of consideration is the motion of the snake robot on a flat surface. The robot's trajectory and control signals are calculated in MATLAB. Different approaches have been presented to show how the robot moves - from 2D plots and 3D animations observed from a computer screen to realistic visualisations displayed in the Virtual Reality headset. The proposed VR simulation will allow watching the simulation results and manipulating simulation parameters from inside of VR.

Keywords: Virtual Reality · Snake robot · Simulations of multi-body systems

1 Introduction

Biomimetic robots are a class of robots that resemble a living organism's shape, appearance, or behaviour. They are designed to use biological principles in engineered systems to behave like a natural being, allowing them to solve specific problems, not possible for standard machines [1]. A snake robot is an example of a biomimetic, hyper-redundant robot with many degrees of freedom. Changes in the internal shape cause the snake robot's motion, similar to the natural biological creatures. Each robot configuration is characterized as a series of angles in joints connecting a series of robot segments [2].

Virtual Reality (VR) provides a platform to visualize the immersive behaviour of objects in a three-dimensional environment. Nowadays, VR is widely used in entertainment to enhance immersive movements. However, VR is also popular among researchers to construct complex environments in simulation and observe the behaviour of objects.

A comprehensive review of virtual reality interfaces for controlling and interacting with robotics is presented in [3]. The authors describe that VR interfaces are not only used for visualization but also for robotic interaction, planning, usability and infrastructure for both expert and non-expert users. However, this

review couldn't include biomimetic robotics. In [4] motion of haptic snakeserpentine shapes is investigated for multiple feedback, i.e. tapping and gesture feedback in virtual reality. However, literature related to snake robots in virtual reality is limited.

The main aim of this paper is to present a VR environment prototype to visualise the snake robot's motion. The proposed system can be used to evaluate different robot control algorithms.

This article is organized as follows. The Introduction section describes the importance of virtual reality simulations in snake robot design. Section 2 describes the snake model used in this paper. Section 3 presented a snake robot simulation in MATLAB. Section 4 presents results and simulations in MATLAB and Unity software. Finally, the conclusions and future works are given

2 Snake Robot Model

The model used in the article is based on widely used equations of snake robot motion on a flat horizontal surface. The dynamical model of the snake robot can be derived based on the torque equilibrium equation. A detailed description of the model can be found in [5] and [6]:

$$\mathbf{M}_\theta \ddot{\boldsymbol{\theta}} + \mathbf{W} \dot{\boldsymbol{\theta}}^2 - l \mathbf{S}_\theta \mathbf{K} \mathbf{f}_{R,x} + l \mathbf{C}_\theta \mathbf{K} \mathbf{f}_{R,y} = \mathbf{D}^T \mathbf{u} \quad (1)$$

where $\mathbf{S}_\theta = \text{diag}(\sin(\theta)) \in \mathbf{R}^{N \times N}$ and $\mathbf{C}_\theta = \text{diag}(\cos(\theta)) \in \mathbf{R}^{N \times N}$ are square matrices with trigonometric functions of link angles at the diagonal and zeros in the remaining elements, and link angles $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N]^T \in \mathbf{R}^N$ in global coordinate system. The vector $\mathbf{u} \in \mathbf{R}^{N-1}$ defines the controllable parameters - actuator torques exerted on successive links. The $\mathbf{f}_{R,x}$ and $\mathbf{f}_{R,y}$ vectors represent components of friction force on the links in global x and y direction.

The movement of the snake robot is possible due to anisotropic friction force. The friction coefficient in the longitudinal direction of each joint is much lower than the coefficient in the lateral direction. This property allows the robot joints to slide in the forward direction.

3 MATLAB Simulations

The robot's equations of motion have been implemented in MATLAB software and solved using the ode solver. MATLAB implementation also included control algorithms that allowed the snake robot head to reach a designated position. It is also possible to track the position of the robot's centre, but it is less useful in trajectory tracking problems.

A simple path-following method by the snake robot is called a Line-of-Sight (LoS) method [7]. According to this method, the robot is tracking a straight line. This approach requires the definition of the global coordinate system $\{x, y\}$ in which the x axis is aligned along with the forward movement. The implementation of LoS algorithm is available at [11, 12]. Full description of the program and implemented snake robot model can be found in [9].

The MATLAB Model Predictive Control Toolbox allowed the implementation and testing MPC algorithm in trajectory tracking. Two MPC variations have been tested - when all joints of the robot could move independently and when all joints were coupled in sinusoidal function. In the first case, the MPC algorithm calculated nine independent variables for ten segment robot. In the second case, there where only three variables - parameters of the function $a_1(\sin(a_2 + a_3 \cdot i))$ where i is the segment's number. Figure 1 shows the comparison of the resultant trajectory of the robot head for different control algorithms. During simulation the robot consisting 10 segments of length 0.2 m the robot had to achieve the following points one by one: (2.5 m, 0 m), (3 m, -1 m), (4 m, 0 m), (6 m, 1 m), (8 m, 1 m).

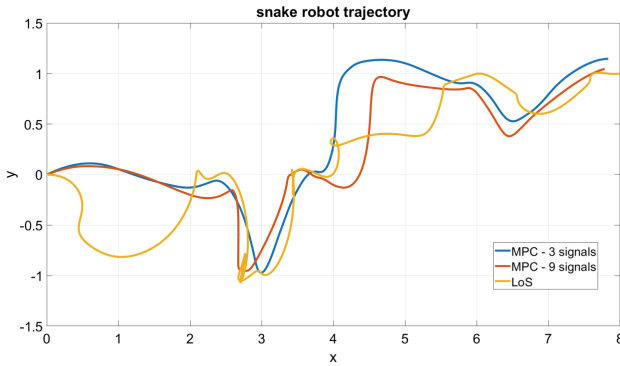


Fig. 1. Plot showing trajectory of the snake robot's head for different control algorithms.

The resultant position and orientation of snake robot links at each moment have been saved to the txt file for each control strategy. Based on this information, visualisation programs can fully restore the snake robot's motion

4 Visualisation of the Snake Robot Motion

4.1 Simulink 3D Animation

Visualization of robot movement is implemented using a MATLAB Simulink 3D Animation toolbox. The geometry of the robot segment is imported from the STL file created in SOLIDWORKS. The details of the segment design were described in [9]. The introduced framework allows simulating the robot's motion for different geometry parameters (e.g., number of joints, weight, friction coefficients), target trajectories (position of points to follow), and control parameters (e.g., head/centre tracking, controller gains, joints' disturbances). The ready-to-use code with comments, helping to run the software, is available in [11] and [12]. The main LiveScript file `start.mlx` contains a detailed description of the model and its implementation. The Simulink file `VRmodel12.slx` enables running 3D visualization of the snake motion shown in Fig. 2.

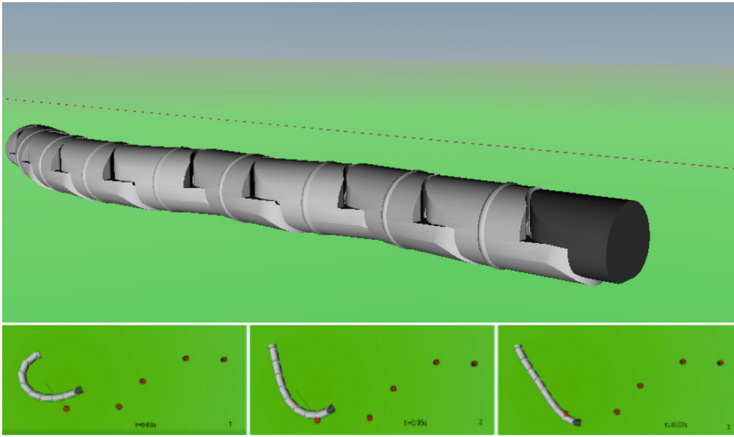


Fig. 2. Visualisation of the snake robot implemented in Simulink 3D Animation.

4.2 3D Simulations in Unity

The following work step was the implementation of the 3D visualisation of the snake robot using software dedicated to graphical animations. We used Unity Engine and Oculus Interaction SDKK [8]. The implemented program allows simulation in Oculus Quest 2 headset [10]. The user can interrupt the program execution using VR controllers. The visualisation program used the same CAD segments as the animation described in Sect. 4.1. The position and orientation of the robot segments in consecutive moments were interpolated using data read from a `txt` file generated by MATLAB. The virtual environment also displays the location of the reference point the snake is currently following (Fig. 3).

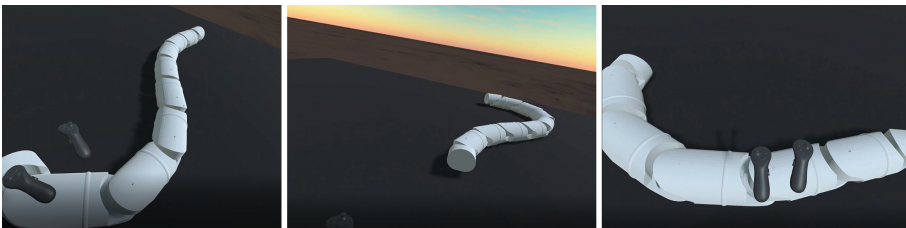


Fig. 3. Visualisation of the snake robot implemented in Unity.

The main advantage of VR visualisation is that the user can observe the snake robot's motion from any distance and perspective. The proposed solution allows users to catch any distortion of robot motion and get an idea of how would the real robot behave.

Aside from the segments position program can also visualise the velocity of each segment. For this purpose, an animation shows a point corresponding with the future segment's position. The current and future coordinates line describes the robot's velocity value and direction. Users can change the selected link at any time during the simulation (Fig. 4).

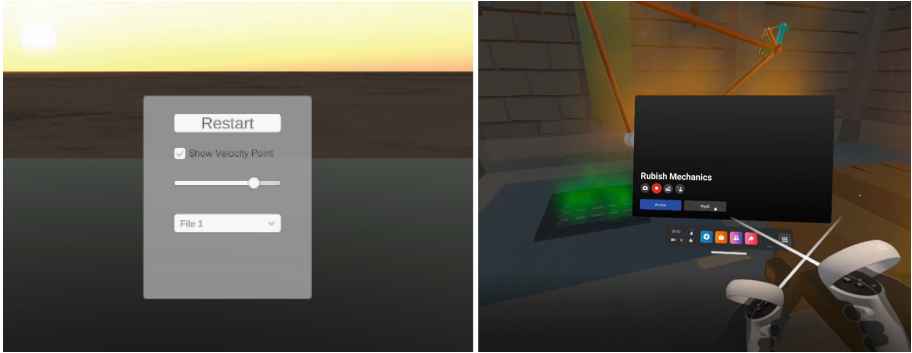


Fig. 4. Graphical User Interface and controllers ray allowing interaction with GUI.

By manipulating the Oculus Quest 2 controllers, the user can change the speed of the simulation, stop it or rewind it. There is also a Graphical User Interface (GUI) shown in VR, which allows manipulating the simulation time, restarting it or running different simulations by changing the file with the positions of robot segments. The GUI provides for changing the selected link for velocity visualisation and includes sliders, push, toggle, and a drop-down list. The user interacts with GUI elements by “ray” going out from the controller.

5 Conclusions

Virtual Reality is a very innovative and quickly developing technology. It offers the possibility of not only observing but also experiencing the virtual environment. Above entertainment, education and socialization, it brings unique opportunities in engineering research. The article shows a novel way of using VR for visualization of the motion of the snake robot. This approach can significantly affect the development of control algorithms for robotic systems.

Future work should concentrate on observing the snake robot's motion and interacting with it from VR. Visualisation in Unity and simulation in MATLAB will run parallel, and there will be established communication between both programs. The program will allow changing of the robot's trajectory from the interior of VR, so the observer wearing a VR headset could analyze the robot's motion and interact with it in real-time. The destination point that the robot should reach would be assigned by a user using the VR controller. The user will

mark the desired position of the robot on the floor and click the trigger button to record it. Unity would send the coordinates of the chosen point to MATLAB. Unit-MATLAB communication will be based on TCP protocols. The control algorithm implemented in MATLAB would calculate the target trajectory based on the received reference point. The computed positions of segments will be sent back to Unity, and the output robot's configuration will be displayed in VR. The new versions of the program GUI will include the algorithm selection and allow for parameter changes.

Acknowledgment. This work was supported from the grant financed by the Dean of Mechatronics Faculty of Warsaw University of Technology (grant number 504/04735/1143/44).

References

1. Siciliano, B., Khatib, O. (eds.): Springer Handbook of Robotics. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-32552-1>
2. Hirose, S.: Biologically Inspired Robots, in Snake-Like Locomotors and Manipulators, p. 1993. Oxford University Press, Oxford (1993)
3. Wonsick, M., Padir, T.: A systematic review of virtual reality interfaces for controlling and interacting with robots. *Appl. Sci.* **10**(24), 9051 (2020). <https://doi.org/10.3390/app10249051>
4. Al-Sada, M., Jiang, K., Ranade, S., Kalkattawi, M., Nakajima, T.: HapticSnakes: multi-haptic feedback wearable robots for immersive virtual reality. *Virtual Reality* **24**(2), 191–209 (2019). <https://doi.org/10.1007/s10055-019-00404-x>
5. Liljebäck, P., Pettersen, K.Y., Stavdahl, Ø., Gravdahl, J.T.: Snake Robots: Modelling, Mechatronics, and Control, pp. 29–37. Springer, London (2013). <https://doi.org/10.1007/978-1-4471-2996-7>
6. Sato, M., Fukaya, M., Iwasaki, T.: Serpentine locomotion with robotic snakes. *IEEE Control Syst.* **22**, 64–81 (2002). <https://doi.org/10.1109/37.980248>
7. Kelasidi, E., Liljebäck, P., Pettersen, K.Y., Gravdahl, J.T.: Integral line-of-sight guidance for path following control of underwater snake robots: theory and experiments. *IEEE Trans. Robot.* **33**, 610–628 (2017). <https://doi.org/10.1109/TRO.2017.2651119>
8. Oculus Integration at Unity Asset Store. <https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022>. Accessed 3 Dec 2021
9. Sibilska-Mroziewicz, A., Możaryn, J., Hameed, A., Fernández, M.M., Ordys, A.: Framework for simulation-based control design evaluation for a snake robot as an example of a multibody robotic system. *Multibody Syst. Dyn.* **55**(4), 375–397 (2022). <https://doi.org/10.1007/s11044-022-09830-3>
10. Meta Quest documentation. <https://developer.oculus.com/documentation>. Accessed 3 Dec 2021
11. GitHub repository. <https://github.com/asibilska/Snake-Robot-Locomotion-MATLAB->. Accessed 3 Dec 2021
12. Sibilska-Mroziewicz, A.: Snake-robot-locomotion-MATLAB, MATLAB central file exchange (2021). <https://uk.mathworks.com/matlabcentral/fileexchange/102910-snake-robot-locomotion-matlab>. Accessed 3 Dec 2021

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Prototype of Virtual Reality Game to Support Post-stroke Recovery in Patients with Spatial Neglect Syndrome

Katarzyna Matys-Popielska^(✉) , Krzysztof Popielski ,
and Anna Sibilska-Mroziewicz 

Department of Mechatronics, Warsaw University of Technology, Warsaw, Poland
katarzyna.matys.dokt@pw.edu.pl

Abstract. Stroke is the second cause of mortality and one of the leading causes of disability in adults. Post-stroke complications involve many different systems through which they involve difficulties in daily life. A very common complication that involves about 25–30% of post-stroke patients is spatial neglect syndrome, which involves the impaired perception of one's body and space. An important aspect of treatment for stroke patients is rehabilitation, both while still in the hospital and later in rehabilitation facilities as well as at home. Many studies have shown effective virtual reality (VR)-based general therapy systems after stroke. In particular, systems for motor function rehabilitation. In the following paper, a game proposal for the rehabilitation of patients with unilateral spatial neglect syndrome is shown. This game takes into account the specific perception and special motor skills of patients with spatial neglect syndrome. The described game was presented to a team of rehabilitation specialists working at the Department of Neurology and Stroke Unit of the University Clinical Hospital in Białystok and was evaluated by these specialists.

Keywords: Rehabilitation · Virtual reality in medicine · Stroke · Spatial neglect syndrome

1 Introduction

A stroke is a life-threatening emergency condition that results in acute symptoms of focal damage to the brain, spinal cord, or retina. It represents a very significant social problem. Every year 90 thousand people in Poland suffer a stroke, while in the world it is 17 million. Moreover, there is an increase in the number of cases in young and middle-aged people. An important side of this condition is also the social aspect. Stroke is the first cause of morbidity, long-term disability, and epilepsy in the elderly [1]. It is also the second cause of dementia and the second cause of death [2]. Complications following stroke include cardiac, pulmonary, gastrointestinal, musculoskeletal, neurological complications, and other not classified complications like fatigue, depression, or fever [3]. A complication that affects about 25–30% of patients is unilateral hemispheric neglect syndrome [4]. This is a disorder of spatial attention in which the perception of and response to stimuli

from the part of the body opposite to the location of the stroke focus are reduced. The consequence of this syndrome is increased length of hospitalization and increased cost of care [5]. This is due to the need for comprehensive rehabilitation [6].

Community-based, outpatient, and home-based rehabilitation enable improved functioning, acquisition of new self-care skills, and partial or complete recovery of independence [7]. The extent of rehabilitation should always be tailored to the needs of the patient. The goal of physiotherapy in a patient with a motor deficit, whether resulting from damage to the primary motor cortex or more complex, is to restore motor skills or compensate for them. Movement deficits can be addressed directly by movement therapy with active patient participation. The guidelines for stroke management mention rehabilitation with virtual reality, indicating that there are high hopes for this form of rehabilitation [6]. Studies show that using immersive VR for the general rehabilitation of stroke patients improves their balance, reduces the risk of falls, and improves the perception of visual verticality [8]. Other studies indicate that VR can reduce upper limb motor disabilities [9, 10] and can encourage physical activity and social participation [11, 12].

2 Methods

2.1 Rehabilitation in Semi-neglect Syndrome

Neurological physiotherapists indicate that during exercises for unilateral atrophy syndrome, various stimuli should be involved: auditory and visual, which should direct the patient's attention to the neglected side. It is important that both limb exercises on the neglected side and visuospatial search training be adapted to the patient's altered or, if possible, corrected midline. The eye or limb movement should be progressively performed from the non-skipped side towards the skipped side (to a line or fixed point/object that is clearly visible to the patient) [13].

The first exercises should be based on grasping objects with the non-neglected side and transferring them, by crossing the midline, to the neglected side. Subsequent exercises can also be based on grasping objects with the neglected side and transferring them to the neglected side - the alternating handwork of crossing the midline is intended to make the brain aware of the neglected side again.

2.2 VR-Based Application for After-Stroke Complications

Based on the above physiotherapists' guidelines, a VR game was developed (Fig. 1.). It was located in a quiet, open forest area, which allows the patient to relax while receiving physiotherapy treatments. The aim of the game is to collect apples placed on the trees and place them in the boxes appearing on the opposite side. It is important to note that red apples can only be picked with the right hand and green apples with the left hand (if you try to grab an apple with the opposite hand, the box will not appear).



Fig. 1. VR game for a patient with hemineglect, unilateral neglect syndrome

Figure 1 shows that the patient's task is to grab a green apple with his left hand and move it to the crate on the right. During this movement, there is a crossing of the midline and a visual search of the space on the neglected side. Additionally, correctly locating the box and hitting it with the apple is reinforced with a haptic stimulus in the form of a vibrating controller. At the same time, the patient may also attempt to grasp the red apples with the other hand and move them to the opposite side. The ability to grasp with both the active and inactive side also helps to train manual dexterity of the neglected side.

During the game, the patient can check his progress by monitoring the statistics: the number of green apples collected, the number of red apples collected, the percentage of apples correctly thrown, and the average time from collecting any/red/green apples to throwing them into the box.

The game created in this way was presented to a team of neurological rehabilitation specialists working at the Department of Neurology and Stroke Unit of the University Clinical Hospital in Białystok, composed of Agnieszka Zieziula, M.Sc. in physiotherapy, Izabela Zalesko, M.Sc. in physiotherapy, and Justyna Karpińska, M.Sc. in physiotherapy (Fig. 2.).



Fig. 2. Game testing by a team of physiotherapists

3 Results and Conclusion

After testing, the game was commented on by a team of physiotherapists. In their opinion, it has potential in the rehabilitation of patients with upper limb paresis as well as a certain group of patients affected by unilateral spatial neglect syndrome. In the case of unilateral atrophy, the target group could be patients in the home rehabilitation stage. Among the advantages of the developed application, physiotherapists emphasize the correct course of movement of the upper limb, as well as the frequent crossing of the center line. In addition to the physiotherapeutic elements, specialists also note the pleasant environment, which can contribute to prolonged physiotherapy time. This is consistent with many other studies that indicate the effectiveness of using VR games in rehabilitating various neurological conditions, such as Parkinson's disease [14], spinal cord injury [15], and phantom pain [16]. In addition to improvements in physical performance, many studies also indicate greater satisfaction with physiotherapy among patients and more willingness to exercise, which translates not only into improvements in physical fitness but also in mental health [17, 18]. In the case of hospital patients, a simpler version of the game would be necessary - exclusion of movement in the game, limitation of control to one hand, and higher color contrast. This is due to the greater limitations of these patients and the need for better adaptation to their requirements.

3.1 Further Steps

It is planned to develop the application so that it is adapted to the needs of patients in hospital neurological and stroke wards. As part of the customization of the app, it

is primarily planned to simplify the game. In the game under development, the patient should not have to move around the environment. In addition, the game should be able to be operated with a single controller. Significantly from the point of view of physiotherapy, the destination - the box, should appear in different locations.

References

1. Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.: Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747–1757 (2006). [https://doi.org/10.1016/S0140-6736\(06\)68770-9](https://doi.org/10.1016/S0140-6736(06)68770-9)
2. Kim, W.-S., et al.: Clinical application of virtual reality for upper limb motor rehabilitation in stroke: review of technologies and clinical evidence. *J. Clin. Med.* **9**, 3369 (2020). <https://doi.org/10.3390/jcm9103369>
3. Kumar, S., Selim, M.H., Caplan, L.R.: Medical complications after stroke. *Lancet Neurol.* **9**, 105–118 (2010). [https://doi.org/10.1016/S1474-4422\(09\)70266-2](https://doi.org/10.1016/S1474-4422(09)70266-2)
4. Gammeri, R., Iacono, C., Ricci, R., Salatino, A.: Unilateral spatial neglect after stroke: current insights. *Neuropsychiatr. Dis. Treat.* **16**, 131–152 (2020). <https://doi.org/10.2147/NDT.S171461>
5. Kot-Bryćko, K., Pietraszkiewicz, F.: *Psychologia w medycynie. Część 2 – rehabilitacja neuropsychologiczna po udarze mózgu*, pp. 344–347 (2012)
6. Błażejewska-Hyżorek, B., et al.: Wytoczne postępowania w udarze mózgu. *Pol. Przegląd Neurol.* **15**, 1–156 (2019). <https://doi.org/10.5603/PPN.2019.0001>
7. Legg, L., Langhorne, P.: Rehabilitation therapy services for stroke patients living at home: a systematic review of randomized trials. *Lancet* **363**, 352–356 (2004). [https://doi.org/10.1016/S0140-6736\(04\)15434-2](https://doi.org/10.1016/S0140-6736(04)15434-2)
8. Cortés-Pérez, I., Nieto-Escamez, F.A., Obrero-Gaitán, E.: Immersive virtual reality in stroke patients as a new approach for reducing postural disabilities and falls risk: a case series. *Brain Sci.* **10**, 296 (2020). <https://doi.org/10.3390/brainsci10050296>
9. Fang, Z., et al.: Effect of traditional plus virtual reality rehabilitation on prognosis of stroke survivors. *Am. J. Phys. Med. Rehabil.* **101**, 217–228 (2022). <https://doi.org/10.1097/PHM.0000000000001775>
10. Rodríguez-Hernández, M., Criado-Álvarez, J.-J., Corregidor-Sánchez, A.-I., Martín-Conty, J.L., Mohedano-Moriano, A., Polonio-López, B.: Effects of virtual reality-based therapy on quality of life of patients with subacute stroke: a three-month follow-up randomized controlled trial. *Int. J. Environ. Res. Public Health* **18**, 2810 (2021). <https://doi.org/10.3390/ijerph18062810>
11. Mekbib, D.B., et al.: Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. *Brain Inj.* **34**, 456–465 (2020). <https://doi.org/10.1080/02699052.2020.1725126>
12. Maggio, M.G., et al.: Virtual reality and cognitive rehabilitation in people with stroke: an overview. *J. Neurosci. Nurs.* **51**, 101–105 (2019). <https://doi.org/10.1097/JNN.00000000000000423>
13. Konkul, M., Drozd, A., Nowacka-kłos, M., Hansdorfer-korzon, R., Barna, M.: Zespół pomijania stronnego u pacjentów po udarze mózgu — przegląd metod fizjoterapeutycznych. *Forum Med. Rodz.* **9**, 405–415 (2015)
14. Lei, C., et al.: Effects of virtual reality rehabilitation training on gait and balance in patients with Parkinson’s disease: a systematic review. *PLoS ONE* **14**, e0224819 (2019). <https://doi.org/10.1371/journal.pone.0224819>

15. de Araújo, A.V.L., Neiva, J.F.D.O., Monteiro, C.B.D.M., Magalhães, F.H.: Efficacy of virtual reality rehabilitation after spinal cord injury: a systematic review. *Biomed. Res. Int.* **2019**, 1–15 (2019). <https://doi.org/10.1155/2019/7106951>
16. Osumi, M., Inomata, K., Inoue, Y., Otake, Y., Morioka, S., Sumitani, M.: Characteristics of phantom limb pain alleviated with virtual reality rehabilitation. *Pain Med.* **20**, 1038–1046 (2019). <https://doi.org/10.1093/pm/pny269>
17. Sveistrup, H.: Motor rehabilitation using virtual reality. *J. Neuroeng. Rehabil.* **1**, 10 (2004). <https://doi.org/10.1186/1743-0003-1-10>
18. Rose, T., Nam, C.S., Chen, K.B.: Immersion of virtual reality for rehabilitation - review. *Appl. Ergon.* **69**, 153–161 (2018). <https://doi.org/10.1016/j.apergo.2018.01.009>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Author Index

A

Ajenaghughrure, Ighoyota Ben 127
Asesh, Aishwarya 26

B

Bäcker, Niklas 139
Balki, Fatih 111
Berdek, Kacper 298
Białorucki, Przemysław 242
Borkiewicz, Paulina 202
Brömmling, Lukas 139
Brzezinski, Piotr 35
Buchem, Ilona 139
Buczek, Aleksandra 213
Buczowski, Przemysław 35

C

C. Firmino De Souza, Debora 127
Cam, Veli 111
Carrasco Limeros, Sandra 93
Chakraborty, Somenath 67
Chaniecki, Zbigniew 213
Cnotkowski, Daniel 202

D

Daszczyński, Tadeusz 298
Du, Huaiyu 45

F

Farooq, Muhammad Umar 12

G

Goibenko, Iryna 76, 103
Grudzień, Krzysztof 213
Grzeszczuk, Maciej 287

H

Hameed, Ayesha 307

J

Jaskulska, Anna 150, 176, 287
Jeanty, Mathias 213

Jędrzejewski, Zbigniew 150
Jóźwiak, Rafał 45, 76, 85, 103

K

Karpowicz, Barbara 176
Kjellberg, Magnus 93
Klimaszewski, Jan 242, 250, 270
Kobyliński, Paweł 202
Kopeć, Wiesław 150, 176, 287
Kornacka, Monika 176, 287
Kozłowski, Marek 35
Kupiński, Kamil 213

L

Lapin, Kristina 191
Leischner, Vojtěch 3
Lorenc, Tomasz 76, 103

M

Majchrowska, Sylwia 93
Maslyk, Rafał 150
Matys-Popielska, Katarzyna 281, 314
Mitura, Jakub 76, 103
Możaryn, Jakub 233, 307
Murali, Beddhu 67
Mykhalevych, Ihor 76, 85, 103

N

Nagajek, Damian 167
Naruszewicz, Dariusz 298
Naydonov, Mykhaylo 159
Naydonova, Lyubov 159

O

Ordys, Andrzej 307
Oseka, Laura 202, 222
Ostaszewska-Lizewska, Anna 250

P

Pabiś-Orzeszyna, Michał 202
Piotrowski, K. 56
Piotrowski, Krzysztof 167

Pochwatko, Grzegorz 150, 159, 176, 202,
222, 287
Popielski, Krzysztof 281, 314

R

Rąpała, M. 56
Rąpała, Michał 167
Raza, Rana Hammad 12
Romanowski, Andrzej 213
Rosén, Anna 93
Różańska-Walczuk, Monika 260

S

Sarp, Salih 111
Sibilska-Mroziewicz, Anna 281, 307, 314
Sjöblom, Lisa 93
Skorupska, Kinga 150, 176, 287
Sobecki, Piotr 76, 85, 103
Sobiech, Franciszek 213
Sultan, Zamra 12

Suvilehto, Juulia 93
Świdrak, Justyna 150, 159, 222

T

Tikka, Pia 127
Tupikowski, Krzysztof 76, 103
Turchan, K. 56
Turchan, Krzysztof 167

V

Volungevičiūtė, Laima 191

W

Walczak, Natalia 213
Wierzbowski, Mariusz 202
Wotoszyn, Kamil 56, 167

Y

Yildirim, Enver 111

Z

Zoubi, Mohamad Khir 93