



**Proceedings of the  
10th Human-Computer Interaction Slovenia  
Conference**

University of Primorska  
Faculty of Mathematics, Natural Sciences  
and Information Technologies  
Koper, Slovenia  
**13 October 2025**

**Proceedings of the 10th Human-Computer Interaction  
Slovenia Conference**

Koper, Slovenia | 13 October 2025

Organised by University of Primorska,  
Faculty of Mathematics, Natural Sciences  
and Information Technologies

Published by University of Primorska Press  
Titov trg 5, 6000 Koper, Slovenia  
Koper | 2025

 © 2025 Authors

Electronic Edition

<https://www.hippocampus.si/ISBN/978-961-293-559-7.pdf>

<https://doi.org/10.26493/978-961-293-559-7>

Kataložni zapis o publikaciji (CIP) pripravili  
v Narodni in univerzitetni knjižnici v Ljubljani  
COBISS.SI-ID 266380291  
ISBN 978-961-293-559-7 (PDF)

## Kazalo vsebine/Table of Contents

|                                                                                                                                                                                                |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Uvodni del zbornika/Front Matter</i> .....                                                                                                                                                  | 1   |
| <i>Glavni govorec/Keynote speaker</i> .....                                                                                                                                                    | 2   |
| <i>Organiztorji/Organisers</i> .....                                                                                                                                                           | 3   |
| <i>Programski odbor/Programme Committee</i> .....                                                                                                                                              | 3   |
| <i>Translation of Situation Awareness Rating Technique Questionnaire in Slovenian</i><br>Kristina Stojmenova Pečecnik and Grega Jakus .....                                                    | 4   |
| <i>Učenje jazz glasbe v navidezni resničnosti</i><br>Helena Jeretina and Matevž Pesek .....                                                                                                    | 16  |
| <i>Razvoj didaktične mobilne igre za vzpodbujanje kreativnosti</i><br>Emir Hodžić and Matevž Pesek .....                                                                                       | 30  |
| <i>Implementation of the Science on a Sphere Visualization System as a Web Application</i><br>Jurij Anžič and Ciril Bohak .....                                                                | 40  |
| <i>Visualization of 3D Ultrasound Uterine Data in Virtual Reality</i><br>Ilija Gavrilović and Ciril Bohak .....                                                                                | 50  |
| <i>Human-Computer Interaction in Slovenia: A Retrospective and Trend Analysis of Local Research</i><br>Ciril Bohak .....                                                                       | 59  |
| <i>Razvoj in uporabniška študija mobilne aplikacije za vadbo poliritmov</i><br>Jaka Kužner, Matevž Pesek and Matija Marolt .....                                                               | 74  |
| <i>Comparison of Unity and FMOD Libraries for Spatial Audio Localization in Virtual Reality</i><br>Gašper Leskovec, Eva Gabersček and Jaka Sodnik .....                                        | 88  |
| <i>Web Implementation of 3-way Chess</i><br>Janez Koprivec, Matija Marolt and Ciril Bohak .....                                                                                                | 97  |
| <i>Real-Time Gesture Transmission with a Robotic Hand: Embodied Signals for Non-Verbal Remote Communication</i><br>Lea Pajnič, Matjaž Kljun, Maheshya Weerasinghe and Klen Čopič Pucihar ..... | 107 |
| <i>Designing the Ideal Political Identity Questionnaire Using Machine Learning and Ideology Scales</i><br>Ana Nikolić, Uroš Sergaš and Marko Tkalčič .....                                     | 118 |
| <i>Optimizing Product Catalogue Design: A Comparative Study of Traditional Photography and 3D Modeling</i><br>Simon Kolmanič, Jan Hrašar, Štefan Horvat and Domen Mongus .....                 | 131 |
| <i>Integration of Hybrid Animation in a 360-degree Environment</i><br>Aleksandar Ilievski, Suzana Žilič Fišer and Simon Kolmanič .....                                                         | 141 |

|                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>NERVIS: An Interactive System for Graph-Based Exploration and Editing of Named Entities</i><br>Uroš Šmajdek and Ciril Bohak .....                                      | 147 |
| <i>Transparent Persona Generation With LLMs: An Evidence-based and Traceable Method for User-centred Design</i><br>Bojan Blažica, Manca Topole and Marko Debeljak .....   | 157 |
| <i>Inferring a Mobile User's Valence and Arousal through On-Screen Text Analysis</i><br>Edita Džubur and Veljko Pejović .....                                             | 171 |
| <i>Exploring the Effects of Multimodal User Interfaces in Autonomous Vehicles</i><br>Kristina Stojmenova, Timotej Gruden, Grega Jakus, Sašo Tomažič and Jaka Sodnik ..... | 181 |
| <i>Sodelujoči avtorji/Contributing Authors .....</i>                                                                                                                      | 187 |

# Uvodni del zbornika/Front Matter



**Klen Čopić Pucihar**

Predsednik konference/Conference chair

Univerza na Primorskem/Università del Litorale/University of Primorska

klen.copic@famnit.upr.si

## Nagovor/Welcome

### Slovensko

Z velikim veseljem gostimo v Kopru konferenco HCI SI 2025, ki je posvečena slovenskemu razvoju področja interakcije človek-računalnik (IČR). Še posebej nas veseli, da letos praznujemo deset let skupnosti, ki se je organizirala v okviru slovenskega poglavja ACM SIGCHI Bled, sodelovanja ter razvoja raziskav IČR v Sloveniji.

Od prvih korakov v okviru multikonference Informacijska Družba do vzpostavitve samostojne konference je HCI SI postal letni zbor raziskovalcev, študentov in strokovnjakov iz Slovenije in tujine. V zadnjem desetletju se je konferenca stalno širila po številu udeležencev in po raznolikosti obravnavanih tem. Predstavljene raziskave so vključevale teme kot so uporabnost, vizualizacija informacij, razširjena resničnost, afektivno računalništvo, prepričljive tehnologije, interakcija z umetno inteligenco, izobraževanje, kulturna dediščina in avtomobilski vmesniki. Tudi letošnji zbornik ohranja raznolikost in odraža skupnost, ki dozoreva, a ostaja živahna in odprta za novosti.

Iskrena zahvala gre vsem avtorjem, organizacijskemu odboru, programskemu odboru, glavnemu govorniku in prostovoljcem, ki s svojo predanostjo vsako leto omogočate izvedbo konference HCI SI. Vaš prispevek bogati dogodek in krepi prepoznavnost slovenske IČR skupnosti.

Dobrodošli na HCI SI 2025.

### English

It is our pleasure to host HCI SI 2025 in Koper. As we gather for another edition of Slovenia's Human-Computer Interaction (HCI) conference, we celebrate a decade of community building within the ACM SIGCHI Chapter Bled, collaboration, and continuous growth of Slovenian HCI research.

Since its early beginnings within the Information Society multiconference and its evolution into a standalone venue, HCI SI has become an annual meeting point for researchers, students, and practitioners across Slovenia and beyond. Over the past ten years, the conference has witnessed steady expansion in both participation and thematic diversity. The research presented included usability, information visualisation, studies using XR, affective computing, persuasive technologies, AI-driven interaction, education, cultural heritage, and automotive UI. This year's proceedings continue this tradition and reflects a community that is both maturing and renewing itself.

We are grateful to the authors, organisation committee, program committee, keynote speaker and volunteers whose commitment makes HCI SI possible each year. Your work strengthens this venue and visibility of Slovenian HCI.

Welcome to HCI SI 2025.

## Glavni govorec/Keynote speaker

### Slovensko



Letošnji osrednji govorec je bil **prof. dr. Jürgen Ziegler**. Jürgen je redni profesor na Fakulteti za računalništvo Univerze Duisburg–Essen, kjer raziskuje na področju inteligentnih interaktivnih sistemov. Njegovi glavni raziskovalni interesi zajemajo interakcijo človek–računalnik, sodelovanje med človekom in umetno inteligenco, priporočilne sisteme, družbena omrežja ter vizualizacijo informacij.

#### Naslov osrednjega predavanja:

Znmanjšanje negotovosti v inteligentnih interaktivnih sistemih

#### Povzetek osrednjega predavanja:

Razmah tehnologij umetne inteligence prinaša nove, uporabne funkcionalnosti interaktivnim sistemom, vendar lahko hkrati poveča negotovost med uporabniki. Medtem ko tradicionalni interaktivni sistemi povzročajo negotovost predvsem zaradi neustrezno zasnovanih funkcionalnosti in uporabniških vmesnikov, inteligentni sistemi negotovost vnašajo že po svoji naravi, saj temeljijo na verjetnostnih metodah, ki pogosto delujejo znotraj netransparentnih modelov »črne skrinjice«. Negotovost lahko občutno vpliva na uporabniško izkušnjo in zaupanje v sistem ter ima škodljive posledice na individualni ali družbeni ravni. V predavanju je Jürgen predstavil tehnične in kognitivne vire negotovosti ter načine za soočanje z njimi z vidika uporabnika. Na primeru priporočilnih sistemov je predlagal načela za zmanjševanje negotovosti v interakciji med uporabnikom in sistemom. Predavanje se je osredotočilo na pristope za boljše razumevanje odločitvenega prostora, omogočanje uporabniškega nadzora in raziskovanja ter na prehod od razlaganja delovanja umetne inteligence k sodelovalnim sistemom človek–agent.

### English

This year's keynote speaker was **Prof. Dr. Jürgen Ziegler**. Jürgen is a senior full professor in the Faculty of Computer Science at the University of Duisburg–Essen where he is conducting research in the field of Intelligent Interactive Systems. His main research interests lie in the areas of human-computer interaction, human-AI cooperation, recommender systems, social media, and information visualization.

#### Keynote title:

Bridging the Gulf of Uncertainty in Intelligent Interactive Systems

#### Keynote abstract:

The surge in AI technologies introduces new, useful functionality to interactive systems but may also increase uncertainty for their users. While traditional interactive systems cause uncertainty for users mainly through inadequately engineered functionality and user interfaces, intelligent systems inherently introduce uncertainty by applying probabilistic methods that mostly operate within opaque black box models. Uncertainty can significantly impair the user experience and trust in the system and have harmful effects on an individual or societal level. In this talk, Jürgen discussed technical and cognitive sources of uncertainty and how to cope with them from a user-centric perspective. Using recommender systems as an example, he proposed principles for bridging the 'gulf of uncertainty' in user–system interaction. Specifically, the talk addressed approaches to promoting user understanding of the decision space, enabling user control and exploration, and progressing from explaining AI functions towards co-operative human-agent systems.

## Organiztorji/Organisers

Predsednik konference/Conference chair:

- Klen Čopič Pucihar

Programski predsednik/Program chair:

- Marko Tkalčič

Web chair/Skrbnik spletne strani:

- Nilukshan Krishnaram

Lokalni organizatorji/Local organizers:

- Karolina Trajkovska
- Michelle Fernando

Uredniki zbornika/Proceedings chairs:

- Arsen Matej Golubovikj
- Kosar Seyyedhosseinzadeh
- Nadun Liyanage

Vodji promocije/Promotion Chairs:

- Nilukshan Krishnaram
- Jordan Aiko Deja

## Programski odbor/Programme Committee

Nuwan Attygalle: Univeristy cathlique de Louvain, Belgium

Bojan Blažica: Jožef Stefan Institute, Slovenia

Ciril Bohak: University of Ljubljana, FRI, Slovenia

Niko, Caluya: Ritsumeikan University, College of Information Science and Engineering, Japan

Cuauhtli Campos: Cosy Lab, Slovenia

Luka Čehovin: University of Ljubljana, FRI, Slovenia

Klen Čopič Pucihar: University of Primorska, FAMNIT, Slovenia

Stellenbosch University, South Africa

Jordan Aiko Deja: De La Salle University, Philippines

Dario Di Dario: University of Salerno, Italy

Arsen Matej Golubovikj: University of Primorska, FAMNIT, Slovenia

Jože Guna: University of Ljubljana, NTF, Slovenia

Andor Gužvanj: Zagreb University of Applied Sciences, Croatia

Grega Jakus: University of Ljubljana, FE, Slovenia

Matjaž Kljun: University of Primorska, FAMNIT, Slovenia

Stellenbosch University, South Africa

Ines Kozuh: University of Maribor, FERI, Slovenia

Kayhan Latifzadeh: University of Luxemburg, Luxemburg

Irena Lovrenčič Držanič: University of Maribor, FERI, Slovenia

Sergej Lugović: Zagreb University of Applied Sciences, Croatia

Geert Lugtenberg: NAIST, Multimedia Lab, Japan

Utrecht University, The Netherlands

Matevž Pesek: University of Ljubljana, FRI, Slovenia

Briane Paul Samson: De La Salle University, Philippines

Kosar Seyyedhosseinzadeh: University of Primorska, FAMNIT, Slovenia

Yuki Shimizu: Nara Institute of Science and Technology, Multimedia Lab, Japan

Kristina Stojmenova Pečecnik: University of Ljubljana, FE, Slovenia

Marko Tkalčič: University of Primorska, FAMNIT, Slovenia

Maheshya Weerasinghe: University of Primorska, FAMNIT, Slovenia

Juri Yoneyama: INRIA, France

# Translation of Situation Awareness Rating Technique Questionnaire in Slovenian

Kristina Stojmenova Pečičnik<sup>1\*</sup> and Grega Jakus<sup>1</sup>

<sup>1</sup> University of Ljubljana, Faculty of Electrical Engineering, Tržaška cesta 25, Ljubljana, 1000, Slovenia

## Abstract

The Situational Awareness Rating Technique (SART) is a questionnaire designed to assess situational awareness when operating a machine. It consists of ten questions corresponding to ten dimensions that can be used to assess an operator's situational awareness when interacting with and operating a machine. Although initially developed for assessment situational awareness of pilots, with the rising complexity of vehicles and introduction of automation, its use has spread significantly in automotive domain in the past decade. In this paper, we present the Slovenian version of the original questionnaire and the process of its development. For the translation of the questionnaire SART into Slovenian, the standard procedures for the translation of research instruments were applied: translation, coordination, back-translation into the source language, comparison of the back-translation with the source text and final coordination. Lastly, it was validated against a reference system in a driving-based environment.

## Keywords

Situational awareness rating technique, situational awareness, in-vehicle human-machine interaction, questionnaire, translation

## 1. Introduction

Situational awareness (SA) or knowing what is going on around you is essential in any dynamic human decision-making process because it provides the level of knowledge needed to make informed decisions and take appropriate actions [1]. There is not a single agreed upon definition of situational awareness, but the three most commonly used definitions (see **Table 1**) all seem to refer to three aspects that constitute a situationally aware operator: gathering information from the environment to obtain a knowledge of the situation, interpreting the perceived information to understand its meaning in relation to the observed system, and being able to plan or project the next step of the system's operation.

All these aspects indicate that SA is a cognitive construct that requires multiple resources rather than a single one. As such, SA is also inextricably linked to other cognitive theories, such as attention, short term and long-term memory, and cognitive workload. Cognitive workload is often defined as a function of the supply and demand of attentional and processing resources [2].

It is constrained by the operator's limited short-term memory (working memory), and processing resources are influenced by the operator's domain knowledge in long-term memory [3]. More experienced operators have a broader range of skills and can therefore process larger or more complex information in their working memory. At the same time, as the workload increases, more attention is required for task performance, leaving fewer resources for situational awareness [2]. In this regard, SA competes with task performance for attentional and processing resources. Assessing situational awareness is therefore not an easy task and requires the observation of several psychological constructs.

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

✉ kristina.stojmenova@fe.uni-lj.si (K. S. Pečičnik); grega.jakus@fe.uni-lj.si (G. Jakus)

ORCID 0000-0001-6584-7147 (K. S. Pečičnik); 0000-0001-9373-7885 (G. Jakus)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

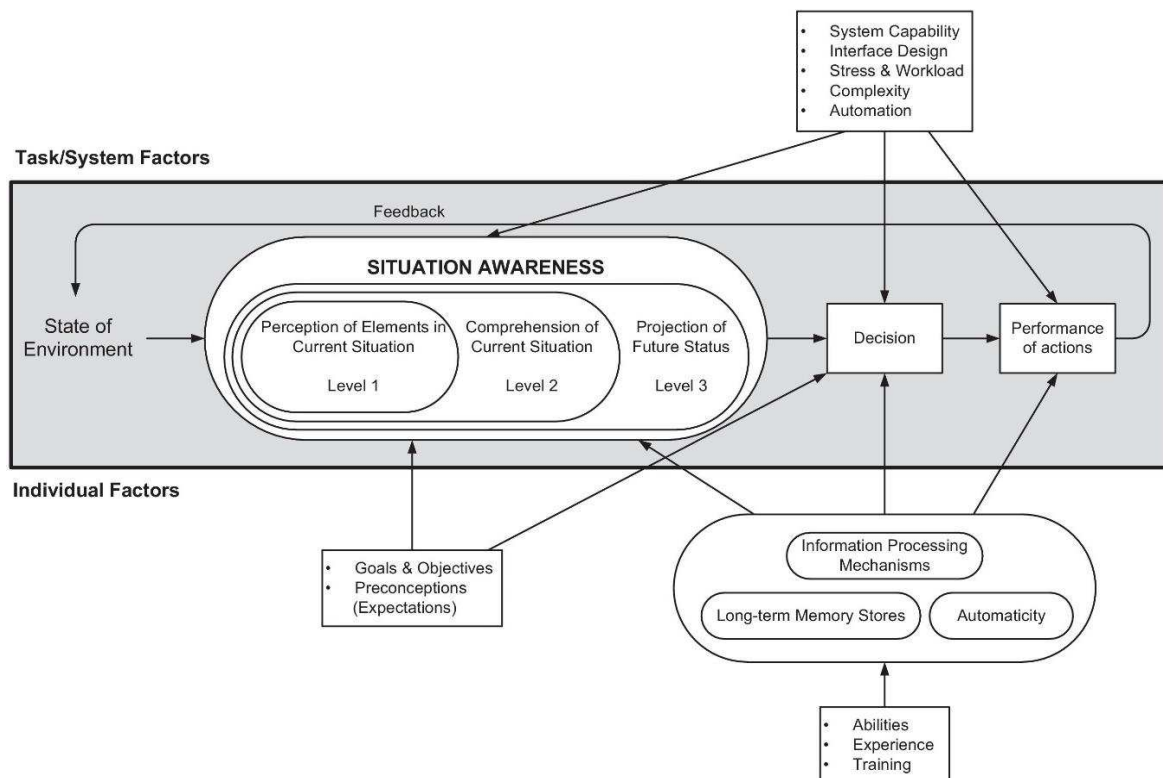
<https://doi.org/10.26493/978-961-293-559-7.1>

**Table 1**

Definitions of situational awareness

| Author                     | SA definition                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Endsley, 1988. [4]         | "Situational awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and a projection of their status in the near future."                                                                                                                                                                                                                                      |
| Smith & Hancock, 1995. [5] | "Situational awareness is the invariant in the agent-environment system that generates the momentary knowledge and behavior required to attain the goals specified by an arbiter of performance in the environment."                                                                                                                                                                                                                         |
| Bedny & Meister, 1999. [6] | "Situational awareness is the conscious dynamic reflection on the situation by an individual. It provides dynamic orientation to the situation, the opportunity to reflect not only the past, present and future, but the potential features of the situation. The dynamic reflection contains logical-conceptual, imaginative, conscious and unconscious components which enables individuals to develop mental models of external events." |

Although there is more than one theory, the most popular theory of situational awareness proposes a three level SA model (**Figure 1**), which suggests that achieving SA requires perceiving the elements of the environment (level 1), understanding on their meaning (level 2), and being able to project their status in the near future (level 3) [2]. Based on this model, numerous assessment procedures have been developed that aim to capture all aspects of SA.

**Figure 1:** Situational awareness model [7].

The three broad groups into which they can be classified are self-assessment, inferential and query procedures. Self-assessment procedures, usually based on questionnaires and scales, survey participants' self-perceptions to look for subjective signs of SA. Inference-based techniques look for implicit clues about the presence or absence of SA in the context of an individual's performance and behavior. Finally, query techniques look for direct evidence of the

content of SA. This approach provides a set of information about the person's perception and understanding of the situation, which is then compared to an established ground truth.

The first and still most widely used query method for assessing SA is the Situation Awareness Global Assessment Technique or SAGAT [8],[9]. SAGAT was originally used to assess SA of industrial machinery operators, but was later used and adapted for numerous other domains. The operator is presented with a simulation of a device they are supposed to operate. Using a frame-freeze technique, the simulation is paused at random critical situations and the operator is asked various machine- and environment-related questions to determine his or her SA at that moment. The operator's answers are then compared to predefined correct answers that attempt to cover all three levels of SA. Since stopping the process for each query can significantly affect the operator's behavior and responses, the frame-freeze query method does not allow realistic observation of a process in its natural environment. As a more suitable alternative to the query-based approach, a real-time-based method have been proposed, such as the Situation Present Assessment Method (SPAM) [10]. SPAM distinguishes the workload from SA by alerting the operator that a question (similar to the ones used in SAGAT) is waiting while they are actually operating the machine. The time to accept the question, the time to answer the question, and the error rate are then evaluated as indicators of SA.

Because of its simplicity and cost-effectiveness, another common method of assessing SA that does not affect the operation of the machine, is self-assessment through the use of questionnaires after a task has been completed. One of the most popular and widely used methods is the Situational Awareness Rating Technique or SART [11],[12], which was originally developed to assess the situational awareness of pilots. However, it has also been used to assess situational awareness of operators in a variety of other domains, as the dimensions of SART are general in nature, and attempt to provide an insight into the three most important domains of SA: attentional demand, attentional supply and understanding of the perceived information.

### **1.1. Situational Awareness in Vehicles**

SA has been recognized as an important factor in the automotive domain as it provides important information on driver behavior, and is helpful in the process of design and evaluation of advanced driver assistance systems (ADAS). Driving as a task can be especially demanding as it consists of more tasks that run concurrently (maintain longitudinal and lateral control, follow traffic rules, adapt to other traffic participants, adapt to weather and road conditions, etc.). With increased automation some of these tasks are controlled by the vehicle, which makes the driving task easier, but at the same time, in moments when the driver has to take over control of the vehicle, makes regain of SA much harder. Some of the ADASs such as parking sensors, back rear camera and blind spot sensor systems were indeed introduced to increase driver's SA. A lot of other ADAS for obtaining longitudinal (adaptive cruise control) and lateral control (lane keep assist) on the other hand, assist the performance of the driving task in terms of safety and comfort by reducing the driver's control of the vehicle. This opens space for performance of non-driving related tasks, which can make the driving process more entertaining and less demanding, however it can also negatively affect driving safety.

Increase of ADASs in vehicles thus reveals the need to identify potential negative side effects as of any new technology introduced to the vehicle and implement better human-machine interaction (HMI) models to exploit their benefits while eliminating (or at least reducing) negative influences on driving safety. Using methods such as SART can contribute to this understanding.

## **2. Situational Awareness Rating Technique (SART)**

SART is a questionnaire-based technique consisting of ten questions that ask the operator about different dimensions of situational awareness: instability, complexity and variability of situation, arousal, spare mental capacity, concentration and division of attention, information quantity and

quality, and familiarity with the situation. The operator answers the questions on a seven-point rating scale, where 1 = Low and 7 = High, to indicate their assessment rating of the situation about the specific dimension. The ten questions correspond to ten dimensions of SA, which are then grouped into the three SA constructs: attentional demand, attentional supply and understanding of the situation as presented in **Table 2**. After the trial, participants are asked to rate each dimension on a seven-point Likert scale (1 being the lowest and 7 being the highest) according to how well they performed on the task analyzed. The ratings are then combined to calculate a measure of the participant's situational dimension using the formula  $SA = U - (D - S)$ , where U is the summed Understanding, D is the summed attentional Demand, and S is the summed attentional Supply. From the formula, it is easy to see that the questionnaire suggests that there is a negative effect on understanding and an overall reduction of SA when demand exceeds supply.

The SART approach is also available in a faster version known as 3D SART. The 3D SART uses a 100-point scale from 0 (low) to 100 (high), for each measure (demand for attentional resources, supply of attentional resources and understanding of the situation). The overall SART score is then calculated similarly to the original scale, i.e.  $SA = U - (D - S)$ , where U is the score for understanding, D is the score for attentional demand and S is the score for attentional supply.

**Table 2**  
SART dimensions [12]

| Domain             | Dimension                    | Definition                                                |
|--------------------|------------------------------|-----------------------------------------------------------|
| Attentional demand | Instability of the situation | Likelihood of situation to change suddenly                |
|                    | Complexity of the situation  | Degree of complication of situation                       |
|                    | Variability of the situation | Number of variables that require attention                |
| Attentional supply | Arousal                      | Degree that one is ready for activity                     |
|                    | Spare mental capacity        | Amount of mental ability available for new variables      |
|                    | Concentration of attention   | Degree that thoughts are brought to bear on the situation |
|                    | Division of attention        | Amount of division of attention in the situation          |
| Understanding      | Information quantity         | Amount of knowledge received and understood               |
|                    | Information quality          | Degree of goodness of value of knowledge communicated     |
|                    | Familiarity with situation   | Degree of acquaintance with situation experience          |

## 2.1. Motivation for This Paper

Standardized questionnaires for assessing SA are an important tool for performing SA research as they provide methods that allow direct comparison of research results conducted in different settings. However, because these methods are only available in English, they can be used by native English speakers or those fluent in English. The latter are still exposed to potential cultural differences that may affect their comprehension, which may reduce the reliability and validity of the methods. Therefore, the main motivation for conducting this research was to expand the applicability of this method also to non-native English speakers, and to help SA researchers

collect data also with non-English speaking communities and users, which can lead to assessment of SA in new domains and areas. To our best knowledge a Slovenian translation of any SA standardized questionnaire does not exist. While there are reports on the use of SART in Slovenia [13], there are no reports on the translation process nor is the translation publicly available. Looking at the available research on SA, we find that in addition to Slovenia, SART has been used for assessment of SA in several non-English speaking countries, such as Portugal [14], Germany [15][16], Sweden [17], Norway [18], Korea [19][20] and China [21]. However, the authors rarely report in which language SART was presented to the study' participants (e.g., [15], [16], [19] and [21]), or if translated, which method was used (e.g., [14] or [18]). An exception is a Korean translation, where the authors used back-translation and expert knowledge to obtain a validated Korean SART – (K-SART) [20]. If the questionnaires are not in the native language of the respondents, this may affect the respondents' comprehension of what they are assessing, and consequently, affect their responses, calling into question the reliability of the results obtained. This problem is exacerbated when standardized methods are used as reference systems when evaluating new methods for assessing SA, as was the case with Satuf et al. [14] and Parka et al. [19]. This is primarily because the benefits of a validated standardized method are lost if not used in the manner in which the method was validated. This makes the interpretation of such results less reliable and raises questions about the actual sensitivity of the newly proposed methods.

### 3. Methodology

#### 3.1. User study

For the translation of the questionnaire SART into Slovenian, the standard procedures for the translation of research instruments (translation, coordination, back-translation into the source language, comparison of the back-translation with the source text, final coordination) were applied [22].

The English version of the SART questionnaire was independently translated into Slovenian by three translators who are native Slovenian speakers. Two of them were professional translators who were not familiar with the concepts that the questionnaire was intended to measure, and the third translator was an expert in the field and fluent in English. According to the recommendations of Tsang et al. [23], it is important that at least one of the translators is familiar with the concepts that the questionnaire is intended to measure in order to produce a translation that is closer to the original instrument. In this way, subtle differences in the original questionnaire could be identified and corrected by the expert group. The independent forward translations were combined into one document along with the original questions in English. The order of the proposed translations was randomized to exclude possible bias toward a particular translator.

An expert panel was then organized to discuss the proposed translations. Seven experts (native Slovenian speakers with at least B2 English proficiency level) from the fields of human-computer interaction, cognition, philosophy, transportation and driving safety were invited to serve on the panel. Three of the participants were from the University of Ljubljana, Faculty of Electrical Engineering, one from Slovenian Academy of Sciences and Arts, two from the University of Maribor, Faculty of Logistics and one was from the Slovenian Traffic Safety Agency.

The panelists discussed the translation of each questionnaire item, considering idiomatic, semantic, conceptual, and terminological aspects, until consensus was reached on a single translation that was acceptable to all panelists. In some cases, the panelists selected one of the three proposed translations without making any changes, but in other cases, they combined parts of two proposed translations to produce the final translation (see **Figure 2**). The panelists then proceeded to the next item on the questionnaire.

Once the Slovenian translation of the questionnaire was agreed upon, three other professional translators (who were not related to those who had translated the questionnaire into Slovenian) were hired to translate the accepted translations back into English to ensure that the meaning

was preserved. The translators had not previously seen the original English text of the questionnaires and were not familiar with the concepts the questionnaire was intended to measure, as recommended by Tsang et al. [23].

**Original: Arousal**

How aroused are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?

**Translation 1: Vznemirjenost**

Kako vas situacija vznemirja? Ste vedno budni in pripravljeni delovati (Visoko) ali je vaša raven budnosti nizka (Nizko)?

**Translation 2: Vznemirjenost**

Kako vznemirjeni ste v tej situaciji? Ali ste budni in pripravljeni na aktivnost (visoka stopnja) ali pa je vaša budnost nizka (nizka stopnja)?

**Translation 3: Vzburjenost**

Kako vzburjeni ste v tej situaciji? Ste zbrani in pripravljeni na dejanja (Visoka) ali je vaša zbranost nizka (Nizka)?

**The agreed upon translation: Vzburjenost**

Kako vzburjeni ste v tej situaciji? Ste zbrani in pripravljeni na aktivnost (Visoka) ali je vaša zbranost nizka (Nizka)?

**Figure 2:** An example of the accepted translation, which consists of a combination of parts of the two proposed translations.

Finally, the working group compared the two English texts and, where necessary, formulated a more appropriate Slovenian form. All final forms of the translation, named SI-SART for simplicity, were adopted with the agreement of all members of the working group and are presented in **Table 3**.

## 4. Questionnaire validation

The concurrent validity of the questionnaire was evaluated with a short user study. As a reference method, the Situation Present Assessment Method (SPAM) was used. The study was conducted in a simulated driving environment in which the operator (driver) operated a vehicle.

**Table 3**

Original SART and translated SART in Slovenian – SI-SART

| SART                                                                                                                                                                                        | SI-SART                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Title</b><br>Situation awareness rating technique                                                                                                                                        | <b>Naslov</b><br>Metoda ocenjevanja situacijskega zavedanja                                                                                                                                                     |
| <b>Instability of situation</b><br>How changeable is the situation? Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?   | <b>Nestabilnost situacije</b><br>Kako spremenljiva je situacija? Ali je situacija zelo nestabilna in se lahko nenadoma spremeni (Visoka) ali pa je zelo stabilna in jasna (Nizka)?                              |
| <b>Complexity of the situation</b><br>How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)?                   | <b>Kompleksnost situacije</b><br>Kako zapletena je situacija? Je situacija kompleksna z mnogimi medsebojno povezanimi elementi (Visoka) ali je enostavna in jasna (Nizka)?                                      |
| <b>Variability of the situation</b><br>How many variables are changing within the situation? Are there a large number of factors varying (High) or are there very few variables changing?   | <b>Spremenljivost situacije</b><br>Koliko dejavnikov se spreminja v okviru situacije? Se spreminja veliko število dejavnikov (Visoka) ali se spreminja zelo malo spremenljivk (Nizka)?                          |
| <b>Arousal</b><br>How aroused are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?                                         | <b>Vzbujenost</b><br>Kako vzburjeni ste v tej situaciji? Ste zbrani in pripravljeni na aktivnost (Visoka) ali je vaša zbranost nizka (Nizka)?                                                                   |
| <b>Concentration of Attention</b><br>How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High) or focused on only one (Low)?           | <b>Koncentracija pozornosti</b><br>V kolikšni meri ste osredotočeni na situacijo? Se osredotočate na več vidikov situacije (Visoka) ali le na enega (Nizka)?                                                    |
| <b>Division of Attention</b><br>How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (High) or focused on only one (Low)?            | <b>Delitev pozornosti</b><br>V kolikšni meri je vaša pozornost v tej situaciji razdeljena? Ste osredotočeni na več vidikov situacije (Visoko) ali le na enega (Nizko)?                                          |
| <b>Spare Mental Capacity</b><br>How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (High) or nothing to spare at all (Low)? | <b>Prosta umska zmogljivost</b><br>Kolikšna je vaša še razpoložljiva umska zmogljivost v tej situaciji? Je imate dovolj, da lahko spremljate več dejavnikov (Visoka), ali je sploh nimate na razpolago (Nizka)? |
| <b>Information Quantity</b><br>How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?            | <b>Količina informacij</b><br>Koliko informacij o situaciji ste pridobili? Ste prejeli veliko informacij in jih razumeli (Visoka) ali zelo malo (Nizka)?                                                        |
| <b>Familiarity with Situation</b><br>How familiar are you with the situation? Do you have a great deal of relevant experience (High) or is it a new situation (Low)?                        | <b>Poznavanje situacije</b><br>V kolikšni meri že poznate situacijo? Imate veliko relevantnih izkušenj s situacijo (Nizko) ali je situacija za vas nova (Visoko)?                                               |

### 4.1. Situation Present Assessment Method (SPAM)

The Situation Present Assessment Method (SPAM) is a real-time-based method [10] that asks the operator questions about their situational awareness while operating a machine. The time to accept the question (request response time), the time to answer the question (answer response time), and the success and error rate are then evaluated as indicators of SA. By providing the option for the operator to choose when to accept a new question request, this method allows for safer and less (critical) interruptions of the completing their primary task (operating a machine). However, this introduces a new factor that has to be considered when analyzing the results on time to answer question. The operator might take the time to accept the request to gain understanding of the elements of the environment and the current situation the machine is, and only then accept the request. As a consequence, they could respond to the question much quicker compared to operators that would accept the request when seeing it, without prior situational assessment of the environment. Both of these factors further impact the success rate of the queries, highlighting the need to observe all three factors when using this method.

For this study, the SPAM method was implemented in the form of a mobile application displayed on a head-down display mounted on the physical dashboard in the driving simulator. It was placed to the right of the steering wheel within reach of the driver so that he or she could easily accept and answer SPAM queries. Ten questions were used in the study to assess the driver's situational awareness, all related to driving and operating the vehicle. As defined by SPAM, the application would first prompt a question request, which the driver could either accept or decline. If accepted, the driver was presented with a question with four possible answers, only one of which was correct. There was also always an "I do not know" option. Two examples are presented in **Figure 3**.

| What is the current speed limit?                           | What was the last road sign for?                               |
|------------------------------------------------------------|----------------------------------------------------------------|
| a) 30 km/h<br>b) 50 km/h<br>c) 60 km/h<br>d) I do not know | a) Speed limit<br>b) Yield<br>c) Stop sign<br>d) I do not know |

**Figure 3:** Example of questions presented with the SPAM method to assess driver's situational awareness.

To avoid anticipation effects, the question requests were presented in random intervals between 60 to 120 seconds. In the case of a declined question request, a new prompt would appear at a random interval of 20 to 30 seconds.

The question request time, the time to answer, and the number of correct answers were recorded and used as reference measures for validating SART.

### 4.2. Study design

13 participants took part in the study. The participants first and primary task was safe driving. The participants' secondary, but equally important, task was to answer to SPAM questions. They were instructed to try to complete the SPAM tasks as quickly and as accurately as possible, without neglecting the operation of the vehicle or engaging in any dangerous driving behavior.

After the drive, the participants were asked to complete the SART questionnaire.

### 4.3. Study results

**Table 4** shows the situational awareness results obtained with SI-SART, while **Table 5** shows the situational awareness results obtained with SPAM. A correlation test was used to test for correlations between the two sets of results; the calculated Pearson correlation is  $r(13) = 0.49, p$

= 0.087, 95% CI [-.08, .82], please see **Figure 4**. Although this result did not reach conventional significance, the correlation exceeds the typical validity threshold of .30, providing evidence for the concurrent validity of the Slovene translation. The wide confidence interval reflects the small sample size, suggesting that further validation with a larger cohort is needed.

**Table 4**  
SPAM results

| Participant ID | Request response time [ms] | Answer response time [ms] | Correct answers [%] |
|----------------|----------------------------|---------------------------|---------------------|
| 1              | M=3553.7, SD=4675.2        | M=13140.7, SD=11472.6     | 80                  |
| 2              | M=7616.0, SD=10230.0       | M=21569.8, SD=11472.4     | 70                  |
| 3              | M=3185.9, SD=614.2         | M=11383.8, SD=5980.9      | 90                  |
| 4              | M=5807.4, SD=11138.1       | M=10523.6, SD=4838.7      | 80                  |
| 5              | M=8007.3, SD=15917.1       | M=14512.1, SD=6579.7      | 70                  |
| 6              | M=3329.1, SD=725.6         | M=10104.0, SD=4823.7      | 60                  |
| 7              | M=8335.6, SD=10754.8       | M=33395.1, SD=21166.5     | 80                  |
| 8              | M=3275.1, SD=2429.2        | M=12315.5, SD=5247.3      | 90                  |
| 9              | M=2389.5, SD=594.9         | M=8599.8, SD=3269.6       | 90                  |
| 10             | M=2672.0, SD=1473.3        | M=10955.6, SD=5985.6      | 70                  |
| 11             | M=7805.4, SD=3816.0        | M=13440.0, SD=5561.8      | 90                  |
| 12             | M=2910.8, SD=2294.0        | M=12878.9, SD=7674.7      | 100                 |
| 13             | M=6212.9, SD=5903.2        | M=8838.8, SD=3685.2       | 60                  |

**Table 5**  
SI-SART results

| Participant ID | SART-U | SART-D | SART-S | Overall score U-(D-S) |
|----------------|--------|--------|--------|-----------------------|
| 1              | -1     | -6     | 10     | 15                    |
| 2              | 2      | 2      | 6      | 6                     |
| 3              | 5      | -7     | 10     | 22                    |
| 4              | -1     | -5     | 4      | 8                     |
| 5              | 5      | -6     | 10     | 21                    |
| 6              | 1      | -4     | 9      | 14                    |
| 7              | 2      | 3      | 4      | 3                     |
| 8              | 4      | -2     | 11     | 17                    |
| 9              | 0      | -2     | 7      | 9                     |
| 10             | 2      | 0      | 3      | 5                     |
| 11             | 4      | 2      | 8      | 10                    |
| 12             | 5      | -6     | 11     | 22                    |
| 13             | 0      | 0      | 3      | 3                     |

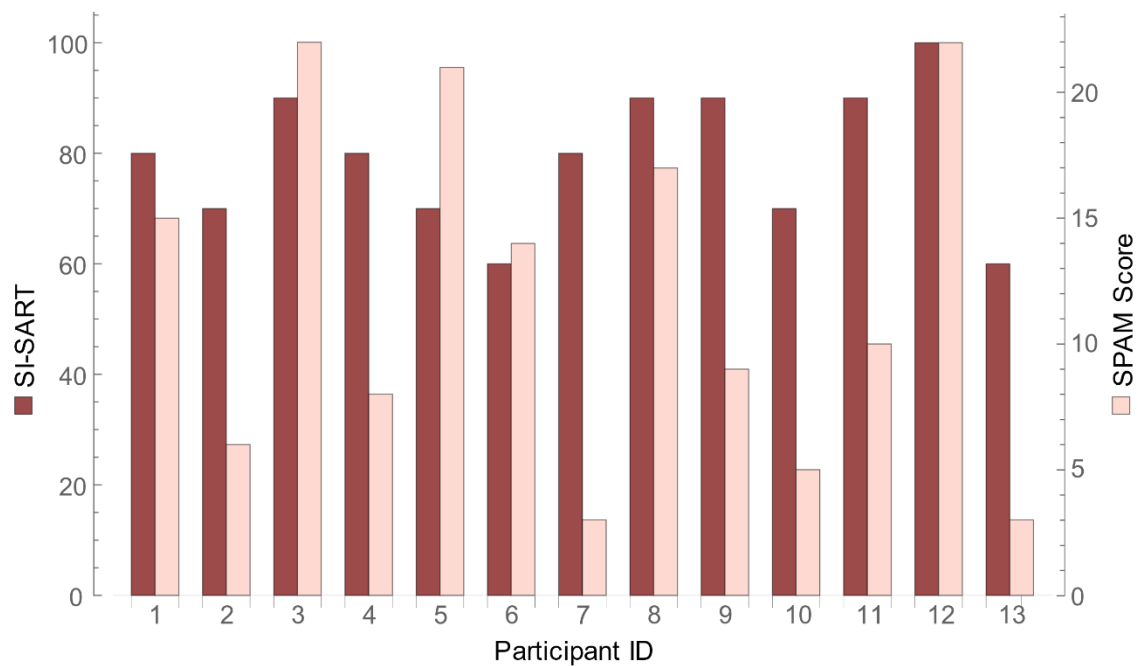
## 5. Discussion

The main goal of this research was to translate the SART questionnaire for assessing SA to enable its use for Slovenian speakers. The multi-step translation process included all of the steps required for the translation of the questionnaire – initial translation, coordination, back translation into the source language, comparison of the back translation with the source text, and final coordination. The translated version SART was also evaluated with the SPAM method, further confirming the adequacy of SI-SART.

SART can be used in multiple domains, but in this specific case we focused on the automotive domain for validation of the translated questionnaire. By involving multiple expert profiles

(linguistics, human-computer interaction, cognition, philosophy, transportation and driving safety) in the translation process, a holistic approach was taken that although primarily focusing on driving, it still attempted to retain the possibility of using SI-SART in multiple domains, as it was intended for the original SART.

The translated version will enable Slovenian researchers to use the SART questionnaire confidently without worries of losing information due to language or cultural specifics and enrich the scientific community on global scale by enable direct comparisons among research conducted in different countries. Moreover, it will enable researchers around the world to include Slovenian speaking participants in their studies, or use data collected by Slovenian researchers expanding the possibility to share data sets and research information.



**Figure 4:** SI-SART and SPAM results per participant

## Acknowledgements

The authors would like to thank all of the professionals that contributed to the translation and coordination steps for all of their work and valuable comments and suggestions. The authors also thank Aleksa Ćirković for developing the mobile app for presenting the SPAM method and all of the volunteers that participated in the user study.

The work presented in this paper was financially supported by the Slovenian Research Agency within the program ICT4QL, grant no. P2-0246 and by the European Union's HorizonEurope research and innovation program for the project FRODDO, grant agreement no. 101147819.

## References

- [1] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. In *Human Factors Journal*, 37(1), 32-64.
- [2] Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness.
- [3] Matthews, M. L., Bryant, D. J., Webb, R. D., & Harbluk, J. L. (2001). Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems. *Transportation research record*, 1779(1), 26-32.

- [4] Endsley, M. R. (1988, October). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society annual meeting (Vol. 32, No. 2, pp. 97-101). Sage CA: Los Angeles, CA: Sage Publications.
- [5] Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human factors*, 37(1), 137-148.
- [6] Bedny, G., & Meister, D. (1999). Theory of activity and situation awareness. *International Journal of cognitive ergonomics*, 3(1), 63-72.
- [7] Endsley, M. R., & Smolensky, M. W. (1998). Situation awareness in air traffic control: The picture.
- [8] Endsley, M. R. (1988, May). Situation awareness global assessment technique (SAGAT). In Proceedings of the IEEE 1988 national aerospace and electronics conference (pp. 789-795). IEEE.
- [9] Endsley, M. R. (2017). Direct measurement of situation awareness: Validity and use of SAGAT. In *Situational awareness* (pp. 129-156). Routledge.
- [10] Durso, F.T. and Dattel, A.R. (2004) 'SPAM: the real-time assessment of SA', in Banbury, S. and Tremblay, S. (Eds.): *A Cognitive Approach to Situation Awareness: Theory and Application*, pp.137-154, Ashgate Publishing Ltd., Hampshire.
- [11] Selcon, S. J., & Taylor, R. M. (1990). Evaluation of the Situational Awareness Rating Technique(SART) as a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 8* (SEE N 90-28972 23-53).
- [12] Taylor, R. M. (1990). Situation awareness rating technique (SART): the development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations* (Chapter 3). France: Neuilly sur-Seine, NATO-AGARD-CP-478.
- [13] Krivec, N. (2020). Situational awareness and group cohesivness as predictors of competition's results in team sports. <https://repozitorij.uni-lj.si/IzpisGradiva.php?id=114471>
- [14] Satuf, E. N., Kaszkurewicz, E., Schirru, R., & de Campos, M. C. M. M. (2016). Situation awareness measurement of an ecological interface designed to operator support during alarm floods. *International Journal of Industrial Ergonomics*, 53, 179-192.
- [15] Dreyer, D., Oberhauser, M., & Bandow, D. (2014, July). HUD symbology evaluation in a virtual reality flight simulation. In Proceedings of the International Conference on Human-Computer Interaction in Aerospace (pp. 1-6).
- [16] Colley, M., Eder, B., Rixen, J. O., & Rukzio, E. (2021, May). Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-11).
- [17] Päivärinne, K. (2018). Using Augmented Reality to Increase a Heavy Vehicle Operator's Situational Awareness. <https://www.diva-portal.org/smash/get/diva2:1183908/FULLTEXT01.pdf>
- [18] Ernstsén, J. & Villanger, D. (2014). Situation Awareness in Disaster Management: A study of a Norwegian collaboration exercise (Master's thesis). <https://www.duo.uio.no/bitstream/handle/10852/39779/MA-thesis-Jrgen-Ernstsén---Daniela-Villanger.pdf?sequence=1&isAllowed=y>
- [19] Parka, J., Kima, Y., & Junga, W. Toward a novel situation assessment (SA) measure. Probabilistic Safety Assessment and Management PSAM 14, September 2018, Los Angeles, CA.
- [20] Jeong, Y. (2022). Verification of Reliability and Validity of K-SART to Assess of Situational Awareness of Patients with Acute Coronary Syndrome. *Journal of the Korea Convergence Society*, 13(4), 603-611.
- [21] Liu, S., Wanyan, X., & Zhuang, D. (2014). Modeling the situation awareness by the analysis of cognitive process. *Bio-medical materials and engineering*, 24(6), 2311-2318.
- [22] Cruchinho, P., López-Franco, M. D., Capelas, M. L., Almeida, S., Bennett, P. M., Miranda da Silva, M., ... & Gaspar, F. (2024). Translation, cross-cultural adaptation, and validation of

- measurement instruments: A practical guideline for novice researchers. *Journal of Multidisciplinary Healthcare*, 2701-2728.
- [23] Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi journal of anaesthesia*, 11(5), 80.

# Učenje jazz glasbe v navidezni resničnosti

Helena Jeretina<sup>1</sup>, Matevž Pesek<sup>1,\*</sup>

<sup>1</sup>University of Ljubljana, Faculty of computer and information science, Večna pot 113, 1000 Ljubljana, Slovenia

## Abstract

Članek opisuje preliminarne rezultate testiranja igre Improviano, zasnovane za učenje jazz improvizacije na klavir v okolju obogatene resničnosti. Udeleženci, ki predhodno niso igrali jazz klavirja, so poročali o napredku pri razumevanju blues lestvic, večji motivaciji in samozvesti pri improvizaciji ter koristnosti vizualnih pomagal, kot so pobarvane tipke in prikaz imen lestvic. Igra jim je omogočila bolj intuitivno povezovanje akordov z ustreznimi lestvicami. Večina je ocenila zahtevnost ponujenih skladb kot ustrezno, udobje uporabe VR očal pa kot zadovoljivo. Posebej priljubljen je bil način improvizacije, ki je omogočil svobodno in kreativno ustvarjanje ter večjo uspešnost kot konvencionalno učenje.

## Keywords

navidezna resničnost, obogatena resničnost, učenje glasbe

## 1. Uvod

V zadnjem desetletju sta navidezna in obogatena resničnost doživeli resnejši preobrat na področju učenja, saj so naprave postale bolj razvite ter lažje dostopne uporabnikom. Navidezna resničnost omogoča obliko učenja, ki je za mnoge posameznike bolj učinkovita od konvencionalnih načinov, saj ponuja možnost samostojnega učenja brez prisotnosti učitelja, časovno in prostorsko fleksibilnost, ter prilagoditev učne poti glede na individualne potrebe in stile učenja. Na trgu obstaja že veliko aplikacij, katerih namen je skozi igro uporabnika naučiti osnov matematike, kemije, jezikov, plesa, glasbene teorije, instrumentov, in mnogo drugih sposobnosti. Navidezna resničnost se je izkazala za učinkovito tudi na področju glasbe – raznovrstne igre v navidezni in obogateni resničnosti igralca skozi igro učijo ritem, melodije, ali pa celo igranja na instrument.

Večina obstoječih izobraževalnih aplikacij, ki se osredotočajo na učenje instrumenta, so namenjene začetnikom in temeljijo na popularnih ali klasičnih žanrih. Žanri, ki potrebujejo več glasbenega predznanja, so zato pogosto prezrti. V okviru te raziskave ovrednotimo uporabo obogatene resničnosti na področju jazz glasbe, oziroma bolj natančno, pri učenju improvizacije na klavir.

V članku najprej analiziramo uveljavljene pristope k učenju glasbe ter njihove prednosti in slabosti, pregledamo primere VR iger na trgu, ki spodbujajo igralca k učenju glasbe in/ali instrumenta, ter opišemo osnove jazz in blues žanra. Na podlagi ugotovitev implementiramo igro Improviano v obogateni resničnosti, ki je namenjena pianistom z že nekaj predhodne glasbeno-teoretične podlage. Osredotoča se na učenje jazz skladb ter na improvizacijo. Uporabnikom z osnovnim znanjem glasbene teorije in igranjem na klavir postavi strukturo, da lažje začnejo z učenjem improvizacije. Z dodatnimi vizualnimi elementi v igri ter sistemom točkovanja točnosti igranja jih dodatno motivira, da se izboljšujejo iz dneva v dan. Po uspešni implementaciji igre izvedemo testiranje s šestimi posamezniki, kjer primerjamo postopek in rezultate pri učenju z obogateno resničnostjo ter brez nje.

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper+, Slovenia*

\*Corresponding author.

✉ [hj2109@student.uni-lj.si](mailto:hj2109@student.uni-lj.si) (H. Jeretina); [matevz.pesek@fri.uni-lj.si](mailto:matevz.pesek@fri.uni-lj.si) (M. Pesek)

🌐 <https://matevzpesek.si/> (M. Pesek)

🆔 0000-0001-9101-0471 (M. Pesek)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.2>

## 2. Pregled področja

Glasbena vzgoja v osnovnih šolah je raznolika glede na kulturno okolje in državno ureditev področja glasbeno-izobraževalnih institucij. Večina držav na zahodu zahteva obvezno glasbeno vzgojo v javnih šolah, seveda pa se učni načrt za predmet glasbe razlikuje glede na odločitev vodstva posameznih šol [1]. Vsebina predmeta običajno zajema učenje in razvoj pevskih sposobnosti, igranje na različna glasbila, predvsem na tista, ki spadajo pod kvalifikacijo Orffovih inštrumentov (zvončki, ksilofon, metalofon, palčici, kraguljčki, ropotulja in podobno), učenje glasbene teorije, poslušanje glasbe in učenje glasbene zgodovine [2]. Učenci se pri predmetu glasbene vzgoje naučijo najbolj osnovnih glasbenih principov. Pouk ni prilagojen glede na posameznikove potrebe, saj poteka v večjih skupinah in je njegov cilj zgolj seznaniti učenca z osnovami glasbe. Obiskovanje glasbene šole sicer omogoča pestro in poglobljeno glasbeno izobraževanje, vendar ni vedno prava izbira za vsakogar. Strogo določen učni načrt lahko omejuje ustvarjalnost profesorjev in učencev, saj zmanjšuje fleksibilnost pri izbiri repertoarja in pedagoških pristopov. Predpisane skladbe in težavnostne stopnje včasih ne ustrezajo posamezniku in zmanjšujejo možnost za prilagoditev skladb glede na osebne interese ali ustvarjalne želje. Učni načrt prav tako velikokrat poudarja tehnične vidike glasbenega učenja in ponuja premalo možnosti, da bi učenec izrazil svojo ustvarjalnost preko glasbe.

Samostojno učenje [3] postaja vse bolj priljubljeno zaradi dostopnosti spletnih tečajev, aplikacij in video vodičev, vendar zahteva visoko stopnjo discipline in sposobnost izbire kakovostnih učnih virov. Čeprav omogoča več svobode in spodbuja ustvarjalnost [4], se učenci pogosto soočajo s težavami pri vzdrževanju motivacije in občutku napredka [5]. Večje težave pri samostojnem učenju predstavljajo pomanjkanje motivacije, občutek, da ni nobenega napredka, negotovost, kako lahko svojo tehniko izboljšati in izbira kvalitetnih virov za učenje. Med drugim se lahko tem težavam izognemo z uporabo različnih digitalnih orodij ali aplikacij, ki nudijo strukturirano učenje, ter vključujejo interaktivne elemente, povratne informacije in zabavo med učenjem. Samostojni pristop k učenju glede na raziskave dokazuje uspešnost in potencial, saj spodbuja samostojno učenje, pospešuje sodelovanje in obogati glasbeno razumevanje [3], ki ga z vodenim, strukturiranim in omejenim učnim načrtom posameznik mogoče ne bi dobil pri pouku glasbene šole. V pomoč pri teh izzivih se vse pogostejše uporabljajo digitalna orodja in interaktivne aplikacije, ki ponujajo strukturirano učenje, povratne informacije in elemente poigritve.

V zadnjih desetletjih so se na področje glasbenega izobraževanja razširile tudi tehnologije navidezne (VR) in obogatene (AR) resničnosti. AR je v izobraževanju prepoznana kot prelomna tehnologija, ki lahko s svojimi interaktivnimi in poglobljenimi funkcijami preoblikuje tradicionalno poučevanje in učenje, saj proces postane bolj živ, intuitiven in privlačen [6]. Izobraževalne aplikacije pa ne povečajo le zanimanja in sodelovanja učencev, temveč z vključevanjem navideznih 3D modelov in dinamičnih simulacij v učni proces tudi bistveno izboljšajo razumevanje in pomnjenje [7]. Med drugim so se razvile tudi aplikacije, ki raziskujejo vpliv navideznega okolja na razvoj posameznikovega znanja in obnašanja. V preiskavi [8] so naslovili kako izpostavljenost navideznemu okolju vpliva na tremo pred nastopom, in ugotovili, da bi lahko glasbene ustanove močno pripomogle k izboljšanju samozavednosti nastopajočih, če bi investirale v navidezne sisteme. Razviti so bili tudi sistemi, ki se osredotočajo na učenje konceptov glasbe, kot so ritem in akustika ter ugotavljanje žanra glasbe [9].

AR omogoča dodajanje navideznih elementov na fizične inštrumente, ta funkcionalnost pa je lahko zelo dobro izkoriščena za motivacijo in izboljšanje učnega procesa. Največ aplikacij je bilo razvitih za učenje klavirja, kitare in bobnov, saj so ti inštrumenti najširše uporabljeni ter z vidika implementacije omogočajo najzanimivejše vizualne nadgradnje (obarvane tipke, strune, ali opna, padajoče note).

## 3. Izdelava igre Improviano

Improviano je namenjena pianistom z osnovnim znanjem, ki želijo osvojiti jazz skladbe in improvizacijo. Igra z vizualnimi namigi (obarvane tipke, prikaz akordov in blues lestvice) znižuje vstopni prag improvizacije, kjer so ključni poznavanje tonalitete, glavnih akordov in zgradbe (heptatonične) molove

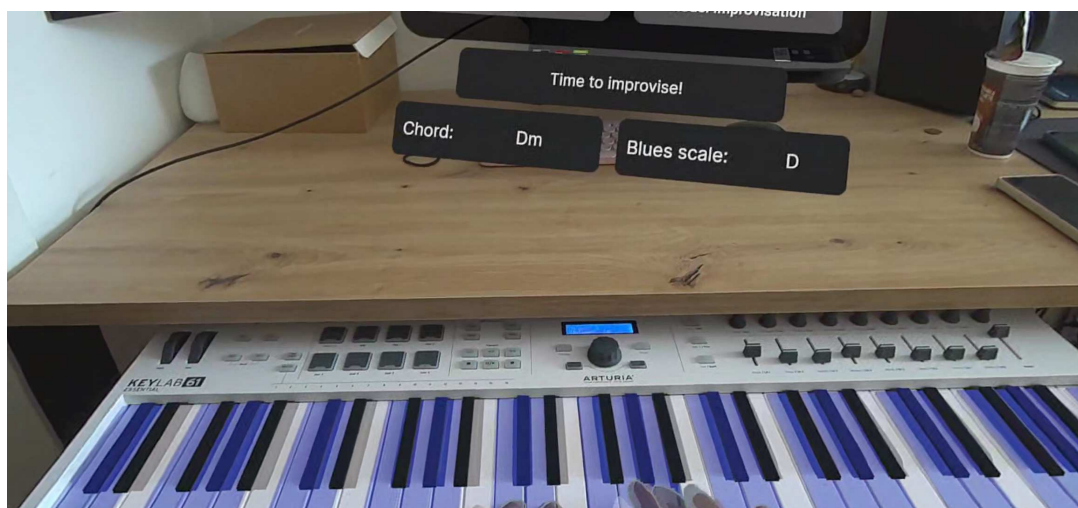
blues lestvice.

Jedro zasnove sledi jazz strukturi *head–solo–head*: najprej učenje melodije s spremljavo, zatem vodena improvizacija na podlagi sproti izračunane blues lestvice, nato ponovitev teme. Implementirani so natančno zajemanje MIDI dogodkov, robustno prepoznavanje akordov ter funkcije za tempo, barvanje tipk in skoke po času.

Način *practice* omogoča vse funkcije brez točkovanja; v *improvisation* se predvaja le spremljava in barvajo tipke lestvice (možen izklop pomoči); *scoring* zaklene tempo/skoke in ločeno ocenjuje melodijo ter spremljavo, medtem ko se pri improvizaciji zgolj beleži razmerje pravih/napačnih tonov. Časovno odstopanje  $\pm 0,25$  s je dovoljeno. Končna ocena je povprečje melodije in spremljave; zvezdice se podelijo pri 70 %, 80 % in 95 %.

Med igranjem je na voljo nadzorna plošča (premor, previjanje, tempo  $\pm$ , spremljava/melodija on/off, barvanje tipk on/off, barvanje lestvice, menjava načina). Pred začetkom se nastavi začetni tempo in odštejeta dva takta; plošča prikazuje tudi trenutni akord in ime predlagane lestvice.

V igri je na voljo pet skladb (*Beginner's Blues*, *Blueberry Hill*, *Blue Bossa*, *Blue Monk* in *Fly Me To The Moon*); prvi dve sta primerni za začetnike (trije akordi, enostavna melodia/spremljava), preostale skladbe pa so zahtevnejše (več akordov, kompleksnejši ritem). Za vsako skladbo so v zaledju igre ustvarjene tri MIDI datoteke (spremljava, melodia, glavni akordi), ki jih pogon igre uporablja za natančnejše ocenjevanje in zanesljivejšo detekcijo akordov. Datoteke so bile ustvarjene posebej za uporabo v igri.



**Figure 1:** Primer obogatene tipkovnice z označenimi tipkami v lestvici.

### 3.1. Oprema, potrebna za igranje igre

Za igranje igre Improviano potrebuje igralec naslednjo opremo, ki jo vidimo na Sliki 2:

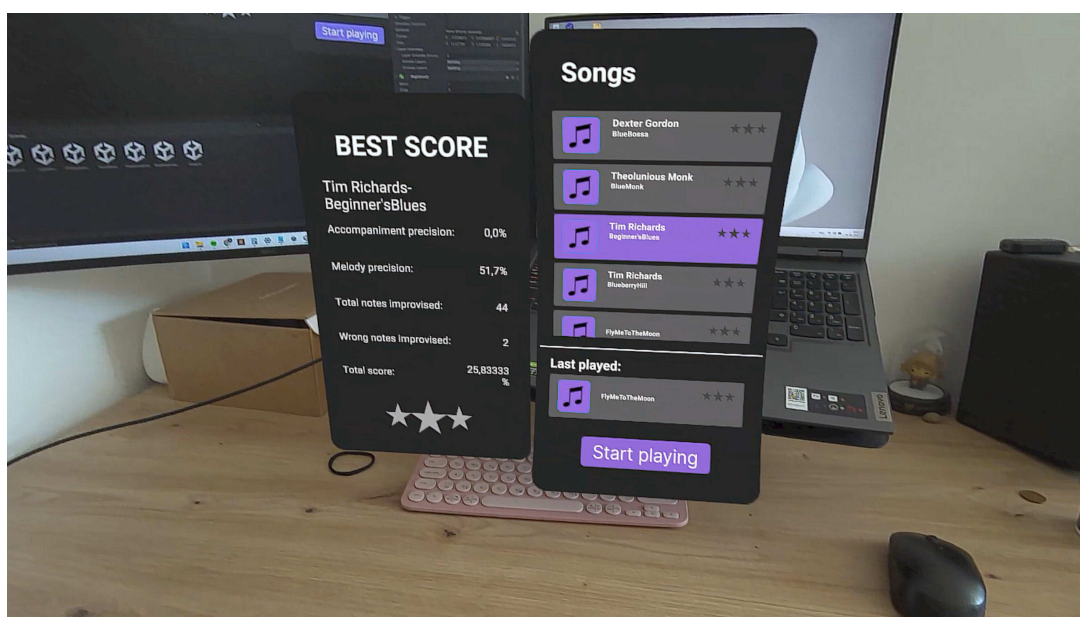
- Računalnik z nameščeno igro Improviano, Oculus PC programom ter programom Pianoteq 8, ki da klaviaturi zvok
- Meta Quest 3 Headset (igra je bila testirana le na tem modelu Oculus naprav)
- Klaviaturo z MIDI OUT izhodom
- Kabel MIDI-OUT v USB-C ali MIDI-OUT v USB-A
- Kabel USB-C v USB-C ali USB-C v USB-A

Na Sliki 4 vidimo primer uporabniškega vmesnika v igri Improviano.

Igralec zažene igro prek računalnika, ter Oculus Meta Quest 3 očala preko kabla priklopi na računalnik. Za najboljšo izkušnjo igre priklopi tudi kabel MIDI v USB-C iz klaviature na računalnik. Igranje igre je možno tudi brez tega kabla, vendar v tem primeru način ocenjevanja pravilnosti ne deluje, saj se MIDI dogodki brez kabla ne morejo zaznati.



**Figure 2:** Oprema, potrebna za igranje igre Improviano.



**Figure 3:** Primer uporabniškega vmesnika v igri Improviano.

### 3.2. Navigacija in interakcija z navideznimi predmeti

Glavna lastnost igre je igranje na klaviaturo, kar pomeni, da ima igralec ves čas zasedene roke. Zato smo se odločili, da bo za to igro bolj primerno, da igralec ne uporablja kontrolerjev, priloženih VR očalom. V Unity smo integrirali *Interaction SDK*, ki omogoči sledenje rokam in interakcijo z navideznimi predmeti s prijemanjem (*angl. grab*), dotikanjem (*angl. poke*) ali usmerjanjem žarka (*angl. raycast*).

### 3.3. Navidezna klaviatura

Navidezna klaviatura, prikazana na Sliki 4, je bila izdelana v programu Blender, njen originalen razpon pa zajema 88 tipk. Ta navideznoresničnostna komponenta je najpomembnejša v tej igri in zahteva natančno poravnavo z uporabnikovo fizično klaviaturo. Med igranjem se na njej obarvajo tipke, ki jih mora igralec pritisniti. S to vizualno pomočjo se uporabnik postopoma in sistematično uči izbrane skladbe, kar omogoča učinkovit in interaktiven proces učenja.



Figure 4: Navidezna klaviatura v igri Improviano.

#### 3.3.1. Premikanje navidezne klaviature

Ko igralec odpre nastavitve, se omogoči premikanje klaviature, katere pozicijo lahko spremeni kar z roko. To funkcijo smo omogočili z *OVR Hand Grab interaction*, za natančnejše pozicioniranje klavirja pa so na voljo dodatne tipke za spremembo rotacije, translacije in skaliranje objekta. Na tem mestu lahko igralec prilagodi tudi število tipk, da se ujema z njegovo klaviaturo. Na izbiro ima 49, 61, 76 ali 88 tipk, kar so standardni razponi klaviatur, ki so na voljo na trgu. Ko se spremembe shranijo, ostane klavirska tipkovnica na istem mestu do nadaljnjega. Da smo zagotovili še bolj precizno in statično postavitve klaviature, smo uporabili *OVRAnchor*, ki predmetu doda prostorsko sidro. Na ta način smo dosegli stabilno pozicijo navidezne klaviature v fizičnem prostoru, pozicija pa se obdrži tudi ob ponovnem obisku igre. Ostale nastavitve, kot so obseg klavirskih tipk in skaliranje tipk po meri, smo shranili v *PlayerPrefs*, v Unity vgrajen sistem za shranjevanje majhnih količin podatkov, ti pa se ob vsakem ponovnem obisku igre ponovno naložijo in uveljavijo na klavirski tipkovnici.

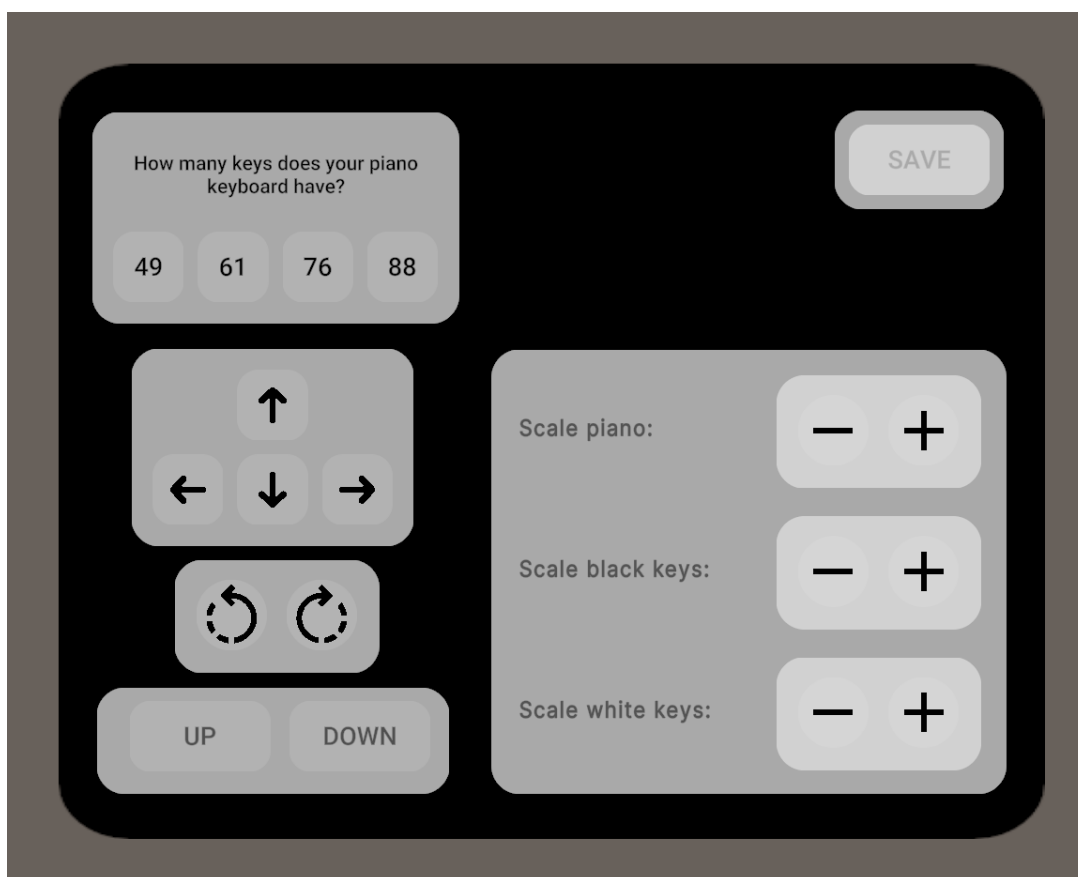
### 3.4. Različni načini igranja

Igra vsebuje tri načine igranja. Igralec začne v načinu za vadbo skladbe, potem pa se lahko prestavi v način improvizacije ali način, kjer je njegovo igranje ocenjeno.

Skladbe upoštevajo *head-solo-head* jazz strukturo. To pomeni, da se najprej predvaja glavna melodija s spremljavo, kateri mora igralec slediti ter jo poskusiti zaigrati na enak način. Sledi *solo* del, v katerem se improvizira na glavne akorde v skladbi. Nazadnje se zopet ponovi del, kjer igralec igra glavno melodijo in spremljavo (*head*).

#### 3.4.1. Vadba (angl. practice)

V načinu za vadbo ima igralec na voljo vse funkcionalnosti igre razen sistema točkovanja. Med predvajanjem melodije in spremljave se istočasno obarvajo klavirske tipke, ki jih mora igralec pritisniti. Tipke spremljave so v drugačni barvi kot tipke melodije, da je igralcu bolj jasno, katero roko mora uporabiti za katere tipke. V drugi iteraciji, kjer nastopi improvizacija, se predvaja samo spremljava in obarvajo se klavirske tipke, ki pripadajo blues lestvici, ki se izračuna glede na trenutni glavni akord skladbe. Igralec tako hitreje razpozna pravilne note, ki jih lahko pritiska med improvizacijo. V tretji iteraciji je postopek enak kot v prvi.



**Figure 5:** Nastavitve za natančno pozicioniranje klaviature.

### 3.4.2. Improvizacija (angl. improvisation)

V načinu za improvizacijo se predvaja samo *solo* del - v ozadju se torej predvaja le spremljava, igralcu pa se obarvajo tipke blues lestvice. Tako se lahko osredotoči samo na improvizacijo, brez da bi se mu ponavljala še melodija. Barvanje tipk lahko izklopi ter se tako preizkusi v samostojni improvizaciji, kjer ima kot pomoč na voljo le prikazano ime akorda in blues lestvice.

### 3.4.3. Ocenjevanje (angl. scoring)

V načinu za ocenjevanje igralec ne more preskakovati po skladbi ali spreminjati tempa med igranjem. Ocenjevanje je razdeljeno na tri dele. Posebej se zaznava pravilnost zaigranih not melodije in spremljave. V drugi iteraciji skladbe, kjer glasbenik improvizira, se beleži le število pravilno pritisnjenih klavirskih tipk ter število napačnih. Ta rezultat se ne upošteva pri končni oceni, saj je improvizacija subjektivna in ne obstaja nikakršnega vzorca, po katerem bi lahko določili kako uspešno je to bilo. Kljub izključitvi tega rezultata od končne ocene pa je igralcu ob koncu skladbe vseeno prikazano število napačno in pravilno pritisnjenih tipk med improvizacijo, da ima boljši občutek kako mu je šlo.

Da je zaigrana nota ocenjena kot pravilno, mora biti zaigrana tudi ob pravem času. Zaradi možnega časovnega zamika, ki se lahko pojavi pri klaviaturah, smo dovolili časovno odstopanje do 0.25 sekunde. Zabeležili smo si število pravilno pritisnjenih not, ter število napačnih. Končna ocena spremljave ali melodije je bila izračunana po naslednji formuli:

$$\frac{\text{pravilneNote} - (\text{napačneNote} / 2)}{\text{številoVsehNot}}$$

Skupna končna ocena je prikazana v odstotkih, ti pa so izračunani kot povprečje ocene melodije in spremljave. Vsaka skladba za lažjo preglednost vsebuje tri zvezdice, ki se pobarvajo glede na dosežke -

ena zvezdica se pobarva ko igralec doseže 70% pravilnost, druga 80% pravilnost, vse tri zvezdice pa se pobarvajo, ko igralec obvlada skladbo s 95% pravilnostjo.

## 4. Eksperiment

Po uspešni implementaciji igre smo izvedli eksperimentalni načrt z namenom analize odziva udeležencev, zbiranja predlogov izboljšav in mnenj o igri. Testirali smo z namenom, da preverimo uporabnost igre v obogateni resničnosti kot učinkovit pristop k učenju improvizacije na klavir, v kakšni meri se razlikuje od navadnega pristopa k učenju, in če ima pozitivne in/ali negativne posledice na udeleženca.

Igra Improviano je namenjena udeležencem z osnovnim znanjem klavirja, ki imajo že nekaj predhodne glasbeno-teoretične podlage. To pomeni, da morajo poznati strukturo osnovnih in naprednih akordov ter jih znati zaigrati na klavir. Prav tako morajo poznati tudi osnovne lestvice. Branje not za to igro ni nujno potrebno, saj vsebuje dovolj vizualnih elementov, ki so primerni tudi za udeležence, ki not ne poznajo. Vse te predpogoje smo upoštevali med iskanjem primernih udeležencev za testiranje, da smo zagotovili čimbolj zanesljive rezultate.

Za potrebe testiranja smo pripravili dva vprašalnika. Prvega je udeleženec rešil tik pred začetkom testiranja, kjer je odgovoril na vprašanja glede njegove zgodovine glasbenega izobraževanja, glasbenih interesov, znanja improvizacije na klavir, ter izkušnje z uporabo VR očal do sedaj.

Po testiranju je udeleženec rešil drugo anketo, kjer je podal odgovore na specifična vprašanja o igri, ter imel možnost podati konstruktivne povratne informacije, na primer kaj se mu je zdelo najbolj uporabno in zabavno ter kaj so mogoče pomanjkljivosti te igre. V obeh anketah je bilo tudi kratko preverjanje znanja blues lestvice, na ta način pa smo preizkusili, če se je lestvico možno naučiti z igro. Odgovarjanje na obe anketi ni bilo časovno omejeno.

V igri je na voljo pet različnih skladb, zato so bile udeležencem med testiranjem ponujene te. Od najmanj zahtevne do najbolj zahtevne so to Beginner's Blues, Blueberry Hill, Blue Bossa, Blue Monk in Fly Me To The Moon. Skladbe smo udeležencem dodelili glede na njihovo predhodno glasbeno-teoretično podlago, izkušnje z igranjem klavirja, pazili pa smo tudi na to, da so jim skladbe všeč in da jih še ne znajo zaigrati. Ena skladba se je uporabila za prvi del testiranja, kjer se je udeleženec moral sam naučiti melodije, spremljave in improvizacije, druga skladba pa se je uporabila v drugem delu med igranjem igre (več o tem v nadaljevanju).

Večina udeležencev, ki so se udeležili eksperimenta, ni imelo nobenih predhodnih izkušenj z uporabo navideznega okolja, ali pa zelo malo. Zato smo jim po reševanju prve ankete dali navidezna očala Meta Quest 3 in jim razložili, kako se jih uporablja, kakšne gibe morajo posnemati, da sistem zazna klike, in tako dalje.

Po seznanjanju z navideznim okoljem smo udeležencu razložili, da ima do 30 minut časa, da se nauči igrati prvo skladbo. Udeležencu smo povedali, naj se loti učenja skladbe na enak način, kot bi se ga doma, nato pa smo opazovali, kakšen način uporablja in kako učinkovit je za učenje. Med samim postopkom smo udeležencu po potrebi dajali dodatna navodila, na primer, "poskusi se osredotočiti še na improvizacijo" in "poskusi si zapomniti glavne akorde te skladbe". Na ta način smo zagotovili, da se je udeleženec v 30 minutah poskusil naučiti in zapomniti čim več o tej skladbi. Med igranjem smo ga včasih prekinili ter preverili njegovo znanje trenutne tonalitete skladbe, kateri akordi so v skladbi in katero blues lestvico bi igrali glede na te akorde.

### 4.1. Učenje skladbe in improvizacije z igro

Ko smo presodili, da se je igralec dovolj natančno naučil prve skladbe, ali pa začel izgubljati interes, smo pripravili okolje v igri Improviano. Da bi bilo zanj na začetku manj novih informacij, ter da bi ga čim hitreje seznanili s potekom igre, sem najprej VR očala nadela prva avtorica, udeleženec pa je sledil igri kar preko ekrana računalnika. Nato smo mu opisali potek igre, kako lahko prestavlja različne navidezne objekte, predvaja skladbe, in podobno.

Za ta del testiranja je oseba v igri izbrala drugo skladbo, ki ji je bila dodeljena na začetku testiranja, potem pa smo opazovali, na kakšen način uporablja igro za učenje, na primer, če veliko ustavlja skladbo,

gleda napisane akorde in ime blues lestvice v kateri igre, kateri način igranja ji je najljubši (način točkovanja, vadbe ali improvizacije), kako pogosto uporablja gumbe za vklop/izklop barvanja tipk in spremljave ali melodije, in podobno. Vsako osebo smo spodbujali naj čim več komentira med igranjem, ter si beležili vsa opažanja, pripombe in pohvale. Po potrebi smo ji vmes postavljali dodatna vprašanja, da smo si zagotovili čim več povratnih informacij.

Ob pričetku testiranja smo osebo usmerili v način ocenjevanja, da smo si zabeležili pravilnost igranja v levi in desni roki ter med improvizacijo. Po tem je igralec imel čas, da igro prosto uporablja, na koncu testiranja pa smo zopet pogledali njegovo pravilnost v načinu ocenjevanja. Točke smo primerjali in opazili, da se je pravilnost not izboljšala, igralci pa so se izboljšali tudi v pravilnosti in pogostosti pritisnjenih not med improvizacijo.

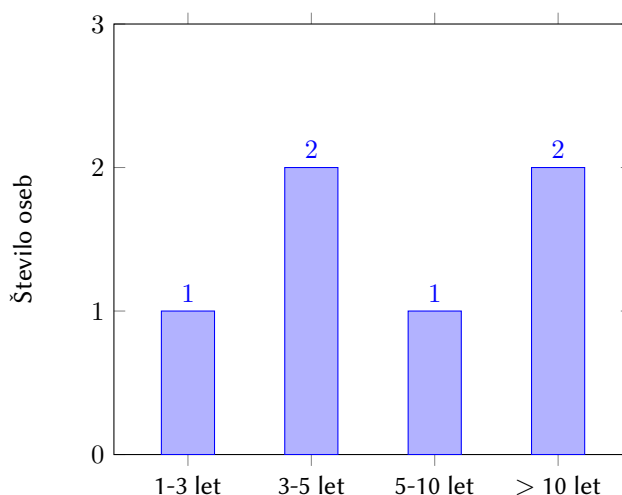
## 5. Rezultati

Testiranje smo izvedli s šestimi študenti različnih strok, starimi med 20 in 24 let. Tri osebe so imele manj predhodnih izkušenj z igranjem klavirja, saj je bilo njihovo glasbeno izobraževanje bolj osredotočeno na drug inštrument. Druge tri osebe so imele napredno znanje igranja na klavir, to pa je bil tudi njihov primarni inštrument. To, da smo imeli ljudi z različnim predhodnim znanjem klavirja nam je omogočilo, da smo lahko primerjali, kako se na igro odzivajo manj izkušeni in bolj izkušeni pianisti. Testiranje je trajalo od 60 do 90 minut na osebo.

### 5.1. Anketa 1

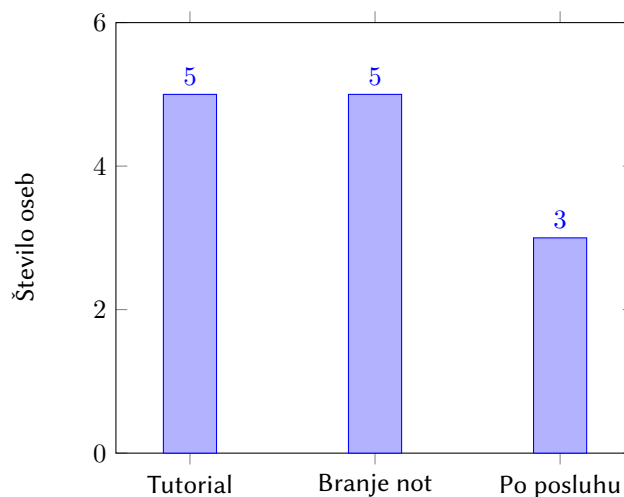
#### 5.1.1. Način in trajanje predhodnega glasbenega izobraževanja

Študente smo povprašali o njihovi zgodovini izobraževanja in učenja igranja na klavir. Rezultati so pokazali, da se je večina oseb učila klavir med obiskovanjem glasbene šole. Čeprav imajo vse osebe podobno starost (med 20 in 24 let), se njihova dolžina pianističnega glasbenega izobraževanja močno razlikuje med seboj. Tri osebe so se učile klavir v glasbeni šoli kot primarni inštrument (te osebe ga igrajo najdlje), dve osebi pa sta obiskovali pouk klavirja kot dodatni, dopolnilni predmet v srednji glasbeni šoli oziroma na glasbeni akademiji. Ena oseba se je učila igranja klavirja samostojno, pred tem pa tudi nikoli ni obiskovala nobenega vodenega programa za učenje glasbe ali inštrumenta.



**Figure 6:** Prikaz števila anketirancev glede na dolžino igranja klavirja v letih.

Eno izmed vprašanj je anketirance povpraševalo o njihovih izkušnjah improvizacije na klavir; improvizacije so se učile le štiri od šestih oseb. Ena oseba je svoje znanje improvizacije ocenila z 10, saj se je izobraževala v jazz smeri, druge tri osebe pa so svoje znanje ocenile v povprečju z 2.5. Podale so tudi svoje mnenje o največjih izzivih pri učenju improvizacije. Odgovori opisujejo, da se jim zdi težko



**Figure 7:** Prikaz števila oseb, ki uporabljajo določen pristop k učenju skladbe.

učenje lestvic in glasbene teorije, sledenje glavnim akordom v skladbi ter strukturi skladbe, sledenju solistom ali spremljevalcem. Najbolj pogost odgovor je bila negotovost, kako z improvizacijo sploh začeti, torej kako se lotiti učenja le-te, kako dobiti občutek ter samozavest.

Vse osebe, ki so sodelovale pri testiranju, so bile zainteresirane v učenje improvizacije. Razlogi za to, da se nekateri niso tega učili že prej ali vztrajali pri tem, so različni, vključujejo pa negotovost, kako se učenja lotiti, pomanjkanje časa, dvom v svoje sposobnosti, ter zavedanje, da verjetno nikoli ne bodo dobili povratnega mnenja o kvaliteti njihove improvizacije, kot bi ga na primer lahko v izobraževalni ustanovi.

Nihče izmed anketirancev ne uporablja navideznih očal redno, saj nimajo dostopa do njih. Najbolj pogost primer uporabe je igranje igre z navideznimi očali v *Woop!* centru.<sup>1</sup> Pomanjkanje izkušenj z uporabo navideznega okolja ni vplivalo na igranje igre, le na začetku je trajalo malo več časa, da se je uporabnik privadil na novo tehnologijo.

## 5.2. Izvedba

V prvem delu testiranja, ko so bili študenti pozvani k učenju skladbe na svoj način, smo si zapisovali kako so k temu pristopili ter če se ta način razlikuje od tega, kar so zapisali v anketi. Izkazalo se je, da so se vsi lotili učenja med testiranjem na enak način kot doma. Pristopi k učenju se med osebami razlikujejo, podrobnosti pa so prikazane v Grafu 7.

Vse osebe so se najprej lotile učenja spremljave in melodije. Dva posameznika, ki sta se učenja skladbe lotila z branjem not, sta spregledala, da je spremljava zapisana v basovskem ključu. Ta študenta sta imela v preteklosti manj izkušenj z igranjem klavirja kot ostali. Druga dva igralca nista imela težav pri branju not in razpoznavi pravega ključa. Preostali dve osebi sta se lotili učenja tako, da sta najprej poslušali posnetke glasbe na *youtube*, potem pa poskusili izvajati skladbo po posluhu na klavir. Obema je to predstavljalo nekaj težav, zato sta si kasneje odprla notni zapis in nadaljevala učenje s kombinacijo poslušanja posnetka ter branja not. Pri učenju melodije in spremljave je bila samo manjšina pozorna na to, kateri akordi so v pesmi glavni. Zavedanje glavnih akordov je pomembno za uspešno improvizacijo, zato smo udeležence po nekaj minutah opomnili, naj bodo pozorni tudi na to, ter naj si poskusijo zapomniti glavne akorde.

Edina oseba, ki se je lotila improvizacije samoiniciativno, je testna oseba šest, ki se je v preteklosti izražala v smeri jazz klavirja. Ostale igralce smo morali spodbuditi, da se začnejo učiti improvizacijskega dela, saj so bili preveč osredotočeni samo na učenje melodije in spremljave.

<sup>1</sup><https://woop.fun/ljubljana-vr/>.

### 5.2.1. Učenje improvizacije

Večina igralcev je imelo težave z improvizacijo, saj niso vedeli kako bi se tega lotili. Na kratko smo jim razložili postopek, po katerem ponavadi poteka improvizacija z uporabo molove blues lestvice ter predlagali naj najprej glede na glavne akorde ugotovijo, katere blues lestvice bodo igrali, ter naj si zapomnijo kateri toni sestavljajo vsako posamezno lestvico. Osebe, ki pred tem niso poznale blues lestvice, so se jo poskusile naučiti z uporabo interneta. Na ta način so bile sposobne zaigrati lestvico počasi in premišljeno. Ti posamezniki so imeli več težav pri improvizaciji kot drugi, ki so imeli blues lestvico že zvađeno, saj niso imeli možnosti svobode zamišljanja melodije, ker so bile omejene z razmišljanjem, kateri toni so sploh pravilni. Poznalo se je, da so se osebe, ki obvladajo klavir, tudi zmožne hitreje naučiti novih tehnik - primer, obe testni osebi s predhodnim klasičnim znanjem igranja na klavir nista imeli večjih težav z učenjem nove lestvice ter transponiranjem glede na spremljevalne akorde.

Kljub razlikam med predhodnim znanjem in izkušnjami so bile vse osebe zmožne vsaj malo improvizirati, tako da so z desno roko igrale melodijo po svoji domišljiji, ki vključuje samo note blues lestvice - seveda v zelo počasnem tempu ter v nepravilnem ritmu.

### 5.3. Uporaba VR očal

Študentom smo pred testiranjem igre Improviano nadeli na glavo navidezna očala, da so se naučili pravilne uporabe okolja, navadili na sledenje rokam, klikanje na gumbe in dotikanje navideznih predmetov. Nobena oseba s tem ni imela večjih težav, zato smo lahko nadaljevali z naslednjim delom testiranja.

V drugem delu testiranja so se osebe premaknile v svet obogatene resničnosti in preizkusile igro Improviano. Vsi igralci so na začetku potrebovali nekaj časa, da so se seznanili z okoljem igre in se navadili na navidezni klavir, ki je bil postavljen čez fizičnega. Veliko časa so namenili natančni postavitvi navidezne klaviature, postavitve pa se jim je zdela zahtevna, saj na splošno niso navajeni prijemanja in klikanja navideznih predmetov, poleg tega pa so pričakovali 100-odstotno natančno prileganje, ki ga je v navideznem okolju skoraj nemogoče doseči.

### 5.4. Priljubljenost različnih načinov igranja v igri

V igri so, kot je bilo že prej omenjeno, na voljo trije različni načini igranja, to so način vadbe, improvizacije, ter ocenjevanja. Vsak igralec je med testiranjem preizkusil vse tri načine, najbolj priljubljen med vsemi pa je bil način improvizacije. V tem načinu je bilo igralcem všeč to, da so se lahko preizkusili v improvizaciji brez da bi potrebovali pri tem veliko razmišljati o sestavi blues lestvice, saj pobarvane tipke močno olajšajo delo. Omogočajo jim prosto igranje z osvetljenimi tipkami, ki bodo definitivno lepo zvenele ob spremljavi, to pa jim da občutek, kot da so že izkušeni jazz glasbeniki.

Način vadbe je bil všeč predvsem osebam z več predhodnimi izkušnjami. Razlog za to je, da so bile že takoj na začetku sposobne dojemati vse informacije, ki so pred njimi. Drugi, manj izkušeni pianisti, so podali mnenje, da je na začetku okrog njih preveč informacij, in da težko sledijo vsemu, kar se pojavlja pred njimi. Obarvane klavirske tipke, glasba v ozadju, imena akordov in pripadajočih blues lestvic, upravljanje gumbov za vklop ali izklop dodatnih informacij - vse to naenkrat je zanje za začetek preveč.

Način, v katerem je bilo igranje ocenjeno, je bil večini udeležencev všeč, saj so na takšen način dobili povratne informacije o njihovem igranju. Služilo jim je kot motivacija, da poskusijo skladbo čim večkrat ponoviti in izboljšati svoj najboljši dosežek, zanimivo pa jim je bilo videti tudi to, koliko pravilnih in napačnih not so pritisnili v načinu improvizacije. Nekateri igralci so mnenja, da lahko v primeru, ko se igralcu napredek ne pozna, to deluje na negativen način. Prav tako nekaterim udeležencem način ocenjevanja ni bil privlačen in so ga videli celo kot dodaten napor.

## 5.5. Anketa 2

Vsi udeleženci so se med igranjem igre naučili nekaj novega. Nekateri so se naučili različne blues lestvice ter pravilnega načina uporabe med improvizacijo, drugi so pridobili motivacijo za učenje improvizacije na jazz skladbe ter večjo samozavest med igranjem. Veliko udeležencev je bilo presenečenih nad tem, da je možno povezovati glavne akorde s tem, katero blues lestvico lahko igraš. Mislili so, da ne obstaja nobene formule, in da se čez celotno skladbo igra samo eno blues lestvico osnovne tonalitete, ali pa da je potrebno ugibati pravo lestvico.

Tri osebe so mnenja, da je ime blues lestvice, ki je prikazano na nadzorni plošči pred igralcem, zelo uporabna informacija med igranjem. Pomaga jim določiti osnovni ton lestvice, saj ta ni bil pobarvan z drugo barvo na klaviaturi, postavi jim osnovo na katero morajo improvizirati, poda možnost koordinacije, ter informacijo, katere stopnje so pomembne in katere bodo sledile. Poleg tega je lažje imeti napisano ime lestvice pred sabo, kot pa da bi si to moral sam zapomniti ali preračunati.

Ostale tri osebe vidijo ta dodaten prikaz po eni strani kot uporabno informacijo, sploh če ob improvizaciji nimajo pobarvanih tipk na klaviaturi. Po drugi strani se jim zdi to kot neka dodatna nepotrebna informacija, ki te lahko zmede, ko se učiš nekaj novega, vendar pa pride prav potem, ko neko skladbo že malo bolje obvladaš. Nekateri, ki so menili, da je preveč podatkov, tega prikaza sploh niso opazili.

Vsem anketirancem se je funkcionalnost barvanja tipk med igro zdela zelo uporabna. S pomočjo pobarvanih tipk so imeli jasno določen okvir, znotraj katerega so lahko po lastni volji improvizirali in poskušali nove stvari, ki vseeno zvenijo privlačno. Poleg tega se jim je zdela bolj intuitivno, kot zgolj učenje prek not in igranje lestvic na pamet. Všeč jim je bilo, v kakšni meri so imeli olajšano možnost improvizacije, kar pa ne bi bilo mogoče brez uporabe navideznih očal. Eden izmed anketirancev pa je zaskrbljeno opomnil, da lahko na dolgi rok igralec postane preveč odvisen od postavljenih meja, ter ga to ne spodbudi k dejanskemu učenju in aktivnemu razmišljanju o lestvicah in akordih med igranjem.

Odgovori na vprašanje o zapisu akordov so bili precej mešani. Udeleženci z večjim glasbenim predznanjem so cenili preprostost zapisa, saj so vajeni okrajšav in vejo, kaj te okrajšave pomenijo. Ostali z manj izkušnjami pa zapisov akordov niso dojeli kot nekaj uporabnega, ali pa se jim je zdel zapis prekratek in ga niso razumeli.

Večini udeležencev (štiri od šestih) se je zdela zahtevnost skladb primerna. Enemu udeležencu, ki ima že od prej izkušnje, se je zdela zahtevnost premajhna, enemu pa se je zdela prevelika (udeleženec z manj predhodnih izkušenj).

Večino udeležencev navidezna očala niso motila. Čeprav niso bili navajeni na nošenje le-teh, niso imeli nobenega občutka slabosti. Enega udeleženca je malo bolela glava, drugega pa vrat zaradi konstantnega pogleda navzdol proti klaviškim tipkam.

Kot že prej omenjeno, so udeleženci eksperimenta dobili v obeh vprašalnikih preverjanje znanja blues lestvice v C. V obeh vprašalnikih so morali iz svojega spomina napisati vse note, ki jih ta lestvica vsebuje. Nekateri igralci so blues lestvico poznali že pred igranjem, zato se pri njih odgovor na to vprašanje med prvo in drugo anketo ni razlikoval in je bil pravilen. Ostali udeleženci, ki so pri prvi anketi podali napačen odgovor, so na drugi anketi napisali pravilne note heptatonične blues lestvice.

Vsak posameznik ima drugačne potrebe za učenje, nekateri potrebujejo čim več informacij okrog sebe, nekateri pa so bolj zadovoljni s preprostim. Zato so bili igralci veseli, da so lahko prilagajali to, kdaj vidijo pobarvane tipke, kdaj se predvaja samo melodija in samo spremljava, ter da so lahko po prosti volji ustavljali in predvajali glasbo. Na ta način so si sami lahko prilagodili količino podatkov, s katerimi so bili obkroženi med igro.

Kot že prej omenjeno, je bil način improvizacije med igralci najbolj priljubljen. Vsi so začeli igrati tako, da so imeli vklopljeno barvanje tipk, ki so del blues lestvice, saj je to privzeta lastnost igre. Ta način jim je omogočil prosto improvizacijo brez razmišljanja o teoriji glasbe. Igralce smo spodbujali, naj si poskusijo zapomniti akorde, ki se predvajajo tekom skladbe, ter pripadajoče blues lestvice. Po nekem času smo vsakega igralca prosili naj izklopi barvanje tipk ter se preizkusi v improvizaciji brez vizualne pomoči. Takrat so se zanašali na zapis lestvice in akordov, ki so bili napisani pred njimi. To jim je tudi bolj pomagalo, in je bila improvizacija na ta način veliko bolj uspešna, kot med igranjem brez

VR naprave.

### 5.5.1. Ocena izkušnje igranja

Po koncu testiranja so v anketi igralci pustili še svojo oceno igranja igre ter komentar. Večini udeležencev se je igra zdela zelo poučna in jo vidijo kot dober dodatek k učenju improvizacije. Rezultati so vidni na grafu 8.

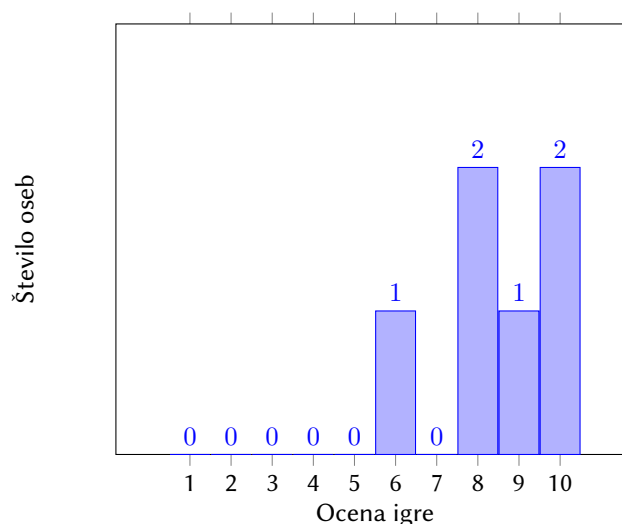


Figure 8: Pregled ocen igre udeležencev.

## 5.6. Predlagane izboljšave za igro

Udeleženci testiranja so podali veliko uporabnih povratnih informacij o igri, med drugim tudi predloge za izboljšave.

**Nastavitev in prikaz navideznega klavirja:** Način postavitve klavirja v igri ni najbolj uporabniku prijazen, saj ne omogoča hitre in natančne postavitve predmeta. Igralec mora sam prilagajati velikost ter postavitev klavirja, kar je lahko precej zamuden postopek. Vsi igralci so mnenja, da bi z bolj preprostim postopkom postavitve klavirja zagotovili manj začetne frustracije igralcu ter posledično večje zanimanje za igro.

Nekateri igralci si navidezne klaviature niso nastavili najbolj natančno čez pravo in jih je med igro motilo, da ne vidijo točnih mej tipk. Eden udeleženec je predlagal, da bi se temu lahko izognili tako, da bi vse bele in črne tipke po nastavitvi pozicije klaviature naredili transparentne. Na ta način bi igralec še vedno videl samo svojo pravo klaviaturo, potem pa bi se obarvale samo tiste tipke, ki jih mora igralec pritisniti med igranjem skladbe.

**Uvedba prikaza padajočih not:** Padajoče note (angl. pianoroll) so način vizualizacije not, ki se pogosto uporablja v glasbenih programih in igrah, na primer Magic Keys in PianoVision. To je prikaz, kjer note "padajo" z vrha zaslona proti dnu, vsaka vodoravna vrstica pa predstavlja eno tipko na klavirju. Ko nota doseže tipko, jo mora igralec pritisniti. Ta vizualizacija je dober način za učenje skladbe, saj igralcu poda informacije o prihodnjih notah, še preden jih mora pritisniti - kot, da bi bral note.

Nekateri udeleženci testiranja (predvsem začetniki) so izrazili, da jim je težko igrati neznano skladbo, če ne vejo že malo vnaprej katero noto morajo pritisniti. Navajeni so branja not, ki omogoča "videnje v prihodnost", v igri pa barvanje tipk ob istem času, kot jih mora igralec pritisniti, tega ne omogoča. Menijo, da bi pianoroll v veliki meri izboljšal izkušnjo učenja skladb.

**Ohranjanje obarvanih tipk ob pritisku na pavzo:** Igra je zasnovana tako, da se ob prekinitvi igranja skladbe barve vseh tipk ponastavijo na osnovno barvo (torej belo ali črno). Eden izmed udeležencev testiranja je med igranjem velikokrat ustavil skladbo, zato da bi videl kateri akord mora

pritisniti, vendar so se ob tem razveljavile tudi pobarvane tipke. Predlagal je, da se ob pavzi ohranijo prejšnje pobarvane tipke, saj bi to omogočilo igralcu, da med pavzo analizira pobarvane tipke ter si tako vzame več časa, da si zapomni prave note.

## 6. Diskusija in zaključek

Rezultati testiranja igre Improviano so dokazali, da je učenje improvizacije skozi navidezno igro lahko za posameznika poučno in zagotavlja hitre rezultate. Sistem točkovanja pravilnosti, ki je povsem izbiren, omogoča uporabniku možnost, da najprej skladbo in improvizacijo zvadi samostojno in se tako brez dodatnega pritiska osredotoča na učenje. Vizualni dodatki v igri mu hitreje pomagajo identificirati pravilne note, možnost prilagajanja hitrosti skladbe in deleža vizualnih podatkov pa mu podajajo večjo svobodo pri nastavitvi okolja, ki mu najbolj ustreza.

Tekom testiranja smo spoznali pomembnost uporabniške izkušnje. Veliko udeležencev je pripomnilo, da bi enostavnejša postavitve klaviature izboljšala njihovo izkušnjo igranja, saj je namen igre čim več časa igrati, ne pa se na to pripravljati. Poudarjali so tudi, da je način za improvizacijo zelo privlačen za igranje, vendar mu manjka poučnih vsebin. Želeli so si vadnic, kjer bi lahko samostojno zvalili posamezno lestvico, ter se naučili njeno zgradbo, ter vadnico, kjer bi glede na podani akord ugotovili, v kateri tonaliteti blues lestvice je potrebno igrati. Prav tako so pripomnili, da brez naše začetne razlage ne bi bili prepričani, kako igro igrati ali krmariti po njej.

Rezultati testiranja so nam podali boljšo predstavo o tem, kaj si glasbeniki in igralci pri učenju skozi igro želijo. Da bi bila aplikacija primerna za nadaljnjo uporabo v širši javnosti, bi morali implementirati uporabniški vmesnik, ki bi bil igralcu že od samega začetka dovolj jase, da bi ga lahko uporabljal sam, brez dodatne pomoči. Dodatna učna vsebina je pri takšni igri zelo pomembna, saj veliko začetnikov blues lestvice in osnovnih jazz načel še ne pozna. Da bi zagotovili pravilen učni postopek, ki zagotavlja najboljše in najhitrejšje rezultate, bi morali premisliti o natančnem zaporedju uvajanja novih informacij, povezanih z jazz in blues teorijo, ter prav tako počasi uvajati nove skladbe, z zahtevnostjo od najlažje do najtežje.

Pomemben del igre je tudi navidezna klaviatura, ki mora ves čas natančno pokrivati tipke fizične klaviature. Igralci so med testiranjem omenili, da je moteče, da vidijo hkrati tipke navidezne in fizične klaviature, zato bi ta element izboljšali tako, da bi nepobarvane tipke naredili transparentne. Tipke leve in desne roke so pobarvane z različnima barvama, kar je bilo igralcem všeč, so pa pripomnili, da enotna barva tipk pri prikazu blues lestvice ni idealna. Prikaz lestvice na navidezni klaviaturi bi izboljšali tako, da bi osnovno noto obarvali z drugo barvo; na ta način bi igralec hitreje vedel tonaliteto blues lestvice.

Priprava na igranje igre Improviano poteka precej veliko časa, saj mora igralec pripraviti računalnik in nanj priklopiti klaviaturo ter navidezna očala. Da bi postopek poenostavili, bi lahko igro izboljšali na tak način, da bi bila popolnoma primerna za samostojno delovanje na Android VR napravi, igralec pa bi lahko MIDI v USB-C kabel priklopil kar direktno iz svoje klaviature na očala. V našem primeru je bila težava v knjižnici DryWetMIDI, ki ni podprta za Android naprave, tako da bi bilo potrebno zamenjati večji del kode z drugo knjižnico, ki to podpira in mogoča enako upravljanje z MIDI podatki.

Vredno je omeniti tudi omejitve navidezne naprave, ki smo jo uporabili za testiranje igre. Meta Quest 3 ima določene prednosti in slabosti. Glavna pomanjkljivost je še vedno precej nizka ločljivost, predvsem kadar se uporablja obogatena resničnost, kjer je poleg navideznih elementov viden tudi prostor, v katerem se uporabnik nahaja (*angl. passthrough*). Poleg tega lahko dolgotrajna uporaba te naprave igralcu povzroči neudobje, saj je precej težka. Kljub temu ima naprava dobre lastnosti, saj jo odlikujeta cenovna dostopnost in enostavna uporaba, ki omogočata dostop do navidezne resničnosti širšemu krogu uporabnikov.

Z uvedbo predlaganih opomb bi uporabniku omogočili več poučnih vsebin ter boljšo izkušnjo igranja, s tem pa zagotovili boljše rezultate, ki bi se poznali na igranju klavirja ter improvizaciji. Tudi brez dodatnih učnih vsebin pa igra in primerjalni eksperiment dokazujeta, da je učenje improvizacije z VR uspešno in uporabniku omogoča lažji vstop v svet improvizacije, saj ima na voljo vodeno vsebino. Omogočena uporaba navideznih dodatkov na začetnike lahko vpliva pozitivno, sistem točkovanja pa

nekateri še dodatno spodbudi, da pri igranju vztrajajo še naprej.

## References

- [1] J. García, K. Dogani, Music in schools across Europe: analysis, interpretation and guidelines for music education in the framework of the European Union, Peter Lang, Frankfurt am Main, 2011, pp. 95–122.
- [2] A. Lamont, K. Maton, Unpopular music: Beliefs and behaviours towards music in education, in: R. Wright (Ed.), *Sociology and Music Education*, Ashgate, Basingstoke, 2010, p. 63–80.
- [3] X. Li, Y. Yang, S. K. W. Chu, Z. Zainuddin, Y. Zhang, Applying blended synchronous teaching and learning for flexible learning in higher education: an action research study at a university in hong kong, *Asia Pacific Journal of Education* 42 (2022) 211–227. doi:10.1080/02188791.2020.1766417.
- [4] H. McQueen, S. Hallam, A. Creech, Teachers' and students' music preferences for secondary school music lessons: reasons and implications, *Music Education Research* 20 (2018) 22–31. URL: <https://doi.org/10.1080/14613808.2016.1238059>. doi:10.1080/14613808.2016.1238059. arXiv:<https://doi.org/10.1080/14613808.2016.1238059>.
- [5] E. Costa-Giomi, P. J. Flowers, W. Sasaki, Piano lessons of beginning students who persist or drop out: Teacher behavior, student behavior, and lesson progress, *Journal of Research in Music Education* 53 (2005) 234–247. doi:10.1177/002242940505300305.
- [6] E. Rovithis, N. Moustakas, A. Floros, K. Vogklis, Audio legends: Investigating sonic interaction in an augmented reality audio game, *Multimodal Technologies and Interaction* 3 (2019). doi:10.3390/mti3040073.
- [7] M. Cook, Augmented reality: Examining its value in a music technology classroom. practice and potential., *Waikato Journal of Education* 24 (2019) 23–38. doi:10.15663/wje.v24i2.687.
- [8] J. Bissonnette, F. Dubé, M. Provencher, M. T. Moreno Sala, Virtual reality exposure training for musicians: Its effect on performance anxiety and quality, *Medical problems of performing artists* 30 (2015) 169–177. doi:10.21091/mppa.2015.3032.
- [9] L. Turchet, R. Hamilton, A. Çamci, Music in extended realities, *IEEE Access* 9 (2021) 15810–15832. doi:10.1109/ACCESS.2021.3052931.

# Razvoj didaktične mobilne igre za vzpodbujanje glasbene kreativnosti

Emir Hodžić<sup>1</sup>, Matevž Pesek<sup>1,\*</sup>

<sup>1</sup>University of Ljubljana, Faculty of computer and information science, Večna pot 113, 1000 Ljubljana, Slovenia

## Povzetek

Članek predstavlja razvoj in začetno vrednotenje prototipa mobilne aplikacije za igrivo glasbeno ustvarjanje. Aplikacija, zgrajena v okolju Flutter, združuje sekvencer in Piano Roll v preprosto, a odziven vmesnik, ki spodbuja eksperimentiranje in glasbeno "čečkanje" tudi brez predznanja teorije. Ključna tehnična rešitev je ločen proces za sintezo zvoka, ki zagotavlja tekoče delovanje in neprekinjen ustvarjalni tok. Uporabniška raziskava potrjuje visoko stopnjo stimulacije, jasnosti in privlačnosti, kar podpira hipotezo o motivacijskem učinku igrive zasnove, čeprav omejitve vzorca (7 udeležencev) zahtevajo obsežnejše nadaljnje študije. V primerjavi z obstoječimi orodji, kot sta GarageBand in Soundtrap, prototip izstopa po preprostosti ter potencialu za sodelovalno in didaktično rabo. Načrtovana nadgradnja vključuje mrežno sodelovanje, poigritvene elemente in testiranje v realnem pedagoškem okolju. Razviti prototip prikaže možnost združevanja dostopnosti za začetnike z možnostjo kompleksnega ustvarjanja ter odpiranja novih poti za sodobno glasbeno izražanje.

## Keywords

kreativnost, glasbeno izražanje, mobilna aplikacija, sodelovalno učenje

## 1. Uvod

Glasbeno izražanje se tradicionalno osredotoča na tehnično dovršenost, interpretacijo obstoječih del in učenje glasbene teorije. Čeprav so ti elementi temeljni za razvoj glasbenih veščin, pogosto ostaja v ozadju ključna komponenta glasbenega udejstvovanja – kreativnost. Spodbujanje učencev k samostojnemu ustvarjanju, improvizaciji in skladanju je pogosto zapostavljeno, kar lahko vodi v zmanjšano motivacijo in omejen razvoj celostnega glasbenega razumevanja [1].

Ta prispevek izhaja iz prepričanja, da je kreativnost temeljna človekova sposobnost, ki jo je potrebno sistematično negovati tudi pri glasbenem pouku. Pojav mobilnih tehnologij odpira nove možnosti za preoblikovanje pedagoških pristopov, saj lahko aplikacije delujejo kot interaktivna in dostopna orodja, ki olajšajo glasbeno ustvarjanje in spodbujajo učenje skozi igro. Na tej podlagi smo zasnovali in razvili prototip mobilne aplikacije, ki kot zbirka skupinskih didaktičnih iger krepi ustvarjalnost pri glasbenem pouku.

V teoretičnem delu članka bomo najprej opredelili pojem kreativnosti in predstavili ključne modele, ki pojasnjujejo ustvarjalni proces. Sledil bo pregled nevroznanstvenih dognanj o tem, kaj se dogaja v možganih med glasbenim ustvarjanjem, s posebnim poudarkom na procesih improvizacije in sodelovanja. V nadaljevanju bomo analizirali vlogo tehnologije, sodelovalnega učenja in obstoječih digitalnih rešitev v sodobnem izražanju. Ta pregled bo služil kot podlaga za identifikacijo vrzeli, ki jih poskušamo zapolniti z novo aplikacijo.

V nadaljevanju predstavimo razvoj mobilne aplikacije za podporo kreativnemu ustvarjanju glasbe pri samostojnem in skupinskem glasbenem izražanju. Z aplikacijo naslavljamo problematiko sodelovalnega učenja v sodobnem izražanju in jo preliminarno evalviramo z namenom nadaljnje uporabe med učitelji glasbe in učenci v pedagoškem procesu.

---

Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia

\*Corresponding author.

✉ eh3501@student.uni-lj.si (E. Hodžić); matevz.pesek@fri.uni-lj.si (M. Pesek)

🌐 <https://matevzpesek.si/> (M. Pesek)

🆔 0000-0001-9101-0471 (M. Pesek)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.3>

## 2. Pregled področja

V akademski skupnosti je splošno sprejeto, da kreativnost pomeni sposobnost ustvarjanja novih in uporabnih idej ter izdelkov [2]. Ta definicija zajema dva ključna elementa: novost in uporabnost. Dinamična definicija opisuje kreativnost kot dejanje, ki izvira iz zaznavanja okolja in prepoznavanja določenega neravnovesja. To vodi v produktivno aktivnost, ki izziva ustaljene miselne procese in norme ter ustvarja nekaj novega – bodisi v obliki fizičnega predmeta, bodisi zamisli ali čustvenega doživetja [3]. Kombinatorna narava kreativnosti se kaže v tem, da nove ideje, koncepti in izdelki nastajajo s povezovanjem obstoječih elementov na nove načine [4, 1], kar poudarja pomen znanja in izkušenj kot temeljnih gradnikov kreativnega izražanja.

Pri opredeljevanju kreativnosti se pogosto poudarja element ustvarjanja nečesa novega. Vendar pa je pomembno razumeti, da se ta novost lahko nanaša na ravni posameznika. Koncept "vsakdanje kreativnosti" [4, 1] priznava, da so lahko kreativna dejanja tudi otroška risba ali študentska pesem, čeprav ta dela morda niso prelomna v globalnem merilu. Za razumevanje kreativnosti v izobraževalnem kontekstu je ključna razmejitev, ki jo je uvedla filozofinja Margaret A. Boden. Razlikuje med dvema temeljnima vrstama kreativnosti: psihološko (P-kreativnost) in historično (H-kreativnost) [5]. **P-kreativnost (psihološka kreativnost)** se nanaša na ideje ali izdelke, ki so novi, presenetljivi in predstavljajo vrednost posamezniku, ki jih je ustvaril. Pri tem ni pomembno, ali je kdo drug v zgodovini že prišel do enake zamisli. Gre za osebni preboj, za trenutek, ko posameznik sam zase odkrije nekaj novega. To je vrsta kreativnosti, ki jo doživlja otrok, ko prvič uspešno sestavi hišo iz kock, ali študent, ki samostojno najde novo rešitev za programerski problem. P-kreativnost je temeljna za učenje, saj predstavlja proces osebnega odkrivanja in osmišljanja sveta. **H-kreativnost (historična kreativnost)** na drugi strani opisuje ideje, ki niso nove le za posameznika, temveč za celotno človeško zgodovino. Gre za prelomna dela in odkritja genijev, kot so Mozartove simfonije ali Einsteinova teorija relativnosti. H-kreativnost je izjemno redka in predstavlja poseben primer P-kreativnosti – vsaka H-kreativna ideja mora biti namreč najprej P-kreativna.

Ta razdelitev je za načrtovano mobilno aplikacijo bistvenega pomena, saj postavlja cilje v realen in pedagoško smiseln okvir. Cilj naše aplikacije ni spodbujanje H-kreativnosti oziroma iskanje novega Mozarta. Takšen cilj bi bil nerealen in večine uporabnikov ne bi motiviral. Namesto tega je naša aplikacija zasnovana kot orodje za sistematično spodbujanje P-kreativnosti. Želimo ustvariti dostopno in varno digitalno okolje, v katerem lahko vsak uporabnik, ne glede na predznanje, doživi lastno izkušnjo glasbenega ustvarjanja. Z omogočanjem preprostega kombiniranja ritmičnih in melodičnih elementov aplikacija uporabnika vabi k eksperimentiranju in igrivemu raziskovanju. Končni izdelek – melodija ali ritem – morda ne bo zgodovinsko prelomen, a za uporabnika bo predstavljal nekaj novega in osebno pomembnega. Prav ta izkušnja osebnega dosežka in "aha" trenutka je ključna za razvoj notranje motivacije, samozavesti in dolgoročnega zanimanja za glasbo, kar je temeljni namen našega didaktičnega orodja.

### 2.1. Kreativnost v glasbenem izobraževanju

Kreativnost predstavlja ključni element sodobnega glasbenega izobraževanja, ki presega tradicionalno osredotočenost na ponavljanje in tehnično dovršenost. V kontekstu glasbenega izobraževanja kreativnost zajema sposobnost učencev, da raziskujejo, eksperimentirajo in ustvarjajo lastne glasbene izraze, pri čemer povezujejo teoretično znanje s praktično uporabo. V kontekstu glasbenega izobraževanja lahko skladanje razumemo kot kreativni proces reševanja problemov, kjer učenci sprejemajo odločitve in raziskujejo različne možnosti, dokler ne najdejo rešitve, ki najbolj ustreza njihovim ustvarjalnim namenom. Burnard ta proces opisuje kot iskanje "najboljše ustreznosti", s poudarkom na tem, kako pomembni so iteracija, eksperimentiranje in izpopolnjevanje pri glasbenem ustvarjanju [6].

Tehnologija je spremenila področje glasbenega izobraževanja in odprla nove možnosti za kreativno raziskovanje. Tehnologija učencem omogoča ukvarjanje z "mikro-fenomeni" zvoka, kar jim daje možnost za podrobno manipulacijo in eksperimentiranje [7]. Orodja, kot so *reverb*, *delay* in digitalne avdio delovne postaje, učencem omogočajo neposredno interakcijo z zvokom ter razširjajo obseg ustvarjalnih

možnosti. Ključen vidik kreativnosti v glasbi je proces raziskovanja, pogosto opisan kot glasbeno "čečkanje". Savage ta pojem opisuje kot neformalno in intuitivno interakcijo z zvokom, kjer "nek srečen slučaj gibanja ujame domišljijo – ritmični utrip, fraza ali sprememba tempa" [7]. To spontano eksperimentiranje spodbuja inovativnost in služi kot osnova za bolj strukturirano ustvarjalno delo.

Razumevanje nevroloških osnov glasbene kreativnosti nam omogoča boljši vpogled v to, kako zasnovati orodja, ki podpirajo naravne kognitivne procese. V tem razdelku se osredotočamo na dogajanje v možganih med ustvarjalnim procesom, s poudarkom na prehodih med spontan in strukturiranim delom. Kreativni procesi, kot so improvizacija, komponiranje in glasbeno eksperimentiranje, vključujejo delovanje več možganskih delov, ki prispevajo k ustvarjanju, ocenjevanju in izvedbi novih idej. Raziskave so pokazale, da so pri teh procesih aktivna tako območja, povezana s spontanostjo, kot tudi območja, ki omogočajo strukturirano razmišljanje.

Improvizacija, na primer, pogosto vključuje dorzolateralni prefrontalni korteks (DLPFC), kar nakazuje, da gre za transformativno vedenje, ki zahteva kompleksno presojo in intelektualni napor [8]. Kljub temu pa se v drugih študijah improvizacija povezuje z delovanjem privzetega omrežja (DMN), ki je bolj aktivno med notranje usmerjenimi miselnimi procesi, kot so domišljija in razmišljanje brez konkretnih ciljev [8]. To kaže, da lahko različni vidiki ustvarjalnosti aktivirajo različna možganska omrežja, odvisno od narave naloge. Prehodi med spontanimi in strukturiranimi oblikami ustvarjalnosti predstavljajo dinamičen proces, v katerem sodelujejo različna možganska omrežja. Spontana ustvarjalnost se običajno pojavi brez zavestnega napora in vključuje privzeto omrežje (DMN) [8]. Nasprotno pa strukturirana ustvarjalnost zahteva zavestno osredotočenost [9] in vključuje omrežja, povezana s kognitivnim nadzorom, kot je dorzolateralni prefrontalni korteks [8]. Prehodi med spontanimi in strukturiranimi procesi vključujejo izmenjavo med privzetim omrežjem in omrežji, ki so odgovorna za pozornost in nadzor. Ta dinamična izmenjava, imenovana tudi "ustvarjalna sistola in diastola" [8], omogoča, da se ustvarjalni procesi prilagajajo spreminjajočim zahtevam nalog.

## 2.2. Obstoječe aplikacije

*Apple GarageBand*, *BandLab* in *Figure* ponujajo tri različne pristope k digitalnemu glasbenemu ustvarjanju, ki pa vsak zase ne zadostijo ciljem našega projekta. *GarageBand* je brezplačna, zelo razširjena aplikacija z bogato knjižnico zvokov in zmogljivim večslednim snemalnikom, a je primarno namenjena produkciji, z zahtevnim vmesnikom in omejenim sodelovanjem [10]. *BandLab* uvaža sodelovalno delo v oblaku in učiteljem omogoča dodeljevanje nalog ter spremljanje napredka, vendar ostaja osredotočen na sodelovalno produkcijo, kar ni optimalno za spontano, igrivo učenje [11]. *Figure* pa s svojo izjemno preprostostjo in vizualno privlačnimi kontrolniki omogoča takojšnje ustvarjanje in stanje zanos, a je namenjen posamezniku in nima didaktične nadgradnje [12].

Analiza sorodnih del jasno pokaže, da obstaja nezapolnjena niša. Na eni strani so zmogljiva, a kompleksna produkcijska orodja (*GarageBand*, *BandLab*), ki so za didaktično rabo v začetnih fazah učenja manj primerna. Na drugi strani imamo izjemno intuitivne in igrive aplikacije (*Figure*), ki pa so omejene na individualno izkušnjo in nimajo pedagoške nadgradnje.

Naša rešitev je zasnovana kot neposreden odgovor na to vrzel, pri čemer njena zasnova zavestno temelji na teoretičnih modelih kreativnosti, predstavljenih v prejšnjih poglavjih. Ne gre zgolj za tehnološko orodje, temveč za pedagoško okolje, ki skuša v praksi udejanjiti spoznanja o tem, kako ljudje ustvarjamo. V naši rešitvi zato združujemo najboljše iz teh pristopov: dostopnost in igrivost *Figure*, strukturiran proces ustvarjanja (npr. z orodjem *Piano Roll*) in sodelovalne funkcije, ki jih je navdihnil *BandLab*, ob hkratnem zavestnem odmiku od kompleksnosti in profesionalne paradigme *GarageBanda*. Tako nastaja pedagoško okolje, ki podpira P-kreativnost, omogoča ravnoesje med divergentnim in konvergentnim mišljenjem ter v realnem času spodbuja sodelovanje, s čimer zapolnjuje vrzel med profesionalnimi orodji in preprostimi, a individualnimi aplikacijami.

V prvi vrsti naša aplikacija naslavlja razlikovanje med P-kreativnostjo in H-kreativnostjo. Zavestno se odklikamo od paradigme profesionalne produkcije, značilne za H-kreativnost, ki jo posebej *GarageBand*. Namesto tega ustvarjamo prostor za P-kreativnost – osebno, vsakdanjo ustvarjalnost. Cilj ni ustvariti zgodovinsko pomembno delo, temveč omogočiti vsakemu uporabniku, da doživi lasten

“aha” trenutek in zadovoljstvo ob ustvarjanju nečesa, kar je novo in vredno zanj osebno. To dosegamo z implementacijo filozofije, ki smo jo spoznali pri aplikaciji Figure, in jo neposredno povezujemo s konceptom zanosa (ang. *flow*). Z nizkim vstopnim pragom, intuitivnim vmesnikom in takojšnjo zvočno povratno informacijo ustvarjamo pogoje, v katerih je izziv v ravnovesju z uporabnikovimi sposobnostmi. To zmanjšuje frustracijo, ki jo povzroča kompleksnost avdio postaj (angl. Digital Audio Workstation – DAW), in preprečuje dolgočasje, v katerega lahko vodijo preveč omejene aplikacije. Uporabnik je povabljen v stanje popolne osredotočenosti, kjer je dejavnost sama sebi nagrada – to je bistvo glasbenega “čečkanja” in igrivega raziskovanja, ki sta temelj za razvoj kreativnosti.

Hkrati naša aplikacija ponuja pot naprej. Medtem ko je Figure osredotočen predvsem na spontano ustvarjanje, naša rešitev z vključitvijo bolj strukturiranih orodij, kot je Piano Roll, podpira celoten kreativni proces po Wallasovem [1, 6] ali Websterjevem [6, 13] modelu. Uporabnik lahko preklaplja med divergentnim mišljenjem (igrivo postavljanje not v sekvencerju) in konvergentnim mišljenjem (natančno urejanje melodije v *Piano Roll*-u), kar ustreza fazam od priprave in iluminacije do končne verifikacije in izpopolnjevanja ideje.

### 3. Aplikacija

Razvita aplikacija je večplatformni prototip za Android/iOS, zgrajen v Flutterju (Dart) z upravljanjem stanja prek Providerja, zasnovan po večplastni arhitekturi (UI - State - Services - Data Models), ki omogoča razširljivost, ločeno odgovornost in odziven uporabniški vmesnik. Jedro predstavljata hibridni avdio pogon in nizkoločljivostne odločitve: SoLoud predvaja prednaložene .wav vzorce z minimalno zakasnitvijo za ritmične sledi, medtem ko Soundfont v ločenem izolatu (Isolate) sintetizira melodične tone iz SoundFont (.sf2) in vrne PCM medpomnilnike za tekoče predvajanje; asinhroni tokovi in async/await skrbijo, da uporabniški vmesnik ostaja neblokiran. Podatkovni model je kompakten in namenski: Track (koraki + akcenti + glasnost), Note (višina, nevezano pozicioniranje s pozicijo in dolžino), Pattern (ritem + melodija z metapodatki: BPM, swing, accent, instrument), Song (aranžma zaporedij vzorcev).

Uporabniški vmesnik je zasnovan kot jasno strukturiran tok ustvarjanja z ViewSelectorjem in tremi pogledi: Sequencer za ritem na Sliki 2, Piano Roll za melodijo na Sliki 3 ter Song Mode za aranžma. Takšna organizacija omogoča osredotočeno delo na enem kognitivnem problemu naenkrat, hkrati pa podpira naravni prehod od igrivega eksperimentiranja k strukturiranemu urejanju. Vizualna metaforika posnema fizične naprave (mreža korakov, rotacijski gumbi, drsniki), kar zmanjšuje kognitivno breme in olajša prenos obstoječih mentalnih modelov v digitalni kontekst.

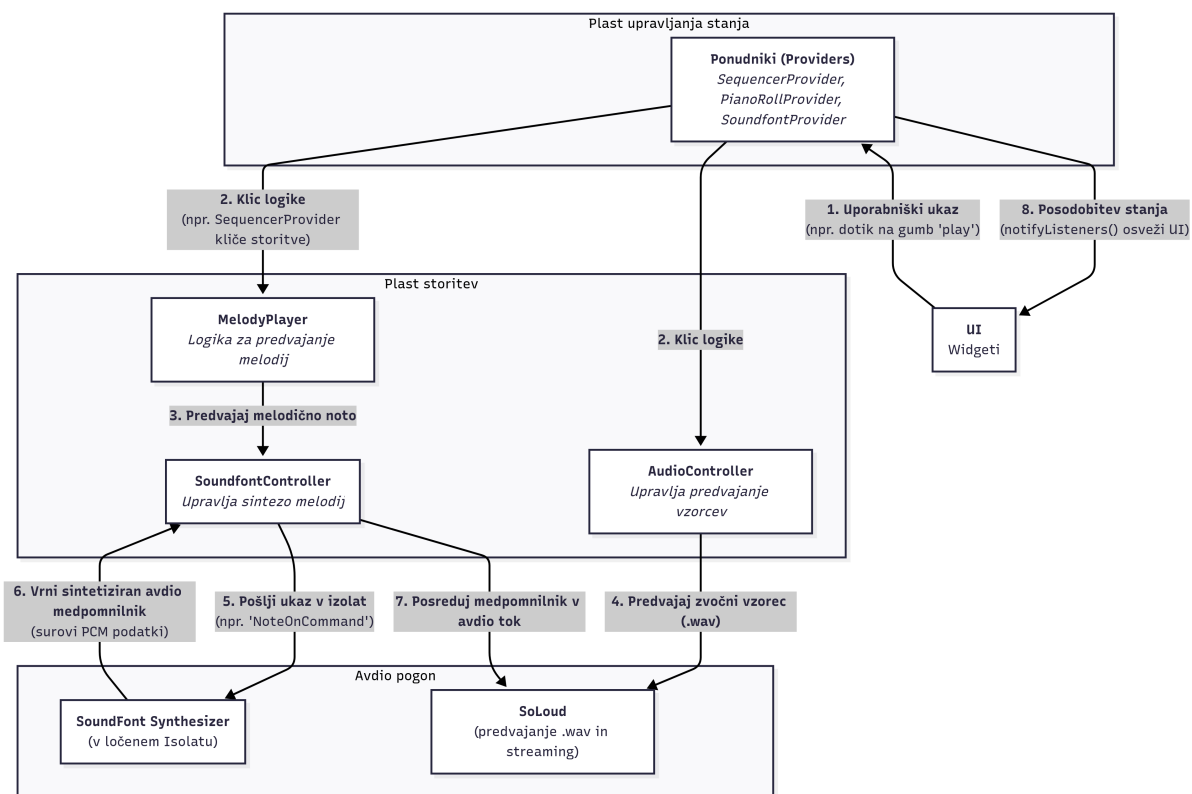
V Sequencerju mreža SequencerGrid operacionalizira “tinkering”: onTap vklopi/izklopi korak, onLongPress doda akcent, aktivni korak je med predvajanjem osvetljen, ob strani pa so drsniki za glasnost po sledeh. V Piano Rollu dvodimenzionalna mreža omogoča dodajanje not z dotikom, njihovo premikanje v času in spreminjanje trajanja z vlečenjem roba; “snap to grid” omogoča preklap med natančnostjo in svobodo, izbirnik instrumenta pa takojšnjo zvočno primerjavo. Song Mode omogoča sestavljanje vzorcev v celoto z minimalnim naborom gest, kar sklene pot od ideje do aranžmaja brez izhoda iz ustvarjalnega toka.

Interakcije temeljijo na GestureDetector in reaktivnem posodabljanju prek Providerja (notifyListeners()), zato se ob spremembah prerišejo le odvisni gradniki, medtem ko začasna vizualna stanja (npr. med vlečenjem note) ohranjajo 60 fps odzivnost. Vmesnik daje takojšnjo vizualno in zvočno povratno informacijo (nizka latenca avdio pogona), uporablja dosledno barvno kodiranje stanj (aktivno/akcentirano/izbrano) in ohranja tipografsko hierarhijo za orientacijo. Skupaj te odločitve ustvarijo okolje z nizkim vstopnim pragom, ki spodbuja hitro iteracijo, učenje skozi igro in prehod od divergentnega h konvergentnemu mišljenju.

#### 3.1. Diagram arhitekture in tok podatkov

Za lažje razumevanje medsebojnega delovanja opisanih plasti smo pripravili diagram arhitekture sistema, ki prikazuje ključne komponente in tok informacij med njimi. Diagram na Sliki 1 ponazarja celoten

proces, od uporabnikove interakcije do generiranja zvoka in osvežitve uporabniškega vmesnika.



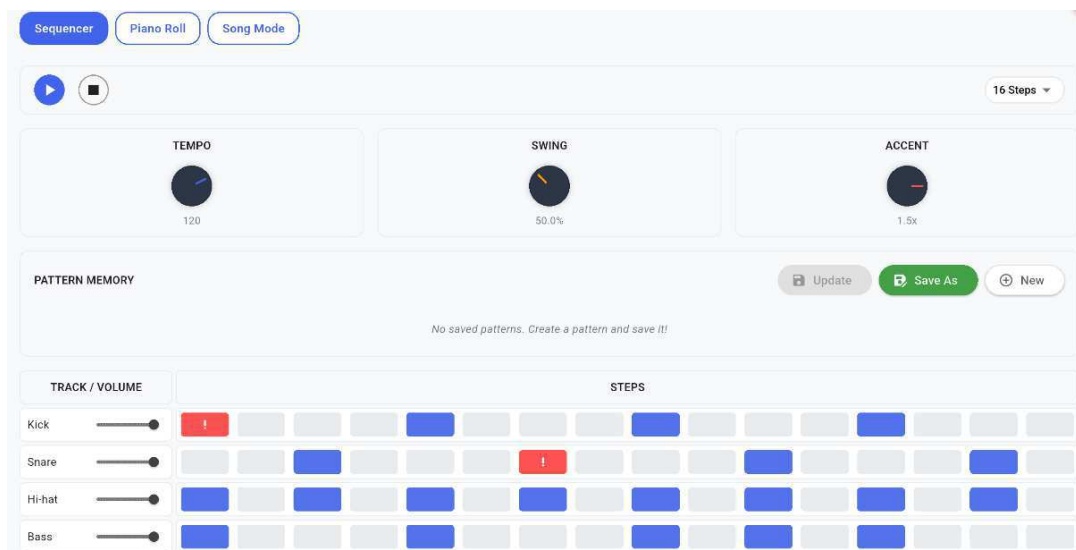
**Slika 1:** Diagram arhitekture sistema, ki prikazuje tok podatkov od uporabniškega vmesnika preko plasti za upravljanje stanja in storitev do avdio pogona. Številke ponazarjajo zaporedje ključnih korakov v procesu.

Celoten proces, ki ga sproži uporabnik, poteka po naslednjih korakih, oštevilčenih na diagramu:

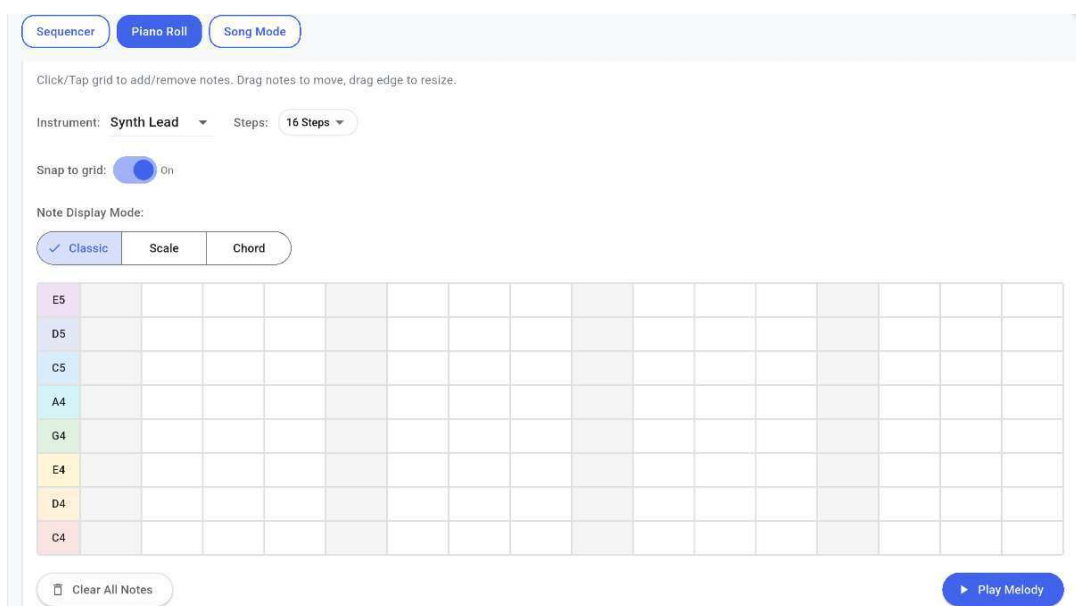
- 1. Uporabniški ukaz:** Proces se začne v plasti uporabniškega vmesnika, ko uporabnik izvede dejanje, na primer pritisne gumb za predvajanje ali doda noto v mrežo. Gradnik (ang. *Widget*) ne vsebuje logike, temveč zgolj pokliče ustrezno metodo v plasti za upravljanje stanja.
- 2. Klic logike:** Ponudnik (npr. *SequencerProvider*) sprejme ukaz in sproži logiko, ki je implementirana v plasti storitev. To pomeni klic metod v ustreznih krmilnikih, kot sta *MelodyPlayer* ali *AudioController*.
- 3. Razdelitev na avdio poti:** Na tej točki se tok razdeli glede na tip zvoka. Če je treba predvajati melodično noto, *MelodyPlayer* sproži ukaz v *SoundfontController*. Če pa je treba predvajati kratek zvočni vzorec (npr. udarec bobna), *AudioController* neposredno komunicira s knjižnico *SoLoud*, ki vzorec predvaja iz pomnilnika z minimalno zakasnitvijo.
- 4. Ukaz za sintezo:** Za melodične note *SoundfontController* zapakira informacije o noti (višina, jakost) v ukazni objekt (npr. *NoteOnCommand*) in ga preko namenskega komunikacijskega kanala pošlje v ločen proces (izolat).
- 5. Sinteza v izolatu:** Izolat, ki teče v ozadju in ne blokira uporabniškega vmesnika, sprejme ukaz. V njem delujoč sintetizator na podlagi datoteke *SoundFont* generira surove zvočne podatke (v formatu PCM) za zahtevano noto in jih zapakira v medpomnilnik. Ta medpomnilnik nato pošlje nazaj na glavni proces.
- 6. Posredovanje avdio toka:** *SoundfontController* sprejme sintetiziran medpomnilnik in ga posreduje knjižnici *SoLoud*, ki ga doda v neprekinjen avdio tok za predvajanje. Ta pristop omogoča tekoče predvajanje kompleksnih melodij brez prekinitev.
- 7. Posodobitev stanja in osvežitev vmesnika:** Po izvedeni logiki ponudnik posodobi svoje stanje (npr. spremeni trenutni korak sekvencerja) in pokliče metodo *notifyListeners()*. S tem obvesti vse

gradnike v UI, ki “poslušajo” ta del stanja, da se morajo osvežiti in prikazati novo stanje. S tem se sklone zanka in uporabniški vmesnik vedno odraža trenutno stanje aplikacije.

Ta arhitektura, ki jasno ločuje odgovornosti in prenaša računsko zahtevne operacije v ozadje, je ključna za zagotavljanje odzivne in tekoče uporabniške izkušnje, kar je pri glasbenih aplikacijah bistvenega pomena.



**Slika 2:** Uporabniški vmesnik pogleda “Sequencer”. Vidna je mreža s sledmi, koraki in kontrolnimi elementi.



**Slika 3:** Uporabniški vmesnik pogleda “Piano Roll” za urejanje melodičnih linij.

## 4. Evalvacija

Za objektivno oceno razvitega prototipa in preverjanje doseganja zastavljenih ciljev je bila izvedena empirična evalvacija z uporabniki. Namen vrednotenja je bil pridobiti vpogled v to, kako uporabniki z različnimi stopnjami predznanja dojemajo aplikacijo, ter na podlagi kvantitativnih in kvalitativnih podatkov identificirati njene ključne prednosti in področja za nadaljnje izboljšave.

## 4.1. Udeleženci

V raziskavi je sodelovalo 7 prostovoljcev, starih med 20 in 30 let. Vzorec je bil heterogen glede na predhodne izkušnje z glasbeno programsko opremo, kar je omogočilo celovitejšo oceno dostopnosti in intuitivnosti aplikacije:

- 3 udeleženci niso imeli predhodnih izkušenj.
- 3 udeleženci so imeli osnovne izkušnje (npr. z aplikacijo GarageBand).
- 1 udeleženec je bil izkušen uporabnik (npr. FL Studio, Ableton Live).

Vsak udeleženec je individualno izvedel sejo, ki je vključevala naslednje korake:

1. **Uvodna predstavitev (2 min):** Udeležencu je bil na kratko predstavljen namen aplikacije. Poudarjeno je bilo, da testiramo aplikacijo in ne njihovih sposobnosti.
2. **Izvedba nalog (10–15 min):** Udeleženec je samostojno rešil niz vnaprej pripravljenih nalog, ki so pokrivalo ključne funkcionalnosti: ustvarjanje ritma v sekvencerju, dodajanje poudarkov, komponiranje melodije v pogledu *Piano Roll* ter shranjevanje in nalaganje vzorcev.
3. **Izpolnjevanje vprašalnika (5–10 min):** Takoj po uporabi aplikacije je vsak udeleženec izpolnil standardiziran vprašalnik uporabniške izkušnje (UEQ) in odgovoril na odprta vprašanja.
4. **Intervju:** Po uporabi smo z udeležencem izvedli intervju.

## 5. Rezultati

### 5.1. Kvantitativni rezultati (UEQ)

Podatki, zbrani od 7 udeležencev, so bili obdelani z uradnim orodjem za analizo UEQ. Tabela 1 prikazuje povprečne vrednosti za vsako od šestih lestvic (na intervalu od -3 do +3) in njihovo primerjalno uvrstitev (ang. *Benchmark*) glede na obsežno bazo podatkov več sto drugih produktov.

**Tabela 1**

Povprečni rezultati analize UEQ (N=7).

| Lestvica     | Povprečna vrednost | Interpretacija |
|--------------|--------------------|----------------|
| Privlačnost  | 1.95               | Dobro          |
| Jasnost      | 1.62               | Dobro          |
| Učinkovitost | 1.48               | Nadpovprečno   |
| Zanesljivost | 1.71               | Dobro          |
| Stimulacija  | 2.15               | Odlično        |
| Novost       | 1.88               | Dobro          |

Kot je razvidno iz grafičnega prikaza na Sliki 4, je aplikacija dosegla visoke ocene na vseh področjih. Izstopa izjemno pozitivna ocena na lestvici **stimulacije**, kar neposredno naslavlja osrednji cilj raziskave – spodbujanje motivacije in kreativnosti.

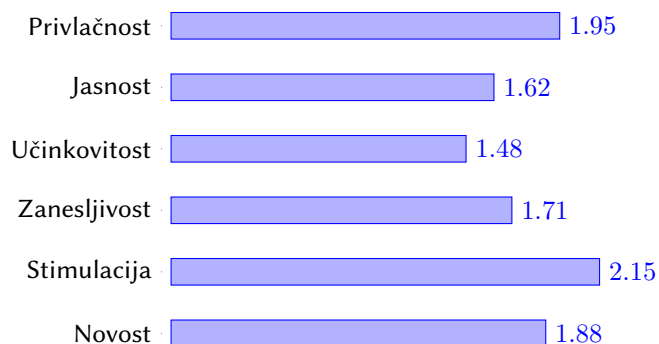
### 5.2. Kvalitativni rezultati

Analiza odprtih vprašanj je ponudila kontekst za kvantitativne rezultate.

#### 5.2.1. Identificirane prednosti aplikacije

Analiza odgovorov je pokazala, da so uporabniki kot ključno prednost soglasno izpostavili neposrednost interakcije in igrivost, kar neposredno potrjuje uspešnost zasnove, ki temelji na konceptu glasbenega "čečkanja". Uporabniki, zlasti tisti brez predhodnih izkušenj, so poudarili, da takojšnja zvočna povratna informacija omogoča hitro učenje in ustvarjanje brez potrebe po branju navodil. Interakcija z mrežo v sekvencerju je bila opisana kot intuitivna in zabavna, kar prispeva k visoki oceni stimulacije. Izkušenejši uporabniki so opazili, da takšna zasnova spodbuja hitro preizkušanje idej in omogoča doseganje stanja ustvarjalnega zanosa (ang. *flow*).

**Slika 4:** Grafični prikaz rezultatov UEQ po lestvicah.



### 5.2.2. Identificirane pomanjkljivosti in težave

Kritike in opažene težave so bile večinoma povezane z vidiki uporabniškega vmesnika, ki bi lahko bili bolj učinkoviti, ter z odkrivanjem naprednejših funkcij. Ponavljajoča se pripomba se je nanašala na odkrivanje funkcij, ki so dostopne preko specifičnih gest, kot je dolgi pritisk za dodajanje poudarkov ("accent"). Uporabniki so to funkcionalnost odkrili pretežno po naključju. Nadalje je bila izpostavljena manjša natančnost pri interakciji v pogledu Piano Roll, kjer je bilo na manjših zaslonih težko ločiti med premikanjem note in spreminjanjem njene dolžine. Tretje področje za izboljšave je jasnost delovnega toka pri upravljanju z vzorci, saj logika shranjevanja in nalaganja uporabnikom brez izkušenj ni bila takoj intuitivna.

### 5.2.3. Predlogi za nadaljnji razvoj

Na podlagi povratnih informacij udeležencev so bile identificirane štiri ključne smeri za nadaljnji razvoj, osredotočene na izboljšanje uporabnosti in razširitev funkcionalnosti v skladu z zastavljenimi didaktičnimi cilji.

- **Podpora za nelinearni potek dela:** Najpogosteje izpostavljena potreba je bila implementacija funkcije za razveljavitev in uveljavitev dejanj (ang. *Undo/Redo*). Odsotnost te temeljne zmožnosti predstavlja tveganje za nenamerno izgubo dela, kar lahko zavira ustvarjalno eksperimentiranje, ki je v središču zasnove aplikacije. Vključitev te funkcionalnosti bi uporabnikom omogočila bolj svobodno in brezskrbno raziskovanje.
- **Izboljšanje uvajanja uporabnikov:** Za lažje odkrivanje funkcij, zlasti tistih, ki so vezane na specifične geste (npr. dolgi pritisk), so udeleženci predlagali vključitev kratkega interaktivnega vodiča ob prvem zagonu. Sistematičen uvodni postopek bi pospešil učno krivuljo, zmanjšal kognitivno obremenitev in zagotovil, da uporabniki razumejo celoten spekter orodij, ki so jim na voljo.
- **Povečanje učinkovitosti delovnega toka:** Izkušenejši udeleženci so izrazili potrebo po naprednejših orodjih za urejanje, kot je kopiranje in lepljenje ritmičnih ter melodičnih fraz med različnimi vzorci. Implementacija takšnih funkcij bi bistveno zmanjšala ponavljajoče se delo in omogočila hitrejšo gradnjo kompleksnejših glasbenih struktur, kar podpira prehod od enostavnega eksperimentiranja k bolj strukturiranemu komponiranju.
- **Implementacija osrednje vizije – sodelovanja:** Vsi udeleženci so pokazali veliko zanimanje za načrtovani sodelovalni način in ga prepoznali kot ključni potencial aplikacije. Ta ugotovitev potrjuje, da je socialna in skupinska dimenzija glasbenega ustvarjanja pomemben motivacijski dejavnik, zato mora biti implementacija mrežnih funkcionalnosti za igro v realnem času prednostna naloga pri nadaljnjem razvoju.

## 6. Diskusija

Cilj članka je bil zasnovati in razviti orodje, ki niža vstopni prag za glasbeno ustvarjanje in spodbuja kreativnost skozi igro in sodelovanje. Razviti prototip ta cilj dosega z več ključnimi značilnostmi. Uporabniški vmesnik, zlasti pogled sekvencerja, s svojo preprosto mrežo omogoča takojšnje eksperimentiranje z ritmi. Ta neposredna interakcija uteleša koncept glasbenega "čečkanja", kjer lahko uporabniki brez predznanja teorije glasbe raziskujejo in odkrivajo glasbene zamisli. S tem se naslavlja problem, kjer tradicionalni pristopi pogosto zahtevajo visoko stopnjo tehničnega znanja, preden dovolijo ustvarjalno izražanje.

V primerjavi s sorodnimi aplikacijami, kot sta GarageBand ali Soundtrap, ki sta močni, a kompleksni produkcijski okolji, naša aplikacija ponuja bolj osredotočeno in igrivo izkušnjo. Medtem ko GarageBand in podobna orodja ponujajo profesionalne zmožnosti, lahko njihova kompleksnost na začetnike deluje zastrašujoče. Naša rešitev se bolj zgleduje po preprostosti aplikacije Figure, a jo nadgrajuje z bolj strukturiranimi pogledi, kot je Piano Roll, in z arhitekturo, ki je pripravljena na prihodnjo implementacijo sodelovalnih in didaktičnih funkcionalnosti.

Arhitekturna zasnova, ki ločuje avdio pogon v ločen proces (izolat), se je izkazala za ključno pri zagotavljanju odzivnega uporabniškega vmesnika. To je pomembno, saj vsaka zakasnitev ali zatikanje med ustvarjalnim procesom prekine tok misli in zmanjša motivacijo. S tehničnega vidika je torej aplikacija robustna osnova za nadaljnji razvoj.

Kljub temu pa ima prototip v trenutni fazi pomembne omejitve. Čeprav je bilo izvedeno začetno vrednotenje uporabniške izkušnje, je treba njegove rezultate interpretirati z zadržkom zaradi pomembne metodološke omejitve. Testni vzorec je bil z zgolj sedmimi udeleženci relativno majhen, kar zadošča za zbiranje prvih kvalitativnih vtisov, ne pa tudi za statistično zanesljive zaključke. Za pridobitev bolj robustnih in posplošenih ugotovitev bi bilo nujno izvesti obsežnejšo raziskavo, ki bi v vsako od treh opredeljenih skupin uporabnikov – začetnike, osnovno izkušene in napredne – vključila vsaj za en velikostni razred (50-100) večje število udeležencev.

Poleg omenjene metodološke omejitve ima prototip tudi funkcionalne in vsebinske pomanjkljivosti. Največja je, da kljub zasnovi, usmerjeni v sodelovanje, dejanske mrežne funkcionalnosti za skupinsko delo še niso implementirane. Aplikacija trenutno deluje kot orodje za enega uporabnika. Nadalje manjka formalna evalvacija v realnem izobraževalnem okolju; testiranje z učenci in učitelji bi bilo nujno za pridobitev povratnih informacij o didaktični vrednosti in motivacijski privlačnosti. Nazadnje je tudi nabor didaktičnih iger zaenkrat le konceptualen; implementiran je osnovni ustvarjalni način, ki pa še ne vključuje specifičnih elementov poigritve, kot so cilji, točkovanje ali vodeni izzivi.

Empirično vrednotenje je potrdilo, da razviti prototip uspešno naslavlja zastavljene cilje in je s strani uporabnikov zelo dobro sprejet. Izjemno visoka ocena na lestvici stimulacije (2.15, "Odlično") neposredno potrjuje osrednjo hipotezo zastavljenega cilja – da igriv in neposreden pristop k ustvarjanju glasbe učinkovito spodbuja kreativnost, motivacijo in doseganje stanja zanosa. Visoke ocene na lestvicah jasnosti in privlačnosti kažejo, da je aplikacija dostopna tudi popolnim začetnikom, kar je bil eden od ključnih ciljev zasnove. Nekoliko nižja, a še vedno nadpovprečna ocena učinkovitosti, skupaj s kvalitativnimi komentarji, jasno nakazuje na najpomembnejša področja za izboljšave: dodajanje funkcije "Undo/Redo", izboljšanje natančnosti vmesnika v pogledu *Piano Roll* in boljše komuniciranje naprednih funkcij. Vrednotenje je pokazalo, da je prototip trdna in obetavna osnova. Z upoštevanjem priporočil ima aplikacija velik potencial, da postane učinkovito in priljubljeno orodje za glasbeno izobraževanje.

## 7. Zaključek

Osrednji prispevek dela je funkcionalni prototip mobilne aplikacije, razvit v ogrodju Flutter. Aplikacija združuje sekvencer za ustvarjanje ritmov in Piano Roll za komponiranje melodij v intuitivnem in odzivnem uporabniškem vmesniku. Ključna tehnična rešitev je bila uporaba ločenega procesa (izolata) za sintezo zvoka, kar zagotavlja tekoče delovanje aplikacije brez zakasnitev, ki bi motile ustvarjalni proces. Arhitektura sistema je zasnovana modularno in razširljivo, kar omogoča preprosto dodajanje

novih funkcionalnosti v prihodnosti.

Z razvojem prototipa smo uspešno dokazali, da je mogoče ustvariti orodje, ki je hkrati dostopno začetnikom in dovolj zmogljivo za kompleksnejše ustvarjanje. Zasnova vmesnika neposredno podpira koncepte, kot sta kombinatorna kreativnost in glasbeno "čečkanje", ter s tem aktivno spodbuja eksperimentiranje in raziskovanje.

Razvita aplikacija odpira številne možnosti za nadaljnji razvoj. Ključni naslednji korak je implementacija mrežnega modula za omogočanje sodelovanja v realnem času, kar bi aplikacijo preoblikovalo v pravo orodje za skupinske didaktične igre. Nadalje bi bilo potrebno razviti specifične mehanizme poigritve (izzive, cilje, povratne informacije), ki bi učence vodili skozi učne vsebine. Nujna bo tudi izvedba sistematičnega testiranja s ciljno skupino – učenci in učitelji glasbe –, da bi pridobili empirične podatke o učinkovitosti našega pristopa in zbrali predloge za izboljšave. Verjamemo, da ima predstavljeni koncept potencial, da postane dragocen pripomoček pri sodobnem glasbenem pouku, ki v ospredje postavlja ustvarjalnost, sodelovanje in veselje do glasbe.

## Literatura

- [1] M. Fautley, J. Savage, *Creativity in Secondary Education*, 2007. doi:10.4135/9781446278727.
- [2] B. Gaut, *The philosophy of creativity*, *Philosophy Compass* 5 (2010) 1034–1046. doi:<https://doi.org/10.1111/j.1747-9991.2010.00351.x>.
- [3] C. Walia, *A dynamic definition of creativity*, *Creativity Research Journal* 31 (2019) 237–247. doi:10.1080/10400419.2019.1641787.
- [4] R. A. Beghetto, *Creative learning: A fresh look*, *Journal of Cognitive Education and Psychology* 15 (2016) 6–23. doi:10.1891/1945-8959.15.1.6.
- [5] M. A. Boden, Chapter 9 - creativity, in: M. A. Boden (Ed.), *Artificial Intelligence, Handbook of Perception and Cognition*, Academic Press, San Diego, 1996, pp. 267–291. doi:<https://doi.org/10.1016/B978-012161964-0/50011-X>.
- [6] P. Burnard, B. Younker, *Problem-solving and creativity: Insights from students' individual composing pathways*, *International Journal of Music Education* 22 (2004) 59–76. doi:10.1177/0255761404042375.
- [7] J. Savage, *Working towards a theory for music technologies in the classroom: how pupils engage with and organise sounds with new technologies*, *British Journal of Music Education* 22 (2005) 167–180. doi:10.1017/S0265051705006133.
- [8] *Cambridge Handbooks in Psychology*, Cambridge University Press, 2018, p. i–ii.
- [9] A. Abraham, *The Neuroscience of Creativity*, *Cambridge Fundamentals of Neuroscience in Psychology*, Cambridge University Press, 2018.
- [10] Apple Inc., *Garageband*, <https://www.apple.com/mac/garageband/>, 2025. Digital audio workstation software.
- [11] BandLab Technologies, *Bandlab*, <https://www.bandlab.com/>, 2025. Online digital audio workstation and social music platform.
- [12] Reason Studios, *Figure*, <https://www.reasonstudios.com/mobile-apps>, 2025. Mobile music-making app.
- [13] P. R. Webster, *Creativity and music education: Creative thinking in music: Advancing a model*, *Creativity and Music Education* 1 (2002) 16.

# Implementation of the Science on a Sphere Visualization System as a Web Application

Jurij Anžič<sup>†</sup>, Ciril Bohak<sup>\*,†</sup>

*University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, SI-1000 Ljubljana, Slovenia*

## Abstract

SOS is a visualization system originally developed by the U.S. National Oceanic and Atmospheric Administration (NOAA) to project dynamic global datasets onto a spherical display, enabling audiences to explore complex Earth system processes in an intuitive and immersive way. While highly effective in museums, science centers, and classrooms, the traditional Science on a Sphere (SoS) hardware installation requires dedicated infrastructure, limiting its accessibility and scalability. This paper presents the design and implementation of a web-based application that reproduces the core functionality of the Science on a Sphere (SoS) system in a browser environment. The application utilizes modern web technologies to render real-time spherical visualizations of global datasets. Users can load, manipulate, and interact with datasets such as atmospheric phenomena, ocean currents, or planetary imagery without the need for specialized hardware. The system also supports interactive controls for rotating, zooming, and overlaying multiple data layers, extending the pedagogical potential of Science on a Sphere (SoS) by enabling personal exploration on laptops, tablets, and mobile devices. By transitioning Science on a Sphere (SoS) from a physical installation to a lightweight, browser-based platform, the proposed solution broadens access to scientific visualizations, promotes remote and classroom learning, and ensures that Science on a Sphere (SoS) content can be integrated into modern online education ecosystems.

## Keywords

Science on a Sphere (SOS), Web-based visualization, WebGL, WebGPU, Geospatial data, Educational technology

## 1. Introduction

Visualization plays a crucial role in communicating complex scientific phenomena to diverse audiences, ranging from experts to the general public. Large-scale datasets describing the Earth's atmosphere, oceans, and climate systems are inherently spatio-temporal and multidimensional, making them challenging to interpret without interactive and intuitive tools. To address this challenge, the National Oceanic and Atmospheric Administration (NOAA) developed the Science on a Sphere (SoS) platform, which projects dynamic datasets onto a spherical display, providing viewers with a global perspective of scientific processes. Since its introduction, SoS has been successfully deployed in museums, science centers, and classrooms worldwide, where it supports science communication, education, and public engagement.

Despite its success, the traditional SoS installation has several limitations. It requires specialized hardware, including high-resolution projectors and a physical sphere, as well as dedicated space and maintenance. These constraints hinder adoption in settings with limited resources and prevent learners from accessing SoS content outside of specialized venues. Furthermore, as educational practices increasingly move toward digital and remote learning environments, there is a growing need to make interactive scientific visualizations accessible on personal devices such as laptops, tablets, and smartphones.

Web technologies provide an opportunity to overcome these limitations. Advances in Web Graphics Library (WebGL), Web Graphics Processing Unit API (WebGPU), and modern JavaScript frameworks

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

<sup>†</sup>These authors contributed equally.

✉ ig80333@student.uni-lj.si (J. Anžič); ciril.bohak@fri.uni-lj.si (C. Bohak)

🌐 <https://lgm.fri.uni-lj.si/ciril> (C. Bohak)

🆔 0000-0002-9015-2897 (C. Bohak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.4>

have enabled real-time rendering of complex 3D graphics directly in a web browser, without the need for plugins or specialized hardware acceleration beyond consumer-grade devices. By leveraging these technologies, it is now possible to implement core SoS functionality as a browser-based application, making global scientific visualizations accessible to a much wider audience.

This paper presents the design and implementation of a web-based application that reproduces and extends the capabilities of the SoS system. The application enables users to interactively explore global datasets on a virtual sphere, including rotating, zooming, and layering multiple data sources. In contrast to the original hardware-bound installation, the web version emphasizes the user-reach, portability, and integration with digital learning platforms. Our goals are threefold: (i) to broaden the reach of SoS visualizations beyond dedicated installations, (ii) to enhance educational opportunities by enabling interactive engagement in remote and classroom settings, and (iii) to provide an open and extensible framework for future scientific visualization applications.

The remainder of this paper is structured as follows. Section 2 reviews related work on spherical visualization systems, educational applications of scientific visualization, and web-based 3D rendering technologies. Section 4 describes the methodology used in implementing the web-based SoS system. Section 5 presents the outcomes of the implementation and performance evaluation. Section 6 discusses the strengths, limitations, and implications of this approach. Finally, Section 7 concludes with a summary of findings and future directions.

## **2. Related Work**

### **2.1. Digital Earth and Virtual Globes**

The concept of a “Digital Earth” has been a driving vision for global-scale visualization systems, emphasizing the integration of geospatial data into accessible, interactive platforms [1]. Virtual globes have become a central tool in this vision, supporting both scientific analysis and public communication of geospatial phenomena [2]. Among them [3], Google Earth<sup>1</sup> has had a transformative impact over the last two decades, widely adopted in research, education, and public outreach. Its accessibility and integration of diverse data sources have demonstrated the potential of virtual globes to democratize access to geospatial information.

### **2.2. Web-based Virtual Globes and Thematic Mapping**

The evolution from desktop to web-based platforms has significantly expanded the reach of virtual globe technologies. Recent work [4] highlights how multiple-view comparisons of geospatial datasets can be effectively implemented in browser environments. Cesium<sup>2</sup>, an open-source web-based virtual globe [5], has become a popular framework for thematic mapping and visualization, supporting applications across science, urban planning, and education. National Aeronautics and Space Administration (NASA) has also contributed important web-accessible visualization platforms such as Global Imagery Browse Services (GIBS) [6] and Worldview [7], providing global imagery and near-real-time environmental data to a broad user base. In addition, Google Earth’s migration to WebAssembly<sup>3</sup> demonstrates how large-scale applications can be efficiently brought to the browser [8], while educational initiatives such as Earth Voyager [9] extend the pedagogical potential of these systems.

### **2.3. 3D Web Visualization Platforms**

The integration of geospatial visualization with web-based 3D graphics has benefited from advances in web technologies such as HTML5, WebGL, and WebGPU. Early explorations [10] demonstrated the feasibility of web-based 3D analysis and visualization of spatial data using HTML5 and WebGL. More

---

<sup>1</sup><https://earth.google.com/>

<sup>2</sup><https://cesium.com>

<sup>3</sup><https://webassembly.org>

recent studies [11] have leveraged game engines to deliver web-based 3D visualization of urban big data, reflecting a convergence of gaming and geospatial domains. A broader review [12] highlights the adaptation of technologies from the gaming industry into geospatial visualization platforms, emphasizing performance and interactivity. Emerging rendering engines such as RenderCore [13] extend this trend by harnessing WebGPU for efficient visualization of scientific datasets, demonstrating the potential of next-generation web technologies for scalable and high-performance rendering.

## 2.4. Geospatial Standards and Coordinate Systems

The effectiveness of global visualization systems depends on robust geodetic standards. The World Geodetic System 1984 (World Geodetic System 1984 (WGS84)<sup>4</sup>) provides the foundation for global positioning, georeferencing, and dataset integration across platforms. Such standards ensure interoperability between different virtual globe implementations, enabling applications to align diverse datasets consistently on a spherical model of the Earth.

Prior work has established a rich ecosystem of virtual globes, web-based visualization frameworks, and rendering engines, making global-scale data more accessible. However, existing systems often require significant computational resources, rely on proprietary platforms, or lack seamless integration into lightweight educational contexts. Our work contributes to this space by adapting the principles of SoS into a browser-based application that combines accessibility, interactivity, and extensibility for educational and scientific visualization.

## 3. Science on a Sphere

SoS is a visualization system developed by NOAA to enhance public understanding of Earth system science through large-scale, immersive displays. The system projects dynamic datasets onto a physical sphere suspended in space, creating the illusion of a rotating planet (see fig. 1). This format allows audiences to intuitively perceive global phenomena such as atmospheric circulation, ocean currents, climate variability, and planetary exploration data. Since its introduction, SoS has been deployed in museums, science centers, and educational institutions worldwide, where it has become a widely used tool for science communication and environmental education.

The strength of the original SoS lies in its ability to combine scientific accuracy with engaging visualization. Data sources include satellite imagery, real-time environmental observations, and model simulations curated by NOAA and partner organizations. Audiences interact with the globe through guided presentations, enabling educators to contextualize complex scientific processes on a global scale. However, the system's reliance on specialized hardware—high-resolution projectors, spherical displays, and dedicated installation space—has limited its availability to institutions with sufficient resources. These constraints motivated the development of alternative implementations, including the web-based version presented in this work, which seeks to broaden access to SoS content by leveraging modern web technologies.

The dataset preparation pipeline involves several steps as illustrated in fig. 2. First, the base layers are defined (usually from NASA image data), next, the educators and partners define the data and scenarios they would like to visualize on the sphere, which translates into a new SoS dataset. In the following step, these datasets are reviewed by NOAA, where the maintenance is also defined, and finally, these datasets are included in the SoS Dataset Catalog. The system is also available as a desktop application SoS Explorer<sup>5</sup>.

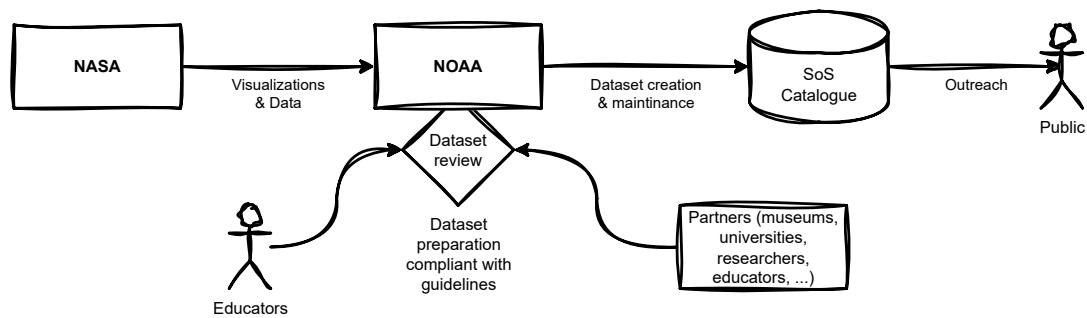
---

<sup>4</sup><https://gisgeography.com/wgs84-world-geodetic-system/>

<sup>5</sup><https://sos.noaa.gov/sos-explorer/>



**Figure 1:** The setup of the original SoS system projecting the content onto a spherical projection globe using 4 projectors.



**Figure 2:** The outline of the dataset preparation pipeline.

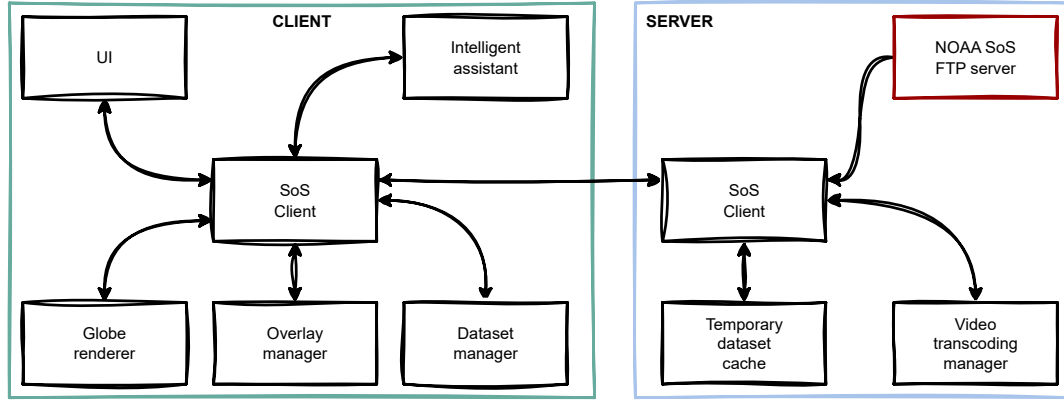
## 4. Methodology

The development of the web-based SoS system followed a modular architecture designed to replicate the core functionalities of the original installation (see fig. 3) while ensuring scalability, portability, and accessibility across devices. The methodology is organized into four main components: data acquisition and preprocessing, rendering pipeline, user interaction design, and system integration.

### 4.1. Data Acquisition and Preprocessing

The web-based system relies on publicly available geospatial datasets provided by organizations such as NOAA and NASA. Data sources include satellite imagery, atmospheric and oceanographic data, and planetary datasets accessible via services such as NASA’s GIBS [6] and Worldview [7]. Acquired datasets are stored in widely used formats (e.g., GeoTIFF, NetCDF, PNG image tiles). Where necessary, preprocessing ensures consistency in resolution, projection, and temporal alignment. Georeferencing is handled using the WGS84 coordinate system<sup>6</sup>, allowing datasets to be mapped accurately onto a spherical model.

<sup>6</sup><https://earth-info.nga.mil/?dir=wgs84&action=wgs84>



**Figure 3:** The architecture of our system is composed of client and server, outlining the functional components.

## 4.2. Visualization Pipeline

For the visualization engine, we use Babylon.js<sup>7</sup>, enabling efficient rendering of spherical projections directly in the browser without external plugins. Textures representing global datasets are projected onto a 3D sphere mesh, which serves as the virtual analog of the physical SOS globe. Tile-based rendering strategies are employed for large datasets to support progressive loading and minimize memory consumption, following approaches used in web-based virtual globes [4, 5].

## 4.3. User Interaction Design

A key design goal was to provide intuitive interaction paradigms that mirror the physical experience of rotating and exploring a spherical display and experience in SoS Explorer. Users can rotate and zoom the globe using standard mouse or touch gestures. Additional interactive features include:

- Layer management for overlaying multiple datasets (e.g., atmospheric circulation over sea surface temperature).
- Temporal controls for exploring time-series datasets, including animation playback.
- Thematic filters, which allow users to highlight or adjust specific data ranges (e.g., temperature thresholds).

Interactive elements are implemented using lightweight JavaScript frameworks, ensuring responsiveness across desktops, tablets, and mobile devices. The UI is presented in fig. 4

## 4.4. System Integration and Deployment

The application was designed as a client-centric web platform requiring no server-side rendering, reducing infrastructure demands and ensuring scalability. Datasets are accessed dynamically through APIs or preprocessed into static tiles hosted on a web server. This architecture minimizes latency and enables offline deployment in controlled environments such as classrooms or exhibitions. The system is fully compatible with modern browsers supporting WebGL and WebGPU, ensuring broad accessibility without additional installation steps.

## 4.5. Intelligent Assistant

The intelligent assistant module is implemented as a lightweight, local language model that runs directly in the browser using the Web-LLM<sup>8</sup> library. Specifically, it employs the model Llama-3.2-3B-Instruct-q4f32\_1-MLC, optimized for on-device inference. The assistant provides concise descriptions

<sup>7</sup><https://www.babylonjs.com>

<sup>8</sup><https://webllm.mlc.ai>



**Figure 4:** The UI of our system, closely mimicking the UI in SoS Explorer.

of datasets and supports geographic orientation by extracting coordinates from its responses, enabling the camera to automatically fly to the corresponding location on the globe.

#### 4.5.1. System Instructions

Before any interaction, the assistant is initialized with a role definition as a Geographic Information System (GIS) helper operating in the WGS84 geodetic reference system. Responses are typically limited to two to four sentences, without filler or meta-commentary. When possible, the assistant concludes with a line in the format `COORDS: <lat>, <lon>` (degrees). All exchanges, both user prompts and assistant replies, are stored in a conversation history, allowing for contextual continuity in dialogue.

#### 4.5.2. Coordinate Extraction

The preprocessing script first attempts to parse explicit coordinates in the format `COORDS: lat, lon`. If no such line is present, the text is normalized (diacritics and punctuation removed), and the regular expressions and a synonym dictionary are used to identify referenced countries. The system then retrieves the geographic center of the country from the local index. Additionally, the system stores suggested coordinates for selected datasets; if the model outputs “COORDS: Unknown”, the response is supplemented with dataset-specific central coordinates, which are then used for the camera jump.

#### 4.5.3. Integration with the Globe

After generating a response, the system attempts to resolve coordinates. If successful, a “Fly to ...” button is appended below the assistant’s reply. When clicked, this button moves the camera to the computed location, directly linking the textual explanation with the spatial visualization. Whenever a user selects a dataset, the assistant automatically provides a short summary along with coordinates, ensuring that the globe view aligns with the described content.

#### 4.5.4. Worker Optimization

To improve responsiveness and stability, the assistant runs in a separate worker thread. The worker initializes a single persistent model instance, while the main thread handles the UI and globe rendering. Streaming generation delivers text fragments incrementally as they are produced, significantly reducing

perceived latency. On the main thread, fragments are combined and displayed within the rendering loop, ensuring smooth integration with rendering.

#### **4.6. UI Integration**

The intelligent assistant is integrated into the system within a new panel, displaying the content generated by the local LLM. Two responses are visible on the right-hand side of the UI shown in fig. 4.

### **5. Results**

The web-based SoS system was preliminarily evaluated with respect to visualization quality, system performance across devices, and educational applicability. We report the outcomes of rendering tests, interaction responsiveness, and demonstrations in classroom settings.

#### **5.1. Visualization Outcomes**

The system successfully reproduced the core functionality of the original SoS installation. Global datasets such as atmospheric circulation, ocean surface temperature, and planetary imagery were projected onto the virtual globe with high visual fidelity. Tile-based rendering allowed smooth exploration of large datasets without noticeable delays, while layer management enabled users to combine and compare multiple datasets interactively. Figures 5a to 5c illustrate examples of rendered datasets, including Earth observation imagery and thematic overlays. Figure 5a shows the age of different parts of the sea floor together with the colormap, fig. 5b shows an animated dataset of flights within one day, and fig. 5c shows an animated dataset of carbon emissions.

#### **5.2. Visualization Evaluation**

Performance tests were conducted on a laptop with an integrated GPU—Lenovo ThinkPad X1 Gen 13 and a Razer Blade Stealth 13, both running Windows with Chrome—and a smartphone—Google Pixel 6a running Android 15 with Chrome. The application runs smoothly (above 60 FPS) on all devices, with the main shortcoming being the lower loading of larger (video) datasets. Since the UI was originally developed for a desktop screen, it is not suitable for comfortable use on a smartphone. All other functionalities work well.

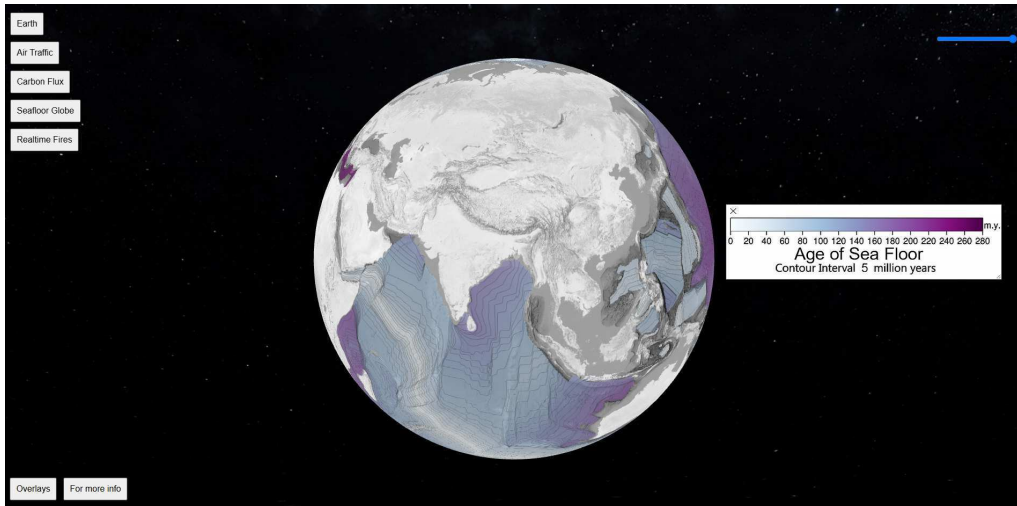
#### **5.3. User Interaction Responsiveness**

Interaction responsiveness was assessed for globe rotation, zooming, and layer management. On all platforms, globe manipulation through touch and mouse gestures was immediate, with no perceptible latency. Temporal controls for time-series datasets supported animation playback sufficient for exploring phenomena such as seasonal variation in sea ice extent. It would be beneficial to also implement some on-client loading and caching of the video datasets to avoid pauses due to loading. The overlays could be improved by exchanging them for a vector or multi-resolution representation, which would allow users to zoom in closer to the surface. The same is true for the tiled globe rendering, which only supports a single level of detail.

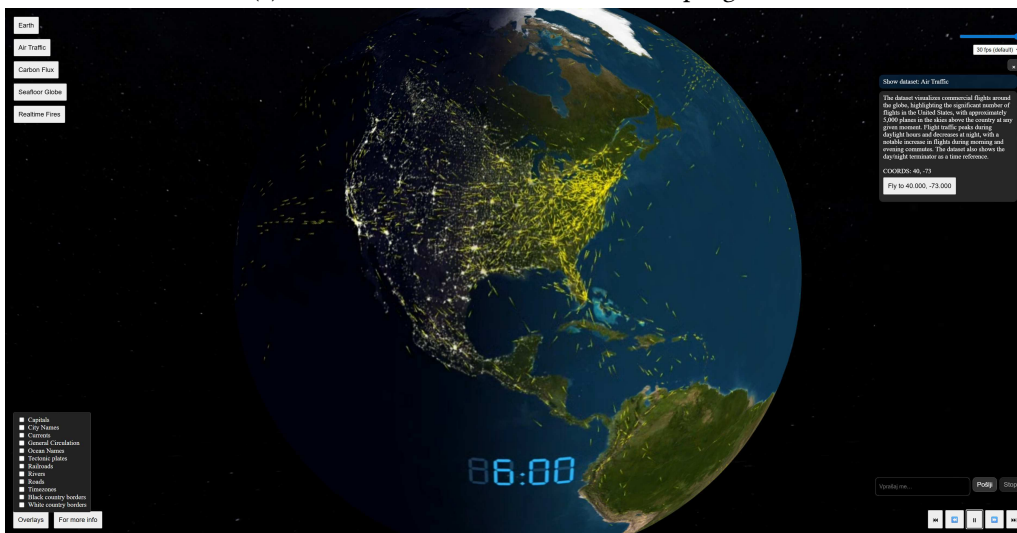
A more thorough system evaluation is planned once more features are implemented.

### **6. Discussion**

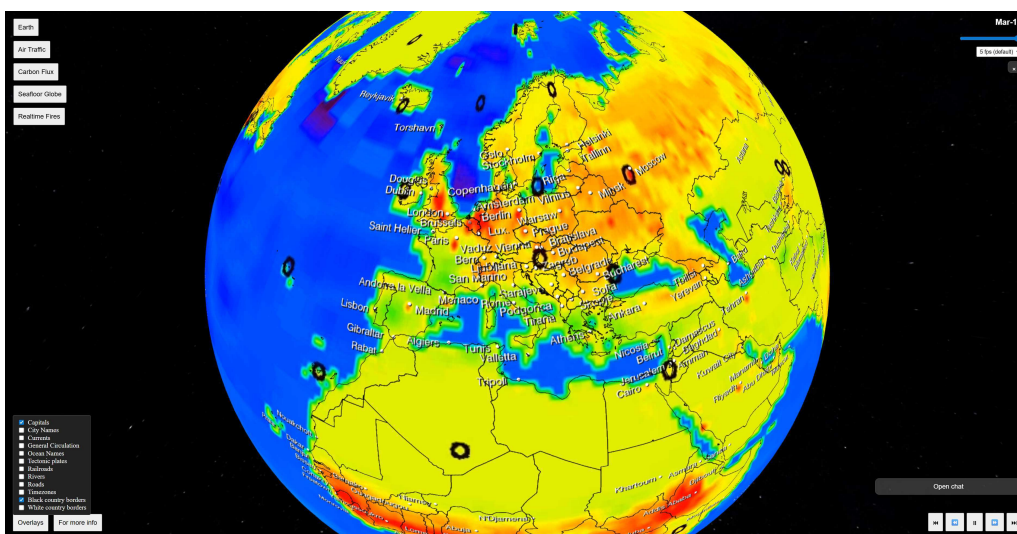
The results demonstrate that a web-based implementation of SoS is both technically feasible and pedagogically valuable. By replicating the essential visualization capabilities of the original hardware-bound system in a browser environment, the platform broadens access to global scientific datasets and enables their use in diverse educational contexts.



(a) Sea floor visualization with color map legend.



(b) Global flights dataset visualization.



(c) Carbon emissions dataset visualization.

**Figure 5:** Examples of datasets visualized in the web-based SoS application.

A primary strength of the system lies in its broad reach and portability. Unlike traditional SoS installations, which require dedicated hardware and physical space, the web-based version runs directly in modern browsers across desktops, tablets, and smartphones. This cross-platform compatibility allows students, educators, and researchers to engage with SoS datasets regardless of their location or available infrastructure. Performance benchmarks confirm that the system maintains smooth interactivity even on mobile devices, demonstrating the maturity of web technologies for real-time 3D rendering.

Another advantage is extensibility. The modular design allows the integration of new datasets and interaction modes with minimal overhead. For example, APIs such as NASA’s GIBS [6] or educational datasets from platforms like Google Earth Voyager [9] could be incorporated seamlessly. This flexibility ensures the system can evolve alongside emerging data sources and pedagogical needs.

### **6.1. Limitations**

Despite these strengths, several limitations must be acknowledged. The browser environment imposes constraints on memory management and parallel processing, limiting scalability for very large spatio-temporal datasets. Additionally, the system currently provides visualization and basic exploratory tools but lacks advanced analytical functionality (e.g., quantitative measurements, cross-sectional analysis), which may limit its use in research-focused scenarios. Finally, user evaluations were limited to informal interviews with a few users; structured usability studies with broader audiences are necessary to validate their educational impact.

### **6.2. Future Work**

Future developments will focus on several key areas. First, the adoption of next-generation rendering engines, such as RenderCore [13], which may significantly improve performance and support more complex visualizations. Second, expanding the set of interaction modes—such as annotation tools, storytelling features, or integration with virtual and augmented reality—could further enhance engagement and learning outcomes. Third, large-scale user studies across different educational settings will be conducted to systematically evaluate the system’s effectiveness in supporting comprehension of global processes. In addition, consultations and interviews with educators and domain experts are planned to better understand pedagogical needs and identify potential improvements for classroom use.

### **6.3. Implications**

The transition of SoS from a hardware-dependent installation to a lightweight web application underscores the potential of modern web technologies to broaden access to scientific visualization. By aligning with established educational initiatives in digital Earth and virtual globe research [1, 2, 3], the system contributes to the ongoing trend of leveraging digital platforms for science communication and education. While the implementation demonstrates technical feasibility and a promising framework for accessible, interactive visualization, its educational and pedagogical value warrants further investigation through more comprehensive user studies and longitudinal evaluations.

## **7. Conclusion**

This paper presented the design and implementation of a web-based version of the SoS visualization system. By leveraging modern web technologies, the platform successfully reproduces the core functionalities of the original hardware installation, including spherical projection of global datasets, interactive manipulation, and multi-layer overlays. Performance benchmarks across laptops and smartphones confirmed that the system delivers responsive and visually accurate experiences on widely available devices, demonstrating the feasibility of extending SoS beyond dedicated exhibition spaces.

The transition of SoS to a browser environment has some additional educational implications. It broadens access to global datasets, allowing learners to explore atmospheric, oceanographic, and planetary

phenomena in classrooms, at home, or in remote learning contexts. Preliminary demonstrations suggest that the system enhances engagement, fosters spatial understanding, and supports interdisciplinary learning by linking diverse datasets within an intuitive spherical model.

Looking ahead, future work will focus on expanding interaction features, integrating additional datasets, and evaluating the platform in large-scale educational settings. Advances in rendering technologies, along with the integration of storytelling and annotation tools, will further increase the system's interactivity and pedagogical value. Overall, the web-based SoS application highlights the potential of lightweight, scalable visualization systems to democratize access to scientific knowledge and strengthen digital education ecosystems.

## Declaration on Generative AI

During the preparation of this work, the author used ChatGPT (GPT-5, OpenAI) for grammar and spelling checks, and sentence rephrasing. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

## References

- [1] M. F. Goodchild, The future of digital earth, *Proceedings of the National Academy of Sciences* 109 (2012) 11088–11094. doi:10.1073/pnas.1202383109.
- [2] T. Blaschke, et al., Virtual globes: Serving science and society, *Information* 3 (2012) 372–390. doi:10.3390/info3030372.
- [3] S. Liang, et al., Applications and impacts of google earth: A decadal review (2006–2016), *ISPRS Journal of Photogrammetry and Remote Sensing* (2018). doi:10.1016/j.isprsjprs.2018.08.019.
- [4] L. Zhu, et al., Multiple-view geospatial comparison using web-based virtual globes, *ISPRS Journal of Photogrammetry and Remote Sensing* (2019). Article on web-based virtual globes.
- [5] M. Gede, Thematic mapping with cesium, in: *ICCGIS 2016*, 2016. URL: <https://cartography-gis.com/docsbca/iccgis2016/ICCGIS2016-29.pdf>.
- [6] Global imagery browse services (gibs) — api documentation, <https://nasa-gibs.github.io/gibs-api-docs/>, 2024.
- [7] Nasa worldview — tool overview, <https://www.earthdata.nasa.gov/data/tools/worldview>, 2025.
- [8] J. Mears, How we're bringing google earth to the web (webassembly case study) (2019).
- [9] Explore earth – google earth voyager, 2025. URL: <https://www.google.com/earth/education/explore-earth/>.
- [10] K. Chaturvedi, Web based 3D analysis and visualization using HTML5 and WebGL, Master's thesis, University of Twente, Faculty of Geo-information Science and Earth Observation (ITC), Enschede, The Netherlands, 2014. URL: [https://hindi.iirs.gov.in/iirs/sites/default/files/StudentThesis/MSc\\_Thesis\\_KanishkChaturvedi.pdf](https://hindi.iirs.gov.in/iirs/sites/default/files/StudentThesis/MSc_Thesis_KanishkChaturvedi.pdf).
- [11] M. Zheng, H. Zhu, Y. Huang, Web-based 3d visualization of urban big data using game engine: A case study of nanjing, china, in: *Proceedings of the 26th International Conference on Geoinformatics*, 2018.
- [12] A. Heller, M. Gede, From the gaming industry to geospatial: A review of 3d web visualization platforms, in: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 43, 2020, pp. 111–118.
- [13] Bohak, Ciril, Kovalskyi, Dmytro, Linev, Sergey, Mrak Tadel, Alja, Strban, Sebastien, Tadel, Mat-evž, Yagil, Avi, Rendercore – a new webgpu-based rendering engine for root-eve, *EPJ Web of Conf.* 295 (2024) 03035. URL: <https://doi.org/10.1051/epjconf/202429503035>. doi:10.1051/epjconf/202429503035.

# Visualization of 3D Ultrasound Uterine Data in Virtual Reality

Ilija Gavrilović<sup>†</sup>, Ciril Bohak<sup>\*,†</sup>

*University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, SI-1000 Ljubljana, Slovenia*

## Abstract

Virtual reality (VR) has emerged as a powerful medium for medical imaging visualization and training, offering immersive and interactive experiences that complement conventional two-dimensional (2D) imaging workflows. In gynecology, ultrasound (US) is the primary imaging modality for assessing uterine morphology, yet its interpretation requires significant spatial reasoning and is challenging for trainees with limited clinical exposure. This paper presents a standalone VR system for interactive visualization of uterine US volumes, designed for deployment on Meta Quest 2. The system integrates three core components: (i) textured surface meshes mapped directly from US intensity values, (ii) per-vertex deviation heatmaps comparing individual anatomy to a population-average uterus, and (iii) orthogonal slice browsing with adjustable transfer functions and lookup tables. The system is developed in Unity for immersive rendering. Preliminary demonstrations indicate that providing outside- and inside-based views improves spatial understanding and provides educational value by contextualizing individual variability. Performance profiling confirms real-time rendering on standalone hardware, ensuring fluid interaction without tethered computing resources. By unifying segmentation-driven shape analysis with immersive visualization, this work highlights the potential of lightweight VR applications to enhance gynecological training and provide accessible platforms for medical education and research.

## Keywords

virtual reality, hysteroscopy, simulation, uterus.

## 1. Introduction

Virtual Reality (VR) has matured into a practical medium for interactive visualization and simulation across various domains, including medical education and training. In gynecology, many diagnostic and therapeutic procedures are invasive, opportunities for repeated hands-on practice with real patients are limited, and the interpretation of two-dimensional (2D) imaging modalities such as Ultrasound (US) requires substantial spatial reasoning. These constraints motivate the development of realistic, repeatable, and safe training environments where learners can explore anatomy, rehearse procedures, and build visuospatial expertise without risk to patients. VR-based simulation meets these needs by enabling controlled scenarios, objective progress tracking, and immersive interaction with complex anatomical data [1, 2, 3].

Hysteroscopy—a minimally invasive endoscopic procedure for examining the uterine cavity—is a representative task that particularly benefits from immersive simulation. Successful navigation of the cervical canal and uterine cavity requires an accurate understanding of individual uterine morphology and spatial orientation, both of which are difficult to master through conventional 2D ultrasound training alone. Conventional training approaches rely heavily on observation and limited hands-on opportunities, which can result in uneven skill acquisition and reduced learner engagement. An immersive visualization environment that fuses volumetric US data with a patient-specific surface representation can help bridge the gap between 2D image slices and three-dimensional (3D) anatomical understanding, offering a cognitively intuitive way to link imaging with real procedural perspectives.

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

<sup>†</sup>These authors contributed equally.

✉ ig80333@student.uni-lj.si (I. Gavrilović); ciril.bohak@fri.uni-lj.si (C. Bohak)

🌐 <https://lgm.fri.uni-lj.si/ciril> (C. Bohak)

🆔 0000-0002-9015-2897 (C. Bohak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.5>

A key educational and clinical challenge is understanding the variability in uterine morphology across patients. Differences in uterine size, shape, and orientation can influence both diagnostic interpretation and the execution of procedures such as hysteroscopy or intrauterine device (IUD) placement. Recognizing and comparing these variations supports the development of spatial reasoning, facilitates individualized treatment planning, and enhances awareness of normal versus pathological anatomical configurations. Therefore, a visualization system that allows comparison between an individual uterus and a population-based morphological reference can serve as a valuable tool for both training and research.

Building on these motivations, this work presents a VR application for interactive visualization of uterine US volumes that integrates: (i) a surface mesh view textured with information derived directly from the US volume, (ii) per-vertex visualization of distances to a population-average uterine shape, and (iii) slice-based exploration of the volume along orthogonal planes, enabling users to cross-reference surface structures with internal volumetric information. The inclusion of slice-based exploration is particularly important for training, as it reinforces the connection between volumetric US data and the spatial geometry visualized on the uterine surface. The system is designed for standalone deployment on Meta Quest 2, emphasizing fluid interaction, spatial understanding, and real-time performance suitable for instructional and clinical educational contexts.

## 2. Related Work

Extended Reality (XR) technologies, encompassing VR, Augmented Reality (AR), and Mixed Reality (MR), have become increasingly prominent in medical education and clinical visualization. Prior work has established the potential of XR in medical visualization [1, 2, 4]. Immersive environments offer unique advantages for exploring complex anatomical structures, improving spatial understanding, and supporting procedural training. In particular, VR enables embodied interaction with volumetric datasets, allowing clinicians and trainees to intuitively navigate anatomy beyond the limitations of flat, screen-based tools. Recent research emphasizes the importance of integrating multiple data representations—specifically, linking slice-based 2D information with immersive 3D views—to reinforce spatial reasoning and diagnostic accuracy [3]. Combining orthogonal plane navigation with volumetric exploration allows learners to understand the correspondence between image slices and three-dimensional form, a critical skill in interpreting ultrasound data.

The preparation of medical data for immersive visualization remains a significant technical challenge. Automatic segmentation of anatomical structures from ultrasound volumes is particularly demanding due to speckle noise, acoustic shadowing, and low tissue contrast. Deep learning architectures such as U-Net [5] and nnU-Net [6] have become the standard for biomedical image segmentation, while hybrid and attention-based variants such as RA-UNet [7] and generalization-focused approaches [8] address domain shifts and noisy acquisitions. Survey work highlights the persistence of imperfect datasets and the importance of developing robust segmentation frameworks for clinical deployment [9]. In gynecology, segmentation and alignment techniques have recently been applied to ultrasound volumes of the uterus, enabling the construction of population-level statistical models that capture inter-patient morphological variability [10]. The open-source dataset *UterUS* provides both data and implementations for these workflows, promoting reproducibility and supporting downstream applications in visualization and training.

Several studies have explored the use of VR in gynecological education and simulation. Early efforts demonstrated the feasibility of web-based VR environments for medical visualization [11], while later systems extended toward interactive navigation of volumetric and patient-specific data [12]. In the context of hysteroscopy, immersive simulation is particularly valuable, as clinicians must reason spatially within a confined anatomical space. Recent work on automatic uterine segmentation and geometric alignment [10] directly supports this domain by providing accurate surface and volumetric models suitable for subject-specific visualization and population-level comparison.

The integration of artificial intelligence into ultrasound imaging pipelines further enhances data

preparation for immersive visualization [13]. AI-driven segmentation facilitates the generation of accurate anatomical models, while XR visualization offers an intuitive environment for validating and interpreting algorithmic results. This synergy enables not only the inspection of individual anatomy but also the exploration of population-level variability—providing a pedagogical bridge between data analysis and experiential learning.

In summary, prior work has established the educational and clinical potential of XR visualization, demonstrated robust AI-based segmentation pipelines, and explored the foundations of population-based uterine modeling. However, few systems integrate segmentation-driven shape analysis with interactive, immersive visualization on standalone VR hardware. The present work addresses this gap by combining surface- and volume-based representations with population-level deviation mapping, offering an accessible and pedagogically oriented visualization framework for gynecological training.

### 3. Methodology

Our methodology follows a modular pipeline that transforms raw US volumes into an interactive VR environment optimized for standalone deployment. The workflow comprises three main components: (i) data preparation, (ii) surface and texture processing, and (iii) immersive visualization and interaction design. Each stage builds on open-source frameworks to ensure reproducibility and extendability.

#### 3.1. Data Preparation

Raw 3D US volumes of the uterus serve as the input data. Segmentation and alignment are performed using the publicly available *UterUS* pipeline [10], which provides preprocessed meshes and population-based shape correspondence. This stage yields polygonal surface representations of individual uteri registered to a population-average model, enabling consistent visualization of morphological variability across subjects. Because the segmentation framework and dataset are described in prior work, this paper focuses on the visualization and interaction stages that build upon these outputs.

#### 3.2. Processing the Surface Mesh

Segmented uterine meshes are refined for real-time rendering. Each mesh is registered to the population-average uterus, and per-vertex deviations are encoded as scalar values that later serve as color-coded indicators of local anatomical differences. In Blender, UV unwrapping and texture baking are performed to map ultrasound-derived intensity values onto the surface. Surface normals are recalculated to ensure coherent lighting for both external and internal perspectives. The processed assets—including meshes, deviation fields, and textures—are exported in optimized formats for Unity to ensure efficient rendering on standalone VR hardware.

#### 3.3. Volumetric Data Integration

To maintain clinical relevance and aid spatial reasoning, orthogonal ultrasound slice views (axial, sagittal, and coronal) are integrated alongside the surface model. Slice stacks are aligned with the registered meshes and stored as 2D textures for interactive browsing within the VR environment. Users can scroll through slices using controller input while simultaneously inspecting the corresponding anatomical regions on the surface model. A transfer-function module enables dynamic intensity and transparency adjustments; detailed visualization parameter settings are discussed in the Results section.

#### 3.4. Immersive Visualization in Unity

The immersive environment was implemented in Unity, leveraging its built-in XR toolkit and native support for the Meta Quest 2 headset. Two complementary visualization modes are provided: (i) an **external view** for inspecting overall uterine morphology and population deviation mapping, and (ii)

an **internal (hysteroscopic) view**, where the virtual camera is positioned inside the uterine cavity to simulate endoscopic navigation.

Interaction within the VR environment combines ray-based User Interface (UI) controls with direct 3D manipulation. Users can grab, rotate, and scale the uterine model using handheld controllers, switch seamlessly between external and internal viewpoints, and move through the cavity using joystick-based locomotion or teleportation. Slice planes can be activated, repositioned, or scrolled interactively to correlate volumetric US slices with the surface model, supporting the development of spatial reasoning between 2D and 3D representations. Menu-based controls allow users to adjust visualization options such as color mapping, transparency, and population deviation overlays.

To maintain real-time performance on standalone hardware, mesh decimation and texture compression were applied to balance geometric fidelity and memory footprint. The final build sustains 72 FPS on Meta Quest 2, providing a stable and low-latency experience suitable for educational use.

A preliminary user evaluation was conducted with eight participants: four gynecology residents, two medical imaging researchers, and two computer science students with prior VR experience. Participants were given a short tutorial on navigation and were then asked to complete three tasks: (1) identify major anatomical landmarks in the external view, (2) navigate from the cervical canal to the fundus in the internal view, and (3) locate corresponding structures in the slice-based visualization. After completing the tasks, participants answered four open-ended questions focusing on usability, perceived educational value, clarity of anatomical representation, and comfort during use. No small-scale questionnaires (e.g., Likert items) were used at this stage to encourage unconstrained qualitative feedback. To minimize experimenter bias, all instructions were standardized, and participants were free to explore the application for up to 10 minutes per session.

Feedback varied according to background. Medical participants emphasized the value of integrating 2D slice information with immersive 3D visualization for understanding uterine orientation and depth. In contrast, computer science participants commented on rendering quality, frame rate, and interface responsiveness. Across all groups, users appreciated the ability to freely switch between internal and external perspectives, noting that it improved comprehension of the spatial relationships between ultrasound slices and surface geometry.

These insights will inform the design of a structured usability study in future work, incorporating quantitative metrics and standardized questionnaires to evaluate learning outcomes and interaction efficiency.

## 4. Results

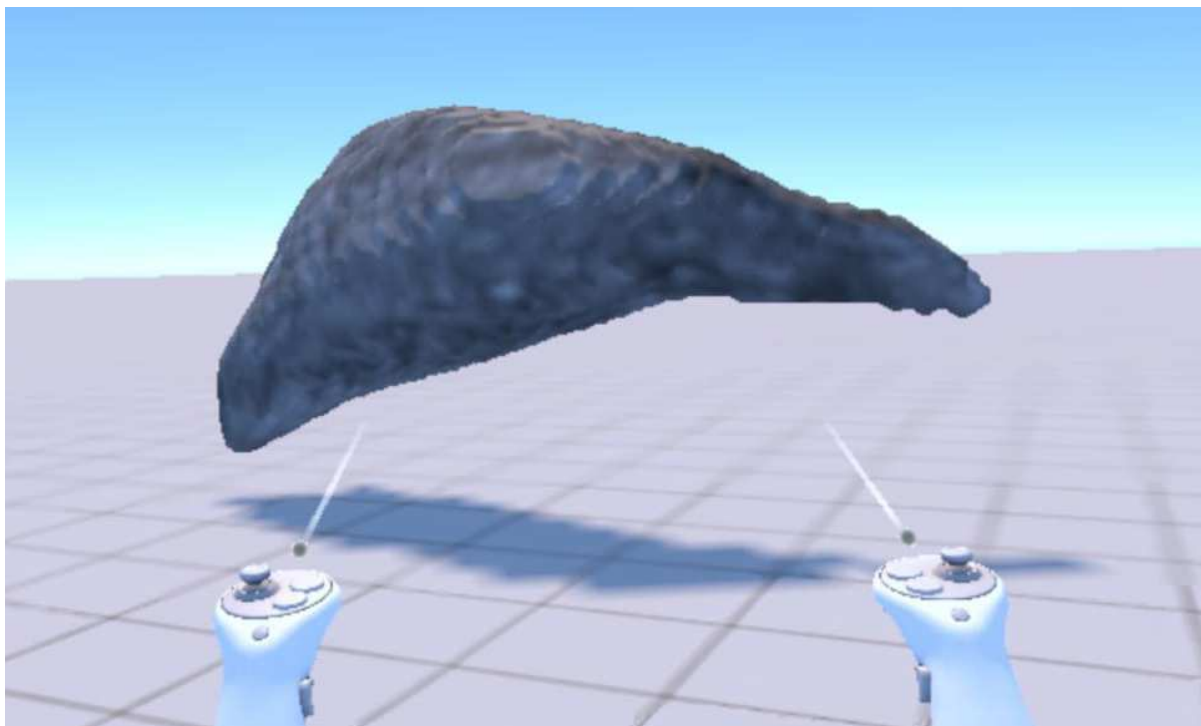
This section presents the qualitative and quantitative evaluation of the proposed VR system for interactive visualization of uterine US volumes. We focus on (i) visual fidelity and correctness of the rendered anatomy, (ii) performance of the standalone application on Meta Quest 2, and (iii) educational affordances observed during preliminary demonstrations.

### 4.1. Qualitative Visualization Outcomes

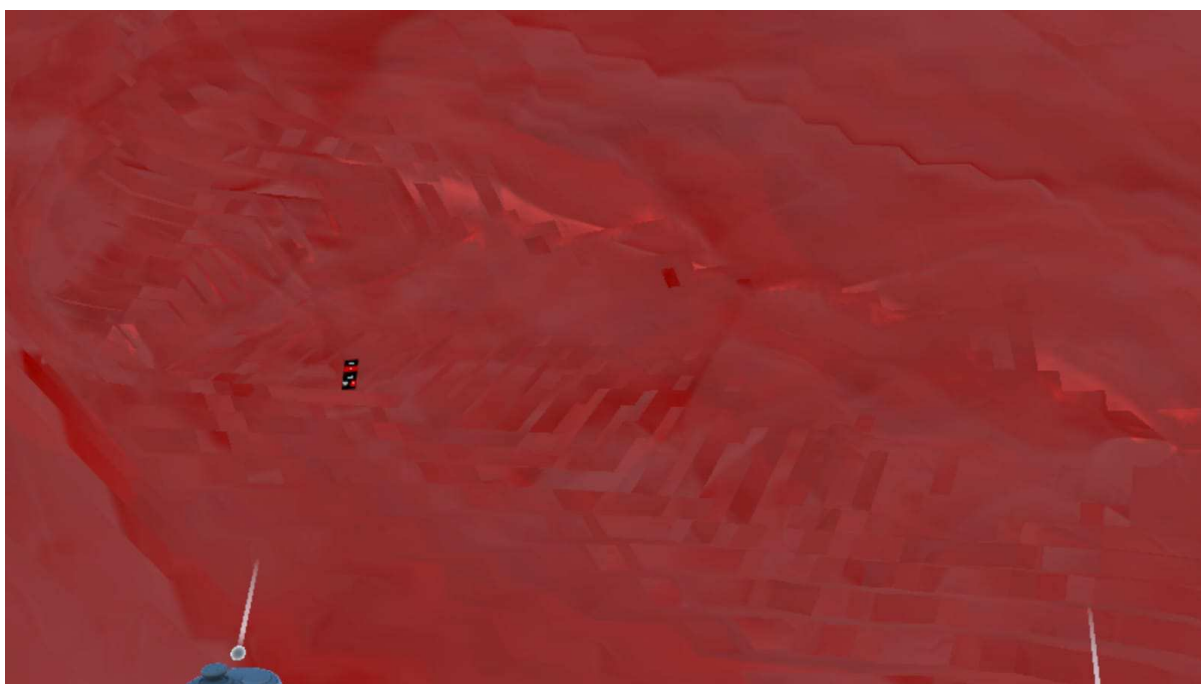
Figure 1 illustrates the external mesh visualization with textures derived from US intensity values. Surface deviations from the population-average uterus are encoded as a color-coded heatmap, enabling immediate identification of morphological differences. In the external view, clinicians can readily inspect the global shape and surface variability, while deviations larger than 3 mm are clearly visible as localized hot spots.

The internal mode (Fig. 2) provides a first-person perspective within the uterine cavity, mimicking hysteroscopic navigation. This view is particularly effective in demonstrating the geometry of the cavity and cervical canal. Dynamic lighting and correct surface normals ensure realistic shading and depth perception, supporting spatial understanding in an immersive context.

Orthogonal slice browsing is shown in Fig. 3. Users can scroll through axial, sagittal, and coronal planes in real time, directly linking volumetric intensities with anatomical landmarks on the surface

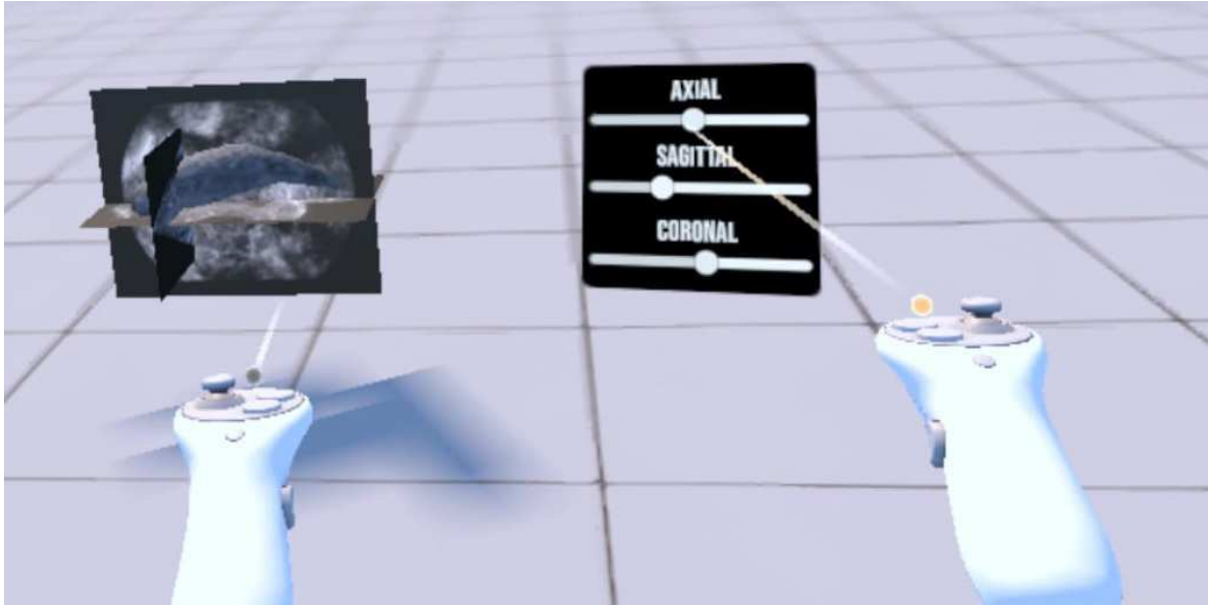


**Figure 1:** External view of the uterus surface mesh: texture-mapped ultrasound intensities.



**Figure 2:** Internal view of the uterus cavity. Clinicians can explore the cervical canal and endometrial cavity in a manner analogous to hysteroscopic navigation.

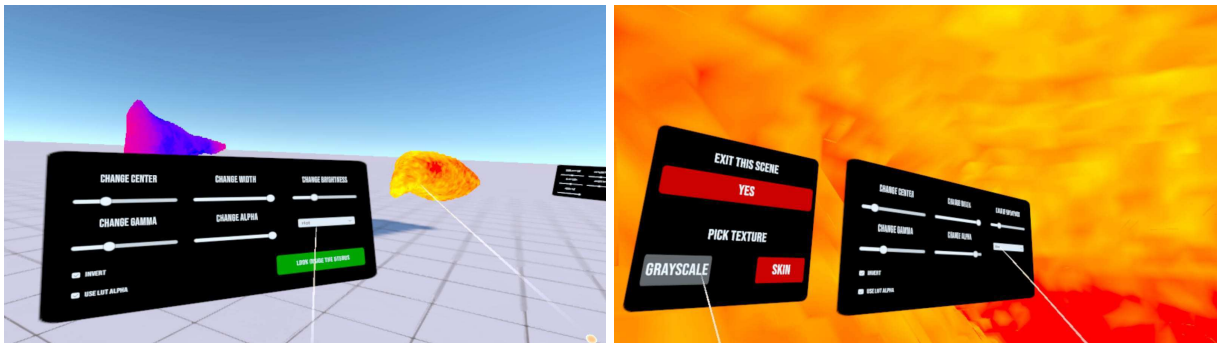
mesh. Windowing, brightness, and gamma adjustments allow clinicians to adapt the visualization to individual preferences, reflecting common workflows in conventional ultrasound analysis.



**Figure 3:** Orthogonal slice views (axial, sagittal, coronal) in VR.

## 4.2. Transfer Function Adjustment

An additional feature of the system is the real-time adjustment of the transfer function in the outside and inside view visualizations. Users can interactively modify window center, window width, brightness, and gamma values, which directly affect the contrast and visibility of structures within the US data. Custom look-up tables with alpha channels enable semi-transparent renderings, allowing subtle anatomical boundaries to be emphasized without occluding surrounding structures. The adjustments are illustrated in Fig. 4 for both inside and outside views.



**Figure 4:** Adjustment of the transfer function for outside (left) and inside (right) views.

## 4.3. Performance on Meta Quest 2

Performance profiling was conducted using Unity's built-in frame diagnostics on Meta Quest 2. The application maintained an average of 71.8 FPS across all tested scenarios, with occasional dips below 70 FPS only when switching between external and internal viewpoints with high-resolution textures enabled. Latency remained below 20 ms, which is acceptable for extended VR sessions without inducing discomfort. Texture resolution (up to 2048<sup>2</sup>) and mesh density (~40k vertices after decimation) proved sufficient to preserve anatomical detail while sustaining real-time performance. The application requested under 2 GB of RAM and ~1.5 GB of VRAM. The usage stabilized throughout the use and did not increase with longer use.

#### **4.4. Educational Affordances**

Preliminary demonstrations with graduate students in biomedical engineering and computer science indicated that the combined slice and mesh visualization improved comprehension of uterine shape and variability. Participants reported that internal viewpoint navigation enhanced their spatial intuition about the cavity and cervical canal, which is difficult to achieve with monoscopic 2D images. The deviation heatmap was highlighted as a valuable feature for understanding anatomical variability across a population, suggesting potential for use in both medical training and research contexts.

### **5. Discussion**

The presented system demonstrates how immersive visualization of uterine US data can be achieved on standalone VR hardware through a modular pipeline integrating segmentation, surface processing, and interactive rendering. This section discusses the educational implications, technical contributions, AI-assisted visualization components, and current limitations, followed by future research directions.

#### **5.1. Educational Impact and Design Rationale**

The primary motivation for this system is to address challenges in gynecological education, particularly the need for safe, repeatable, and spatially coherent training tools. By combining textured surface meshes, per-vertex deviation heatmaps, and interactive slice exploration, the application bridges conventional 2D ultrasound interpretation with immersive 3D anatomical understanding. This multimodal visualization enables trainees to directly correlate features visible in 2D slices with their spatial location on the uterus, reinforcing visuospatial reasoning—an essential skill for hysteroscopic procedures and diagnostic imaging. Early user feedback suggests that this dual representation improves engagement and helps learners contextualize anatomical variability in relation to population norms.

#### **5.2. Technical and Hardware Contributions**

A key strength of the approach is the use of standalone VR hardware, specifically Meta Quest 2, to deliver high-fidelity visualization without reliance on external computing resources. This portability lowers technical and logistical barriers to adoption in educational or clinical settings. Consistent real-time performance above 70 FPS ensures smooth, low-latency interaction, reducing motion sickness and maintaining user comfort during extended training sessions. Performance optimization through mesh decimation and texture compression confirms that complex medical visualization tasks can be efficiently executed on accessible consumer-grade devices.

#### **5.3. AI-Integrated Visualization Advantages**

The integration of segmentation-driven shape modeling provides an additional layer of pedagogical value. By leveraging population-level shape correspondence derived from automated segmentation pipelines [10], the system situates each patient-specific uterus within a broader statistical model. This allows learners to visually compare individual morphology with typical anatomical variations, fostering a deeper understanding of normal versus atypical cases. Unlike most XR-based visualization systems that focus solely on volumetric rendering [1, 2, 3], the presented framework explicitly links per-patient data to explainable, population-level information—aligning with current trends in interpretable medical AI [13].

#### **5.4. Limitations**

Despite promising results, several limitations remain. First, the accuracy of the visualization pipeline is constrained by segmentation quality. Although architectures such as U-Net [5] and nnU-Net [6] provide robust baselines, errors in automated segmentation may propagate to the resulting meshes,

potentially misrepresenting fine anatomical details. Second, the current dataset is limited in size and diversity, which restricts the generalizability of findings. Third, user evaluation has been limited to preliminary qualitative feedback rather than structured usability or clinical validation studies. Finally, while the system achieves real-time performance on standalone hardware, future iterations with higher-resolution models or dynamic segmentation may challenge device capacity. Future work will address these limitations through expanded datasets, improved preprocessing, and larger-scale evaluations.

## **5.5. Future Work**

Planned extensions will focus on both educational validation and technical enhancement. First, structured clinical studies involving gynecologists and medical students will assess usability, engagement, and learning outcomes using standardized tools such as the System Usability Scale (SUS) and NASA-TLX. Second, integrating procedural simulation modules—such as hysteroscopic navigation and biopsy training—will expand the application from visualization to active skill acquisition. Third, incorporating multi-modality data (e.g., MRI or CT) alongside US could enrich spatial context and improve cross-modality comprehension. Finally, advances in lightweight neural rendering and real-time segmentation could enable adaptive model updates directly on standalone devices, streamlining data preparation workflows.

## **5.6. Implications**

This work underscores the feasibility of deploying advanced medical visualization systems on accessible standalone VR hardware. By uniting volumetric, surface-based, and population-level data representations, the system offers a scalable framework for both medical education and research into anatomical variability. More broadly, it demonstrates how open-source, reproducible pipelines—spanning 3D Slicer, Blender, and Unity—can be leveraged to translate complex medical data into pedagogically meaningful immersive experiences, ultimately supporting broader adoption of XR tools in healthcare edu

## **6. Conclusion**

This work introduced a standalone VR system for interactive visualization of uterine US volumes, integrating surface meshes, deviation heatmaps, and orthogonal slice exploration within a unified environment. By leveraging an open-source pipeline that combines segmentation, mesh processing, and immersive rendering, we demonstrated that clinically relevant visualizations can be delivered in real time on accessible hardware such as Meta Quest 2. The system enables learners and clinicians to bridge conventional 2D imaging with immersive 3D exploration, fostering improved spatial understanding of uterine morphology and anatomical variability.

Preliminary demonstrations highlight the educational potential of combining volumetric intensity data with population-based shape analysis, particularly for training scenarios where exposure to real patients is limited. At the same time, limitations remain in terms of dataset size, segmentation accuracy, and the absence of structured clinical evaluation. Future work will focus on extending the framework with procedure-specific simulation modules, larger-scale user studies, and integration of multimodal imaging data.

Overall, the results underline the feasibility and promise of deploying immersive medical visualization on lightweight, standalone hardware. Such systems have the potential to democratize access to advanced training tools, reduce reliance on tethered setups, and open new avenues for research and education in gynecology and beyond.

## References

- [1] F. Pires, C. Costa, P. Dias, On the use of virtual reality for medical imaging visualization, *Journal of Imaging* 4 (2021) 1034–1048. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8455774>. doi:10.1007/s10278-021-00480-z.
- [2] J. Yuan, S. S. Hassan, J. Wu, C. R. Koger, R. R. S. Packard, F. Shi, B. Fei, Y. Ding, Extended reality for biomedicine, *Nature Reviews Methods Primers* 3 (2023) 2662–8449. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10088349>. doi:10.1038/s43586-023-00208-z.
- [3] Q. Liu, S. Qiu, Y. Wang, et al., Coordinated 2d-3d visualization of volumetric medical data in xr with multimodal interactions, 2025. URL: <https://arxiv.org/html/2506.22926v1>. arXiv:2506.22926.
- [4] M. Lang, S. Ghandour, B. Rikard, E. K. Balasalle, M. R. Rouhezamin, H. Zhang, R. N. Uppot, Medical extended reality for radiology education and training, *Journal of the American College of Radiology* 21 (2024) 1583–1594. doi:<https://doi.org/10.1016/j.jacr.2024.05.006>.
- [5] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. URL: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28). doi:10.1007/978-3-319-24574-4\_28.
- [6] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, K. H. Maier-Hein, nnu-net: a self-configuring method for deep learning-based biomedical image segmentation, *Nat Methods* (2021). URL: <https://doi.org/10.1038/s41592-020-01008-z>. doi:10.1038/s41592-020-01008-z.
- [7] Q. Jin, Z. Meng, C. Sun, H. Cui, R. Su, Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans, *Frontiers in Bioengineering and Biotechnology Volume 8 - 2020* (2020) 156–165. URL: <https://doi.org/10.3389/fbioe.2020.605132>. doi:10.3389/fbioe.2020.605132.
- [8] X. Yang, H. Dou, R. Li, X. Wang, C. Bian, S. Li, D. Ni, P.-A. Heng, Generalizing deep models for ultrasound image segmentation, in: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing, 2018, pp. 497–505. doi:10.1007/978-3-030-00937-3\_57.
- [9] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, X. Ding, Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, *Medical Image Analysis* 63 (2020) 101693. doi:10.1016/j.media.2020.101693.
- [10] E. Boneš, C. Bohak, M. Gergolet, Žiga Lesar, M. Marolt, Automatic segmentation and alignment of uterine shapes from 3d ultrasound data, *Computers in Biology and Medicine* 178 (2024) 108794. URL: <https://doi.org/10.1016/j.compbimed.2024.108794>. doi:10.1016/j.compbimed.2024.108794.
- [11] Žiga Kokelj, C. Bohak, M. Marolt, A web-based virtual reality environment for medical visualization, in: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, 2018, pp. 0299–0302. URL: <https://ieeexplore.ieee.org/document/8400057>. doi:<https://doi.org/10.23919/MIPRO.2018.8400057>.
- [12] Žiga Kokelj, Implementacija navidezne resničnosti v spletno vizualizacijsko ogrodje Med3D, Master’s thesis, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2018. URL: <https://core.ac.uk/download/pdf/162020494.pdf>.
- [13] F. Moro, M. T. Giudice, M. Ciancia, D. Zace, G. Baldassari, M. Vagni, H. E. Tran, G. Scambia, A. C. Testa, Application of artificial intelligence to ultrasound imaging for benign gynecological disorders: systematic review, *Ultrasound in Obstetrics & Gynecology* (2025) 295–302. URL: <https://doi.org/10.1002/uog.29171>. doi:10.1002/uog.29171.

# Human-Computer Interaction in Slovenia: A Retrospective and Trend Analysis of Local Research

Ciril Bohak\*

*Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia*

## Abstract

This paper presents a retrospective and trend analysis of Human-Computer Interaction (HCI) research in Slovenia, with a focus on publications from national Human-Computer Interaction (HCI) conference venues between 2014 and 2024. Drawing on a dataset of 84 papers authored by 138 distinct contributors, we examine the evolution of research topics, author participation, and institutional involvement over time. The results show a relatively stable number of accepted papers per year, accompanied by fluctuations in author diversity, with 2022 marking a peak in community engagement. The analysis highlights both the persistence of core Human-Computer Interaction (HCI) themes—such as usability, interaction design, and visualization—and the gradual inclusion of emerging areas, including immersive technologies and data-driven design. By situating local contributions within broader international developments, the study provides an overview of the Slovenian Human-Computer Interaction (HCI) landscape, identifies patterns of collaboration and dissemination, and reflects on challenges and opportunities for strengthening the community in the future.

## Keywords

Human-Computer Interaction, Slovenia, research trends, retrospective analysis, publication analysis

## 1. Introduction

Human-Computer Interaction (HCI) has established itself as a dynamic and interdisciplinary field that bridges computer science, design, psychology, and the social sciences. Over the past decades, HCI has grown from focusing on usability and ergonomics towards encompassing advanced technologies such as Extended Reality (XR), Artificial Intelligence (AI), and multimodal interaction. Alongside global developments, local research communities have emerged to contribute to and contextualize HCI within their cultural, educational, and industrial environments.

In Slovenia, the Human-Computer Interaction Slovenia Conference (HCI-SI) (in the first few iterations promoted under the name HCI in Information Society – HCI IS) has served as the primary venue for presenting and discussing local HCI research for nearly a decade. Since its inception, the conference has provided a platform for researchers, practitioners, and students to exchange ideas, showcase prototypes, and strengthen the national and regional HCI community. The conference proceedings, published annually, represent a valuable record of how research

---

*HCI SI 2025: Human-Computer Interaction Slovenia 2025, , 2025, Koper, Slovenia*

\*Corresponding author.

✉ [ciril.bohak@fri.uni-lj.si](mailto:ciril.bohak@fri.uni-lj.si) (C. Bohak)

🌐 <https://lgm.fri.uni-lj.si/ciril> (C. Bohak)

🆔 0000-0002-9015-2897 (C. Bohak)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.6>

interests, methods, and collaborations in Slovenia have evolved over time.

This paper provides a retrospective and trend analysis of ten years of HCI-SI. The goal is twofold: first, to document and reflect on the development of HCI research in Slovenia as captured by the conference proceedings; and second, to identify emerging themes, methods, and collaborations that illustrate the trajectory of the local community. In doing so, we aim to contribute to a deeper understanding of how a small but vibrant HCI community evolves, while also situating Slovenian HCI research within the wider international context.

The contributions of this paper are threefold:

- A descriptive overview of the HCI-SI across ten years, including its growth, participation, and structural changes.
- A thematic and bibliometric trend analysis of research topics, methods, and collaborations represented in the proceedings.
- A reflection on the role of HCI-SI in shaping the Slovenian HCI community and opportunities for its future development.

## 2. Background

The field of HCI has a long tradition of organizing conferences, workshops, and symposia that bring together researchers and practitioners from diverse domains. Flagship international venues such as the ACM CHI Conference on Human Factors in Computing Systems, ISMAR, PERCOM, UIST, VR, NordiCHI, and CHIItaly have played a central role in establishing and advancing the discipline. These venues not only showcase state-of-the-art research but also foster the creation of research communities that contribute to shaping the direction of HCI worldwide.

In Slovenia, the HCI-SI was established with the goal of providing a dedicated forum for local researchers, educators, and practitioners to present their work, network, and strengthen the national HCI community. Over the years, HCI-SI has evolved into a recognized venue where research prototypes, case studies, and methodological innovations are shared. The conference has also served as an entry point for students and early-career researchers to engage with the broader research community, often acting as a stepping stone toward publishing at larger international venues.

Meta-analyses and retrospective studies of HCI research communities (such as [1]) have been conducted in different contexts. For example, bibliometric analyses of CHI proceedings have provided insights into the evolution of research topics, methodological trends, and patterns of collaboration within the international community. Similar studies have been undertaken for regional conferences, shedding light on how local contexts influence research agendas and community development. These works highlight the value of analyzing conference proceedings as a means of understanding how a research community evolves over time.

This paper contributes to this line of inquiry by providing the first systematic retrospective analysis of the HCI-SI. By examining ten years of proceedings, we document the historical development of the conference, analyze research trends, and position Slovenian HCI within the global research landscape.

### 3. Methodology

This section details our data sources, inclusion criteria, preprocessing pipeline, and the analytical methods used to conduct the retrospective and trend analysis of a decade of the HCI-SI community and related Slovenian HCI venues.

#### 3.1. Data Sources and Coverage

We compiled a corpus from publicly available conference proceedings spanning 2014–2024, covering the *Information Society – HCI* tracks (2014, 2016, 2019, 2020) and the *HCI-SI* CEUR proceedings (2021–2024). The dataset comprises complete bibliographic records—including titles, authors, page ranges, venue metadata, and URLs where available—compiled from the provided references. These span early HCI-SI contributions [2, 3, 4, 5, 6, 7, 8, 9], health- and visualization-focused studies [10, 11, 12, 13, 14], applied XR and interaction research [15, 16, 17, 18], as well as work addressing mobile, privacy, and persuasive technologies [19, 20, 21, 22, 23, 24].

From 2020 onwards, Slovenian HCI research diversified across several thematic areas. In **data visualization and novel interfaces**, work explored advertising personalization, budget visualization, voice interaction, speech synthesis, tangible User Interface (UI)s, anamorphic projections, and digital art interaction [25, 26, 27, 28, 29, 30, 31, 32, 33, 34]. **Games, XR, and immersion** were investigated through speculative music, cultural heritage applications, projected-surface games, tangible programming with Augmented Reality (AR), eye-tracking depth-of-field, reinforcement learning, immersive soundscapes, contextual Non-Playable Character (NPC)s, and educational AR systems [35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45]. **Behavioral and psychological studies** addressed personality-tailored notifications, gullibility prediction, hedonic and eudaimonic experiences, physiological responses, and music-related belief systems [46, 47, 48, 49, 50, 51, 52]. **Automotive and safety research** examined driver disengagement, cognitive load, thermographic monitoring, situational awareness, and parking assistance [53, 54, 55, 56, 57, 58, 59]. **Education and learning technologies** included programming support, algorithm teaching with robots, cognitive recall, and AR-based training [60, 61, 62]. Finally, **social and cross-domain studies** focused on authentication, autoencoder-based behaviour inference, lifestyle and happiness, fake news and tribalism, and sports analytics [63, 64, 65, 66, 67, 68, 69].

In addition to the core dataset, we explicitly acknowledge a set of contributions not included in the groupings above. These include early works on mobile and web application design [9], medical and health-oriented interaction systems [10, 11, 12], safety and environmental visualization [13, 14, 70], as well as applied XR and embodied interaction prototypes [17, 15, 71, 72, 16, 18]. While not all of these papers shaped later research threads directly, they illustrate the breadth of experimentation in the formative years of the HCI-SI community and demonstrate early attention to health, safety, mobile interaction, and tangible/immersive technologies. Their inclusion ensures that the retrospective accounts for the diversity of exploratory projects that contributed to the foundation on which subsequent thematic trends were built.

Beyond these, additional contributions further highlight the diversity of Slovenian HCI research across the years. These include interactive learning pathways and gamification in education [73, 74, 75], blockchain-enabled in-game economies [76], advances in audio-visual immersion [77], personalised and artistic XR interactions [32, 78], user-friendly digital services

[79], safety and context-aware mobile notifications [70], retention-enhancing interactive displays [80], and gesture-based recognition prototypes leveraging radar and motion tracking [81]. While thematically diverse, these contributions underscore the community’s sustained interest in blending technical innovation with user-centered interaction design.

**Time Window:** We analyze eight conference years distributed across the period 2014–2024, reflecting the evolution from the *Information Society* HCI track toward a dedicated HCI-SI venue.

**Document Types:** We include regular and short papers as listed in the proceedings (no posters without papers, prefaces, or front matter).

### 3.2. Data Extraction and Normalization

We extract the following fields: year, title, authors, pages, venue/series, city, month, and url. To ensure analytical consistency:

1. **Encoding:** We transcoded all the text to Unicode.
2. **Author Disambiguation:** We canonicalize author strings via (Last, First) normalization, trim initials spacing, and merge obvious variants (e.g., hyphenation, middle initials) by a deterministic key (casefolded, accent-stripped comparison), while retaining the original form for reporting.
3. **Multilingual Titles:** For titles with translations or bilingual forms, we keep the original title as primary and store translations (if present in the record, otherwise we translated the titles using the Google Translate service) as auxiliary fields.
4. **Pagination:** We parse page ranges (e.g., 5–8) as integer spans; missing ranges are flagged and excluded from page-count statistics but retained for all other analyses.

### 3.3. Coding and Thematic Annotation

We combine a light-weight theme coding with an automated Large Language Model (LLM) pipeline using ChatGPT to process the dataset:

- **Seed Codebook:** We begin with an HCI theme codebook (e.g., *UX/usability*, *XR/VR/AR*, *tangible/embodied*, *health*, *education*, *industry/IoT*, *privacy/ethics*, *recommenders/behavior*, *visualization*, *AI/ML in HCI*).
- **Keyword Extraction:** We apply lemmatization and noun-phrase mining on titles to suggest candidate labels mapped to the codebook.

### 3.4. Positioning and Comparative Context

Where appropriate, we qualitatively contrast local trends with broader HCI currents (e.g., rise of XR, persuasive technologies, and AI/Machine Learning (ML)-enabled interaction), using representative local examples across the decade [5, 7, 15, 19, 25, 37, 60, 63, 58, 41] to anchor observations.

### 3.5. Data Availability

The data is provided in the form of a curated BibTeX source<sup>1</sup>.

## 4. Analysis

Throughout the years, a total of **84 papers** were published by **138 distinct (co)authors**. The conference was organized **six times in Ljubljana** and once in both **Maribor** and **Koper**, reflecting its stable presence in the Slovenian research environment with occasional regional diversification.

Table 1 summarizes the yearly distribution of publications and contributing authors. The data indicate that the years **2020** and **2022** were the most productive in terms of accepted papers, each featuring 13 contributions. In addition, **2022** stands out as the year with the highest number of unique (co)authors (42), suggesting a broader engagement of the research community compared to other years.

Figure 1 provides a visual comparison between the number of papers and the number of authors per year. The number of papers varies to some extent throughout the period, ranging from six to thirteen, whereas the number of authors fluctuates even more. The peak in 2022 reveals a particular expansion in the conference’s reach, both in terms of collaboration and community participation. Overall, the results demonstrate a consistent level of scholarly output, accompanied by periodic increases in the diversity of contributors.

The conference skipped its annual repetition on two occasions. In 2015 and in 2017–2018. Since then, it has been consistently present in the Slovenian research community.

|              | 2014 | 2016 | 2019 | 2020      | 2021 | 2022      | 2023 | 2024 | Total       |
|--------------|------|------|------|-----------|------|-----------|------|------|-------------|
| # of Papers  | 8    | 11   | 12   | <b>13</b> | 10   | <b>13</b> | 11   | 6    | <b>84</b>   |
| # of Authors | 17   | 28   | 22   | 32        | 23   | <b>42</b> | 24   | 14   | <b>138*</b> |

**Table 1**

Publications and unique authors per year (\* some authors published on several conferences).

## 5. Descriptive Overview of the Conferences

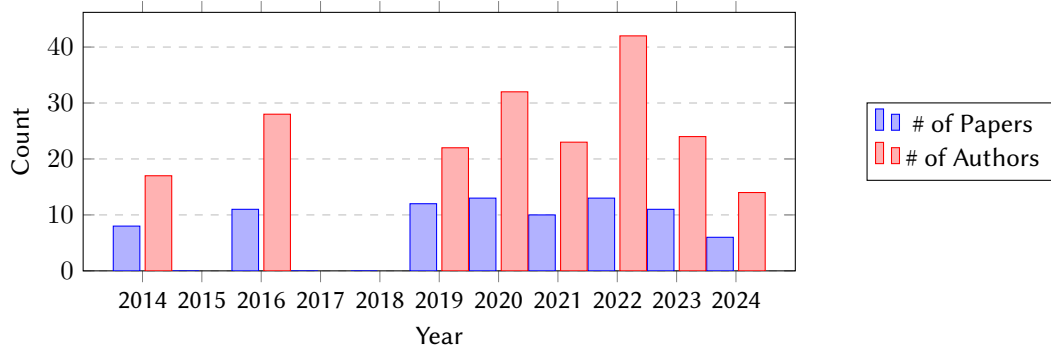
This section provides a descriptive account of the ten years of Slovenian HCI conferences analyzed in this study, focusing on participation, publication volume, and structural aspects of the venues.

### 5.1. Conference Years and Venues

The corpus spans two distinct phases:

---

<sup>1</sup><https://github.com/CirilBohak/HCI-SI-2025---Retrospective>



**Figure 1:** Number of accepted papers and distinct authors throughout the conference history.

1. **Information Society – HCI Tracks (2014–2020):** These early contributions were published as part of the annual *Information Society Multiconference*, within dedicated HCI tracks (2014, 2016, 2019, 2020). The structure was embedded in a larger multidisciplinary event, providing visibility but limiting the scale of HCI-specific sessions due to registration fee and limitations on the size.
2. **HCI-SI Conference (2021–2024):** Starting in 2021, HCI-SI established itself as a standalone CEUR-WS proceedings series, with its own identity, program committee, and thematic breadth. The venue has since been hosted in Koper, Maribor, and Ljubljana, strengthening its national recognition and enabling cross-institutional collaborations.

## 5.2. Authorship and Collaboration

The average number of authors per paper increased steadily over the decade. Early contributions often involved two to three authors, while recent years commonly feature four or more, reflecting both larger research teams and an increasingly collaborative culture. Co-authorship networks reveal a core of recurring contributors, complemented by a continuous influx of new authors, indicating healthy community renewal.

## 5.3. Topical Breadth

The thematic scope has likewise expanded. Early years emphasized usability, UX, and visualization (e.g., [2, 5]), while later conferences highlighted XR, persuasive and affective technologies, and data-driven methods (e.g., [25, 38, 41]). This reflects a gradual shift from applied prototypes to more methodologically rigorous studies with broader connections to global HCI trends.

## 6. Discussion

Building on the descriptive overview, this section examines how research themes, methods, and collaborations in Slovenian HCI evolved across the ten conference years.

## 6.1. Thematic Evolution

The coding scheme described in Section 3 reveals a clear evolution of topics over time:

- **Early years (2014–2016):** Dominated by usability, UX design, and visualization of complex data (e.g., [5, 7, 10]).
- **Mid-phase (2019–2020):** Increasing emphasis on mobile interaction, persuasive technologies, and XR applications in education and cultural heritage (e.g., [19, 36, 37]).
- **Recent years (2021–2024):** Broader diversification into affective computing, behavior change interventions, physiological sensing, and AI/ML-driven interaction (e.g., [60, 63, 66, 41]).

## 6.2. Methodological Shifts

Analysis of methodological tags reveals a transition from exploratory prototyping and case studies in the early years toward more systematic user studies and quantitative evaluations in later conferences. For example:

- *2014–2016:* Frequent use of proof-of-concept prototypes (e.g., [4, 15]).
- *2019–2020:* User-centered evaluations of mobile, persuasive, and mixed-reality systems (e.g., [82, 83]).
- *2021–2024:* Experimental protocols with physiological sensors, behavioral data, and machine learning integration (e.g., [64, 56, 58, 52]).

This methodological maturation follows international HCI research, with local research increasingly adopting more formal evaluation strategies.

## 6.3. Collaboration Networks

Co-authorship analysis highlights the emergence of stable clusters around leading Slovenian research groups, while maintaining openness to new collaborators. Notably:

- Several recurring research teams contribute consistently across the decade, often centered around specific labs or institutions.
- International co-authorship appears more prominently in later years, particularly in CEUR-WS volumes, signaling stronger outward collaboration.

Network measures such as degree centrality and betweenness indicate that a small number of highly active researchers act as bridges across subcommunities, contributing to the cohesion of the HCI-SI network.

## 6.4. Emergent Topics

Keyword dynamics reveal emerging interest in:

- *Affective and persuasive computing* (e.g., [66, 51]);
- *Embodied and tangible interaction* (e.g., [38, 34, 61]);

- *Physiological sensing for interaction* (e.g., [56, 58, 52]);
- *AI/ML integration into HCI* (e.g., [64, 41]).

Trend analysis highlights a trajectory of growth: from foundational usability and visualization studies toward diversified, data-rich, and globally relevant research programs.

## 6.5. Community Building

The transition from embedded *Information Society* tracks to a dedicated HCI-SI conference reflects deliberate community-building. Regular publication series, persistent digital identifiers, and a recognizable conference identity have increased the visibility and accessibility of Slovenian HCI research, since it is now accessible in online proceedings. The establishment of HCI-SI has thus played a key role in consolidating the community and positioning it for stronger international engagement.

## 6.6. Challenges and Future Directions

Despite progress, several challenges remain:

- **Sustainability of the venue:** Ensuring consistent participation, institutional support, and publication quality will be vital for long-term success.
- **Diversity of themes:** Some areas of HCI, such as accessibility and industrial IoT, are underrepresented in comparison to other themes, and could benefit from targeted calls or special sessions.
- **Visibility:** Increasing the international recognition of HCI-SI through indexing, stronger digital presence, and joint events with regional or global venues would enhance impact.

## 7. Conclusion

This paper presented a retrospective and trend analysis of ten years of Slovenian HCI conferences. By compiling and analyzing bibliographic metadata, thematic trends, methodological approaches, and collaboration networks, we identified several key insights.

First, the community has matured from small-scale, embedded sessions to a standalone conference with consistent identity, stable proceedings, and increasing international visibility. Second, thematic developments reveal a balance between global HCI trends—such as XR, persuasive technologies, and AI/ML-enhanced interaction—and local applications in education, cultural heritage, and health. Third, methodological practices have evolved toward more rigorous, data-driven studies, reflecting growing research capacity. Finally, collaboration networks highlight both the cohesion of Slovenian research groups and the gradual increase of international co-authorship.

Looking ahead, the main challenges lie in sustaining the growth and visibility of the HCI-SI venue, broadening thematic coverage, and strengthening international partnerships. Addressing these challenges will ensure that Slovenian HCI not only continues to serve local needs but also contributes more visibly to the global HCI discourse.

## Declaration on Generative AI

During the preparation of this work, the author used ChatGPT (GPT-5, OpenAI) for basic data processing, grammar and spelling checks, and sentence rephrasing. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

## References

- [1] A. Kaltenhauser, G.-L. Savino, N. von Felten, J. Schöning, CHI's Greatest Hits: Analyzing the 100 Most-Cited Papers in 43 Years of Research at ACM CHI, *Interactions* 32 (2025) 28–33.
- [2] J. Guna, E. Stojmenova, M. Pogačnik, UX – From Theory to Practical Application, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, p. 5.
- [3] K. Istenič, L. Čehovin, D. Skočaj, Multi-Touch Surface Based on RGBD Camera, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 6–9.
- [4] A. Černivec, C. Bohak, Using Kinect for Touchless Interaction with Existing Applications, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 10–13.
- [5] S. Pečnik, D. Žlaus, D. Mongus, B. Žalik, An Improved Visualization of LiDAR Data Using Level of Details and Weighted Color Mapping, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 14–17.
- [6] M. Konecki, V. Mrkela, Students' Acceptance of Animated Interactive Presentation of Sorting Algorithms, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 18–21.
- [7] B. Blažica, Use of UX and HCI Tools Among Start-Ups, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 22–25.
- [8] M. Ristič, F. Novak, Decision Support in Emergency Call Service, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 26–29.
- [9] M. Konecki, Mobile and Responsive Web Applications, in: *Proceedings of the 17th International Multiconference INFORMATION SOCIETY – IS 2014 - Volume H*, Ljubljana, Slovenia, 2014, pp. 30–33.
- [10] C. Bohak, P. Lavrič, M. Marolt, Remote Interaction in Web-Based Medical Visual Application, in: *Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E*, Ljubljana, Slovenia, 2016, pp. 5–8. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
- [11] B. Blažica, F. Novak, A. Biasizzo, C. Bohak, 3D Serious Games for Parkinson's Disease Management, in: *Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E*, Ljubljana, Slovenia, 2016, pp. 9–12. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
- [12] P. Novak, B. Koroušić Seljak, F. Novak, Designing Visual Interface for Nutrition Tracking of Patients with Parkinson's Disease, in: *Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E*, Ljubljana, Slovenia, 2016, pp. 13–16. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
- [13] V. Blažica, J. J. Cerar, A. Poredoš, Redesign of Slovenian Avalanche Bulletin, in: *Proceedings of*

- the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 17–20. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
- [14] M. Pesek, A. Isaković, G. Strle, M. Marolt, Improving the Usability of Online Usability Surveys with an Interactive Stripe Scale, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 21–24. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [15] B. Gombač, M. Zemljak, P. Širol, D. Deželjin, K. Čopič Pucihar, M. Kljun, Wizard of Oz Experiment for Prototyping Multimodal Interfaces in Virtual Reality, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 29–32. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [16] E. Šimer, M. Kljun, K. Čopič Pucihar, I Was Here: A System for Creating Augmented Reality Digital Graffiti in Public Places, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 40–43. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [17] A. Malečkar, M. Kljun, P. Rogelj, K. Čopič Pucihar, Evaluation of Common Input Devices for Web Browsing: Mouse vs Touchpad vs Touchscreen, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 25–28. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [18] J. Štrekelj, B. Kavšek, Interactive Video Management by Means of an Exercise Bike, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 44–47. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [19] T. Knez, M. Gjoreski, V. Pejović, Analiza vpliva težavnosti računalniške igre na izmerjene vrednosti fizioloških signalov, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 5–8. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [20] A. Tošič, J. Vičič, M. Burnard, Privacy preserving indoor location and fall detection system, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 9–12. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [21] M. Vrancich, M. Matetic, Exploratory data analysis of stream data in sports medicine domain, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 13–16. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [22] L. Vranješ, J. Žabkar, Sledenje pogledu s spletno kamero, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 17–20. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [23] L. Zorč, K. Čopič Pucihar, M. Kljun, Prepričljive tehnologije za spodbujanje pravilne drže telesa pri sedenju, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 21–24. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [24] C. Campos, C. A. Martínez Sandoval, Nuni-A case study: a platform to distribute digital content to analog television data towards enhancing quality of life of senior citizen in Mexico, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 25–28. URL: <http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/>

IS2019\_Volume\_H%20-%20HCI.pdf.

- [25] A. Martinovic, V. Pejović, Investigating the Role of Context and Personality in Mobile Advertising, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 5–8. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [26] T. Tušar, Interaktivna vizualizacija proračuna Republike Slovenije s Sankeyevim diagramom, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 9–12. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [27] J. Zupančič, M. Štravs, M. Mlakar, MightyFields Voice: Voice-based Mobile Application Interaction, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 13–16. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [28] J. Žganec Gros, M. Romih, T. Šef, eBralec 4: hibridni sintetizator slovenskega govora, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 17–20. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [29] J. Žganec Gros, Žiga Golob, S. Dobrišek, Učinkovita predstavitev slovarskih jezikovnih virov pri govornih tehnologijah, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 45–48. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [30] G. Sotlar, P. Roglej, K. Čopič Pucihar, M. Kljun, Predmetnik: oprijemljiv uporabniški vmesnik za informiranje turistov, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 29–32. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [31] R. Cej, F. Solina, Anamorflična projekcija na poljubno neravno površino, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 41–44. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [32] M. Weerasinghe, K. Čopič Pucihar, M. Kljun, P. Coulton, Playing with the Artworks: A Personalised Artwork Experience, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 74–79. URL: <https://ceur-ws.org/Vol-3054/paper8.pdf>.
- [33] Špela Bricman, I. Kožuh, User Experience and Interface Design for a Digital Pet Adoption Platform, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 32–44. URL: <https://ceur-ws.org/Vol-3657/paper4.pdf>.
- [34] F. Sprostan, M. Kljun, K. Čopič Pucihar, Manipulating and Augmenting Digital Content Through Physical Counterpart, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 45–58. URL: <https://ceur-ws.org/Vol-3657/paper6.pdf>.
- [35] J. A. Deja, N. Attygale, K. Čopič Pucihar, M. Kljun, Sound 2121: The Future of Music is Natural, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 21–24. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [36] M. Plankelj, N. Lukač, S. Rizvić, S. Kolmanič, Ohranjanje kulturne dediščine s pomočjo navidezne in obogatene resničnosti, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 25–28. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [37] P. Škrlj, M. Lochrie, M. Kljun, K. Čopič Pucihar, StreetGamez: detection of feet movements on the projected gaming surface on the floor, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 37–40. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).

- //library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\_Volume\_H%20-%20HCI.pdf.
- [38] K. Trajkovska, K. Čopič Pucihar, M. Kljun, J. A. Deja, M. Weerasinghe, pART Blocks: Programming with Physical Tangible Blocks and AR, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 5. URL: [https://ceur-ws.org/Vol-3300/short\\_5931.pdf](https://ceur-ws.org/Vol-3300/short_5931.pdf).
  - [39] M. A. Berends, J. A. Deja, N. T. Attygalle, M. Kljun, K. Čopič Pucihar, GazeHD: Towards Measuring Effect of Depth of Field Controlled by Eye Tracking in 3D Environments, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 5. URL: [https://ceur-ws.org/Vol-3300/short\\_6981.pdf](https://ceur-ws.org/Vol-3300/short_6981.pdf).
  - [40] J. Markovska, D. Šoberl, Deep Reinforcement Learning Compared to Human Performance in Playing Video Games, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 5. URL: [https://ceur-ws.org/Vol-3300/short\\_8285.pdf](https://ceur-ws.org/Vol-3300/short_8285.pdf).
  - [41] T. Kostovski, M. Kljun, K. Čopič Pucihar, The Role of Energetic Music in a Video Game: Analyzing its Effect on Immersion, Perception and Performance, in: Proceedings of the 9th Human-Computer Interaction Slovenia Conference (HCI-SI 2024), Ljubljana, Slovenia, 2024, pp. 49–63. URL: <https://ceur-ws.org/Vol-3866/paper49.pdf>.
  - [42] G. Radež, C. Bohak, Integrating Environmental Awareness Into NPCs: Contextual Conversational Interaction in Games, in: Proceedings of the 9th Human-Computer Interaction Slovenia Conference (HCI-SI 2024), Ljubljana, Slovenia, 2024, pp. 11–29. URL: <https://ceur-ws.org/Vol-3866/paper11.pdf>.
  - [43] S. Štor, J. A. Deja, I. Pucihar, K. Čopič Pucihar, M. Kljun, Teach Me How to Improvise: Co-designing an Augmented Piano Training System for Improvisation, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 89–94. URL: <https://ceur-ws.org/Vol-3657/paper10.pdf>.
  - [44] M. Weerasinghe, L. Ribič, M. Kljun, K. Čopič Pucihar, I. Devetak, Augmented Reality Training System Fusing the Triple Nature of Chemical Concepts, in: Proceedings of the 9th Human-Computer Interaction Slovenia Conference (HCI-SI 2024), Ljubljana, Slovenia, 2024, pp. 64–70. URL: <https://ceur-ws.org/Vol-3866/paper64.pdf>.
  - [45] A. Shishkov, S. Kolmanič, Progressive Education: Augmented Reality Visualization of Human Anatomy on HoloLens 2, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 12–19. URL: <https://ceur-ws.org/Vol-3657/paper2.pdf>.
  - [46] A. Jankovič, T. Kolenik, V. Pejović, The Role of Personality-Tailored Notifications in Mobile-Based Behavior Change Intervention, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 18–22. URL: <https://ceur-ws.org/Vol-3054/paper2.pdf>.
  - [47] M. Jovanović, V. Groznik, M. Tkalčič, Predicting the Gullibility of Users from their Online Behaviour, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 35–44. URL: <https://ceur-ws.org/Vol-3054/paper4.pdf>.
  - [48] S. Hrustanović, B. Kavšek, M. Tkalčič, Recognition of Eudaemonic and Hedonic Qualities from Song Lyrics, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 45–53. URL: <https://ceur-ws.org/Vol-3054/paper5.pdf>.
  - [49] E. Motamedi, M. Tkalčič, Prediction of Eudaimonic and Hedonic Movie Characteristics From Subtitles, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 54–61. URL: <https://ceur-ws.org/Vol-3054/paper6.pdf>.
  - [50] G. Strle, A. Košir, E. A. Oğuz, U. Burnik, Predicting User Engagement in Video Advertisement: Insights from Pupillary Response and Heart Rate, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 10. URL: [https://ceur-ws.org/Vol-3300/paper\\_6284.pdf](https://ceur-ws.org/Vol-3300/paper_6284.pdf).
  - [51] E. Spirova, A. M. Golubovikj, M. Tkalčič, Music and Myth: The Relationship Between Music

- Preference and Unverified Beliefs, in: Proceedings of the 9th Human-Computer Interaction Slovenia Conference (HCI-SI 2024), Ljubljana, Slovenia, 2024, pp. 3–10. URL: <https://ceur-ws.org/Vol-3866/short1.pdf>.
- [52] N. Miletić, M. Kljun, K. Čopič Pucihar, UX Design: the Impacts on Physiological Responses, in: Proceedings of the 9th Human-Computer Interaction Slovenia Conference (HCI-SI 2024), Ljubljana, Slovenia, 2024, pp. 30–38. URL: <https://ceur-ws.org/Vol-3866/paper30.pdf>.
- [53] G. Strle, A. Košir, K. S. Pečecnik, J. Sodnik, The Effects of Driving Disengagement on Response Time in Transition to Manual Driving Mode, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 10. URL: [https://ceur-ws.org/Vol-3300/paper\\_4485.pdf](https://ceur-ws.org/Vol-3300/paper_4485.pdf).
- [54] A. Košir, U. Burnik, J. Zaletelj, S. Jean, P. Janjušević, G. Strle, Estimation of Mathematical Anxiety Using Psycho-Physiological Data, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 10. URL: [https://ceur-ws.org/Vol-3300/paper\\_4701.pdf](https://ceur-ws.org/Vol-3300/paper_4701.pdf).
- [55] S. Aloui, R. Morvillier, C. Prat, J. Sodnik, C. Diaz-Piedra, F. Angioi, L. L. D. Stasi, Driver Monitoring Systems in Automated Interactions: A Realtime, Thermographic-based Algorithm, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 7. URL: [https://ceur-ws.org/Vol-3300/short\\_392.pdf](https://ceur-ws.org/Vol-3300/short_392.pdf).
- [56] T. Gruden, K. S. Pečecnik, G. Jakus, J. Sodnik, Quantifying Drivers' Physiological Responses to Take-Over Requests in Conditionally Automated Vehicles, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 9. URL: [https://ceur-ws.org/Vol-3300/short\\_3798.pdf](https://ceur-ws.org/Vol-3300/short_3798.pdf).
- [57] K. S. Pečecnik, J. Sodnik, Increasing Driver's Situational Awareness in Semi-automated Vehicles Using a Head-up Display, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 9. URL: [https://ceur-ws.org/Vol-3300/short\\_5226.pdf](https://ceur-ws.org/Vol-3300/short_5226.pdf).
- [58] K. S. Pečecnik, M. Kofol, J. Sodnik, Method for Inferential Continuous Assessment of Driver's Situational Awareness, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 20–31. URL: <https://ceur-ws.org/Vol-3657/paper3.pdf>.
- [59] M. Klopčič, K. S. Pečecnik, Accurate Proximity Sensor for Parking Assistance, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 102–109. URL: <https://ceur-ws.org/Vol-3657/paper12.pdf>.
- [60] J. M. Santiago III, G. Nodalo, J. Valenzuela, J. A. Deja, Explore, Edit, Guess: Understanding Novice Programmers' Use of CodeBlocks for Regression Experiments, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 3–17. URL: <https://ceur-ws.org/Vol-3054/paper1.pdf>.
- [61] P. Jolakoski, J. A. Deja, K. Čopič Pucihar, M. Kljun, Teaching Shortest Path Algorithms With a Robot and Overlaid Projections, in: Proceedings of the 9th Human-Computer Interaction Slovenia Conference (HCI-SI 2024), Ljubljana, Slovenia, 2024, pp. 39–48. URL: <https://ceur-ws.org/Vol-3866/short4.pdf>.
- [62] K. Trajkovska, M. Weerasinghe, K. Čopič Pucihar, M. Kljun, Name and Face Recall Cognitive Failure: Presenting a Short Literature Review and System Design, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 59–67. URL: <https://ceur-ws.org/Vol-3657/paper7.pdf>.
- [63] A. Krašovec, V. Pejović, Investigating Sensor Modality Informativeness and Stability for Behavioural Authentication, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 62–73. URL: <https://ceur-ws.org/Vol-3054/paper7.pdf>.
- [64] A. Kristan, D. Pellarini, V. Pejović, Not Deep Enough: Autoencoders for Automatic Feature Extraction in Wireless Cognitive Load Inference, in: Proceedings of the 6th Human-Computer

- Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 23–34. URL: <https://ceur-ws.org/Vol-3054/paper3.pdf>.
- [65] A. Dziuba, U. Sergaš, Cross-Country Analysis on Connection Between Financial Lifestyle and Happiness, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 2–11. URL: <https://ceur-ws.org/Vol-3657/paper1.pdf>.
  - [66] U. Sergaš, H. Kalkan, M. Tkaličič, Tribalism and Fake News: Descriptive and Predictive Models on How Belief Influences News Trust, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 9. URL: [https://ceur-ws.org/Vol-3300/short\\_5576.pdf](https://ceur-ws.org/Vol-3300/short_5576.pdf).
  - [67] A. M. Golubovikj, B. Kavšek, M. Tkaličič, Imputing Missing Answers in the World Values Survey, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 8. URL: [https://ceur-ws.org/Vol-3300/short\\_1345.pdf](https://ceur-ws.org/Vol-3300/short_1345.pdf).
  - [68] G. Grbec, N. Bašić, M. Tkaličič, Ranking Footballers with Multilevel Modeling, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 95–101. URL: <https://ceur-ws.org/Vol-3657/paper11.pdf>.
  - [69] F. Sobiech, N. Walczak, A. Buczek, M. Jeanty, K. Kupiński, Z. Chaniecki, A. Romanowski, K. Grudzień, Exploratory Analysis of Users' Interactions with AR Data Visualisation in Industrial and Neutral Environments, in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 5. URL: [https://ceur-ws.org/Vol-3300/short\\_8744.pdf](https://ceur-ws.org/Vol-3300/short_8744.pdf).
  - [70] D. Vilar, V. Pejović, B. Blažica, GoraNiNora: Kontekstno-odvisno obveščanje za varen obisk gora (GoraNiNora: Context-Dependent Dissemination of Mountaineering Safety Information), in: Proceedings of the 7th Human-Computer Interaction Slovenia Conference (HCI-SI 2022), Ljubljana, Slovenia, 2022, p. 12. URL: [https://ceur-ws.org/Vol-3300/paper\\_713.pdf](https://ceur-ws.org/Vol-3300/paper_713.pdf).
  - [71] Ž. Kopušar, F. Novak, Towards the Improvement of Guard Graphical User Interface, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 33–36. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [72] G. Pavlin, M. Pavlin, Towards Affordable Mobile Crowd Sensing Device, in: Proceedings of the 19th International Multiconference INFORMATION SOCIETY – IS 2016 - Volume E, Ljubljana, Slovenia, 2016, pp. 37–39. URL: [https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016\\_Volume\\_E%20-%20HCI.pdf](https://is.ijs.si/wp-content/uploads/zborniki/zborniki/2016/IS2016_Volume_E%20-%20HCI.pdf).
  - [73] B. Kolar, M. Kljun, K. Čopič Pucihar, Umestitev interaktivnih elementov in elementov igrifikacije na vnaprej zastavljeni učni poti, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 29–30. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [74] P. Širol, K. Čopič Pucihar, M. Kljun, Interakcija z umetniškimi deli preko množičnega ocenjevanja, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 31–32. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [75] T. Jesenko, M. Kljun, K. Čopič Pucihar, Igrifikacija virtualnega obiska učne poti Škocjan z uporabo mobilnih tehnologij in 360-stopinjskih posnetkov, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 33–34. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [76] G. Moderc, A. Tošić, J. Vičič, In-Game Economy Based on Blockchain, in: Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019 - Volume H, Ljubljana, Slovenia, 2019, pp. 39–42. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_H%20-%20HCI.pdf).
  - [77] R. Prislan, The Fundamentals of Sound Field Reproduction Using a Higher Order Ambisonics

- System, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 49–51. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [78] P. Kocjančič, M. Kljun, K. Čopič Pucihar, Primerjava Vnosa Besedila v Virtualnem Okolju na Različnih Postavitvah Tipkovnice, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 80–86. URL: <https://ceur-ws.org/Vol-3054/paper9.pdf>.
- [79] E. Huskanović, B. Blažica, E-Računi: Razlogi za Uporabo in Uporabniška Prijaznost, in: Proceedings of the 6th Human-Computer Interaction Slovenia Conference (HCI-SI 2021), Koper, Slovenia, 2021, pp. 87–90. URL: <https://ceur-ws.org/Vol-3054/paper10.pdf>.
- [80] N. Kovačević, J. A. Deja, M. Weerasinghe, K. Čopič Pucihar, M. Kljun, Retzzles: Do Jigsaw Puzzle Actions on Interactive Display Maps Increase the Retention of Map Information?, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 68–77. URL: <https://ceur-ws.org/Vol-3657/paper8.pdf>.
- [81] N. T. Attygalle, U. Vuletic, M. Kljun, K. Čopič Pucihar, Towards Hand Gesture Recognition Prototype Using the IWR6843ISK Radar Sensor and Leap Motion, in: Proceedings of the 8th Human-Computer Interaction Slovenia Conference (HCI-SI 2023), Maribor, Slovenia, 2023, pp. 78–88. URL: <https://ceur-ws.org/Vol-3657/paper9.pdf>.
- [82] M. Zorko, M. Debevc, I. Kožuh, Razvoj in Ocenjevanje Prototipa Mobilne Aplikacije z Elementi Igrifikacije in Mešane Resničnosti, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 33–36. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).
- [83] K. S. Orehek, V. Dolnicar, S. H. Touzery, The use of eCare services among informal carers of older people and psychological outcomes of their use, in: Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020 - Volume H, Ljubljana, Slovenia, 2020, pp. 52–54. URL: [http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020\\_Volume\\_H%20-%20HCI.pdf](http://library.ijs.si/Stacks/Proceedings/InformationSociety/2020/IS2020_Volume_H%20-%20HCI.pdf).

# Razvoj in uporabniška študija mobilne aplikacije za vadbo poliritmov

Jaka Kužner<sup>1</sup>, Matevž Pesek<sup>1,\*</sup> and Matija Marolt<sup>1</sup>

<sup>1</sup>University of Ljubljana, Faculty of computer and information science, Večna pot 113, 1000 Ljubljana, Slovenia

## Abstract

V članku predstavimo zasnovano, razvoj in evalvacijo mobilne aplikacije za učenje in vadbo poliritmov, pri čemer je bil poseben poudarek namenjen uporabi polizvezde kot vizualnega pripomočka in integraciji elementov poigritve. Evalvacija, izvedena z uporabniškimi testi in vprašalniki UEQ-S in TAM, je pokazala pozitiven potencial aplikacije ter nakazala priložnosti za izboljšave, zlasti pri algoritmu za ocenjevanje, optimizaciji za različne naprave in razširitvi funkcionalnosti. Izdelana aplikacija predstavlja osnovo za nadaljnji razvoj, ki bo usmerjen v možnosti dinamičnega spreminjanja poliritmov in taktovskega načina, v dodatne načine točkovanja ter integracijo v platformo Trubadur, namenjeno vaji glasbenih konceptov.

## Keywords

učenje glasbe, ritmična igra, poliritem, pedagoška orodja, uporabniška izkušnja, vizualizacija ritma

## 1. Uvod

Ritmične in glasbene igre predstavljajo pomembno stičišče med zabavo, učenjem in digitalno interakcijo. V zadnjih desetletjih so se uveljavile kot priljubljena oblika rekreativne dejavnosti, ki s pomočjo poigritve – vključevanjem mehanizmov, kot so točkovanje, napredovanje po nivojih in vizualne povratne informacije, uporabnike spodbujajo k razvijanju motoričnih spretnosti, koordinacije, pozornosti in različnih glasbenih znanj [1, 2]. Hkrati kažejo tudi velik potencial kot alternativna pedagoška orodja, saj glasbeno izobraževanje prek iger, tako učencem kot učiteljem, odpira nove možnosti za bolj dostopno, motivirano in učinkovito učenje [3, 4].

Kljub širokemu naboru ritmičnih iger na trgu pogosto primanjkuje vsebin, ki bi uporabnika vodile onkraj osnovnega sledenja utripu in ga postopno uvedle v bolj kompleksne glasbene pojave. Eden izmed takih pojavov so **poliritmi** – sočasno izvajanje dveh ali več različnih metričnih vzorcev (npr. v razmerju 3:2 ali 4:3), ki zahteva prefinjen občutek za čas, notranjo koordinacijo in sposobnost poslušanja več ritmičnih slojev hkrati [5, 6]. Sodobne digitalne tehnologije odpirajo nove priložnosti za njihovo predstavitev: interaktivne vizualizacije, natančna povratna informacija in elementi poigritve uporabniku omogočajo, da se s tem zahtevnim konceptom sreča na igriv, raziskovalen način, kjer se prepletata zabavna dinamika ritmičnih iger in pedagoška vrednost poglobljenega glasbenega učenja.

Pomembnost tovrstnega pristopa izhaja iz dejstva, da tradicionalno učenje poliritmov poteka bodisi z vodenim igranjem pod mentorstvom bodisi s pomočjo notnih primerov ali ustnega prenosa. Takšne metode zahtevajo neposreden stik z mentorjem, daljši časovni vložek in pogosto tudi finančna sredstva, kar lahko za marsikoga predstavlja oviro. Digitalno orodje, ki združuje vizualne, slušne in interaktivne elemente, lahko ta proces naredi dostopnejši, hitrejši in bolj prilagodljiv. S tem se zmanjšajo začetne frustracije, poveča angažiranost uporabnika ter omogoči, da se s kompleksnimi ritmi spoprime širši krog ljudi.

V članku predstavljamo aplikacijo, ki je namenjena raznolikemu spektru uporabnikov: glasbenim pedagogom kot dopolnitev učnih metod, učencem in dijakom za samostojno vadbo, samoukim glasbenikom, ki želijo razširiti svoje sposobnosti, ter ljubiteljem ritmičnih iger, ki iščejo nove izzive. Nadaljnje želimo

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

✉ jk1087@student.uni-lj.si (J. Kužner); matevz.pesek@fri.uni-lj.si (M. Pesek); matija.marolt@fri.uni-lj.si (M. Marolt)

🌐 <https://matevzpesek.si/> (M. Pesek); <https://musiclab.si/> (M. Marolt)

🆔 0000-0001-9101-0471 (M. Pesek); 0000-0001-9101-0471 (M. Marolt)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.7>

z aplikacijo nasloviti manjko aplikacij, ki na alternativne načine predstavljajo koncept interaktivne vizualizacije ritma in lahko v prihodnje služijo za različne ustvarjalce glasbe pri ustvarjanju novih del ali kot eksperimentalna platforma za raziskovanje sveta ritmičnih struktur. Prototip razvite aplikacije evalviramo na skupini uporabnikov z raznolikim glasbenim predznanjem v obliki vprašalnikov TAM in UEQ-S in vprašalnika odprtega tipa, ki služijo kot podlaga za nadaljnji razvoj in integracijo aplikacije v obstoječo platformo Trubadur.

## 2. Pregled področja

Ritmične igre so v zadnjih desetletjih postale stalnica v svetu videoiger. Njihova privlačnost temelji na povezovanju glasbe, gibanja, koordinacije in hitrega odzivanja. Ena prvih komercialno uspešnih ritmičnih iger je bila **PaRappa the Rapper** (1996), kjer igralec v vlogi raperskega psa ponavlja ritmične vzorce v pravem tempu in zaporedju. Igra je izšla za konzolo Sony PlayStation in hitro postala kulturna uspešnica zaradi svojega edinstvenega vizualnega sloga, humorja in nalezljive glasbe.

Naslednji pomemben mejnik je bila igra **Dance Dance Revolution** (1998), ki je z uporabo plesne plošče kot vhodne naprave igralcem omogočila telesno aktivnost, spodbujala gibanje ter kreativno oblikovanje lastnih koreografij. Igralci so sledili vizualnim puščicam in v pravem trenutku stopali na ustrezna mesta plošče. Zaradi svoje fizične komponente je DDR hitro postal kulturni fenomen, še posebej v Aziji in v arkadni skupnosti.

V sredini 2000-ih sta priljubljenost ritmičnih iger okrepili seriji **Guitar Hero** in **Rock Band**, ki sta predstavili koncept uporabe poenostavljenih inštrumentov kot kontrolerjev. Igralec ni le gledal zaslona, temveč je dejansko igral na plastično kitaro, kar je ustvarilo močan občutek identifikacije z glasbenikom in populariziralo igranje na odru med širšo publiko.

S prihodom pametnih telefonov so se ritmične igre prilagodile na nove platforme. Medtem ko mobilne igre navadno niso toliko fizično vključujoče kot npr. Guitar Hero ali Dance Dance Revolution, so zelo dostopne in raznolike, kar omogoča igranje praktično kjerkoli in kadarkoli. Takšne igre pogosto ponujajo različne težavnostne ravni, vizualne in zvočne učinke ter interaktivne elemente, ki lahko spodbujajo kognitivne in motorične spretnosti. **Tap Tap Revenge** (2008) je izkoristil zaslone na dotik in postal ena prvih množično igranih glasbenih iger na mobilnih napravah, pri čemer je igralcem omogočil enostaven vstop v ritmične igre brez posebne opreme.

Z razvojem tehnologij, kot so **navidezna resničnost (VR)**, **obogatena resničnost (AR)** in **napredna senzorska tehnologija**, se pojavijo tudi novi načini interakcije v ritmičnih igrah. Ena takšnih je VR-igra **Steady the drums!** [7], ki igralca postavi v vlogo bobnarja v domišljajskem svetu, kjer z udarci po bobnih vodi svojo vojsko.

Ena najbolj popularnih VR ritmičnih iger je igra **Beat Saber**. Igralec se v igri znajde v nadrealističnem neonskem svetu, kjer z laserskimi meči po ritmu pesmi, seka prihajajoče bloke v pravi smeri in tempu. Igra združuje gibanje telesa, natančnost in hitrost, kar ustvarja intenzivno in zabavno izkušnjo. Poleg tega omogoča igralcem, da se potopijo v glasbo na povsem nov način, hkrati pa spodbuja telesno aktivnost in koordinacijo. Zaradi svoje enostavne, a hkrati zahtevne mehanike, je Beat Saber postal priljubljena izbira med ljubitelji VR in ritmičnih iger po vsem svetu. Tak pristop ne vključuje le ritmične natančnosti, temveč tudi prostorsko orientacijo, gibalno odzivnost in občutek telesne prisotnosti.

Poleg digitalnih pristopov pa obstajajo tudi **analogne in nekonvencionalne ritmične igre**. Sem sodijo razne didaktične igre, kot so:

- hoja ali udarjanje po ritmu z roko/stopalom (uporablja se v Orffovem pristopu [8]),
- igranje ritmičnih vzorcev po telesu (angl. body percussion - telo kot inštrument),
- kartice z ritmičnimi zapisi, ki jih učenec prepozna ali reproducira,
- skupinsko bobnanje in "call-and-response" igre v razredu.

Te metode so pogosto priljubljene pri uvajanju otrok v glasbeni svet, saj spodbujajo telesno učenje, skupinsko dinamiko in intuitivno razumevanje ritma.

Kombinacijo mobilnih naprav in učenja glasbe najbolj predstavljajo izobraževalne aplikacije, kot so Yousician (za učenje kitare, klavirja, bas kitare in petja), Rocksmith (z električno kitaro kot vhodno napravo), Piano Maestro (za začetno učenje klavirja). Te aplikacije združujejo strukturirane lekcije, realnočasovno ocenjevanje izvedbe ter poigritvene elemente (točke, dosežki, lestvice), kar omogoča motivirano in samostojno učenje.

Digitalna pedagogika, zlasti na področju glasbe, pogosto naleti na omejitve obstoječih orodij. Platforme, kot je Moodle, se večinoma uporabljajo zgolj za shranjevanje, akumulacijo in distribucijo gradiv, redko pa ponujajo okolje za aktivno interaktivno učenje ali razvoj zahtevnejših orodij [9]. Platforma **Trubadur** si prizadeva to vrzel zapolniti z igrificiranim pristopom in razvijanjem digitalnih učnih pripomočkov za glasbene pedagoge in učence. V slovenskem prostoru je ena redkih odprtokodnih izobraževalnih platform na področju glasbe. Učiteljem nudi orodja za pripravo vaj, avtomatizirano generiranje ritmičnih in melodičnih nalog ter analitiko uspešnosti. Z razvojem prototipne aplikacije poliritmične igre, ki jo nameravamo vključiti v platformo, se platforma razširja z novim tipom nalog, ki presega običajno klikovno interakcijo in uporabniku omogoča igranje zahtevnejših ritmičnih struktur na interaktiven način.

## 2.1. Teoretična in pedagoška osnova poliritmov

Poliritmija je glasbeni pojav, pri katerem sočasno izvajamo dva ali več različnih ritmov znotraj istega osnovnega pulza [10, 11]. Primer je 3:2 poliritem (imenovan tudi "trojka nad dvojko" ali "tri proti dve"), kjer ima en ritem tri enake enote v času, ko ima drugi dve enoti [11]. Splošno gledano tvorita poliritem frazi z  $m : n$  razdelitvami (npr. 3:2, 4:3, 5:4), kar pomeni sočasno igranje  $m$  enakomernih delčkov prvega ritma in  $n$  delčkov drugega. V glasbenem zapisu poliritme pogosto prikažemo kot različne delitve istega tempa (npr. v istem taktu), vendar jih lahko tudi razumemo kot simetrično usklajevanje dveh med seboj povezanih tempov [12].

Poliritmi so pogosti v številnih glasbenih tradicijah. Zasledimo jih v afriških in indonezijskih tradicijah (gamelan idr.), pa tudi v sodobni zahodni umetni glasbi – tako pri skladateljih kot sta Charles Ives in Elliott Carter, kot tudi v jazzu in rocku [10]. Uporaba poliritma skladbi doda notranjo kompleksnost: različno delitev ritma hkrati ustvarja bogato ritmično teksturo in kontrast. Pri analizi poliritmov glasbena teorija pogosto uporablja najmanjši skupni večkratnik (angl. Least Common Multiple, krajše LCM) razdelitvenih enot, da določi skupni cikel (npr. za 3:2 je LCM 6, kar pomeni šest enakih pod-pulzov v ciklu).

Učenje poliritmov prinaša posebne izzive. Tradicionalni koncept neodvisnosti rok se izkaže za manj učinkovit, saj v resnici poliritem zahteva zaznavanje celotnega koordinacijskega vzorca [12]. Swafford (2023) zato predlaga, da se učenci osredotočijo na internalizacijo celotnega vzorca, ne pa izolirane vadbe rok. Poleg tega Pinkl in Cohen (2023) predlagata uporabo ritmičnih jezikovnih fraz kot učnega orodja: na primer angleške fraze, kot sta "hot cup of tea" za poliritem 3:2 ali "what atrocious weather" za 3:4, ki odražajo ustrezno razmerje ritmov in pomagajo študentom lažje zapomniti si strukturo [11].

Učitelji lahko začnejo z osnovnim poliritmom (npr. 3:2), ga učencu pomagajo premostiti v občutenje osnovnega utripanja, nato pa postopoma uvajajo zahtevnejša razmerja (3:4, 5:4 itd.). Yokuş in Yokuş (2015) predlagata metodo organizacije ritmičnih struktur s sintezo, kjer se kompleksne vzorce deli na manjše enote v skupnem ciklu – to znatno izboljša razumevanje in izvedbo [13]. Podobno lahko vaje v gibalni orientaciji (po vzoru euritmike) pomagajo, saj gibanje okrepi povezavo med slušnim zaznavanjem in motorično izvedbo. Ključno je torej razviti sposobnost hkratnega zaznavanja osnovnega in sekundarnega ritma kot celoto, da ju znamo tudi pravilno motorično izvesti. Pri razvoju aplikacije bomo, z vizualno in sonično podporo, igralce spodbudili, da poliritme tudi motorično izvedejo ter da lahko z orodjem razvijajo in urijo sposobnost ločevanja osnovnega in sekundarnega ritma.

### 3. Razvoj aplikacije

#### 3.1. Predstavitev vizualnega oblikovanja aplikacije

Aplikacija “Poliritmična igra” (angl. **Polyrhythm Game**) je izobraževalna mobilna in namizna aplikacija, namenjena učenju in vadbi poliritmov preko interaktivnih vizualizacij in zvočnih povratnih informacij. Uporabnik izbere nivo, ki določa število točk (udarcev) na posamezni polizvezdi, nato pa z dotiki zaslona ali pritiskanjem tipk sledi ritmu, ki ga aplikacija generira in vizualizira.

#### 3.2. Funkcionalnosti aplikacije

Glavne funkcionalnosti vključujejo:

- Izbor različnih poliritmičnih nivojev, kjer vsak nivo določa število točk na posamezni polizvezdi (npr. 3:2, 4:3, 5:4).
- Vizualizacija ritma s pomočjo animiranih polizvezd, kjer vsaka točka predstavlja udarec v ciklu.
- Zvočna informacija ob vsakem udarcu, ki jo generira zvočni pogon SoLoud.
- Merjenje natančnosti dotikov uporabnika glede na generirani ritem in prikaz povratne informacije o uspešnosti.
- Prilagajanje hitrosti ritma (BPM) z drsnikom.
- Podpora za mobilne naprave (dotik zaslona) in namizne računalnike (tipki A in F).

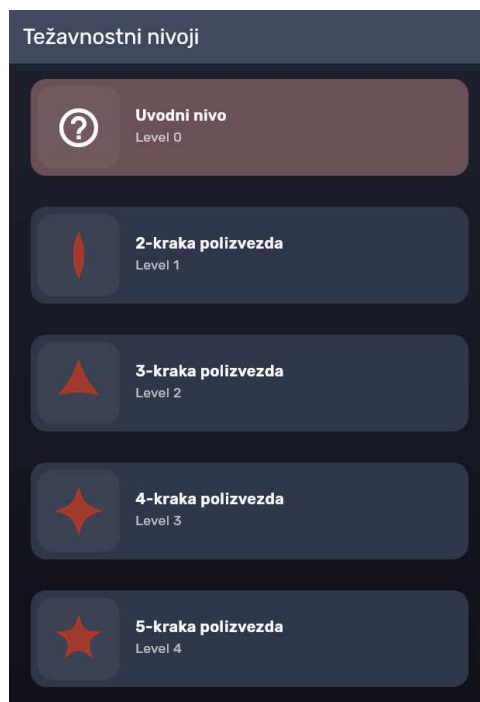
#### 3.3. Arhitektura aplikacije

Aplikacija je zasnovana v okolju Flutter. Aplikacija je zasnovana na šestih osnovnih pogledih. Vstopna točka aplikacije inicializira globalne ponudnike (zvok), lokalizacijo, tematski videz in navigacijo med zaslone (angl. routing). Tukaj so definirani nivoji (`PolyrhythmLevel`) in prikazani v obliki seznama. Nadaljnje osrednja komponenta uporabniškega vmesnika za izvajanje in vizualizacijo poliritmičnih vaj. Vsebuje razred `PolyrhythmGame`, ki prikazuje animirane polizvezde, omogoča interakcijo uporabnika, izračunava končno natančnost in prikazuje rezultate. Podpira tudi prikaz LCM (najmanjšega skupnega večkratnika) zvezde za lažjo predstavo kompleksnih poliritmov in BPM drsnik za prilagajanje tempa. Vsebuje tudi uvodni nivo (angl. Onboarding Level). Pogled izrisa vaje skrbi za animacijo polizvezde, izračun pozicije animacije, zaznavanje dotikov, izračun njihovih natančnosti in izris zvezde. Uporablja tudi komponente metronoma za časovno sinhronizacijo animacije in zvoka. Upravljanje zvočnega pogona SoLoud prednaloži zvočne datoteke in omogoča predvajanje različnih zvokov ob udarcih. Komponenta `Mmtronom` generira časovne intervale za udarce in omogoča sinhronizacijo animacije ter zvoka. V ozadju komponenta strukturnega pogleda vključuje tematske nastavitve (`theme.dart`), definicije nivojev (`polyrhythm_level.dart`), ter pomožne komponente za prikaz in interakcijo.

Nivoje smo glede na težavnost uredili v logično zaporedje:

- **Uvodni nivo**, ki uporabniku razloži koncept poliritmov, predstavi glavni element igre - polizvezdo in kako z njo interagirati ter nekaj splošnih napotkov za učenje poliritmov in uporabo vmesnika.
- **2-, 3-, 4- in 5-krako polizvezdo**, ki uporabniku predstavi različne osnovne načine deljenja ritma.
- Zaporedje poliritmičnih nivojev, ki sledijo glasbenemu priročniku Petra Magadinija za osvajanje poliritmov [14], in sicer **3:2, 4:3, 5:4, 7:4, 11:4 in 13:4**.
- Sledi še 6 dodatnih nivojev, ki ustrezno razširijo nabor nivojev. Vključujejo poliritme **7:3, 7:5, 11:3, 13:3, 11:5 in 13:5**.

Za vsak dotik, katerega natančnost presega 90%, se od `_ratingMetric` odšteje  $1/\text{numberOfPoints}$ . V nasprotnem primeru, če uporabnik ne zadosti vsaj 75% natančnosti, se `_ratingMetric` prišteje 1. Igra se konča, ko `_ratingMetric` doseže število ciklov, določeno s spremenljivko `cycles`. Tako igralca kaznujemo za slabo izvedbo in nagradimo, ko ritmu natančno sledi.



**Figure 1:** Seznam težavnostnih nivojev.

Ob aktivaciji igre s pritiskom na gumb “Start” se sproži animacija polizvezde. Prvi cikel animacije je namenjen kalibraciji uporabnika na izbrani poliritem in se zato ne vključuje v proces ocenjevanja. Šele po zaključenem uvodnem ciklu se polizvezda obarva, s čimer naznani vrednotenje izvajanja. Prikaz aplikacijo z izrisanimi polizvezdami je prikazan na Sliki 2.



**Figure 2:** Primer igre 7:4 poliritma. Na levi strani zaslona je drsnik za BPM, v sredini animirani polizvezdi s kazalcem in vizualno povratno informacijo o točnosti, na desni gumb za začetek/pavzo.

Polizvezda služi več funkcijam:

- je **vizualna ponazoritev ritma** z enakomerno razporejenimi konicami,
- je **navodilo za izvajanje ritma**, saj konice predstavljajo točke udarca, število konic pa število delitev,
- je **interaktivni element**, ki reagira na uporabniške dotike,

- omogoča **sprotno evalvacijo** prek vizualne in numerične povratne informacije.

Polizvezda ima organsko obliko, ki ritmično pulzira in za vsak udarec belo bliskne, po robu zvezde pa kroži animirana pika, podobno kot gibanje dirigentske palice. Polizvezda tako deluje kot glasbeno navodilo (število krakov predstavlja število udarcev v enem ciklu) in predmet interakcije, kar poenostavi grafični vmesnik.



**Figure 3:** Pojavno okence z rezultati ritmične natančnosti, predstavljene kot krožni indikator (CircularPercentIndicator) in numerično izražene v odstotnih točkah.

Razvita aplikacija omogoča izvajanje poliritmov in spremljanje natančnosti izvedbe v realnem času. Testna različica je delovala brez povezave (angl. offline) in ne vključuje naprednih igrifikacijskih funkcionalnosti, kot so shranjevanje napredka ali lestvica igralcev, z namenom evalvacije uporabniškega vmesnika brez morebitnega vpliva poigritvenih elementov na uporabnikovo izkušnjo.

## 4. Evalvacija

Osnovni namen evalvacije je bil preverjanje delovanja posameznih funkcionalnosti; nadaljnje smo tudi želeli oceniti, ali se aplikacija razvija v pravo smer. Želeli smo razumeti, kako jo dojemajo uporabniki, kako ocenjujejo njeno uporabnost, intuitivnost in pedagoško vrednost ter katere izboljšave bi lahko povečale njeno učinkovitost in prijaznost do uporabnika.

Testirali smo naslednje funkcionalnosti in scenarije uporabe:

- **Navigacija** – preprosto gibanje med glavnimi zasloni aplikacije (izbira nivoja, vadba).
- **Polizvezda** – osnovna vizualizacija poliritmov (npr. 3:4 poliritem, predstavljen s trikrako in štirikrako zvezdo).
- **Zvočna podpora** – sinhronizirani ritmični vzorci, ki jih uporabnik poskuša reproducirati s tapkanjem.
- **Povratna informacija** – prikaz natančnosti izvedbe v realnem času z barvno kodiranimi indikatorji in odstotnimi vrednostmi nad vsako polizvezdo.
- **Nastavitve vadbe** – izbira težavnostne stopnje (poliritma) in prilagoditev tempa z BPM drsnikom.
- **Rezultati** – prikaz končne ocene in odstotka ujemanja za igran nivo.

Na ta način smo lahko neposredno povezali posamezne funkcionalnosti aplikacije s povratnimi informacijami in metrikami uspešnosti, kar je omogočilo celovito oceno prototipa. Pri evalvaciji smo se osredotočili na naslednje raziskovalno vprašanje: *Ali razviti prototip aplikacije, ki temelji na interaktivni vizualizaciji poliritmov s pomočjo polizvezde, omogoča uporabnikom enostavno učenje, razumevanje in učinkovito vadbo poliritmov, hkrati pa zagotavlja pozitivno uporabniško izkušnjo?*

Dodatni cilji evalvacije so bili:

- Preveriti razumljivost uporabniškega vmesnika brez dodatnih navodil.
- Ovrednotiti natančnost povratne informacije o ritmu.
- Ugotoviti tehnične omejitve na različnih Android napravah, s posebnim poudarkom na odzivnosti uporabniškega vmesnika in interaktivnih gradnikov igre

#### 4.1. Potek evalvacije

Testiranje je bilo prostovoljno, samostojno in brez aktivne internetne povezave. Gradivo za testiranje je bilo posredovano preko socialnih omrežij in elektronske pošte širšemu krogu sodelavcev Radia Študent, glasbenim pedagogom, študentom glasbe, učencem glasbenih šol in ljubiteljem ritmičnih iger. Poštne naslove glasbenih pedagogov smo pridobili iz javno dostopnih spletnih strani posameznih glasbenih šol. Udeležba je bila prostovoljna.

Za zbiranje podatkov smo uporabili vprašalnik **UEQ-S** – Kratka različica standardnega vprašalnika glede uporabniške izkušnje (angl. User Experience Questionnaire – Short). Sestavljen je iz 8 vprašanj, pri katerih uporabniki na sedemstopenjski Likertovi lestvici ocenijo položaj izdelka med paroma nasprotnojučih si pridevnikov (npr. star – nov). Glede na kakovosti, ki jih vprašanja merijo, lahko vprašanja delimo v dve dimenziji: pragmatična kakovost (enostavnost uporabe, učinkovitost, zanesljivost) in hedonična kakovost (stimulativnost, novost). Vprašalnik je preveden v številne jezike, tudi v slovenščino, zato je primeren za hitro oceno uporabniške izkušnje naše aplikacije [15].

Drugi uporabljen vprašalnik je **TAM** – Model sprejetosti tehnologije (angl. Technology Acceptance Model), ki ocenjuje, kako uporabniki sprejemajo novo tehnologijo. V našem primeru smo uporabili prilagojeno različico, ki je vsebovala 11 vprašanj in je zajemala tri ključne dimenzije: zaznana uporabnost – vprašanja, ki ocenjujejo, ali aplikacija pripomore k izboljšanju veščin in učinkovitosti pri glasbeni vadbi; zaznana enostavnost uporabe – vprašanja, ki ocenjujejo enostavnost uporabe in intuitivnost aplikacije; in namerna uporaba – vprašanja, ki ocenjujejo pripravljenost uporabnika za nadaljnjo uporabo in priporočanje aplikacije. Vsako vprašanje je bilo ocenjeno na petstopenjski Likertovi lestvici (*se ne strinjam* - *strinjam se*), kar je omogočilo kvantitativno merjenje sprejemanja aplikacije in napovedovanje, kako bodo uporabniki aplikacijo dejansko uporabljali [16]. Nazadnje smo uporabili vprašanja odprtega tipa - zbrali smo kvalitativne povratne informacije o pričakovanjih, težavah in predlogih.

Potek uporabe aplikacije za udeležence je bil sledeč:

1. Udeleženci so prejeli **Android različico aplikacije in navodila za namestitev in uporabo**.
2. Naloga udeležencev je bila, da aplikacijo **preizkusijo samostojno**, brez dodatne pomoči, in izvedejo nekaj vadbenih sej.
3. Po uporabi so izpolnili spletni vprašalnik (Google Forms), ki je vseboval UEQ-S, TAM in vprašanje odprtega tipa.

### 5. Rezultati

Zbrane kvantitativne podatke smo analizirali statistično, da ocenimo povprečja in razpršenost ocen, medtem ko so bili kvalitativni odgovori vsebinsko kodirani, da izpostavimo ključne izkušnje in predloge uporabnikov. Za analizo UEQ-S smo uporabili uradni analizni obrazec [17], ki omogoča tudi vizualizacijo rezultatov z ustreznimi grafi. Za analizo TAM vprašalnika pa smo razvili lastno Python skripto, s katero smo izračunali povprečne vrednosti, standardne odklone ter izrisali ustrezne grafe.

#### 5.1. Analiza testnih uporabnikov

Evalvacije se je udeležilo devet uporabnikov različnih starostnih skupin in z različnim glasbenim predznanjem, kar je omogočilo raznolike poglede na aplikacijo:

**Table 1**

Starostna struktura testnih uporabnikov.

| Starostna skupina | Število uporabnikov |
|-------------------|---------------------|
| 18–30 let         | 5                   |
| 31–50 let         | 2                   |
| 50+ let           | 2                   |

**Table 2**

Glasbeno predznanje testnih uporabnikov.

| Stopnja glasbenega predznanja | Število uporabnikov |
|-------------------------------|---------------------|
| Brez                          | 1                   |
| Začetnik                      | 5                   |
| Srednji nivo                  | 3                   |

**Table 3**

Vloga testnih uporabnikov.

| Vloga         | Število uporabnikov |
|---------------|---------------------|
| Glasbenik     | 4                   |
| Učenec        | 2                   |
| Občan         | 1                   |
| Tonski tehnik | 1                   |
| Bivši učenec  | 1                   |

**Table 4**

Izkušnje z ritmičnimi aplikacijami in igrami.

| Vprašanje                                      | Da | Ne |
|------------------------------------------------|----|----|
| Si že uporabljal/-a aplikacije za vadbo ritma? | 3  | 6  |
| Si že igral/-a kakšne ritmične igre?           | 3  | 6  |

## 5.2. Rezultati UEQ

Za ocenjevanje uporabniške izkušnje je bil uporabljen *Short User Experience Questionnaire* (UEQ-S), ki meri dve glavni dimenziji uporabniške izkušnje: **pragmatično kakovost** (uporabnost, preglednost, enostavnost uporabe) in **hedonično kakovost** (privlačnost, zanimivost, inovativnost). Lestvica ocen sega od  $-3$  (zelo slabo) do  $+3$  (zelo dobro), pri čemer vrednosti med  $-0.8$  in  $0.8$  predstavljajo nevtralno oceno, vrednosti nad  $0.8$  pozitivno oceno, vrednosti pod  $-0.8$  pa negativno oceno.

Analiza po dimenzijah, kot jo omogoča uradno orodje UEQ, kaže naslednje rezultate:

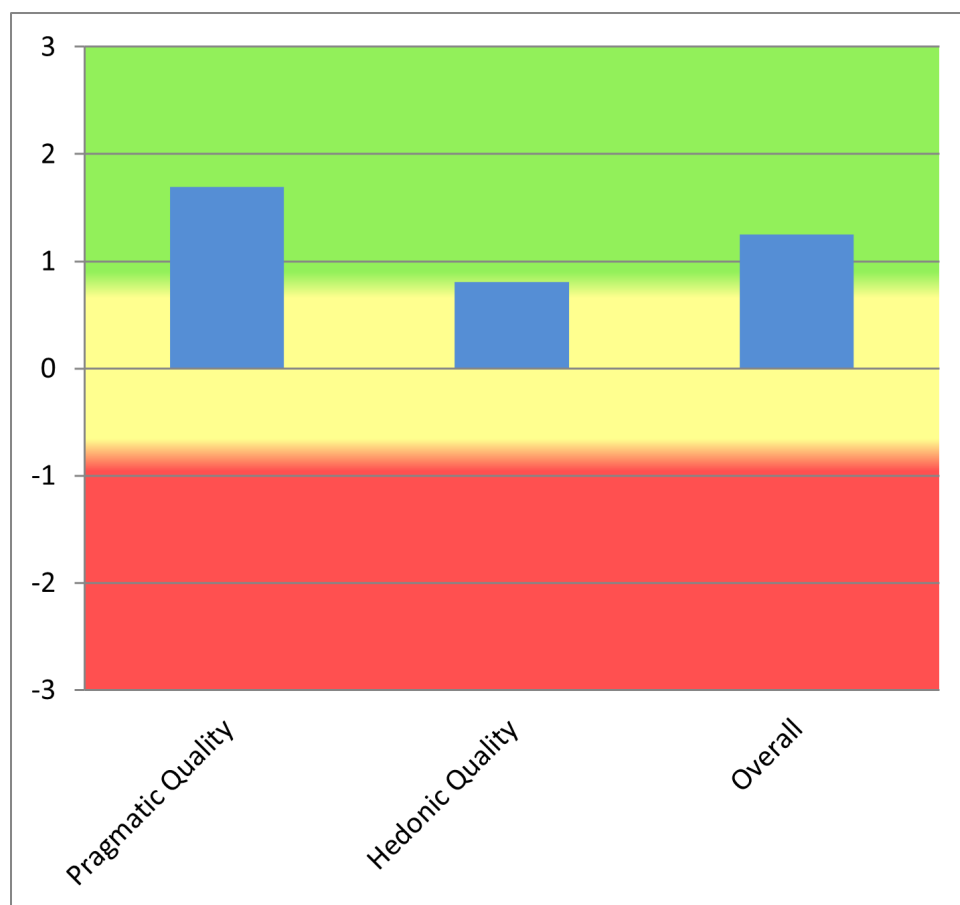
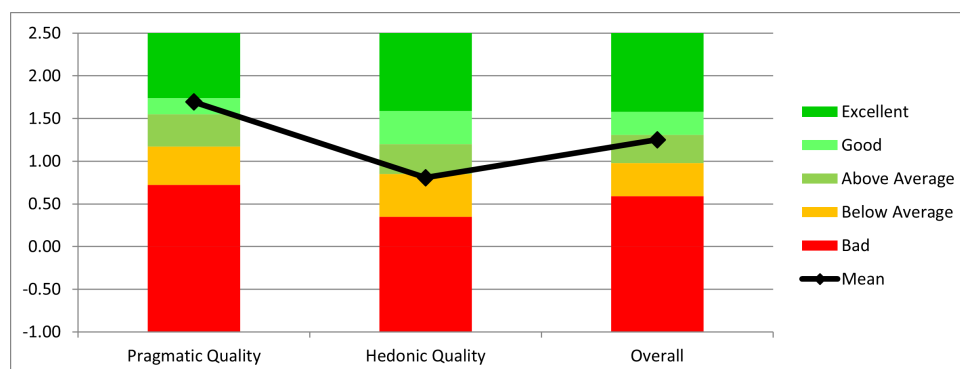
- **Pragmatična kakovost:**  $M = 1.69$  — kaže na visoko pozitivno oceno uporabnosti in preglednosti aplikacije. Vsi pripadajoči elementi so nad pragom  $0.8$ , kar pomeni, da so uporabniki izrazito zadovoljni z osnovnimi funkcionalnimi vidiki.
- **Hedonična kakovost:**  $M = 0.81$  — ocenjena tik nad pragom pozitivne izkušnje, kar pomeni, da aplikacija uporabnikom deluje zanimiva in estetsko privlačna, a z večjo variabilnostjo odgovorov (višje standardne deviacije).
- **Skupna ocena:**  $M = 1.25$  — splošna ocena uporabniške izkušnje je pozitivna, predvsem zaradi visoke pragmatične kakovosti.

Rezultati nakazujejo, da je aplikacija učinkovita in prijazna za uporabo, pri čemer uporabniki prepoznavajo njeno funkcionalno vrednost in preglednost kot največji prednosti. Hedonični vidiki (npr. inovativnost, zabavnost) so sicer pozitivni, a z nekoliko večjo razpršenostjo ocen, kar kaže na možnost nadaljnjih izboljšav vizualne privlačnosti in zabavnosti uporabe.

**Table 5**

Rezultati vprašalnika UEQ.

| Trditev                                     | Povprečje | Std. dev. |
|---------------------------------------------|-----------|-----------|
| Se ne da upravljati – Se z lahkoto upravlja | 1.78      | 1.09      |
| Kompliciran – Enostaven                     | 2.00      | 1.50      |
| Ni učinkovit – Učinkovit                    | 1.11      | 1.05      |
| Ustvarja zmedo – Pregleden                  | 1.89      | 1.27      |
| Dolgočasen – Napet                          | 0.67      | 1.58      |
| Nezanimiv – Zanimiv                         | 1.22      | 1.72      |
| Star – Nov                                  | 0.78      | 1.48      |
| Zastarel – Modern                           | 0.56      | 1.33      |

**Figure 4:** Povprečne vrednosti dimenzij uporabniške izkušnje glede na vprašalnik UEQ-S.**Figure 5:** Primerjava rezultatov aplikacije s primerjalno (angl. benchmark) UEQ bazo podatkov.

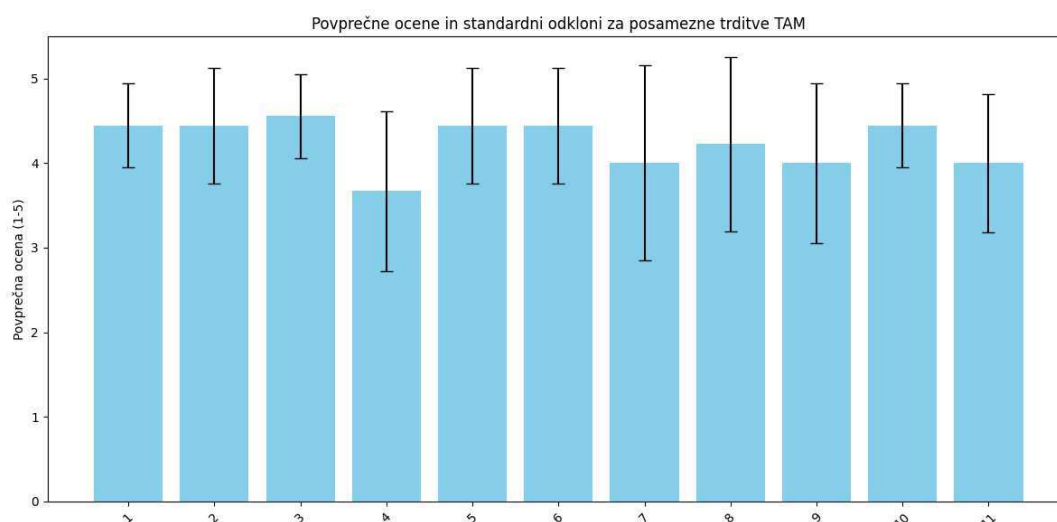
### 5.3. Rezultati TAM

Za oceno sprejemanja aplikacije med uporabniki smo uporabili vprašalnik na podlagi modela Technology Acceptance Model (TAM), ki vključuje 11 trditev o zaznani uporabnosti, enostavnosti uporabe in nameravani uporabi aplikacije. Uporabniki so odgovarjali na lestvici od 1 do 5, kjer je 1 pomenilo močno nestrinjanje, 5 pa močno strinjanje.

**Table 6**

Rezultati vprašalnika TAM: povprečja in standardni odkloni.

| Št. | Trditev                                                           | Povprečje | Std. odklon |
|-----|-------------------------------------------------------------------|-----------|-------------|
| 1   | Aplikacija mi pomaga izboljšati razumevanje ritma.                | 4.1       | 0.96        |
| 2   | Z uporabo aplikacije lahko hitreje napredujem v glasbeni vadbi.   | 3.9       | 1.07        |
| 3   | Aplikacija pripomore k bolj učinkoviti vadbi ritma.               | 4.2       | 0.92        |
| 4   | Aplikacija mi lahko pomaga tudi pri drugih glasbenih dejavnostih. | 3.4       | 1.22        |
| 5   | Aplikacijo je enostavno uporabljati.                              | 4.1       | 1.05        |
| 6   | Navodila in cilji posameznih nalog so bili jasni.                 | 3.9       | 1.11        |
| 7   | Uporaba aplikacije mi ni povzročala težav.                        | 3.9       | 1.26        |
| 8   | Vmesnik je intuitiven in pregleden.                               | 4.0       | 0.95        |
| 9   | V prihodnosti bi želel/-a še naprej uporabljati to aplikacijo.    | 3.8       | 1.24        |
| 10  | Aplikacijo bi priporočil/-a drugim.                               | 4.0       | 1.15        |
| 11  | Aplikacijo bi vključil/-a v svoj proces učenja ali poučevanja.    | 3.8       | 1.21        |



**Figure 6:** Povprečne ocene in standardni odkloni za posamezne trditve vprašalnika TAM.

Rezultati kažejo, da uporabniki aplikacijo na splošno ocenjujejo pozitivno. Najvišje ocene so dobile trditve, ki se nanašajo na izboljšanje razumevanja ritma ( $M=4.1$ ), učinkovito vadbo ( $M=4.2$ ) ter intuitiven in pregleden vmesnik ( $M=4.0$ ). Uporabniki tudi izražajo pripravljenost za nadaljnjo uporabo in priporočilo aplikacije drugim.

Nekoliko nižje povprečje ( $M=3.4$ ) je bilo zabeleženo pri vprašanju o uporabi aplikacije pri drugih glasbenih dejavnostih, kar kaže na prostor za širitev funkcionalnosti ali jasnejšo komunikacijo o možnostih uporabe.

Standardni odkloni so v večini primerov zmerni, kar kaže na relativno enotno mnenje med uporabniki.

V okviru evalvacije smo zbirali tudi odprte komentarje uporabnikov, ki so omogočili bolj poglobljen vpogled v uporabniške izkušnje in možne izboljšave aplikacije. Povzemamo najpogostejše in najbolj relevantne odzive:

Uporabniške povratne informacije so se osredotočile na več vidikov nadgradnje uporabniške izkušnje,

ki jih lahko strnemo v štiri glavne sklope. Prvič, predlagane so bile izboljšave interakcije in natančnosti: povečanje območja zaznave dotika zvezd (tudi do polovice zaslona), dodajanje začetnega odštevanja in možnosti, da se vaja začne šele ob naslednjem ciklu po pritisku gumba, ter vključitev kalibracije ob prvem zagonu zaradi opaženih zamikov med vizualnim, zvočnim in registriranim klikom, zlasti pri višjih BPM in poliritmih. Drugič, uporabniki so izpostavili potrebo po večji zanesljivosti zvočne povratne informacije, vključno z izbiro različnih zvokov (npr. različni bobni), razlikovanjem leve in desne strani (»tik« in »tak«) ter možnostjo predvajanja zgolj osnovnega metronoma. Tretji sklop zadeva estetske in uporabniške prilagoditve, kot so umestitev LCM zvezde v sredino vmesnika, sodobnejši grafični dizajn ter jasna vizualna oznaka začetka vaje, dopolnjena s podrobnimi navodili za začetnike. Četrty sklop zajema razširitev funkcionalnosti in spremljanja napredka, denimo beleženje najboljših rezultatov na posamezni ravni, možnost neprekinjenega ponavljanja vaj ter shranjevanje nastavitve tempa med nivoji. Skupaj ti predlogi ponujajo celovit načrt za tehnično in oblikovno nadgradnjo aplikacije, ki bi izboljšala natančnost, uporabnost in motivacijo uporabnikov.

Ti odprti komentarji kažejo, da kljub pozitivnim ocenam aplikacija ponuja prostor za izboljšave predvsem na področju uporabniške prijaznosti, odzivnosti sistema ter prilagoditev vmesnika za različne profile uporabnikov.

## 6. Diskusija

Zaradi poletnega časa je bil odziv uporabnikov na anketi precej skromen – prejeli smo le devet veljavnih odzivov. Zaradi tega je treba rezultate interpretirati previdno, saj majhno število vzorcev omejuje statistično moč in posplošljivost ugotovitev.

Kljub temu rezultati UEQ-S kažejo, da uporabniki visoko cenijo pragmatične vidike aplikacije, predvsem njeno uporabnost, preglednost in enostavnost uporabe. Hedonična kakovost je ocenjena kot pozitivna, vendar z večjo razpršenostjo, kar nakazuje možnosti za izboljšave v smislu inovativnosti in vizualne privlačnosti. Rezultati TAM podpirajo ugotovitve UEQ, saj uporabniki aplikacijo ocenjujejo kot učinkovito orodje za izboljšanje razumevanja ritma in glasbene vadbe ter kot intuitivno in lahko za uporabo. Pripravljenost za nadaljnjo uporabo in priporočilo aplikacije kaže na dober sprejem med uporabniki. Poleg kvantitativnih odgovorov so odprta vprašanja prinesla dragocene vpogled v konkretne izzive in želje uporabnikov. Ti odgovori so razkrili potrebo po dodatnih pedagoških pojasnilih, večji fleksibilnosti pri vizualizaciji ritmov ter željo po vključitvi avdio povratnih informacij in različnih načinov vadbe. Te ugotovitve so bile ključne pri oblikovanju predlogov za nadaljnje izboljšave, ki so podrobneje predstavljene v naslednjem poglavju.

Med potekom testiranja smo naleteli tudi na tehnične omejitve, ki so vplivale na možnost vključitve širšega nabora uporabnikov. Ena izmed bolj izrazitih je bila povezana z določenimi modeli pametnih telefonov znamke Xiaomi, pri katerih poslovni model močno favorizira lastno distribucijsko platformo in vključuje prikaz oglasov, zaradi česar je namestitev aplikacij iz zunanjih virov otežena. Ta omejitev je v posameznih primerih upočasnila testni proces ter onemogočila sodelovanje nekaterih uporabnikov, kar je dodatno zmanjšalo velikost vzorca.

Ker je vzorec zelo omejen, so te ugotovitve predvsem smernice, ki jih je priporočljivo potrditi z večjim številom uporabnikov v prihodnjih raziskavah. Prav tako bi bilo koristno razširiti analizo s statističnimi metodami, kot so preverjanje zanesljivosti vprašalnikov (npr. Cronbachovo alfa) in povezav med dimenzijami uporabniške izkušnje in sprejemanja.

Pri nadaljnjem razvoju aplikacije je smiselno razmisliti o vključevanju dodatnih funkcionalnosti, kot so razširjeni nabor vaj, naprednejše nastavitve tempa in zvočnih shem ter, z integracijo v platformo Trubadur, predvsem podpora za shranjevanje rezultatov, lestvice igralcev in spremljanje napredka. Kljub temu pa je treba pri tem ohraniti preprostost in minimalistično zasnovo aplikacije, saj preobremenjen uporabniški vmesnik lahko zmanjša preglednost, poveča kognitivno obremenitev uporabnika ter zasenči osnovni namen orodja – intuitivno in učinkovito učenje poliritmov.

## 6.1. Nadaljnji izzivi razvoja

Rezultati evalvacije so pokazali, da ima aplikacija trdne temelje, hkrati pa ponuja več priložnosti za nadaljnji razvoj. Ena izmed pomembnejših ugotovitev je, da pri višjih vrednostih BPM obstoječa metoda ocenjevanja točnosti ne daje enako zanesljivih rezultatov. To kaže na potrebo po podrobni tehnični analizi algoritma za merjenje časovne natančnosti ter njegovi prilagoditvi, da bo ustrezno deloval v celotnem razponu možnih tempov. Pri tem je treba upoštevati tudi naravno manjšo natančnost igralcev pri hitrejših tempih, kar zahteva uravnotežen pristop, ki bo hkrati tehnično natančen in prilagojen realnim zmogljivostim uporabnikov.

V prihodnosti je smiselno oblikovati tudi celovitejši igralni tok, ki uporabnika na intuitiven način vodi skozi posamezen nivo. Ena izmed možnosti je zasnova zaporedja, kjer uporabnik po izbiri nivoja najprej sliši demonstracijo poliritma in si lahko ogleda različne vizualne predstavitve (npr. animiran notni zapis ali interaktivni grafični prikaz). Šele ko se počuti pripravljenega, začne z dejanskim igranjem oziroma ocenjevanjem. Ena izmed potencialnih nadgradenj bi bila tudi možnost, da igralec med izvajanjem (angl. "on the fly") menja poliritme ob ohranjanju osnovnega pulza. Prav tako bi lahko mentorji oblikovali nivoje, pri katerih se poliritmi dinamično spreminjajo med potekom igre, kar bi dodatno razširilo pedagoške in ustvarjalne možnosti uporabe aplikacije. Na tem področju so možnosti skoraj neomejene, saj bi tak pristop omogočal prilagoditev igre različnim učnim slogom in ciljem.

Tak pristop bi omogočal postopno stopnjevanje zahtevnosti: pri uvodnih nivojih bi bil zvok poliritma ves čas prisoten kot referenca, pri zahtevnejših nivojih pa bi bila zvočna podpora poliritma izključena. Igralec bi v teh primerih slišal le osnovni pulz, vizualno pa bi mu bila še vedno na voljo polizvezda kot tiha referenca. Pri tem je mišljeno, da bi moral igralec za popolno osvojitve posameznega nivoja ta nivo premagati dvakrat – najprej z zvočno podporo poliritma kot referenco in nato še zgolj z osnovnim pulzom, pri čemer je poliritmična struktura prisotna le vizualno (npr. z animacijo polizvezde), brez zvočne podpore. Na ta način bi se sistematično krepila sposobnost notranjega zaznavanja ritma in samostojnega izvajanja kompleksnejših ritmičnih struktur.

## 7. Zaključek in nadaljnje delo

V pričujočem članku je bila predstavljena zasnova, razvoj in evalvacija mobilne aplikacije za učenje in vadbo poliritmov, pri čemer je bil poseben poudarek namenjen uporabi polizvezde kot vizualnega pripomočka ter integraciji principov poigritve. Izhodišče za delo je bila prepoznana potreba po orodju, ki bi uporabnikom na intuitiven in dostopen način približalo kompleksne ritmične strukture ter omogočilo njihovo postopno obvladovanje.

Evalvacija aplikacije, izvedena z uporabniškimi testi in standardiziranimi vprašalniki UEQ-S in TAM, je kljub omejenemu številu udeležencev (devet) pokazala pozitiven potencial rešitve. Uporabniki so visoko ocenili enostavnost uporabe, preglednost in uporabnost aplikacije, ob tem pa v odprtih odgovorih nakazali na številne možnosti za izboljšave, ki so skladne z rezultati standardiziranih vprašalnikov. Povratne informacije potrjujejo, da aplikacija že v trenutni obliki podpira glasbeno vadbo, še posebej pri razumevanju in izvajanju poliritmov, a hkrati jasno nakazuje, da gre za prototip z velikim potencialom za nadgradnjo.

Na osnovi ugotovitev so bile oblikovane smernice za nadaljnji razvoj, med katerimi izstopajo: izboljšava algoritma za ocenjevanje natančnosti, uvedba točkovanja glede na kompleksnost poliritma in tempo, oblikovanje dvostopenjskega sistema osvajanja nivojev (z in brez zvočne podpore), možnost menjave poliritmov med igranjem ob ohranitvi osnovnega pulza ter oblikovanje dinamičnih nivojev, ki jih lahko pripravi npr. glasbeni pedagog ali igralec sam. S tem bi bilo mogoče predstaviti še povsem nov koncept – menjave ritmičnega metra oziroma metrske strukture med samim izvajanjem. Tak pristop bi uporabniku omogočil prehajanje med različnimi metri (na primer iz 4/4 v 7/8), kar predstavlja še zahtevnejši izziv od izvajanja poliritmov samih, saj zahteva visoko raven notranje stabilnosti tempa ter sposobnost hitrega prilagajanja novi ritmični logiki.

Treba je poudariti, da je spekter potencialnih izboljšav aplikacije izjemno širok in v praksi skoraj neomejen. Kljub temu razvoj v okviru te raziskave ni stremel k vključevanju celotne magnitude možnih

vsebinskih nadgradenj, temveč k postavitvi trdnih temeljev za stabilno in delujočo igro ter k osebnemu izzivu preizkusiti se v razvoju ritmične igre. Poleg vsebinskih usmeritev so pomembne tudi tehnične izboljšave, kot so optimizacija za različne naprave in dimenzije zaslonov, testiranje na iOS in namiznih računalnikih ter sodelovanje z izkušenimi glasbeniki pri preverjanju uporabnosti in pedagoške vrednosti aplikacije. V tem kontekstu velja izraziti posebno spoštovanje do vseh razvijalcev iger, zlasti glasbenih, saj gre za izjemno zahtevno področje, kjer je tudi z večjimi ekipami in daljšimi roki težko ustvariti izdelek, ki je hkrati tehnično brezhiben, uporabniško prijazen in pedagoško učinkovit.

Čeprav aplikacija še ni popolna, predstavlja trdno osnovo za nadaljnje raziskave in razvoj interaktivnih glasbeno-izobraževalnih orodij. Njena prihodnja integracija v platformo Trubadur bi lahko omogočila širšo dostopnost in dodaten pedagoški doseg, s čimer bi prispevala k sodobnejšemu, interaktivnemu in motivacijskemu poučevanju ritmike. Rezultati tega dela tako niso dokončni zaključki, temveč izhodišče za nadaljnje delo, nadgradnje in preizkušanje novih pristopov, ki bi lahko še bolj približali poliritme širšemu krogu glasbenikov in učiteljev.

## References

- [1] J. Gee, What video games have to teach us about learning and literacy, *Computers in Entertainment* 1 (2003) 20. doi:10.1145/950566.950595.
- [2] N. Whitton, *Digital games and learning: Research and theory*, Taylor & Francis, 2014. doi:10.4324/9780203095935.
- [3] J. Savage, Working towards a theory for music technologies in the classroom: How pupils engage with and organise sounds with new technologies, *British Journal of Music Education* 22 (2005) 167–180. doi:10.1017/S0265051705006133.
- [4] S. Tobias, J. D. Fletcher, Reflections on "a review of trends in serious gaming", *Review of Educational Research* 82 (2012) 233–237. doi:10.3102/0034654312450190.
- [5] J. London, *Hearing in Time: Psychological Aspects of Musical Meter*, Oxford University Press, 2012. URL: <https://doi.org/10.1093/acprof:oso/9780199744374.001.0001>. doi:10.1093/acprof:oso/9780199744374.001.0001.
- [6] G. Toussaint, *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?*, CRC Press, 2013. URL: <https://books.google.si/books?id=ZlKfh83AJEEC>. doi:10.1080/17513472.2014.906116.
- [7] M. Pesek, N. Hirci, K. Žnideršič, et al., Enhancing music rhythmic perception and performance with a vr game, *Virtual Reality* 28 (2024) 118. URL: <https://doi.org/10.1007/s10055-024-01014-y>. doi:10.1007/s10055-024-01014-y.
- [8] E. Orford, K. Bissell, D. Hall, Orff approach, in: *The Canadian Encyclopedia*, Historica Canada, 2022. URL: <https://thecanadianencyclopedia.ca/en/article/orff-approach-emc>, (pridobljeno 20.8.2025).
- [9] M. Pesek, Z. Vučko, P. Šavli, A. Kavčič, M. Marolt, *Troubadour: A gamified e-learning platform for ear training*, *IEEE Access* 8 (2020) 97090–97102. doi:10.1109/ACCESS.2020.2994389.
- [10] *Encyclopædia Britannica*, Polyrhythm, 2023. URL: <https://www.britannica.com/art/polyrhythm>, dostopano: julij 2025.
- [11] J. Pinkl, M. Cohen, Vr drumming pedagogy: Action observation, virtual co-embodiment, and development of drumming "halvatar", *Electronics* 12 (2023). URL: <https://www.mdpi.com/2079-9292/12/17/3708>. doi:<https://doi.org/10.3390/electronics12173708>.
- [12] C. Swafford, *Polyrhythmic Pathways: Using Bimanual Coordination Research to Develop a New Framework for Practice, Performance, and Pedagogy*, Ph.d. dissertation, University of Kentucky, 2023. URL: [https://uknowledge.uky.edu/music\\_etds/220](https://uknowledge.uky.edu/music_etds/220). doi:<https://doi.org/10.13023/etd.2023.176>, theses and Dissertations–Music.
- [13] H. Yokuş, T. Yokuş, Polyrhythmic tapping: Examining the effectiveness of the strategy of organizing rhythmic structures through synthesis, *Educational Sciences: Theory and Practice* 15 (2015) 239–252. URL: <https://files.eric.ed.gov/fulltext/EJ1057480.pdf>. doi:10.12738/estp.2015.1.1917.

- [14] P. Magadini, *Polyrhythms for the drumset*, Belwin, 1995. URL: <https://books.google.si/books?id=iB05h6z144UC>.
- [15] M. Schrepp, A. Hinderks, J. Thomaschewski, Design and evaluation of a short version of the user experience questionnaire (ueq-s), *International Journal of Interactive Multimedia and Artificial Intelligence* 4 (2017) 103. doi:10.9781/ijimai.2017.09.001.
- [16] Q. Ma, L. Liu, *The Technology Acceptance Model*, volume 16, IGI Global Scientific Publishing, 2004, pp. 59–72. doi:10.4018/joeuc.2004010104.
- [17] User experience questionnaire, 2025. URL: <https://www.ueq-online.org/>.

# Comparison of Unity and FMOD Libraries for Spatial Audio Localization in Virtual Reality

Gašper Leskovec<sup>1,\*†</sup>, Eva Gaberšček<sup>1†</sup> and Jaka Sodnik<sup>1†</sup>

<sup>1</sup>University of Ljubljana, Faculty of Electrical Engineering, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia

## Abstract

Spatial audio is central to immersion and situational awareness in virtual reality (VR), but the performance differences between different audio engines have not been sufficiently explored. This study compares Unity's integrated spatial audio system and the FMOD engine in a controlled VR sound localisation task. Thirty volunteers completed three levels of 16 trials each while wearing a head-mounted display (Oculus Quest). They were randomly assigned to one of the two systems and asked to identify the direction of an active loudspeaker among several virtual sources. We measured localisation accuracy, reaction time, spatial distance error and cumulative head rotation; learning effects were assessed by comparing the first and third stages. Participants using FMOD achieved slightly higher accuracy (93% vs 87% overall) and turned their heads less, suggesting that FMOD made it easier for users to localise sounds in space, while Unity users showed greater variability and a slightly greater learning gain. About two-thirds of participants improved their accuracy over time, but there were significant individual differences. These results suggest that both engines perform similarly well on basic localisation tasks, although FMOD has slight advantages in spatial accuracy. The work highlights the importance of evaluating spatial audio engines in user-centred studies and emphasises the need to consider personal factors and task complexity in future research.

## Keywords

Virtual Reality, Spatial Audio, Sound Localization, Unity, FMOD, Immersion, User Experience,

## 1. Introduction

Spatial audio also known as three-dimensional or 3D audio aims to reproduce sound in such a way that it corresponds to the natural listening experience. By reproducing audio data from different directions and distances, the illusion is created that sound sources occupy specific positions in space[1]. Spatial sound allows users to perceive the origin of events without the need for high-fidelity graphics[1], and therefore plays an indispensable role in modern VR applications.

Although many frameworks offer spatial audio features, the underlying rendering algorithms can differ significantly. Unity's native audio engine provides basic spatialisation and distance attenuation, while FMOD is a third-party library that provides advanced features such as head-related transfer functions (HRTFs) and occlusion modelling. However, humans are inaccurate at localising sound sources, especially with multiple sources, due to limitations in human auditory perception, such as angle discrimination errors and front-back confusion [2]. When multiple sources compete with each other localisation blur is increased [3]. The selection of a suitable spatial audio engine can therefore influence the performance and comfort of the user.

This study addresses three research questions:

1. Does localization accuracy differ between FMOD and Unity engine?
2. Do users improve over repeated exposures (learning effect), and does this depend on the engine?
3. How large are individual differences (including outliers) in localization performance?

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

†These authors contributed equally.

✉ leskovec@gmail.com (G. Leskovec); eg0458@student.uni-lj.si (E. Gaberšček); Jaka.Sodnik@fe.uni-lj.si (J. Sodnik)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.8>

## 2. Related work

VR technology is a relatively new field of computer science, and therefore there is not yet much work dealing with the specifics and effectiveness of spatial sound in virtual reality.

In 1994, Begault et al. [4] first published a book entitled "3-D Sound for Virtual Reality and Multimedia", which covers the basic principles of acoustics in 3-D space and how they relate to virtual reality systems. It was, and in some ways still is, a very influential work in this field.

The 2005 article "Improving spatial perception through sound field simulation in VR" by Faria et al. [5] deals with audio immersion in virtual environments, especially CAVE systems, and proposes a rating scale to evaluate the quality of spatial sound immersion in VR configurations.

In "Spatial Sound for Computer Games and Virtual Reality" by Murphy et al. [6] from 2011, the authors also investigate how spatial audio improves immersion in virtual environments and games. They examine technologies ranging from simple stereo to advanced Head-Related Transfer Functions (HRTFs), emphasising how sound design enriches the user experience and creates more believable and engaging virtual scenes through sophisticated sound techniques. The authors also point out hardware limitations and future trends in audio implementation.

A more recent study by Fialho et al. from 2021, "Soundscape VR: Spatial Navigation Using Sound in Virtual Reality" [7], takes a slightly different approach. It investigated how people navigate in virtual environments using only auditory cues. In the study, blindfolded participants completed navigation tasks by localising sound sources and avoiding obstacles, with the level of difficulty adjusted to the complexity of the route. The results showed that spatial learning improved over time and was closely linked to cognitive performance.

## 3. Methodology

### 3.1. Participants

Thirty volunteers participated (the sample consisted of 23 males, 5 females and 2 other participants. Most participants were between 20 and 25 years old (73.3%), with an overall age range from under 20 to 40 years). All participants reported normal or corrected-to-normal vision and hearing. Prior experience with VR varied: 13 had tried VR once or twice, 6 had never used it, 6 had used it several times and 5 reported frequent use. In terms of spatial audio, 13 participants had tried spatial headphones once or twice, 9 had used them frequently and 8 had never used them. Participants were randomly assigned to one of two groups: Unity Audio ( $n = 15$ ) or FMOD ( $n = 15$ ). Participants also provided informed consent.

### 3.2. Experimental procedure

The spatial audio testing programme was developed in Unity 2021.3.16f1. Participants wore a head-mounted display (Oculus Quest) with integrated headphones and used a controller to select the sound sources. The application allowed either the standard spatialiser from Unity or FMOD Studio. The experiment was conducted in a closed room with only the experimenter and the individual participant present. The room was silent to minimise external interference.

The participants were in a virtual room with loudspeakers placed in a circle around them as the distance of 5 meters. The virtual loudspeakers presented in the scene served only as visual representations of sound-source locations; the actual auditory stimuli were rendered as idealized point sources without physical dimensions. During each test, a continuous tone was emitted from a loudspeaker; the participants selected the perceived source using a laser pointer via a controller. The experiment comprised three stages:

1. **Level 1** – 8 loudspeakers evenly spaced on a horizontal circle (45° intervals).
2. **Level 2** – 16 loudspeakers in two vertically separated circles of 8 with an offset by 22.5°.
3. **Level 3** – 8 loudspeakers identical to Level 1 to assess learning effects.

Each level consisted of 16 trials. The same visual scene and task instructions were used for both libraries.

Participants did not undergo any prior preparation or training session before the experiment. This approach was chosen intentionally to observe their initial, untrained responses to the stimuli and to evaluate potential learning effects or performance changes that occurred through repeated exposure during the task.

### 3.3. Dependent variables

**Accuracy.** For each participant  $i$  and level  $l$  we computed the proportion of correctly identified sources:

$$\text{Accuracy}_i^{(l)} = \frac{n_{\text{correct},i}^{(l)}}{n_{\text{trials},i}^{(l)}}, \quad (1)$$

where  $n_{\text{trials},i}^{(l)} = 16$  trials per level.

**Time.** The average completion time per trial was defined as

$$\text{Time}_i^{(l)} = \frac{1}{n_{\text{trials},i}^{(l)}} \sum_{j=1}^{n_{\text{trials},i}^{(l)}} \text{Time}_{ij}^{(l)}. \quad (2)$$

**Spatial distance error.** For each trial  $j$ , the spatial distance error was computed as the Euclidean distance between the chosen and the actual loudspeaker positions:

$$\text{Distance}_{ij} = \|\mathbf{r}_{\text{selected},ij} - \mathbf{r}_{\text{true},ij}\|_2. \quad (3)$$

The average distance error per level was

$$\overline{\text{Distance}}_i^{(l)} = \frac{1}{n_{\text{trials},i}^{(l)}} \sum_{j=1}^{n_{\text{trials},i}^{(l)}} \text{Distance}_{ij}. \quad (4)$$

**Head rotation.**  $\psi_{ijk}$  was the yaw angle at the  $k$ -th sample in trial  $j$ ,  $k = 1, \dots, T_{ij}$ . The cumulative yaw rotation within trial  $j$  was

$$\text{HeadRotationSum}_{ij} = \sum_{k=2}^{T_{ij}} |\psi_{ijk} - \psi_{ij(k-1)}|. \quad (5)$$

The overall head-rotation metric for a participant was the mean across rounds:

$$\text{HeadYawTotalChangeMean}_i = \frac{1}{n_{\text{trials},i}} \sum_{j=1}^{n_{\text{trials},i}} \text{HeadRotationSum}_{ij}. \quad (6)$$

**Learning effect (Level 3 – Level 1).** To capture improvement from the first to the third level we defined:

$$\text{LearnAccDiff}_i = \text{Accuracy}_i^{(3)} - \text{Accuracy}_i^{(1)}, \quad (7)$$

$$\text{LearnTimeDiff}_i = \text{Time}_i^{(3)} - \text{Time}_i^{(1)}, \quad (8)$$

$$\text{LearnDistDiff}_i = \overline{\text{Distance}}_i^{(3)} - \overline{\text{Distance}}_i^{(1)}, \quad (9)$$

$$\text{LearnYawDiff}_i = \text{HeadYawTotalChangeMean}_i^{(3)} - \text{HeadYawTotalChangeMean}_i^{(1)}. \quad (10)$$

A positive  $\text{LearnAccDiff}$  indicates improved accuracy; a negative  $\text{LearnTimeDiff}$  indicates faster completion; a negative  $\text{LearnDistDiff}$  indicates a smaller *spatial distance error* (m); and a negative  $\text{LearnYawDiff}$  indicates reduced head rotation ( $^\circ$ ).

**Table 1**

Average localization accuracy by engine and level.

| Library | Level 1 [%] | Level 2 [%] | Level 3 [%] | All Levels [%] |
|---------|-------------|-------------|-------------|----------------|
| FMOD    | 0.93        | 0.90        | 0.97        | 0.93           |
| Unity   | 0.89        | 0.77        | 0.94        | 0.87           |

**Per-subject aggregation and reporting.** All measures are computed per participant and per level ( $l \in \{1, 2, 3\}$ ) from raw trial logs.

## 4. Results and Analysis

In this section, the results are presented in five steps. First, we describe the statistical procedure used to test the differences between the engines (Section 4.1). We then compare the localisation accuracy between the different engines and task levels (Section 4.2). We then quantify the learning effects between Level 1 and Level 3 for accuracy and time (Section 4.3). We extend the analysis to continuous measures- spatial distance error and cumulative head rotation (Section 4.4). Finally, we analyse individual differences and outliers (Section 4.5).

### 4.1. Statistical analysis

The assumption of normality was evaluated separately for Unity and FMOD using the Shapiro–Wilk test. Between-engine comparisons were conducted with Welch’s two-sample  $t$ -test when both groups met the normality assumption; otherwise, the Mann–Whitney  $U$  test was employed. All tests were two-tailed with a significance threshold of  $\alpha = 0.05$ . To adjust for multiple between-engine comparisons, the Benjamini–Hochberg false discovery rate (FDR) procedure was applied, and adjusted  $p$ -values are reported.

### 4.2. Comparison of Localization Accuracy Between the Libraries

One of the main objectives of the study was to compare the accuracy of sound localisation between two spatial audio libraries, FMOD and Unity.

For each participant, we calculated the average localisation accuracy across the three levels of the task. Accuracy was defined as the proportion of correctly identified sound source locations at a given level. Based on these values, we compared the average results for each library separately.

As shown in Table 1, participants achieved higher average localization accuracy with FMOD compared to Unity. This difference was observed across all task levels, suggesting greater effectiveness of FMOD in spatial sound reproduction. These findings are consistent with prior reports that advanced spatialization techniques (e.g., HRTFs and occlusion) enhance perceived realism [1].

The assumption of normality, tested with the Shapiro–Wilk test, was not satisfied; therefore, two-sided Mann–Whitney  $U$  tests were used to compare engines ( $n = 15$  per group). For localization accuracy, the false discovery rate (FDR)-adjusted  $p$ -values were: Level 1  $p = 0.148$  (not significant), Level 2  $p = 0.011$  (significant at  $\alpha = 0.05$ ), Level 3  $p = 0.627$  (not significant). This pattern suggests that FMOD provided a measurable advantage at intermediate task difficulty, whereas no significant differences emerged at the easiest or hardest levels.

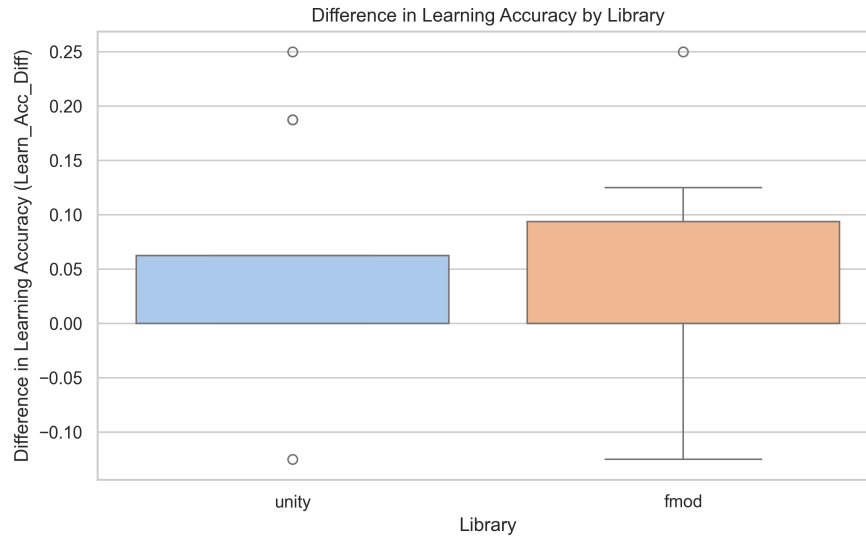
### 4.3. Learning Effect Across Task Levels

The second research question was whether users improve with repetition of the task - both in terms of accuracy and speed of localisation. We analysed this by measuring the difference between the 1st and 3rd repetition for each test case.

**Table 2**

Learning effect: mean change (Level 3 – Level 1) in accuracy and completion time by engine. Negative  $\Delta$ Time indicates faster completion.

| Library | Avg. Accuracy Change [%] | Avg. Time Change [s] |
|---------|--------------------------|----------------------|
| FMOD    | 0.03                     | -1.34                |
| Unity   | 0.05                     | -1.05                |



**Figure 1:** Distribution of learning accuracy difference  $\Delta Acc = Acc_3 - Acc_1$  by engine. Boxes represent interquartile ranges; whiskers extend to  $1.5 \times IQR$  and points denote outliers.

For accuracy, we defined the variable *LearnAccDiff*, which represents the difference in accuracy between the 3rd and 1st levels. Similarly, we defined *LearnTimeDiff*, which indicates the change in completion time. A positive value means higher accuracy or slower completion, while a negative value means faster completion or lower accuracy.

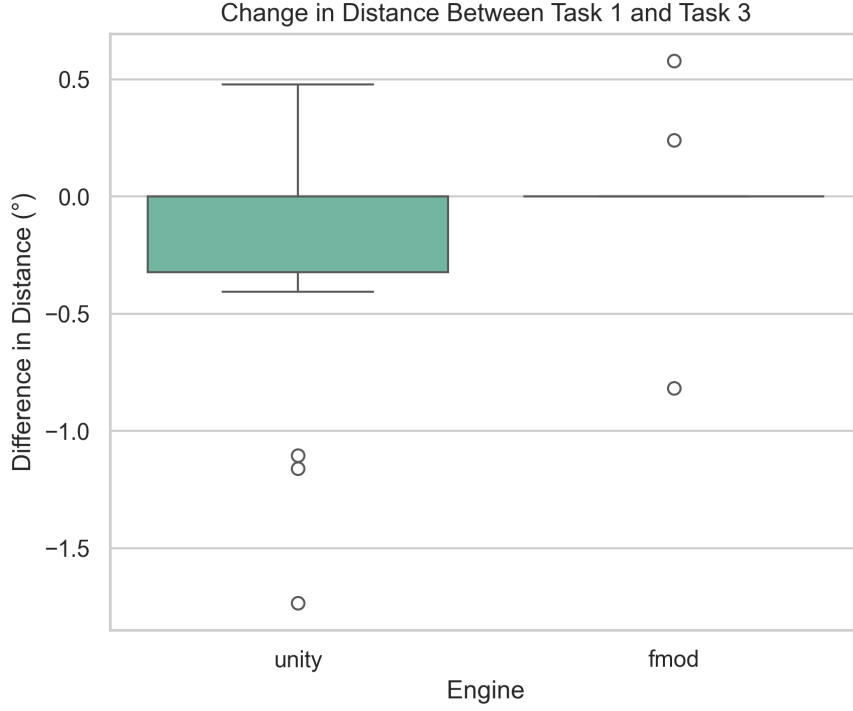
As shown in Table 2, accuracy improved slightly for most participants, while the learning effect was more pronounced for task speed. On average, users in both libraries needed less time for the third stage, which can be attributed to familiarisation with the task and better orientation in the virtual environment. The time taken to complete the tasks was reduced by around 1–1.5 seconds, which can be attributed to a learning effect.

The learning trend is evident for the majority of participants, although some individuals displayed distinctly negative or positive differences, which may indicate fatigue, exceptional adaptation, or a mismatch between the task and expectations. These individuals are shown as outliers in the boxplot in Figure 1, as their changes in accuracy exceed the typical range (defined by the interquartile range). Figure 1 illustrates the distribution of Acc values for both engines. Most participants show small positive changes, indicating slight improvements, but Unity exhibits more variability and a few outliers. Such deviations may reflect various factors such as comprehension, attention, or motivation.

To examine potential differences in accuracy improvement between the Unity and FMOD libraries, we analyzed the values of the variable *LearnAccDiff*, which represents the change in accuracy between the first and third tasks.

Normality (Shapiro–Wilk) was violated for accuracy and distance learning effects, so we used a non-parametric test for those: *LearnAccDiff* (Unity  $p = 0.047$ , FMOD  $p = 0.036$ ) and *LearnDistDiff* (Unity  $p = 0.0019$ , FMOD  $p < 0.001$ )  $\Rightarrow$  Mann–Whitney  $U$ . For *LearnTimeDiff*, normality held (Unity  $p = 0.347$ , FMOD  $p = 0.597$ )  $\Rightarrow$  Welch’s  $t$ -test.

Between-engine comparisons showed no significant differences for accuracy improvement *Lear-*



**Figure 2:** Boxplot of changes in spatial distance error between the first and third task, separated by audio library (Unity vs. FMOD). Negative values indicate improved localization accuracy. Unity participants show greater variance and more outliers, whereas FMOD results remain closer to zero, suggesting more consistent performance across users.

$nAccDiff$ :  $p = 0.416$  (ns), or task time improvement  $LearnTimeDiff$ :  $p = 0.740$  (ns). For distance improvement, the result approached but did not reach statistical significance  $LearnDistDiff$ :  $p = 0.054$ , suggesting a possible trend toward greater learning effects with FMOD in one engine, although this cannot be confirmed at the conventional  $\alpha = 0.05$  level. Overall, these results indicate that while participants exhibited clear within-engine learning effects, the magnitude of improvement did not differ reliably between Unity and FMOD.

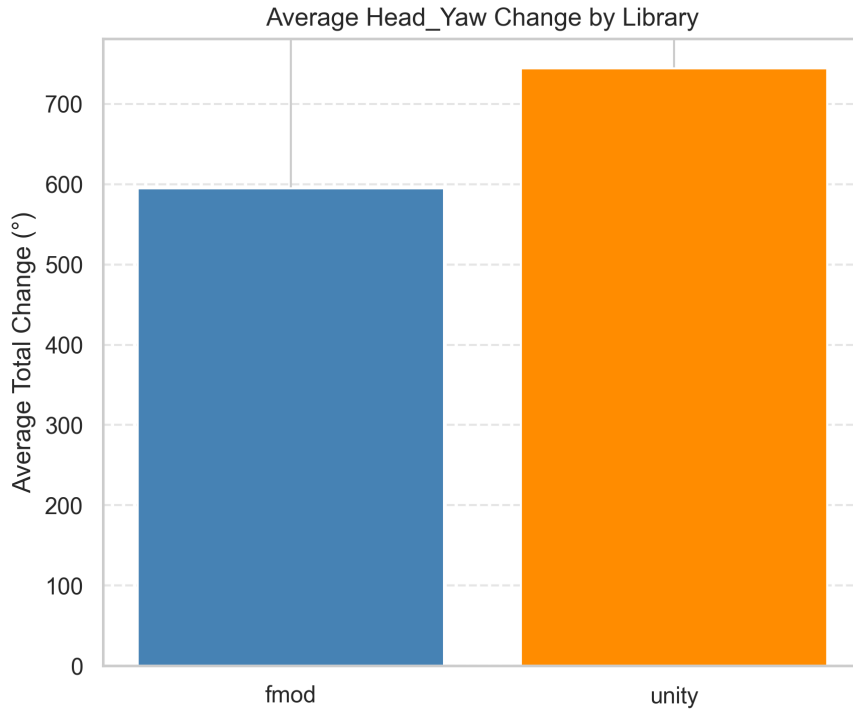
#### 4.4. Analysis of Spatial distance error and Head Rotation

To gain a more detailed understanding of user behavior and localization accuracy, we analyzed two additional variables:

- **Spatial distance error** expressed in meters, representing the difference between the perceived and actual loudspeaker location.
- **Cumulative head rotation** expressed in degrees, measuring the total yaw angle a participant rotated their head while searching for the sound source in each trial. Larger values may indicate higher uncertainty or difficulty in localization.

Spatial distance error provides a continuous estimate of localization precision. We computed the change  $\Delta Dist = Dist_3 - Dist_1$ . On average, Unity users showed a modest reduction in distance error (mean  $-0.29 \pm 0.58$  m;  $n = 15$ ), whereas FMOD users showed essentially no change (mean  $0.00 \pm 0.28$  m;  $n = 15$ ). Negative values indicate improved localization. Figure 2 illustrates the distribution of  $\Delta Dist$ .

Cumulative head rotation Yaw further highlights differences in user strategies. FMOD participants turned their heads less (mean  $595^\circ$ ) than Unity participants (mean  $717^\circ$ ), suggesting that FMOD provided more intuitive spatial cues Figure 3 displays average yaw angles. These trends align with the hypothesis that advanced spatial rendering aids orientation and reduces exertion [1].



**Figure 3:** Average sum of head rotation angles (yaw) by engine. FMOD users rotated their heads less on average, suggesting more intuitive spatial cues.

Overall, FMOD users tended to exhibit slightly smaller spatial distance errors and required less head rotation, pointing toward more efficient and intuitive localization. These trends support the hypothesis that advanced spatial rendering improves orientation and reduces effort.

For the change in spatial distance error ( $\Delta\text{Dist} = \text{Dist}_3 - \text{Dist}_1$ ), normality was violated in both groups (Shapiro–Wilk: Unity  $p = 0.0019$ , FMOD  $p < 0.0001$ ), so we used a two-sided Mann–Whitney  $U$  test. The between-engine difference indicated a non-significant trend in favour of Unity (Unity mean  $-0.294 \pm 0.585$  m, FMOD mean  $0.000 \pm 0.275$  m);  $U = 73.0$ ,  $p = 0.054$ , not significant.

For cumulative head rotation (yaw, degrees), both groups were approximately normal (Shapiro–Wilk: Unity  $p = 0.126$ , FMOD  $p = 0.372$ ), so we applied Welch’s two-sample  $t$ -test. FMOD showed lower rotation on average (FMOD  $595^\circ$  vs. Unity  $717^\circ$ ), but the difference was not significant ( $t = 1.63$ ,  $p = 0.114$ ). This suggests a potential tendency for FMOD to facilitate more efficient orientation, although the evidence is insufficient to confirm a reliable effect.

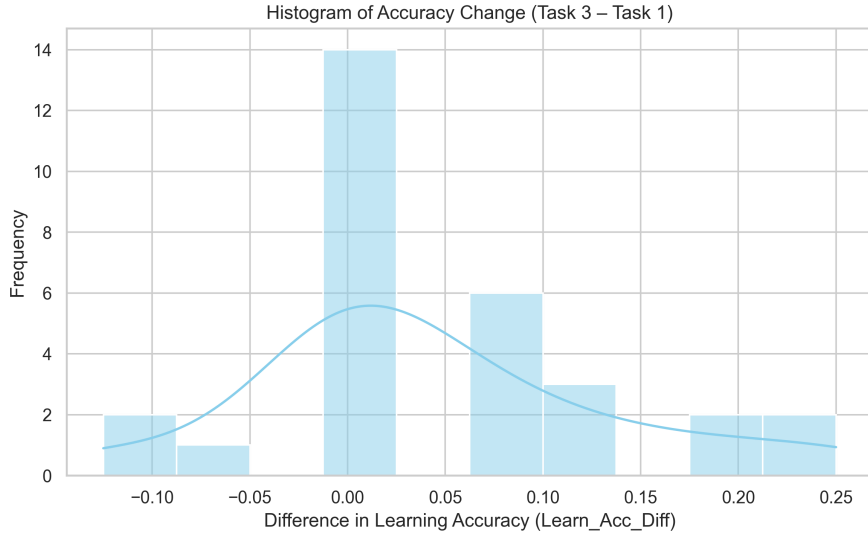
#### 4.5. Individual Differences and Outlier Analysis

To gain a better insight into individual differences, we analyzed the distribution of the variable *LearnAccDiff*, which represents the change in sound localization accuracy between the first and third tasks. The histogram in Figure 4 shows how the results varied depending on the participants’ initial performance.

Most participants demonstrated small to moderate improvements in localization accuracy, confirming the observed learning effect. The majority of participants had differences in the range of 0.00 to 0.05, indicating slight improvement.

A smaller proportion of participants showed more substantial improvement (above 0.15), while some even performed worse (negative values). These outlier cases may reflect various influences such as fatigue, lack of attention, or random errors during task execution.

The presence of a positive shift in most cases suggests that participants were able to adapt to the virtual environment and improve their orientation and perception of spatial audio signals. These patterns suggest that cognitive factors such as spatial abilities, attention and fatigue strongly influence



**Figure 4:** Histogram of changes in localization accuracy between the first and third task ( $\Delta Acc = Acc_{Task3} - Acc_{Task1}$ ). Most participants show small positive differences around zero, indicating slight improvement, while a few outliers demonstrate larger gains or declines in performance. The distribution suggests a general learning effect with variability across individuals.

localization performance. Similar variability has been noted in previous work on spatial awareness tasks [2].

Out of 30 participants, 19 (63%) improved their accuracy between the first and third task (positive *LearnAccDiff* values), 4 (13%) showed no change (difference equal to 0), while 7 (23%) exhibited a decline in accuracy. This further confirms the prevailing positive learning effect, despite some individual negative deviations.

## 5. Discussion

Our results suggest that FMOD offers a slight advantage over Unity in localisation accuracy and head movement, while Unity users showed more pronounced learning effects in accuracy. These results confirm previous observations that advanced spatial rendering can improve realism and reduce localisation blur [3]. Nonetheless, FMOD’s reduced demands on head rotation could lead to less fatigue during prolonged VR experiences.

The third research question concerned the distribution of learning effects. Approximately a quarter of participants deteriorated over time, while most improved modestly. This heterogeneity emphasises the need to take individual differences into account when evaluating spatial audio systems. Future research could include the collection of additional data on cognitive abilities and the inclusion of more complex scenarios with multiple or moving sound sources, as human localisation errors become more pronounced under these conditions [2].

While this study primarily relied on quantitative measures to ensure consistency and comparability across participants, we acknowledge that incorporating complementary qualitative observations could provide additional context for interpreting individual differences and outlier behaviour.

In the present study, the spatial distance error – expressed as the Euclidean distance between the selected and actual virtual loudspeaker – was considered the most appropriate and interpretable metric for this type of categorical localisation task, where participants selected among predefined sound-source positions rather than estimating a continuous direction. Nonetheless, angular error measures may be more informative in future experiments involving continuous or directional localisation responses.

## 6. Conclusion

We presented a controlled comparison of two popular spatial audio engines, Unity and FMOD, in a VR sound localisation task. Using four dependent variables: localisation accuracy, completion time, spatial distance error and head rotation- we evaluated performance on three levels and found learning effects. FMOD generally resulted in higher accuracy and required less head movement. Participants improved modestly with repeated exposure, and individual variability was substantial. We recommend that VR developers consider both objective performance and user comfort when selecting spatial audio libraries.

## Acknowledgments

We thank Miha Malenšek for his help in developing the Unity app, and we are grateful to the volunteer participants for their time. This work was supported by the Research Agency programme ICT4QoL – Information and Communications Technologies for Quality of Life (P2-0246).

## Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (GPT-5, OpenAI) to support translation (Slovenian to English), grammar and spelling checks, and sentence rephrasing. After using this tool, the authors reviewed and edited the content as needed, and take full responsibility for the publication's content.

## References

- [1] What is spatial audio?, Interaction Design Foundation, 2025. URL: <https://interaction-design.org/literature/topics/spatial-audio?sr=Istid=AfmBoRGKxyHuBhyWs'TIWTGy7TDdVCIN8-6wRvqz9AAEqTq5ScnQ>, accessed: 22 Aug. 2025.
- [2] H. Cho, A. Wang, D. Kartik, E. L. Xie, Y. Yan, D. Lindlbauer, Aptomize: Optimal placement of spatial audio cues for extended reality, in: Proceedings of the 37th ACM Symposium on User Interface Software and Technology (UIST '24), ACM, 2024. doi:10.1145/3654777.3676424, the authors propose relocating spatial audio sources to minimize localization confusion due to angular discrimination and front-back errors.
- [3] M. Ramírez, Toward sound localization testing in virtual reality to aid in the screening of auditory processing disorders, Frontiers in Hearing (or similar journal name if known) (2024). Reports that virtualization (e.g. headphone-based VR) increases localization blur compared to loudspeaker setups.
- [4] D. R. Begault, L. J. Trejo, 3-D sound for virtual reality and multimedia, Technical Report, 2000.
- [5] R. R. A. Faria, M. K. Zuffo, J. A. Zuffo, Improving spatial perception through sound field simulation in vr, in: IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2005., IEEE, 2005, pp. 6–pp.
- [6] D. Murphy, F. Neff, Spatial sound for computer games and virtual reality, in: Game sound technology and player interaction: Concepts and developments, IGI Global Scientific Publishing, 2011, pp. 287–312.
- [7] L. Fialho, J. Oliveira, A. Filipe, F. Luz, Soundspace vr: spatial navigation using sound in virtual reality, Virtual Reality 27 (2023) 397–405.

# Web Implementation of 3-way Chess

Janez Koprivec, Matija Marolt and Ciril Bohak\*

*Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia*

## Abstract

This paper presents the design and implementation of a web-based platform for three-player chess, a non-standard variant developed by Dario Varga that until now has only existed in physical form. The work addresses several human-computer interaction challenges: intuitive visualization of a hexagonal chessboard, synchronization of three players in real time, and support for learning unfamiliar rules. The platform was developed using Node.js, React, Mantine, Socket.IO, and MongoDB, providing user account management, both single-device and online multiplayer modes, and access to game history. A key contribution lies in creating a user-friendly and responsive interface that lowers the entry barrier for new players while ensuring robust performance in real-time gameplay. Evaluation confirmed that the application meets functional requirements and offers minimal latency for online play, with results highlighting the impact of server proximity on player experience. The project demonstrates how thoughtful interaction design and modern web technologies can support the digital transformation of analog games, broadening accessibility and engagement with alternative chess variants. The system is available at: <https://chess3.musiclab.si/>.

## Keywords

human-computer interaction, chess, online games,

## 1. Introduction

Chess is one of the most enduringly popular strategy games worldwide, with millions of players engaging through physical boards and digital platforms. Over the past two decades, online platforms such as Chess.com<sup>1</sup> and Lichess.org<sup>2</sup> have made chess more accessible than ever, blending competitive play with learning tools and community features. Alongside standard chess, a wide variety of variants have been developed to expand gameplay and introduce new challenges. Such variants<sup>3</sup> include multiple types of games on a similar  $8 \times 8$  two-color chessboard, while others also introduce new chessboard designs. Among the latter is also Varga's 3-way chess, an alternative version designed by Slovenian actor Dario Varga in the late 20th century. Played on a hexagonal board with 96 cells, it introduces a third color, a third set of pieces, and new movement rules, offering a fundamentally different dynamic of alliances, competition, and strategy.

Despite its unique appeal, Varga's 3-way chess has remained largely confined to physical boards. More about the Varga's version of the chess is presented in [1]. This limitation restricts its reach: learning the rules is difficult without direct guidance, and physical play requires access to a specially designed board and pieces. As a result, the game has never achieved broad popularity. A digital implementation could overcome these barriers by enabling easier rule learning, remote multiplayer play, and global accessibility.

Developing such a platform, however, introduces significant Human-Computer Interaction (HCI) challenges. First, the non-standard chess logic and rule set must be represented in a way that players can easily understand, even if they are unfamiliar with the variant. Second, the visualization of a non-traditional hexagonal chessboard demands careful design to ensure clarity and usability across devices.

---

*HCI SI 2025: Human-Computer Interaction Slovenia 2024, October 13th, 2025, Koper, Slovenia*

\*Corresponding author.

✉ [jk7870@student.uni-lj.si](mailto:jk7870@student.uni-lj.si) (J. Koprivec); [matija.marolt@fri.uni-lj.si](mailto:matija.marolt@fri.uni-lj.si) (M. Marolt); [ciril.bohak@fri.uni-lj.si](mailto:ciril.bohak@fri.uni-lj.si) (C. Bohak)

🌐 <https://lgm.fri.uni-lj.si/matic> (M. Marolt); <https://lgm.fri.uni-lj.si/ciril> (C. Bohak)

🆔 0000-0002-0619-8789 (M. Marolt); 0000-0002-9015-2897 (C. Bohak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

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

<sup>2</sup><https://lichess.org>

<sup>3</sup>[https://en.wikipedia.org/wiki/List\\_of\\_chess\\_variants](https://en.wikipedia.org/wiki/List_of_chess_variants)

<https://doi.org/10.26493/978-961-293-559-7.9>

Third, the platform must synchronize three players in real time, managing fairness and responsiveness while preserving an engaging user experience.

In this paper, we present the design and implementation of a web-based platform for three-player chess that addresses these challenges. The system was developed using modern web technologies and was evaluated with respect to usability, responsiveness, and functional completeness. Beyond its technical contributions, the platform demonstrates how digital interaction design can preserve and popularize niche analog games, broadening access to novel cultural and recreational practices.

## 2. Related Work

Research on digital chess platforms spans education, usability, and multiplayer interaction design. Early work by Picussa et al. [2] explored online educational environments for chess, showing how digital platforms can support teaching and learning by providing real-time feedback and interactive explanations. Similarly, Guid et al. [3] demonstrated the potential of intelligent tutoring systems in chess endgames, focusing on adaptive explanations and player progress tracking. These contributions highlight the importance of user-centered approaches when digitizing traditional board games.

Several studies have investigated technical and interaction challenges of web-based chess implementations. Palomäki [4] examined cross-device human-versus-human play in browser-based chess, emphasizing synchronization and latency issues. Hostettler and Boner [5] focused on digitizing chess scorecards, underlining the role of digital tools in preserving and extending traditional practices. More recently, Vasiljević [6] proposed adapting AlphaZero for three-player hexagonal chess, addressing the computational challenges posed by non-standard rules and board geometries.

Mainstream online chess platforms provide important benchmarks for design and usability. Chess.com is the largest commercial chess service, with over 100 million registered users<sup>4</sup> and was the most downloaded iOS game in 28 countries in 2023<sup>5</sup>. Its infrastructure leverages modern distributed technologies<sup>6,7</sup>. In contrast, Lichess.org is fully open-source<sup>8</sup>, free of ads, and community-driven<sup>9</sup>, integrating Stockfish<sup>10</sup> through WebAssembly<sup>11</sup>. These platforms exemplify best practices in real-time communication, user experience, and scalability, offering design principles relevant to emerging variants.

Three-player chess has been studied less extensively. Historical overviews document numerous hexagonal and triangular board variants<sup>12</sup>, yet few digital implementations exist. One notable example is a community-driven platform of a different game variant that allows browser-based play with friends<sup>13</sup>. However, such implementations typically lack matchmaking, persistent accounts, or large-scale adoption, illustrating the gap between experimental variants and mainstream digital platforms.

## 3. Background

Varga’s 3-way chess is a non-standard variant of chess. Unlike traditional chess, which is played between two players on an  $8 \times 8$  square grid, this variant is played by three participants on a hexagonal board consisting of 96 fields and is illustrated in fig. 1. Each player controls a standard set of chess pieces—pawns, rooks, knights, bishops, a queen, and a king—positioned along one of the three sides of the hexagon. The rules extend classical chess while introducing new dynamics of competition and alliance.

---

<sup>4</sup><https://en.wikipedia.org/wiki/Chess.com>

<sup>5</sup><https://www.chess.com/news/view/chesscom-number-one-in-ios-app-store>

<sup>6</sup><https://www.redpanda.com/case-study/chess-com>

<sup>7</sup><https://himalayas.app/companies/chess-com/tech-stack>

<sup>8</sup><https://github.com/lichess-org/lila>

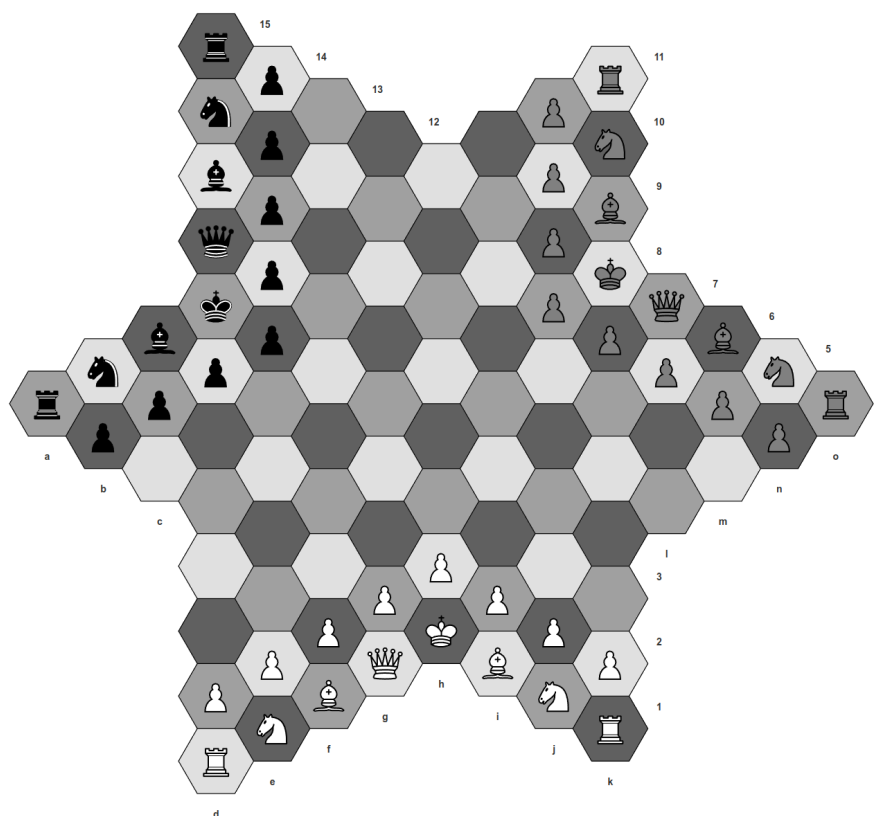
<sup>9</sup><https://lichess.org/about>

<sup>10</sup><https://stockfishchess.org>

<sup>11</sup><https://webassembly.org>

<sup>12</sup>[https://en.wikipedia.org/wiki/Three-player\\_chess](https://en.wikipedia.org/wiki/Three-player_chess)

<sup>13</sup>[https://www.reddit.com/r/chessvariants/comments/116dnuu/i\\_made\\_a\\_website\\_to\\_play\\_three\\_player\\_chess\\_with/](https://www.reddit.com/r/chessvariants/comments/116dnuu/i_made_a_website_to_play_three_player_chess_with/)



**Figure 1:** Varga's 3-way chess board has 96 fields divided between three colors, and 16 chess pieces per player also in three colors.

### 3.1. Movement Rules

Most pieces retain their standard movement patterns, but these are adapted to the geometry of the hexagonal board:

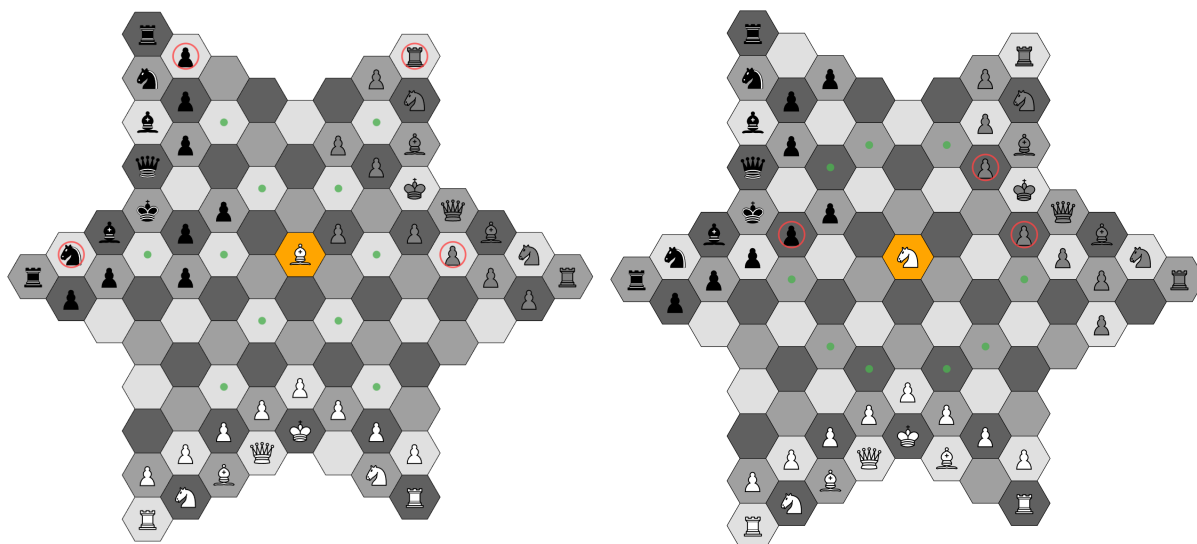
- **Pawns** move one step forward and capture diagonally forward, with promotion possible upon reaching the farthest rank relative to their starting position.
- **Rooks** move along straight lines radiating outward from the hexagon.
- **Bishops** move diagonally across the hexagonal grid, which allows them to cover new movement paths not present in standard chess.
- **Knights** retain their characteristic "L"-shaped movement, adapted to the hexagonal geometry.
- **Queens** combine the movement of rooks and bishops.
- **Kings** move one step in any direction and may castle under conditions analogous to standard chess.

### 3.2. Turn Order and Gameplay

Turns proceed sequentially in a clockwise order around the board, giving each of the three players equal opportunity to move. The game begins with white, followed by the next player in rotation. This structure introduces strategic complexities, as each move must account for two opponents rather than one.

### 3.3. Check, Checkmate, and Elimination

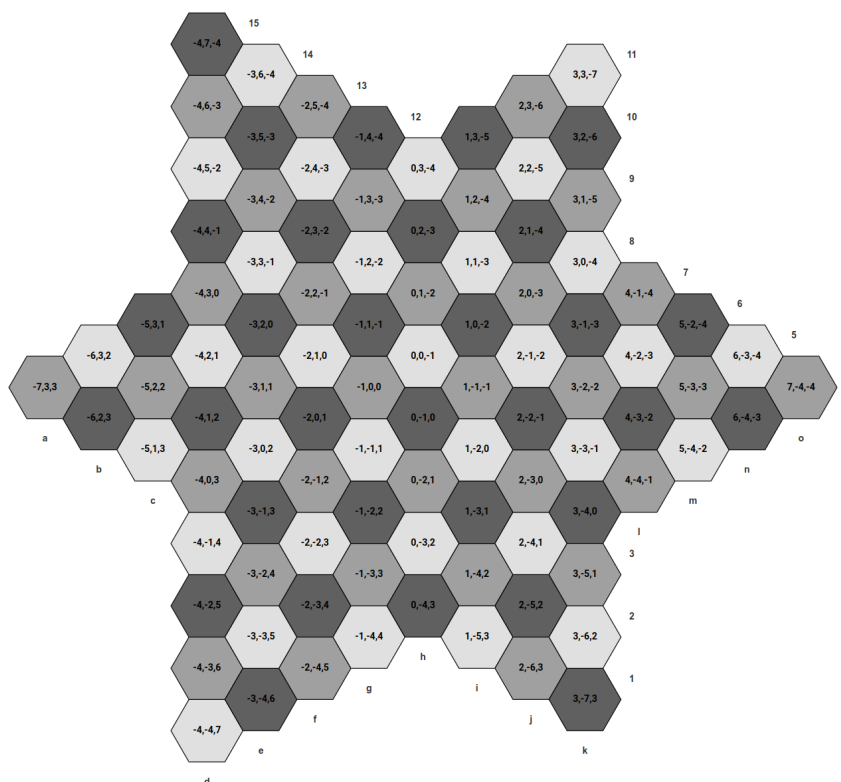
A king placed under threat is considered to be in check, as in standard chess. If a player is checkmated, their pieces are removed from the board, effectively clearing the space for the remaining players. The game continues until one player successfully checkmates the final opponent, thereby claiming victory.



**Figure 2:** Valid moves for white bishop the left image and for white knight in the right image.

### 3.4. Strategic Implications

The introduction of a third player creates possibilities for temporary alliances, betrayals, and shifting balances of power. Unlike standard chess, where play is strictly adversarial, three-player chess often requires situational cooperation to prevent one player from gaining an overwhelming advantage. This dynamic makes the variant highly unpredictable and strategically distinct from its two-player counterpart.



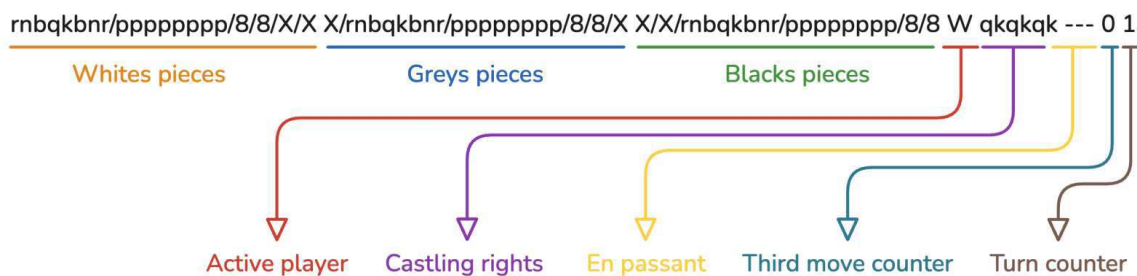
**Figure 3:** Coordinate system for hexagonal chessboard.

### 3.5. Coordinate System

To represent the hexagonal board computationally, a coordinate system was defined that assigns a unique label to each of the 96 fields as shown in fig. 3. Unlike the square  $8 \times 8$  grid of traditional chess, the hexagonal board requires a two-dimensional axial coordinate system, where each field is described by a pair of indices corresponding to its row and diagonal alignment. This approach ensures that piece movement can be calculated consistently, regardless of orientation. For human readability, the system was adapted into an alphanumeric notation similar to algebraic chess notation: files are labeled with letters (A–L) and ranks with numbers (1–16), omitting unused coordinates. This design allows players to communicate moves unambiguously while supporting efficient internal data structures for move validation.

### 3.6. HexFEN Notation

In order to store and exchange game states, we developed an extension of the standard Forsyth–Edwards Notation (FEN), which is widely used in digital implementations of classical chess. The adapted format, referred to as *HexFEN*, encodes the positions of all three sets of pieces on the 96-field board as shown in fig. 4. Each rank of the hexagonal board is represented as a string, with numbers indicating consecutive empty fields and letters denoting pieces, differentiated by color. The notation also encodes additional information such as the player on turn, castling rights, and en passant possibilities. For example, the initial setup of Varga’s 3way chess can be expressed as a single HexFEN string, enabling straightforward storage, retrieval, and reconstruction of board states. This notation proved essential for implementing features such as game history replay, server–client communication, and debugging of rule enforcement.



**Figure 4:** The structure of the HexFEN format

## 4. Methodology

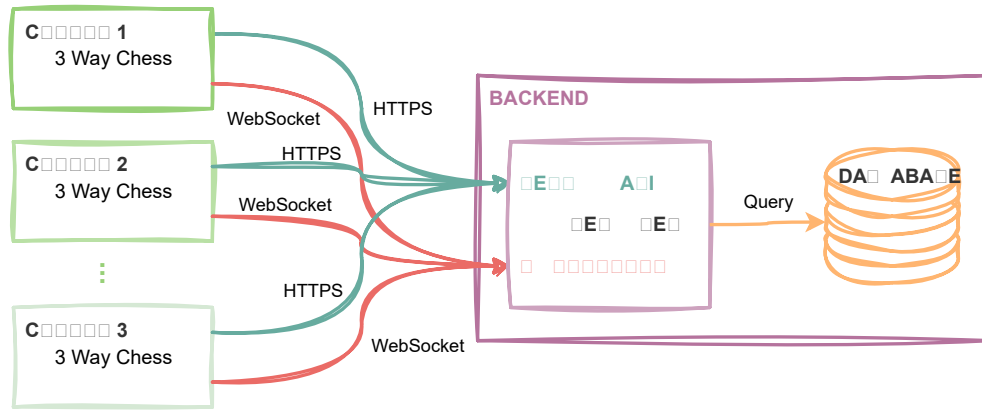
The development of the platform followed a user-centered design process, complemented by iterative technical prototyping. Our primary aim was to translate the physical version of Varga’s three-player chess into a digital format that would be both accessible and intuitive for new players. The methodology consisted of three phases: requirements gathering, system design and implementation, and evaluation.

### 4.1. Requirements Gathering

We began by analyzing the unique characteristics of three-player chess, including its hexagonal board layout, expanded movement rules, and three-way turn structure. This analysis revealed several HCI challenges: (1) learning support for unfamiliar rules, (2) visualization of a non-standard board, and (3) synchronization of three players in real-time gameplay. To address these, we reviewed established online chess platforms Chess.com and Lichess.org, extracting best practices in usability, responsiveness, and real-time interaction. Additionally, we studied community-driven implementations of three-player chess, which helped identify common limitations such as a lack of matchmaking, incomplete rule support, and low accessibility.

## 4.2. System Design and Implementation

Based on the identified requirements, we designed the platform as a web application accessible from any modern browser. The system architecture consists of two main components: (1) a backend server handling authentication, game logic, synchronization and database for; and storing accounts, game states, and histories (2) a front-end interface providing visualization and user interaction.



**Figure 5:** Diagram of the system architecture.

The back-end was implemented using Node.js<sup>14</sup> and Express.js<sup>15</sup> to provide a scalable environment for managing requests and routing. Real-time gameplay was supported by Socket.IO<sup>16</sup>, ensuring low-latency synchronization of moves between players. Game state validation was performed using a WebAssembly<sup>17</sup> module adapted from existing research on three-player chess engines, enabling efficient rule enforcement in the browser.

The front-end was built using React<sup>18</sup>, selected for its modular, component-based architecture, and extended with Mantine<sup>19</sup> to accelerate development of responsive, accessible UI components. To support reliability and consistency across both the client and server, we adopted TypeScript<sup>20</sup> for static type checking. The unconventional chessboard visualization was implemented as a dynamic SVG-based grid, capable of rotating perspectives for different players to reduce cognitive load.

For data storage, we used MongoDB<sup>21</sup> as a NoSQL database, providing flexibility for managing heterogeneous data structures such as user profiles and move sequences. Mongoose<sup>22</sup> was integrated to enforce schema validation, ensuring the integrity of stored records.

## 4.3. Evaluation

To validate the platform, we conducted two types of evaluation. First, a functional evaluation tested compliance with the identified requirements, confirming that features such as account management, online multiplayer, and game history replay worked as intended. Second, a performance evaluation measured latency in real-time synchronization, comparing server locations in Europe and the United States. These tests highlighted the importance of server proximity, with local servers reducing move

<sup>14</sup><https://nodejs.org/en/about>

<sup>15</sup><https://expressjs.com/>

<sup>16</sup><https://socket.io/>

<sup>17</sup><https://webassembly.org/>

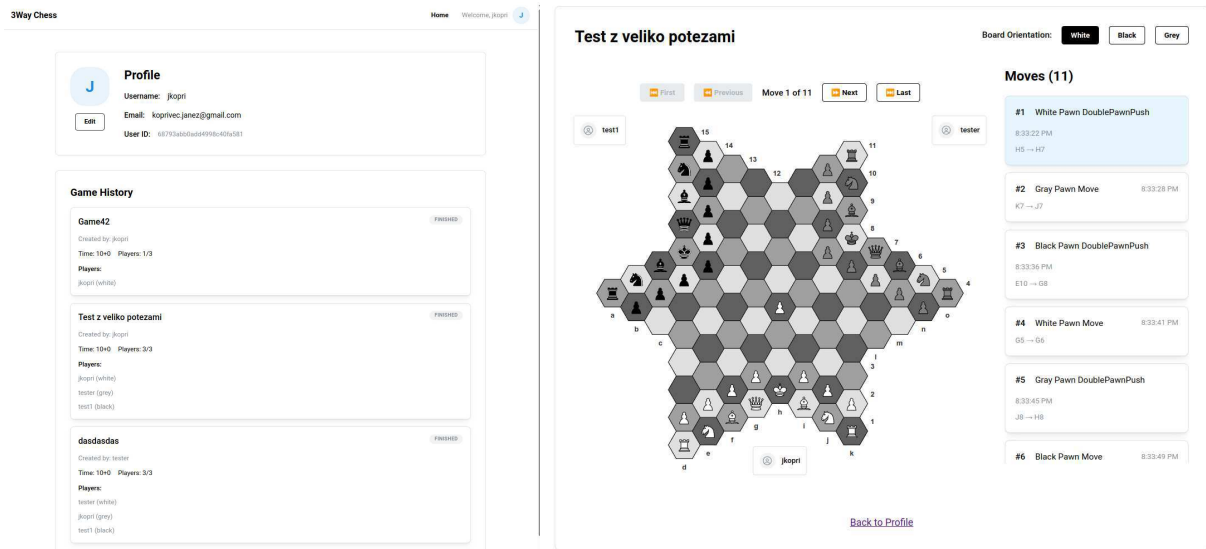
<sup>18</sup><https://react.dev/>

<sup>19</sup><https://mantine.dev/>

<sup>20</sup><https://www.typescriptlang.org/>

<sup>21</sup><https://www.mongodb.com/>

<sup>22</sup><https://mongoosejs.com/>



**Figure 6:** UI implementation of user profile (left) and game review screen (right).

transmission delay by more than an order of magnitude. While primarily technical, these evaluations provide insights into the relationship between system responsiveness and perceived fairness in multiplayer environments, an essential HCI consideration.

## 5. Results

The platform was evaluated with respect to functionality, usability, and technical performance. The results demonstrate that the system successfully enables three-player chess to be played online while addressing the challenges of rule enforcement, board visualization, and real-time synchronization. The deployment of the system is available at: <https://chess3.musiclab.si/>.

### 5.1. Functional Results

All core features were implemented and verified through extensive playtesting. Players were able to create and manage accounts, initiate both local (same-device) and online multiplayer matches, and access their game history. The platform correctly enforced the rules of Varga's 3-way chess, including piece movements on the hexagonal board and the sequential turn system for three participants. The replay feature provided a complete record of past matches, supporting both learning and post-game analysis.

### 5.2. Usability Outcomes

From an interaction design perspective, the dynamic SVG-based chessboard proved effective in visualizing the unconventional geometry. Color-coded move indicators, turn notifications, and perspective rotation helped reduce player confusion during gameplay. Informal usability testing with novice players indicated that most users were able to learn the rules of the variant within a few rounds, suggesting that the interface supports learnability despite the unfamiliar mechanics. Participants also reported that the design of the interface was clear and consistent across devices, which can be attributed to the responsive layout and accessible component library provided by Mantine. The UI implementation of user profile view and game review

### 5.3. Performance Evaluation

System responsiveness was measured by recording round-trip times for move transmission under different server configurations. When hosted on a European server, latency remained consistently below 30 ms, which was not perceptible to players. However, when connecting to a U.S.-based server from Europe, latency increased by up to ~1 s, occasionally disrupting the flow of play. These results, also shown in table 1, confirm that server proximity has a direct impact on user experience in real-time multiplayer games. The integration of Socket.IO successfully minimized delays under optimal network conditions, but global scalability will require distributed server deployment.

**Table 1**

Performance evaluation of online play using servers at different locations.

| Server Location          | Average Latency (ms) | Maximum Latency (ms) | Player Experience                 |
|--------------------------|----------------------|----------------------|-----------------------------------|
| Europe (Ljubljana)       | 21.32                | 29.5                 | Smooth, no disruptions            |
| United States (Virginia) | 736.76               | 2116.0               | Noticeable delays, disrupted flow |

Overall, the evaluation confirmed that the platform meets the functional and interaction requirements identified during the design phase. It not only enables accessible play of a niche chess variant but also demonstrates how careful integration of visualization, interaction design, and network performance considerations can enhance the usability of digital board games.

## 6. Discussion

The development and evaluation of the platform demonstrate that translating niche board games into digital environments requires more than a direct technical implementation. It demands thoughtful consideration of HCI principles to ensure that new players can learn the rules, navigate unfamiliar visualizations, and maintain engagement in real-time multiplayer contexts.

### 6.1. Balancing Familiarity and Novelty

One of the central challenges was the visualization of the hexagonal chessboard. Traditional chess interfaces rely on a highly standardized square grid that players recognize immediately. In contrast, the three-player variant introduces new geometries and movement patterns that are not part of players' prior experience. By incorporating perspective rotation, turn indicators, and color-coded feedback, the platform reduced cognitive load and helped players adapt more quickly. This illustrates the broader HCI principle that unfamiliar mechanics should be introduced through scaffolding that bridges the gap between old and new interaction paradigms.

### 6.2. Responsiveness and Perceived Fairness

Performance evaluation highlighted the close relationship between network latency and user experience in multiplayer environments. Even small delays in move synchronization can undermine the perception of fairness, especially in a strategic game where timing and precision are crucial. While Socket.IO effectively minimized delays on local servers, cross-continental play revealed the limits of centralized hosting. This finding aligns with broader HCI research showing that technical performance is deeply entangled with user trust and engagement in online systems. Future work could explore distributed or peer-to-peer architectures to further reduce latency and improve global accessibility.

### 6.3. Usability and Learnability

Although the rules of three-player chess are inherently more complex than those of standard chess, informal usability testing suggested that most players were able to understand the basics within a few

rounds. This indicates that the interface design supported learnability by making rules visible through interactive feedback, rather than relying solely on textual explanations. Integrating tutorials, tooltips, or adaptive hints could further lower the entry barrier and make the game more approachable to a wider audience.

#### **6.4. Broader Implications**

Beyond the specific case of three-player chess, this work contributes to ongoing discussions in HCI about the digital transformation of cultural practices. Many traditional games and variants remain locked in physical formats due to their unconventional rules or non-standard equipment. By showing how modern web technologies can be combined with user-centered design, this project demonstrates a pathway for preserving and revitalizing such practices in digital form. The insights gained—particularly regarding visualization of non-standard boards, synchronization of more than two players, and scaffolding for rule learning—may be applied to other games and domains where digital adoption has been limited.

Overall, the study highlights the importance of aligning technical design choices with usability principles. A focus on responsiveness, clarity, and learnability allowed us to create not only a functioning multiplayer application but also an interaction environment that invites players to explore, understand, and enjoy a novel chess variant.

### **7. Conclusion**

This paper presented the design and implementation of a web-based platform for Varga’s 3-way chess, a niche board game variant that has remained largely inaccessible in digital form. By combining modern web technologies with a user-centered design approach, we demonstrated how challenges of non-standard board visualization, rule enforcement, and real-time synchronization can be addressed in a way that supports usability and engagement.

The results showed that the platform successfully enables both local and online multiplayer play, with functional features such as account management, game history replay, and responsive visualization of the hexagonal chessboard. Informal usability testing indicated that players were able to learn the unfamiliar rules within a few sessions, suggesting that interaction design elements such as color coding, turn indicators, and perspective rotation effectively reduce cognitive load. Performance evaluation further highlighted the importance of server proximity for maintaining fairness and flow in multiplayer interaction.

Beyond the specific case of three-player chess, this work illustrates how thoughtful application of web technologies can broaden access to alternative game variants and preserve cultural practices through digital transformation. Future work will expand on this foundation by integrating structured usability studies, distributed server architectures for global play, and interactive tutorials that further support learning. More broadly, the study underscores the potential of HCI-driven design to make experimental and unconventional games more accessible, inclusive, and engaging in digital environments.

### **Acknowledgments**

We would like to express our gratitude to Dario Varga, the creator of three-player chess, for his originality and contribution in designing this unique variant of the game. His innovative approach to extending the classical rules of chess to a three-player format provided the inspiration and foundation for this work. Without his creativity and dedication, the development of a digital platform for three-player chess would not have been possible.

### **References**

- [1] J. Koprivec, An online platform for three-player chess, BSc thesis, 2025. BSc thesis.

- [2] J. Picussa, L. S. Garcia, J. Bueno, M. V. Ferreira, A. I. Direne, L. C. de Bona, F. Silva, M. A. Castilho, M. S. Sunye, A user-interface environment solution for an online educational chess server, in: 2008 Second International Conference on Research Challenges in Information Science, IEEE, 2008, pp. 179–186.
- [3] M. Guid, M. Možina, C. Bohak, A. Sadikov, I. Bratko, Building an intelligent tutoring system for chess endgames, in: International Conference on Computer Supported Education, volume 2, SciTePress, 2013, pp. 263–266.
- [4] A.-P. Palomäki, Web browser based online chess, human versus human games with multiple end point devices (2017).
- [5] M. Hostettler, L. Boner, Digitalization of chess score cards, Zurich University, master thesis (2022).
- [6] J. Vasiljević, Adapting AlphaZero for Three-Player Hexagonal Chess, Magistrska naloga, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, 2024. V pripravi.

# Real-Time Gesture Transmission with a Robotic Hand: Embodied Signals for Non-Verbal Remote Communication

Lea Pajnič<sup>1,\*</sup>, Matjaž Kljun<sup>1,2</sup>, Maheshya Weerasinghe<sup>1</sup> and Klen Čopič Pucihar<sup>1,2</sup>

<sup>1</sup>University of Primorska, FAMNIT, Glagoljaška 8, SI-6000 Koper, Slovenia

<sup>2</sup>Stellenbosch University, Department of Information Science, Stellenbosch, South Africa

## Abstract

This work explores how computer vision and robotics can support remote, gesture-based embodied signals for expressing presence and emotion in remote communication. We present an initial proof-of-concept in which users interact through robotic hands placed on their desks: one user's hand gestures are captured in real time by a camera, transmitted over a network, and reproduced by a robotic hand at the remote location. The prototype uses the InMoov robotic hand and MediaPipe Hands for gesture tracking across varied lighting conditions, viewing angles, and backgrounds. Our preliminary tests demonstrate that gestures can be reliably recognised and consistently reproduced through stable network communication. While still at an early stage, this project illustrates the potential of combining affordable robotics with computer vision to create accessible alternatives to voice communication and new forms of remote communication.

## Keywords

robotic hand, gesture transmission, embodied signals, non-verbal communication, remote communication, computer vision,

## 1. Introduction

Communication is fundamental to human interaction, and while speech is the primary channel, non-verbal communication enriches and extends it. Hand gestures, in particular, represent one of the frequently used non-verbal modalities, capable of reinforcing, substituting, or even surpassing spoken words in expressiveness [1, 2, 3]. They can supplement speech to enhance meaning, or replace it entirely when verbal communication is not possible. In everyday life, people naturally employ a repertoire of gestures that expand the communicative capacity of language and provide a direct and often more expressive way of conveying thoughts and intentions.

These characteristics have long inspired research into gestures' role in human-computer interaction (HCI). Compared to traditional input devices, gestures support a more natural form of interaction, eliminating the barriers posed by keyboards, mice, or controllers [4]. Advances in computer vision, radar systems and similar technologies, have further accelerated this research by enabling robust real-time analysis of hand movements, allowing gesture recognition systems to be deployed in increasingly unconstrained environments. Gesture-based interaction has proven effective in a variety of domains, from immersive environments to assistive technologies [5].

Motivated by the communicative power of gestures and their technological feasibility, we set out to design a proof-of-concept prototype for remote non-verbal communication between individuals. By transmitting gestures through robotic embodiment, our project aims to extend hand gestures into new technological contexts for remote interaction. Similar to biofeedback remote signals such as breathing to foster empathy, awareness, and connectedness among individuals across distance [6], our approach highlights how gestural embodied signals can become a medium for expressing presence and emotion in remote communication [7].

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

✉ lea.pajnic@gmail.com (L. Pajnič); matjaz.kljun@upr.si (M. Kljun); maheshya.weerasinghe@famnit.upr.si (M. Weerasinghe); klen.copic@famnit.upr.si (K. Čopič Pucihar)

ORCID 0000-0002-6988-3046 (M. Kljun); 0000-0003-2691-601X (M. Weerasinghe); 0000-0002-7784-1356 (K. Čopič Pucihar)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.10>

## 2. Related Work

This section focuses on two areas related to the paper: (i) hand gesture recognition, and (ii) embodied signals and biofeedback in remote communication.

### 2.1. Hand Gesture Recognition

Hand gesture recognition is generally divided into two basic categories: sensor-based and vision-based methods [8]. Each follows a different technical approach and presents advantages and limitations.

Sensor-based methods rely on wearable or ambient sensors. An example of the former are gloves that can integrate multiple sensors (e.g., flex, accelerometer, or tactile sensors) to capture fine-grained three-dimensional information about hand posture. These systems have demonstrated high recognition performance in areas such as sign language translation, virtual reality interaction, and assistive technologies, with accuracy often exceeding 90% [9, 10]. Their main drawback is the need for specialised equipment, which limits everyday usability.

Ambient sensors include, among others, electromyography (EMG) [11], Wi-Fi [12] and radars [5]. EMS records the electrical activity of muscles using electrodes placed on the skin. This enables detection of subtle hand motions and has achieved recognition rates above 80%, depending on the classifier (e.g., support vector machines, random forests) [13, 14]. However, the requirement for electrodes and calibration can reduce user comfort and accessibility. Wi-Fi and radar-based approaches use signal variations (e.g., Doppler shifts or changes in signal strength) to infer gestures. Systems such as WiSee [15] and Mudra [16] have reported recognition rates above 90%, showing promise for unobtrusive sensing. However, these methods remain highly dependent on environmental conditions and signal propagation quality.

Vision-based methods use cameras to capture gestures without requiring wearables. Advances in monocular, stereo, and depth-sensing technologies have enabled robust skeletal tracking and real-time gesture interpretation. Widely adopted platforms such as Microsoft Kinect, Leap Motion, and Intel RealSense illustrate the potential of camera-based approaches. While they provide a contactless interaction mode, their accuracy can be compromised by external factors such as lighting, occlusion, and background clutter [17]. Given that our prototype places robotic hands on users' desks during computer use, we selected a vision-based method, as standard laptops, as well as desktop computers, are typically equipped with cameras.

### 2.2. Embodied Signals and Biofeedback in Remote Communication

The integration of robotics and computer vision to transmit embodied signals for remote interaction is based on telerobotics and embodiment, particularly work on conveying expressive non-verbal cues [7, 18]. Telepresence robots are designed to make a remote operator feel perceptually co-located, and physical embodiment plays a crucial role in this process, shaping how collaborators perceive presence and social engagement.

Unlike virtual agents, social robots rely on physical embodiment to communicate non-verbal behaviours. For example, the MeBot platform [18] combined audio and video channels with expressive arm and head movements, demonstrating that adding gestural degrees of freedom improved participants' sense of involvement, cooperation, and enjoyment during interaction. Studies comparing different forms of embodiment suggest that physically embodied interactions are favoured over virtual ones and remote teleconference ones, with a co-located physical robot being perceived as both more watchful and more enjoyable than a remote physical robot or a simulated agent [7]. This work strongly indicates that telepresence systems benefit from allowing users to express non-verbal behaviour beyond audio and video, particularly in applications requiring deep engagement and cooperation.

Complementary to gesture-based embodiment, research in biofeedback has investigated how physiological signals can enhance emotional expression and connectedness in remote communication [6]. Biofeedback makes internal processes such as breathing perceptible, supporting interpersonal awareness

and empathy. Breathing, in particular, has been explored as a communicative signal due to its strong links with emotional state, coordination, and social bonding at a distance. Extending these ideas, we set to use a robot hand to externalise gesture-based embodied communication at a distance.

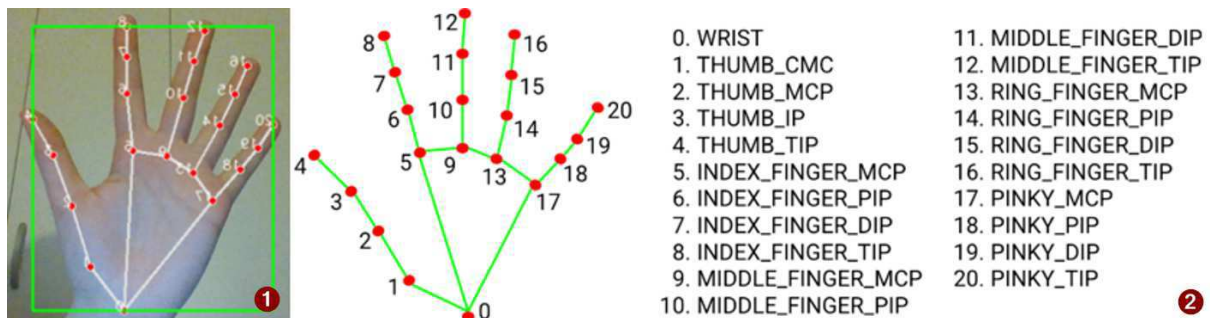
### 3. Design and Implementation of the Prototype

The steps of the design and implementation involved selecting a computer vision (CV) framework, building a robotic hand, selecting a platform to control the hand movements, selecting gesture vocabulary for evaluation, defining communication workflow and writing the code for it. Each of these steps is presented in subsections below.

#### 3.1. Tracking Gestures with Computer Vision

Computer vision (CV) is based on a range of techniques for extracting meaningful information from images and videos, from simple heuristics to advanced machine learning pipelines [19]. Similar to how humans first perceive basic shapes before interpreting a scene, CV systems detect edges and forms and then build higher-level representations [20]. In our prototype, this process begins with a real-time hand tracking. By continuously comparing successive frames from the camera, the prototype can detect motion and changes in gesture, allowing for accurate mapping of the user's hand movements to the robotic hand.

To implement this, we used MediaPipe Hands framework, a robust CV framework designed for high-precision hand detection. The framework identifies 21 landmarks on the hand to form a hand graph as visible in Figure 1. This allowed us robust estimation of finger positions and overall hand orientation across continuous video frames.



**Figure 1:** ① Hand landmarks detection. ② Hand landmarks detection guide, via Google AI for Developers [21].

#### 3.2. Robotic Hand

For the robotic hand, we 3D-printed the right hand and wrist components of the InMoov robot [22]. InMoov, created by French sculptor Gaël Langevin, is an open-source humanoid robot released under a Creative Commons license, with freely accessible models and assembly guides. It has become a widely adopted platform for robotics education and prototyping, supporting integration of speech, vision, and sensing capabilities, which makes it highly adaptable for research and development.

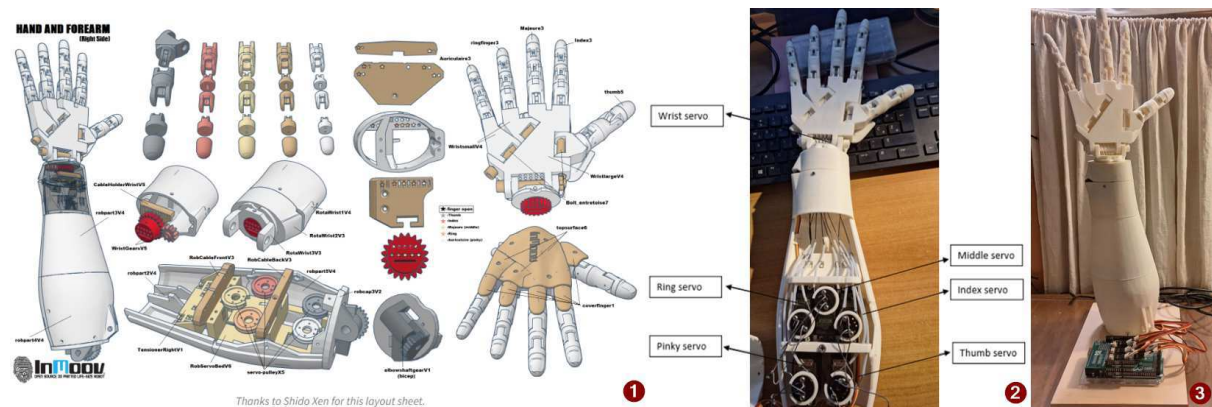
For the purpose of this project, we restricted the implementation to the right hand of the robot since our focus was on hand gestures. Although the right hand was used, the left hand can easily be implemented as well. In addition, all the gestures selected (see next subsection) can be easily completed by both hands. At the time of construction, we used the available version of the hand model visible in Figure 2 ①, although newer 3D components were later released.

Assembly began with the fingers, each constructed from multiple knuckle elements to allow detailed articulation. Selected joints were glued or adjusted to ensure smooth closing, and the palm was subsequently added with additional parts for the thumb, ring, and little fingers. Printed cylinders served

as connectors to provide full mobility. The fingers were actuated using five braided fishing lines of equal length, chosen for their strength and low elasticity.

The wrist, which houses a sixth servo for lateral palm movement (see Figure 2 ②), was then assembled. A connecting cylinder was positioned between palm and wrist to avoid obstructing the finger lines, which were threaded through dedicated channels to prevent tangling during rotation. The forearm was built from components designed to accommodate five servos, each equipped with a circular gear to secure the finger lines. After assembling the main structure, elastic sleeves and springs were added to protect and stabilise the lines. Each line was aligned with its corresponding servo as well as tightened and fixed with screws on the outer ring of the servo gear seen in Figure 2 ②. This ensured that rotating the servo left would tighten one line while releasing the opposing one and vice versa. This adjustment provided reliable finger motion.

Final steps included attaching palm covers, fingertip pieces, and labelling servo cables. The wrist and fingers were actuated using an Arduino Mega 2560 Rev3, controlling six servo motors. For stability, the assembled hand and Arduino board were mounted on a wooden base shown in Figure 2 ③. A demonstration of the working prototype is available at <https://youtu.be/RLTetVgm4bI>. The video also shows the speed of finger movements and prototype's latency without including the network.



**Figure 2:** ① Hand and Forarm pieces from InMoov [22]. ② Assembled hand with connected lines to servos. ③ Whole hand mounted on the wooden board.

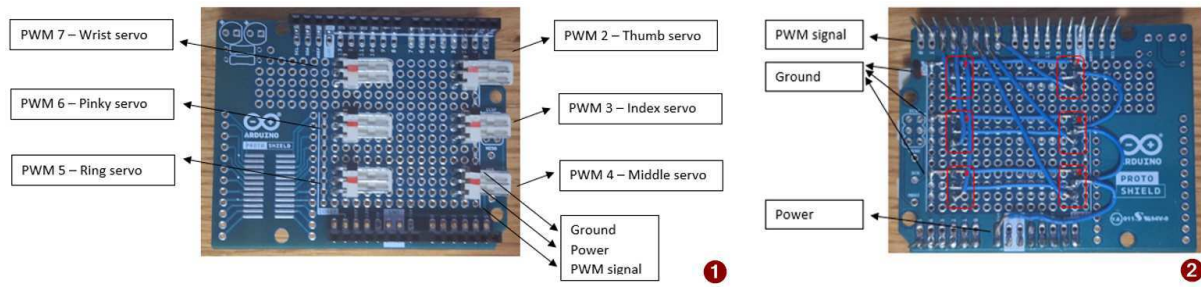
### 3.3. Arduino-Based Control System

The Arduino Mega 2560 Rev3 served as the central controller of the robotic hand, receiving hand gesture data via a WebSocket connection (described later) and converting them into predefined commands for six servo motors. Five of these servos actuated the lines, tightening or releasing them to open and close the fingers, while the sixth servo controlled lateral wrist movement.

To enable simultaneous control of the wrist and fingers, we integrated an Arduino ProtoShield, which provided six dedicated connection pins for the servo wires. Each connection included a ground, power, and PWM (pulse-width modulation) interface, which were matched to the corresponding servo. Specifically, the brown wire of each servo was linked to ground, the red wire to Vin (power), and the yellow wire to the PWM pin. Assigning each servo to a separate PWM channel allowed individual motor control. The ProtoShield also simplified wiring and improved modularity by making servo connections detachable, allowing easy replacement if necessary. The Arduino board was powered via USB and generated PWM pulses to set servo angles. Standard timing values were used, with a 1 ms pulse corresponding to 0°, 1.5 ms to 90°, and 2 ms to 180°. To avoid overdriving the motors, the operational range was limited by 10°, such that relaxed fingers corresponded to 10° and fully closed fingers to 170°.

### 3.4. Code Implementation

We implemented three applications in Python. The first programme supporting communication between two remote systems, was implemented as a Python-based WebSocket server using Ngrok. The two



**Figure 3:** ① Front side of Arduino ProtoShield. ② Back side of Arduino ProtoShield.

systems are the client for detecting gestures and the server with the hand attached for executing gestures. The connection is secured with SSL encryption, ensuring safe data transfer. Once active, the server listens for incoming requests and, upon connection, creates a secure communication channel between two users. Gesture IDs detected on the client side are transmitted over this channel to the server, which then forwards them to the Arduino through a serial interface. Proper communication is maintained via a specified COM port and baud rate.

Next is the CV program that integrates the MediaPipe Hands library and OpenCV for real-time hand gesture recognition running on the client computer. When started, the program attempts to establish a WebSocket connection via Ngrok. If the request fails, an error is displayed. Otherwise, the camera is activated. A short countdown is presented to allow the user to position their right hand in a neutral position, after which the prototype begins tracking 21 hand landmarks and displays the recognised gesture in the corner of the interface.

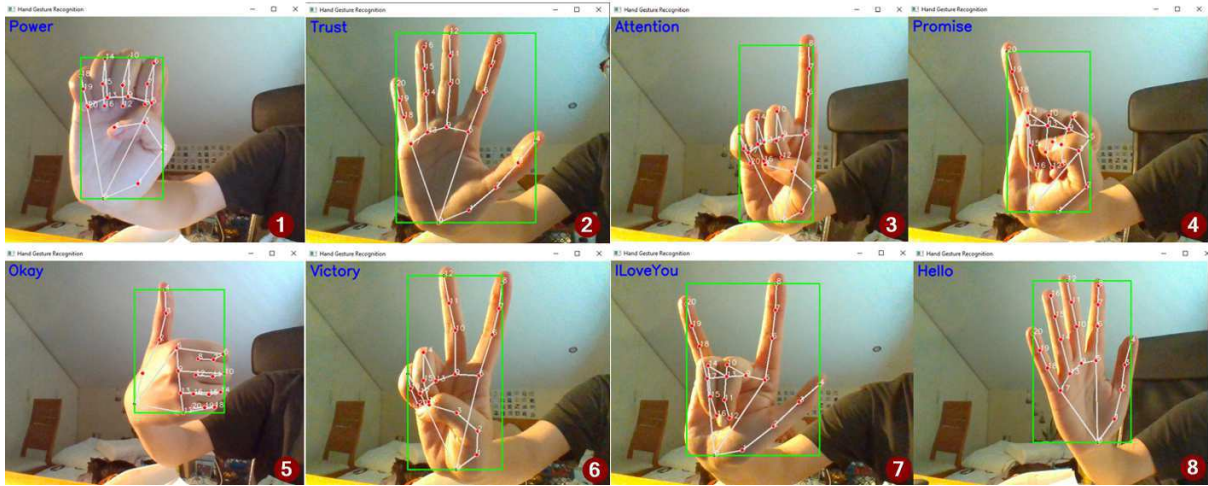
And lastly we implemented the Arduino program, which first establishes serial communication at designed baud rate to receive data. It then attaches servos to Arduino pins 2, 3, 4, 5, and 6, and moves the fingers to the initial neutral open palm position with additional sixth servo on pin 7 that controls the wrist. The loop function continuously checks if data is available. When a new gesture ID is received, the program first validates whether it signals a valid gesture from a predefined set presented in the next subsection. To ensure all servos move correctly and to prevent tangling of lines and fingers, the function uses a small delay after each finger or wrist movement.

### 3.5. Selected Hand Gestures and Their Meanings

Hand gestures range from movements that simply accompany speech rhythm to those that convey explicit meaning. Over time, many gestures have acquired socially recognised significance shaped by historical and cultural contexts [23]. As a result, gestures function not only as physical actions but also as communicative signals. While some gestures are understood almost universally, others remain closely tied to their cultural origins.

For the purpose of our proof-of-concept prototype we decided to select eight (8) distinct single-handed gestures that are widely recognised in Western cultures. Hand gestures can otherwise be also two handed (e.g., forming a hand heart) or combined with other body parts (e.g., facepalm). The selected gestures, as detected by the computer vision framework, are illustrated in Figure 4. This set can be extended in future work. A description of each chosen gesture follows.

- 1.) **Power** - The clenched fist is a widely recognised symbol of resistance, power, and solidarity. Often used against oppression, it has appeared both as a physical gesture and in media graphics, becoming familiar worldwide through repeated historical use [24, 25].
- 2.) **Trust** - The open palm is used to signify many different meanings. But in our project, we have an open palm facing forward, which is used as a psychological and subconscious behaviour in body language to communicate trust, openness and compliance [26].
- 3.) **Attention** - Index finger pointing upwards can mean many things. In many Western countries it is



**Figure 4:** Selected gestures for our project: ① Power, ② Trust, ③ Attention, ④ Promise, ⑤ Okey, ⑥ Victory, ⑦ ILoveYou, and ⑧ Hello.

used for indicating direction, drawing attention, or signalling success or achievement. However, in the Middle East, holding up the index finger is considered rude and can be interpreted as an insult or a threat [27].

**4.) Promise** - The pinky promise, primarily used by children, involves interlocking pinkies to signify a commitment. Originating in Japan as “yubikiri”, meaning “finger cut off”, it was allegedly used by the Yakuza to test loyalty. A related tradition in Japan and China, the “red thread,” ties couples’ pinkies to symbolize their destiny. Overall, the gesture represents a promise that should never be broken [28].

**5.) Okay** - The thumbs-up gesture has a varied history. In ancient Rome, it was linked to gladiatorial judgments, while in medieval England it indicated readiness among archers. By the 20th century, it signaled “all is well” for World War I soldiers and pre-flight checks in World War II. Today, it is widely recognised as a positive sign of approval [29, 30].

**6.) Victory** - The V sign, or victory sign, is made by raising the index and middle fingers with the palm facing outwards. It symbolised victory in World War II, later represented peace during the 1960s hippie movement, and is used in Asia to convey cuteness in photos [31]. If the palm faces inward, the gesture may be offensive in some countries.

**7.) ILoveYou** - The American Sign Language “I love you” sign combines the thumb and index to form an “L” and lifts the little finger to form an “I.” Known as “I-L-Y,” it is a widely recognised symbol in the deaf community and has spread globally, also used to say goodbye or thank you [32].

**8.) Hello** - Waving, commonly used to greet or say goodbye, also serves to acknowledge presence, call for silence, or deny someone. Its origins trace back to the 18th century as a form of saluting, evolving from medieval knights lifting helmet guards to show identity and peaceful intent. By the 1780s, European armies formalised the salute, and waving became a standard way to address others, especially in military contexts [33].

### 3.6. Prototype Communication Workflow

To test functionality of communicating at a distance, two people are needed. For simplicity, only one-direction communication is described. Person-1 runs the CV program and performs hand gestures in front of a camera, while Person-2 runs the socket program on a computer connected to the Arduino via USB. Person-2 must start the socket program first to allow connection from Person-1. It then receives and enters the Ngrok URL to establish the connection. Both users receive an alert once connected. Next, gestures are being tracked in real time, recognised by the gesture module, and assigned one of eight gesture IDs. These IDs are transmitted via the WebSocket to the Arduino, which controls servos to

move fingers accordingly. Small delays may occur due to sequential servo motion. Tracking ends when Person-1 presses “q”, stopping data transmission and closing the socket connection, notifying Person-2 that gesture tracking has finished.

As previously mentioned, we used MediaPipe Hands to track hand gestures. Table 1 shows the classification of eight gestures as well as a neutral position used for the initial interaction with the robotic hand. For the neutral position we selected an open palm, since this is the robotic hand’s default starting posture as well as position that indicates trust as one of the strong positive gestures in our set. Two things need to be noted. To classify the thumb as closed, its fingertip must be positioned below the nearest knuckle joint. For the waving gesture, the hand must repeatedly move left and right until the “Hello” gesture is displayed. This requirement was introduced to prevent false positives from incidental wrist movements.

**Table 1**  
Gesture labels and their corresponding IDs

| ID | Gesture                            | Label     |
|----|------------------------------------|-----------|
| 0  | Neutral                            | Neutral   |
| 1  | Closed fist                        | Power     |
| 2  | Open palm                          | Trust     |
| 3  | Index up                           | Attention |
| 4  | Pinky up                           | Promise   |
| 5  | Thumb up                           | Okay      |
| 6  | Index and middle up                | Victory   |
| 7  | Thumb, index and little up         | ILoveYou  |
| 8  | Open palm, wrist moves left, right | Hello     |

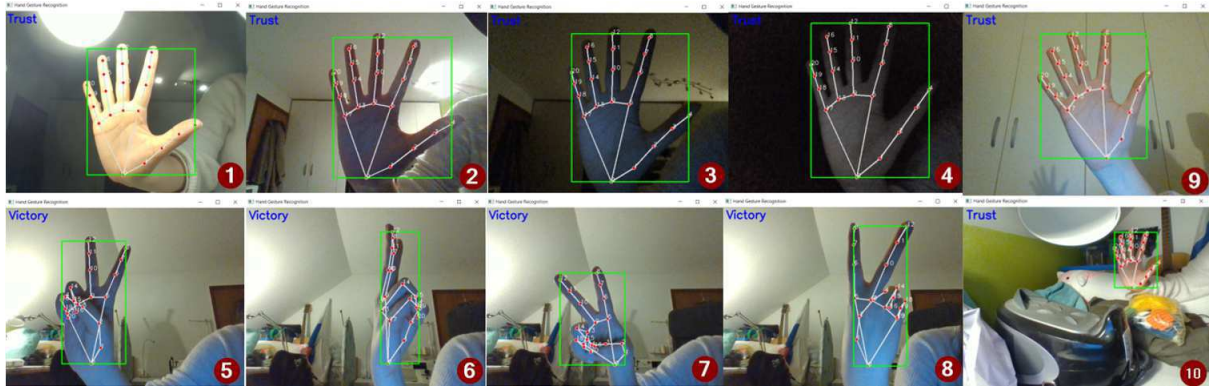
After starting the CV program, a ten-second countdown allows the user to position their right hand in front of the camera in the neutral position, the same as the robotic hand. It is also important to note that hand tracking does not work if the left hand is in the frame. Once the requirements are met, landmarks appear, and gestures are tracked. Gestures are accepted only after two seconds to reduce false changes. The wrist can be rotated left or right, but rotation is limited to the opposite direction after a turn.

## 4. Testing and results

In this section we present evaluation of camera gesture tracking and testing gesture execution.

### 4.1. Camera Gesture Tracking

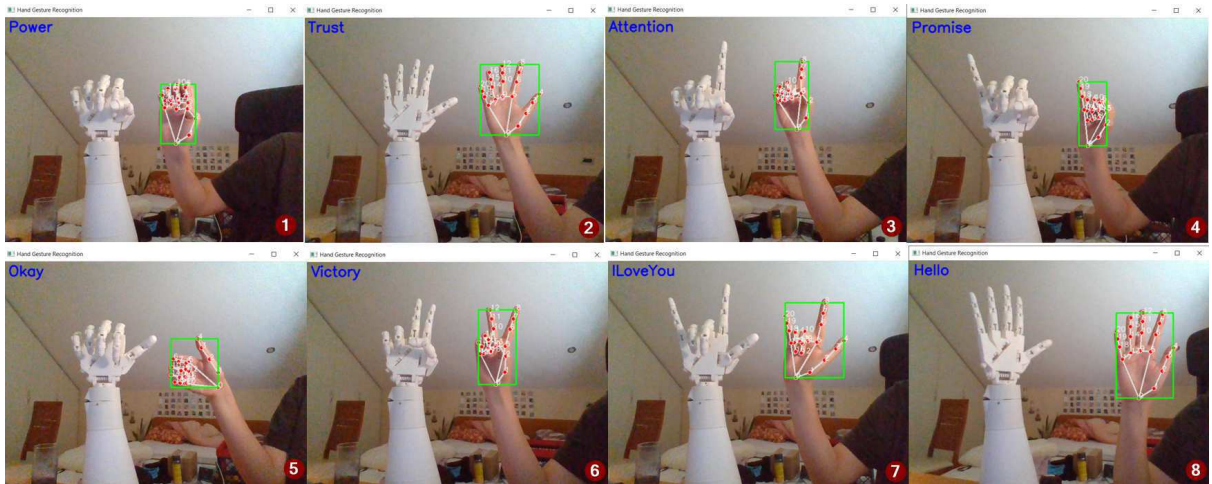
We performed a series of tests to evaluate prototype performance and identify its limitations. The first set of experiments examined the effect of lighting conditions, as illumination strongly influences the detection of all 21 hand landmarks. We then investigated how varying hand orientations affected recognition accuracy, identifying angles at which the program failed to track landmarks reliably. Lastly, we assessed the prototype’s robustness in cluttered environments by introducing multiple background objects. The gesture recognition system demonstrated reliable performance across various lighting conditions, with failure occurring only in complete darkness. Recognition was also robust across different hand orientations, except when the hand was tilted sharply upwards. In extreme cases, such as fully vertical or horizontal orientations, the 21 landmarks overlapped and could not be distinguished, and these scenarios were excluded from the project scope. The system performed best against plain backgrounds but also managed cluttered scenes, provided the hand remained fully visible. For accurate detection, all 21 landmarks had to be unobstructed and free from occlusion by other objects.



**Figure 5:** Gesture tracking in different lightning conditions, angles and backgrounds: ① Bright lighting, ② Normal lighting, ③ Dark conditions, ④ Only screen lighting, ⑤ Front view, ⑥ Side view, ⑦ Tilted downwards, ⑧ Turn backwards, ⑨ Plain background, ⑩ Cluttered background.

## 4.2. Testing Gesture Execution

The first step was to verify that the robotic hand correctly executes each predefined gesture ID, ensuring that the fingers aligned with the servo angles assigned to that gesture.



**Figure 6:** Testing hesture execution: ① Power, ② Trust, ③ Attention, ④ Promise, ⑤ Okey, ⑥ Victory, ⑦ ILoveYou, ⑧ Hello.

Testing showed that the robotic hand consistently executed predefined gestures, with both fingers and wrist moving to the correct positions. Servo response time was measured as the delay between transmitting a gesture ID and completing the motion sequence. Commands were executed within a few seconds, with minor delays resulting from the intentional sequential control of servos to prevent interference. Repeated trials confirmed reliable precision, as the hand consistently reached the expected positions. The only issues observed were small gaps caused by loosely connected servo lines when fingers were released, which were subsequently corrected.

## 5. Discussion, limitations and future work

While this work successfully demonstrates a real-time gesture transmission prototype, several limitations regarding the prototype's scope and cultural applicability must be addressed to ensure scalability and communicative effectiveness in diverse, real-world scenarios.

## 5.1. Gesture Vocabulary

The current proof-of-concept prototype supports a limited gesture vocabulary of eight distinct single-handed gestures primarily drawn from Western cultures. While this selection was sufficient to demonstrate the technical pipeline, it restricts the prototype’s communicative potential and inclusivity. Gestures are shaped by historical and cultural norms. Some, such as waving or pointing, have relatively universal interpretations, while others are culture-specific and may be misunderstood or even offensive in different regions as explained in subsection 3.5. Relying on a small, Western-centric set introduces the risk of cultural ambiguity and limits the generalisability of the prototype.

Future work should explore customisable or extensible gesture sets, such as enabling users to define their own gestures relevant to their contexts. This would make the prototype adaptable to diverse cultural environments and specialised domains (e.g., technical fieldwork, sign languages, or task-specific commands). However, scaling the gesture vocabulary will also require careful management of actuators. The current prototype uses sequential control, requiring a small delay after each finger or wrist movement to prevent line tangling and servo interference. Although these commands are executed within a few seconds and the resulting delays are minor with the current eight gestures, larger or more complex gesture sets requiring simultaneous, multi-servo motion may present a challenge for the current prototype.

## 5.2. Evaluation

The present evaluation focused primarily on technical feasibility rather than human-centred performance. We report early metrics such as gesture recognition accuracy and servo response time, but have not yet conducted user studies to investigate how the prototype supports remote communication, user experience, social presence, or ease of interaction. Consequently, while the prototype demonstrates the technical feasibility of gesture-based robotic embodiment, its communicative effectiveness and usability in real-world scenarios remain unverified. Moreover, the evaluation is mostly qualitative and lacks end-to-end quantitative analysis (between two users) of key aspects such as (i) overall prototype latency from recognition to actuation over the network and different network connectivity, (ii) gesture recognition accuracy under challenging conditions over the network, and (iii) actuation repeatability or reliability over time. Future evaluations should combine technical metrics with human-subject experiments using validated scales for social presence, usability, and user experience.

## 5.3. Alternative approaches

This work would also benefit from comparative analysis with alternative modalities for remote gestural communication. Virtual avatars and wearable haptic devices are commonly explored to convey presence and intent in telecommunication contexts. Contrasting our robotic embodiment with these alternatives could clarify its unique advantages, such as providing visible, physically co-located gestures that can be understood without specialised hardware on the user’s side. A clear articulation of when and why robotic embodiment outperforms purely virtual or haptic solutions would help guide future adoption and inform design trade-offs.

## 6. Conclusion

In this project, we successfully implemented a camera-based hand gesture tracking prototype that transmits gesture commands over a secure network to a robotic hand, which replicates the detected gestures. Testing under various conditions confirmed that the prototype performed reliably in nearly all scenarios. The results demonstrated real-time gesture recognition with high accuracy, with only a minor delays in robotic hand movement due to the intentional sequential operation of the servos to prevent interference during execution. During development we encountered several problems, such as twisting and stretching of the lines, excessive servo rotation, and others, which were progressively

addressed and resolved. In the future we plan to evaluate the prototype in user studies to explore the effects of embodied signals for non-verbal remote communication.

## 7. Acknowledgments

This research was funded by the Slovenian Research Agency, grant number P5-0433, IO-0035, N2-0354, J5-50155 and J7-50096. This work has also been supported by the research program CogniCom (0013103) at the University of Primorska.

## References

- [1] A. Kelmaganbetova, S. Mazhitayeva, B. Ayazbayeva, G. Khamzina, Z. Ramazanova, S. Rahymberlina, Z. Kadyrov, "the role of gestures in communication", *Theory & Practice in Language Studies (TPLS)* 13 (2023).
- [2] M. Studdert-Kennedy, Hand and mind: What gestures reveal about thought., *Language and Speech* 37 (1994) 203–209.
- [3] M. L. Knapp, J. A. Hall, T. G. Horgan, *Nonverbal Communication in Human Interaction*, Cengage Learning, 2013.
- [4] V. I. Pavlovic, R. Sharma, T. S. Huang, Visual interpretation of hand gestures for human-computer interaction: A review, *IEEE Transactions on pattern analysis and machine intelligence* 19 (2002) 677–695.
- [5] K. Čopič Pucihar, N. T. Attygalle, M. Kljun, C. Sandor, L. A. Leiva, Solids on soli: Millimetre-wave radar sensing through materials, *Proceedings of the ACM on Human-Computer Interaction* 6 (2022) 1–19.
- [6] J. Frey, M. Grabli, R. Slyper, J. R. Cauchard, Breeze: Sharing biofeedback through wearable technologies, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.
- [7] J. Wainer, D. J. Feil-Seifer, D. A. Shell, M. J. Mataric, The role of physical embodiment in human-robot interaction, in: *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, 2006, pp. 117–122.
- [8] J. Qi, L. Ma, Z. Cui, Y. Yu, Computer vision-based hand gesture recognition for human-robot interaction: a review, *Complex & Intelligent Systems* 10 (2024) 1581–1606.
- [9] M. Kim, J. Cho, S. Lee, Y. Jung, Imu sensor-based hand gesture recognition for human-machine interfaces, *Sensors* 19 (2019) 3827.
- [10] G. Rodriguez, N. Jofre, Y. Alvarado, J. Fernández, R. Guerrero, Gestural interaction for virtual reality environments through data gloves, *Advances in Science, Technology and Engineering Systems Journal* 2 (2017) 284–290.
- [11] S. Ni, M. A. Al-qaness, A. Hawbani, D. Al-Alimi, M. Abd Elaziz, A. A. Ewees, A survey on hand gesture recognition based on surface electromyography: Fundamentals, methods, applications, challenges and future trends, *Applied Soft Computing* 166 (2024) 112235.
- [12] F. Miao, Y. Huang, Z. Lu, T. Ohtsuki, G. Gui, H. Sari, Wi-fi sensing techniques for human activity recognition: Brief survey, potential challenges, and research directions, *ACM Computing Surveys* 57 (2025) 1–30.
- [13] M. Vuskovic, S. Du, Classification of prehensile emg patterns with simplified fuzzy artmap networks, in: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, volume 3, IEEE, 2002, pp. 2539–2544.
- [14] Y. Wu, S. Liang, L. Zhang, Z. Chai, C. Cao, S. Wang, Gesture recognition method based on a single-channel semg envelope signal, *EURASIP Journal on Wireless Communications and Networking* 2018 (2018) 35.
- [15] Q. Pu, S. Gupta, S. Gollakota, S. Patel, Whole-home gesture recognition using wireless signals,

- in: Proceedings of the 19th annual international conference on Mobile computing & networking, 2013, pp. 27–38.
- [16] O. Zhang, K. Srinivasan, Mudra: User-friendly fine-grained gesture recognition using wifi signals, in: Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies, 2016, pp. 83–96.
  - [17] Q. De Smedt, H. Wannous, J.-P. Vandeborre, Skeleton-based dynamic hand gesture recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 1–9.
  - [18] S. O. Adalgeirsson, C. Breazeal, Mebot: A robotic platform for socially embodied telepresence, in: 2010 5th ACM/IEEE international conference on human-robot interaction (HRI), IEEE, 2010, pp. 15–22.
  - [19] N. Babcock, Computer vision pipeline architecture: A tutorial. *Toptal Developers*, <https://www.toptal.com/computer-vision/computer-vision-pipeline>, n.d.
  - [20] IBM, What is computer vision?, <https://www.ibm.com/think/topics/computer-vision>, 2021.
  - [21] Google AI for Developers, Hand landmarks detection guide, [https://ai.google.dev/edge/mediapipe/solutions/vision/hand\\_landmarker](https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker), 2025.
  - [22] InMoov, Project, <https://inmoov.fr/project/>, 2013.
  - [23] A. Kendon, *Gesture: Visible Action as Utterance*, Cambridge University Press, 2004.
  - [24] People’s History Museum, The raised fist: a history of the symbol, <https://phm.org.uk/blogposts/the-raised-fist-a-history-of-the-symbol/>, 2023.
  - [25] J. B. Kyle Davidson, Semiotic analysis of the raised fist emoji as a sign of resilience, Researcher Gate (2018).
  - [26] E-Book, How to read others’ thoughts by their gestures page 6, <https://www.iwaha.com/ebook/index.php?p=6>, n.d.
  - [27] D. Thompson, Gesture gaffes: Avoiding cultural faux pas with these 5 hand gestures. *Fork Union Military Academy*, <https://www.forkunion.com/gesture-gaffes-the-art-of-avoiding-cultural-faux-pas/>, 2025.
  - [28] RyeBread, What is the origin of the phrase “pinky promise”? *Stack Exchange*, <https://english.stackexchange.com/questions/125205/what-is-the-origin-of-the-phrase-pinky-promise>, 2013.
  - [29] Pollice Verso, The gladiator and the thumb, [https://penelope.uchicago.edu/~grout/encyclopaedia\\_romana/gladiators/polliceverso.html](https://penelope.uchicago.edu/~grout/encyclopaedia_romana/gladiators/polliceverso.html), n.d.
  - [30] M. Fabry, Thumbs-up, thumbs-down meaning: Where do gestures come from?. *TIME*, <https://time.com/4984728/thumbs-up-thumbs-down-history/>, 2017.
  - [31] Icons of England, The v-sign, <https://web.archive.org/web/20080621122852/http://www.icons.org.uk/theicons/collection/the-v-sign/a-harvey-smith-to-you/the-asian-v-sign-in-progress>, n.d.
  - [32] lingvano, “i love you” in sign language, <https://www.lingvano.com/asl/blog/i-love-you-in-sign-language/>, 2020.
  - [33] Wikipedia, Waving, <https://en.wikipedia.org/wiki/Waving>, 2014.

# Designing the Ideal Political Identity Questionnaire Using Machine Learning and Ideology Scales

Ana Nikolić<sup>1</sup>, Uroš Sergaš<sup>1</sup> and dr. Marko Tkalčič<sup>1</sup>

<sup>1</sup>University of Primorska, FAMNIT, Glagoljaška 8, SI-6000 Koper, Slovenia

## Abstract

Political ideology shapes beliefs, behavior and attitudes toward society. Many existing questionnaires for measuring ideology are lengthy, repetitive, and misaligned with self-perception. This paper investigates whether a shorter, reliable, two-dimensional political identity questionnaire can be created using machine learning and psychometric methods. Sixty participants completed four ideological instruments (MFQ, SDO, RWA and 8values). Lasso regression and Random Forest with nested cross-validation identified predictive items, while psychometric evaluation included CFA and Cronbach's alpha. Random Forest outperformed Lasso. Internal reliability was excellent and factor loadings supported a two-factor structure despite moderate model fit. Findings show that ideology can be measured efficiently with reduced items, supporting applications in research, digital platforms and political psychology.

## Keywords

political ideology, questionnaire design, machine learning, psychometrics

## 1. Introduction

Political ideology shapes how individuals think, act and relate to society. It plays a key role in explaining attitudes toward morality, authority and state control. To measure ideology, psychology and political science rely on questionnaires such as the Moral Foundations Questionnaire (MFQ)[2], the Social Dominance Orientation scale (SDO)[3] and the Right-Wing Authoritarianism scale (RWA)[1]. Each of these focuses on different aspects: morality, authority or social hierarchies. Many people are also familiar with popular online quizzes, such as the 8values test or Political Compass, which aim to place users on a political map in a simplified and engaging way. In practice, however, these instruments are often long, redundant and not always aligned with how people perceive their own political orientation.

Recent work in computational social science suggests that machine learning can be used to shorten and refine such questionnaires, while psychometric methods ensure their reliability and validity. This approach has been successfully applied to instruments like the SOAPP-R (The Screener and Opioid Assessment for Patients with Pain-Revised) [7], where machine learning-based pruning was followed by traditional psychometric validation, confirmatory factor analysis (CFA), internal consistency measures and item-total correlations. Yet few studies have attempted to combine multiple ideological instruments and reduce them into a concise tool that still captures the two widely accepted dimensions of ideology: the liberal-conservative axis and the libertarian-authoritarian axis.

This paper addresses this gap by testing whether four ideological instruments (MFQ, SDO, RWA and a subset of 8values) can be reduced into a shorter, valid and reliable two-dimensional political identity questionnaire. We integrate machine learning and psychometric analysis to explore more effective methods for measuring political ideology. The goal is to create a more concise and practical tool that can be used in surveys, academic research or online platforms. The result is not just a shorter questionnaire, but also a better understanding of which questions matter most and why.

The analysis is guided by the following research questions (RQ):

**RQ1** To what extent do political identity questionnaires align with participants' self-identified political ideology?

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

✉ 89211089@student.upr.si (A. Nikolić); uros.sergas@upr.si (U. Sergaš); marko.tkalcic@famnit.upr.si (dr. M. Tkalčič)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.11>

- RQ2** How strongly are the liberal-conservative and libertarian-authoritarian dimensions captured by each questionnaire?
- RQ3** Do demographic factors influence participants' ideological scores or questionnaire preferences?
- RQ4** Can a concise, valid, and reliable two-dimensional political ideology questionnaire be constructed by combining the most predictive and non-redundant items from our questionnaires?
- RQ5** Which specific items are most predictive of self-identified ideology as determined through machine learning?

These questions are designed to move step by step from understanding how current questionnaires perform to exploring how they can be improved. The first two ask whether the instruments actually reflect how people see themselves and how strongly they capture the two main ideological dimensions. The third looks at whether factors like gender play a role in shaping results. Finally, the last two questions focus on building something new: a shorter, more practical questionnaire that still works well, and identifying which items really carry the most weight.

## 2. Methods

In this section, we describe the methodology employed in this paper, focusing on data collection, the use of machine learning techniques to refine existing political ideology questionnaires and the subsequent psychometric evaluation of the reduced item sets.

### 2.1. Participants and Instruments

To investigate how well different ideological questionnaires capture political self-identification, we designed a custom online survey. Sixty adult participants completed an online survey that began with a short demographic section and then included four ideological instruments: the 30-item Moral Foundations Questionnaire (MFQ), the 16-item Social Dominance Orientation scale (SDO), the 20-item Right-Wing Authoritarianism scale (RWA) and 37 items from the government axis of the 8values test, a widely used online questionnaire. Together, these produced 103 predictors. Participants also self-placed on two seven-point scales: one for liberal-conservative orientation and another for libertarian-authoritarian orientation. This allowed us to later compare participants' self-placement with their score-based ideological profiles. When choosing which questionnaires to include, the aim was to combine trusted scientific tools with at least one widely recognised online questionnaire.

### 2.2. Predictive Modelling

We modelled political self-identification as a continuous outcome and applied two algorithms: Lasso regression and Random Forest regression. Lasso was selected for its ability to perform feature selection by shrinking irrelevant coefficients to zero, while Random Forest was chosen for its capacity to capture non-linear relationships and provide robust variable importance rankings. Before modelling, all responses were converted to numeric values. Reverse-coded items were corrected and scales were aligned so that higher scores consistently represented stronger endorsement of the relevant ideological stance. Items were then standardised within each training fold since models were evaluated using nested cross-validation and replicated across twenty random seeds to ensure reliable and consistent results. To preserve interpretability, we avoided dimensionality-reduction techniques such as principal component analysis and refrained from aggressive imputation. This approach made the most of our limited data and helped avoid overly positive performance results by clearly separating model training from testing. Model performance was evaluated with explained variance ( $R^2$ ), mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE). MFQ and SDO were used as predictors for the liberal-conservative axis, while RWA and 8values were used for the libertarian-authoritarian axis.

### 2.3. Psychometric Evaluation

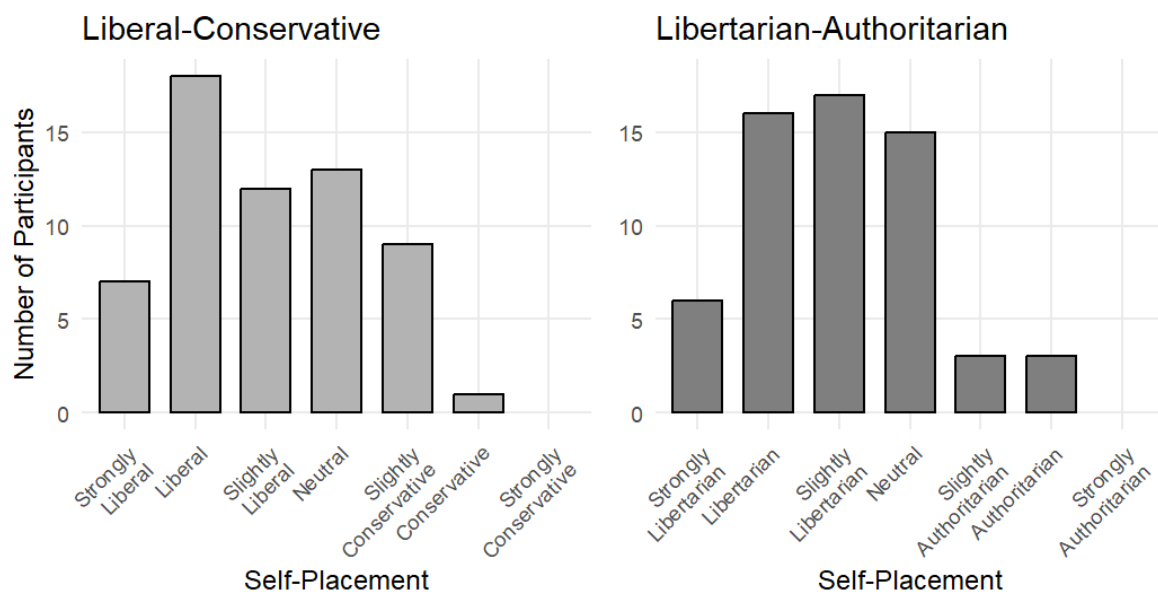
The reduced item sets obtained from modelling were evaluated using Confirmatory Factor Analysis (CFA) and internal consistency testing. CFA assessed whether the items fit the expected two-factor ideological structure, while Cronbach's alpha measured the internal consistency of the reduced scales within each dimension. Items with negative loadings or strong cross-loadings were candidates for exclusion.

## 3. Results

This section presents the results, structured around the five research questions. Each subsection addresses one question, with the final subsection combining RQ4 and RQ5 due to their conceptual overlap. The results include statistical analyses, predictive modelling, psychometric validation and group comparisons, with key findings highlighted for each research question.

### 3.1. RQ1: Alignment with self-identified ideology

The first research question asked how well the four questionnaires aligned with participants' self-identified political orientation. To assess this alignment, Spearman's rank correlation coefficients were calculated between each questionnaire's total score and the corresponding self-placement scale. Before testing correlations, we examined the distribution of self-placement on the two axes. As shown in Figure 1, most participants identified as *Liberal* and *Libertarian* and it is important to mention that no participants selected the *Strongly Conservative* or *Strongly Authoritarian* options, indicating a lack of representation at the extreme ends of the ideological spectrum.



**Figure 1:** Distribution of self-identified ideological positions along the liberal-conservative and libertarian-authoritarian axes.

For the liberal-conservative axis, the MFQ ideology score (binding minus individualising foundations) showed a strong positive correlation with self-placement ( $\rho = 0.73$ ,  $p < .001$ ). SDO also correlated positively with liberal-conservative self-placement ( $\rho = 0.56$ ,  $p < .001$ ), indicating that individuals who preferred hierarchy and inequality were more likely to identify as conservative. MFQ and SDO were themselves correlated ( $\rho = 0.70$ ), which suggests that although the two scales measure different concepts, they move together.

For the libertarian-authoritarian axis, RWA scores strongly correlated with authoritarian self-placement ( $\rho = 0.68, p < .001$ ). The 8values government scale showed the expected negative correlation ( $\rho = -0.72, p < .001$ ), with higher scores linked to lower authoritarian identification. RWA and 8values were strongly negatively correlated ( $\rho = -0.87$ ), confirming that they represent opposite ends of the same underlying dimension of authority and state power.

### 3.2. RQ2: Capturing the ideological axes

The second research question examined how strongly each questionnaire predicted participants' self-identified position on the two dimensions. We trained Random Forest regression models separately for each instrument and for the theoretically grounded combinations (MFQ + SDO on the liberal-conservative axis, RWA + 8values on the libertarian-authoritarian axis). To get more reliable results, we ran each model using multiple random seeds and then averaged the results.

#### 3.2.1. Liberal-Conservative axis

For each model, the dataset was repeatedly split into training (75%) and testing (25%) sets using a stratified split based on participants' self-identified liberal-conservative placement. Stratified splits preserved the original ideological distribution, reducing bias and improving generalizability. Stratification helped maintain diversity and avoid biased or unbalanced models. The MFQ-based model achieved an average  $R^2$  value of 0.43, indicating that moral foundation scores explained approximately 43% of the variance in participants' self-identified political orientation, along the liberal-conservative axis. The SDO-only model performed slightly lower, with an average test  $R^2$  of 0.32, suggesting that preferences for group-based hierarchy and inequality also contribute meaningfully, though somewhat less strongly, to ideological self-placement. When both instruments were combined, the predictive performance improved to an average  $R^2$  value of 0.47, indicating that the two scales capture complementary aspects of the liberal-conservative spectrum. Additional model performance metrics are summarised in Table 1.

**Table 1**

Predictive performance of models for liberal-conservative axis

| Model     | Test $R^2$ | MAE  | MSE  | RMSE |
|-----------|------------|------|------|------|
| MFQ       | 0.43       | 0.74 | 0.83 | 0.90 |
| SDO       | 0.32       | 0.81 | 0.98 | 0.98 |
| MFQ + SDO | 0.47       | 0.70 | 0.77 | 0.87 |

#### 3.2.2. Libertarian-Authoritarian axis

The RWA-only model achieved an average  $R^2$  value of 0.36, indicating that attitudes related to submission, norm adherence and perceived threat serve as moderate predictors of authoritarian self-placement. The 8values model, which measured participants' agreement with government control and state power, resulted in an average  $R^2$  of 0.42. When the two instruments were combined, predictive performance improved further to an average  $R^2$  value of 0.43. This small improvement suggests that although RWA and 8values approach authoritarianism from different perspectives, they largely measure overlapping aspects of the libertarian-authoritarian dimension. Additional model performance metrics are summarised in Table 2.

**Table 2**

Predictive performance of models for libertarian-authoritarian axis

| Model         | Test $R^2$ | MAE  | MSE  | RMSE |
|---------------|------------|------|------|------|
| RWA           | 0.36       | 0.76 | 0.93 | 0.95 |
| 8values       | 0.42       | 0.74 | 0.85 | 0.91 |
| RWA + 8values | 0.43       | 0.74 | 0.83 | 0.90 |

The combination of RWA and 8values produced the highest predictive performance for the libertarian-authoritarian axis, though the gain over using 8values alone was minimal. This suggests that while psychological and value-based dimensions offer slightly complementary insights, much of the explanatory power is already captured by 8values.

### 3.3. RQ3: Demographic effects

The third research question examined whether demographic factors influenced participants' ideology scores, self-identification or questionnaire preferences. Most participants came from Serbia, with a small number from Bosnia and Herzegovina. The sample consisted of individuals in their university years, with most of them being current students or recent graduates. Participants were recruited using a convenience sampling method that targeted university students through peer networks. Because the sample was relatively homogeneous in nationality, age and education, only gender differences were analysed (45 women, 15 men).

#### 3.3.1. Ideological scores

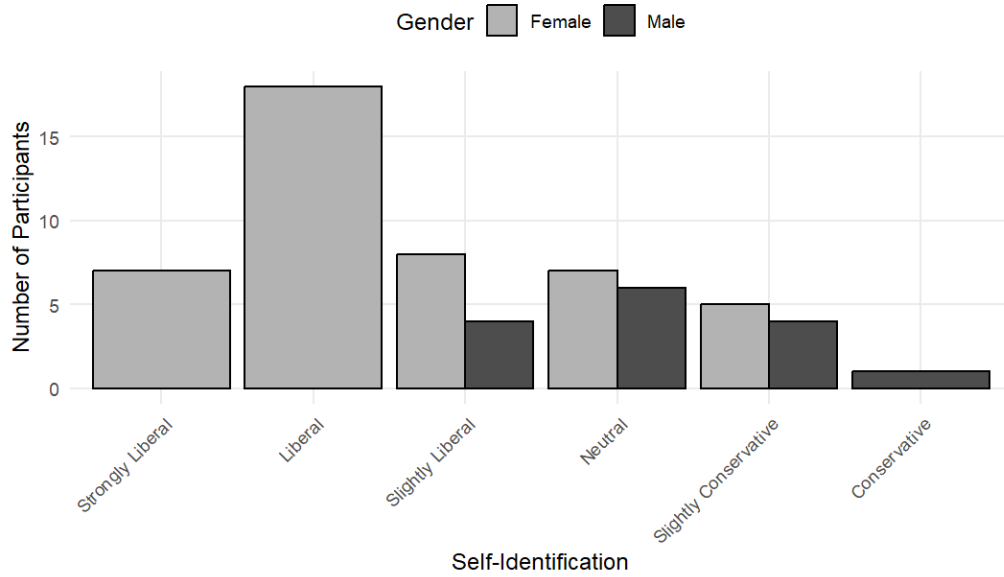
To assess whether men and women differed in their responses to the ideological questionnaires, we performed independent-samples t-tests on the total scores for MFQ, SDO, RWA and 8values. The results revealed statistically significant gender differences on three out of the four instruments, as shown in Table 3. Women scored more liberal on MFQ, less authoritarian on RWA and more libertarian on the 8values government axis. No significant gender difference was found for SDO.

**Table 3**  
Gender differences in political ideology scores across measures

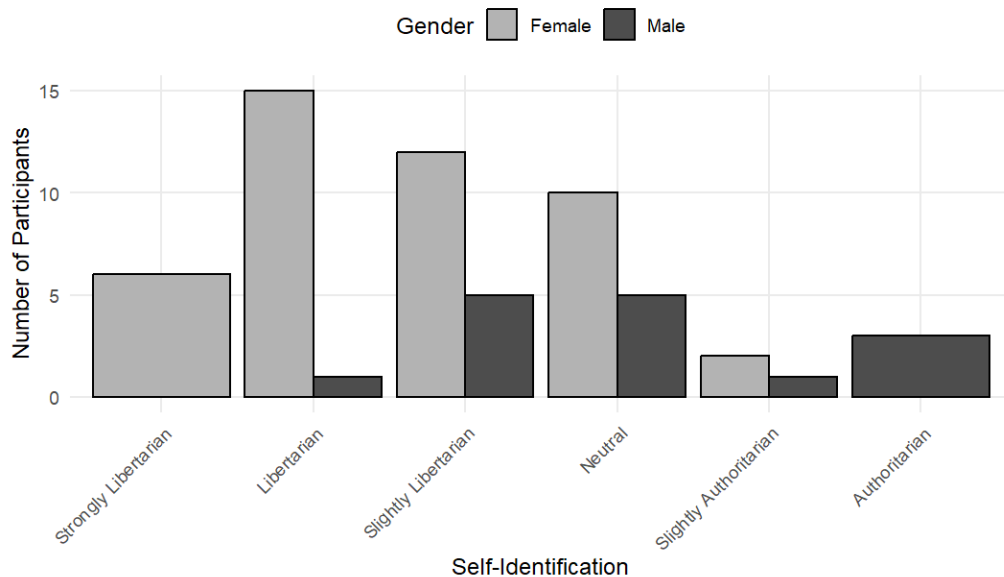
| Measure             | $t(df)$       | $p$ -value | Mean (Women) | Mean (Men) | Interpretation                                        |
|---------------------|---------------|------------|--------------|------------|-------------------------------------------------------|
| MFQ                 | -3.75 (32.04) | 0.0007     | -1.37        | -0.58      | Women scored significantly more liberal               |
| SDO                 | -1.57 (29.16) | 0.128      | 2.58         | 3.07       | No significant difference                             |
| RWA                 | -2.13 (23.17) | 0.034      | 70.49        | 89.40      | Women scored significantly lower (less authoritarian) |
| 8values (gov. axis) | 3.69 (26.60)  | 0.001      | 56.52        | 46.20      | Women scored significantly higher (more libertarian)  |

#### 3.3.2. Political self-identification

To assess whether gender was associated with differences in political self-identification, we conducted Fisher's Exact Tests separately for each ideological axis. Women were more likely to place themselves as *Liberal* and *Strongly Liberal* on the liberal-conservative axis, while men were more likely to select *Conservative*. This pattern is also evident in Figure 2, which shows the distribution of self-placements by gender across the liberal-conservative dimension. On the libertarian-authoritarian axis, women predominantly identified as *Libertarian* or *Strongly Libertarian*, whereas three men identified as *Authoritarian*. This distribution is shown in Figure 3. Both effects were statistically significant (lib-con:  $p = .0009$ ; lib-auth:  $p = .011$ ). Overall, women in this sample were more likely to identify as *Liberal* and *Libertarian* compared to men. As with the questionnaire scores, the uneven gender distribution limits the generalizability of these findings and future studies with larger and more balanced samples are needed to validate these patterns in political self-identification.



**Figure 2:** Distribution of self-identified liberal-conservative ideology by gender.



**Figure 3:** Distribution of self-identified libertarian-authoritarian ideology by gender.

### 3.3.3. Questionnaire preferences

To examine whether participants' preferences for the ideological questionnaires differed by gender, we conducted Fisher's Exact Tests separately for each political dimension. No significant gender differences were found in preferences between MFQ and SDO ( $p = .340$ ). However, preferences between RWA and 8values differed significantly by gender ( $p = .035$ ). However, to understand exactly how men and women differed in their preferences, a more detailed breakdown is needed. Because the sample was small and uneven (45 women and 15 men), these results should be interpreted with caution.

### 3.4. RQ4 & RQ5: Constructing a concise two-dimensional questionnaire

The fourth and fifth research questions asked whether a shorter two-dimensional questionnaire could be created and which specific items were most predictive of self-identification. We applied Lasso and

Random Forest models to select items and then validated the reduced scales psychometrically.

### 3.4.1. Item selection and predictive performance

Lasso regression selected 21 items for the liberal-conservative axis (14 MFQ, 7 SDO) and 25 items for the libertarian-authoritarian axis (9 RWA, 16 8values). Predictive performance was moderate. Summary of selected item counts and performance metrics for each axis is presented in Table 4.

**Table 4**

Model performance metrics across ideological axes

| Axis                      | Questionnaires Used | No. of Items | $R^2$ | MAE   | MSE   | RMSE  |
|---------------------------|---------------------|--------------|-------|-------|-------|-------|
| Liberal-Conservative      | MFQ + SDO           | 21           | 0.321 | 0.866 | 1.130 | 1.050 |
| Libertarian-Authoritarian | RWA + 8values       | 25           | 0.384 | 0.740 | 0.915 | 0.940 |

Random Forest achieved stronger results with similar item counts: 20 items (8 MFQ, 12 SDO) for the liberal-conservative axis and 24 items (14 RWA, 10 8values) for the libertarian-authoritarian axis. These models explained 45% and 47% of the variance respectively, outperforming Lasso and reducing error metrics. A summary of selected item counts and performance metrics from both Random Forest models is provided in the Table 5.

**Table 5**

Random Forest model performance across ideological axes

| Axis                      | Questionnaires Used | No. of Items | $R^2$ | MAE   | MSE   | RMSE  |
|---------------------------|---------------------|--------------|-------|-------|-------|-------|
| Liberal-Conservative      | MFQ + SDO           | 20           | 0.454 | 0.740 | 0.936 | 0.946 |
| Libertarian-Authoritarian | RWA + 8values       | 24           | 0.473 | 0.503 | 0.450 | 0.663 |

Random Forest got better results than Lasso across all performance metrics for both ideological dimensions. It achieved higher  $R^2$  values, indicating stronger explanatory power. Additionally, Random Forest produced lower error rates (MAE, MSE and RMSE), suggesting more accurate and stable predictions. Because political self-placement is shaped by many overlapping factors, predictive fits in this area are typically modest;  $R^2$  values around 0.10-0.50 are commonly considered reasonable for human-behaviour data[5]. These results suggest that Random Forest outperformed Lasso because it can better capture complex, non-linear patterns in participants' ideological responses. While Lasso regression is more appropriate for identifying a sparse set of linear predictors, its linear assumptions may have limited its effectiveness in this context, where non-linear relationships and interactions between questionnaire items likely play an important role.

### 3.4.2. Most predictive items

Using Random Forest feature importance (averaged across folds and 20 seeds, rescaled so that the top item equals 100), the strongest predictors reflected the theoretical focus of each instrument. For clarity and consistency, we refer to each question by its item code (e.g., MFQ15, SDO7, RWA12, VAL26). These codes correspond to the variable names used in the statistical analysis, while the full wording of each item is provided in the Appendix.

On the liberal-conservative axis, Random Forest selected 20 items in total:

- MFQ items: MFQ9, MFQ15, MFQ20, MFQ23, MFQ24, MFQ25, MFQ27, MFQ30
- SDO items: SDO1, SDO3, SDO4, SDO5, SDO6, SDO7, SDO8, SDO9, SDO10, SDO12, SDO13, SDO16

On the libertarian-authoritarian axis, 24 items were selected:

- RWA items: RWA1, RWA2, RWA3, RWA4, RWA8, RWA10, RWA11, RWA12, RWA13, RWA14, RWA15, RWA17, RWA19, RWA20
- 8values items: VAL3, VAL11, VAL23, VAL25, VAL26, VAL28, VAL29, VAL32, VAL36, VAL37

### 3.4.3. Psychometric validation

Following the machine learning item selection procedures, we used Confirmatory Factor Analysis (CFA) to test whether the reduced item set aligned with the hypothesised two-dimensional ideological structure.

The initial set of 44 items was derived from Random Forest feature selection on MFQ + SDO (liberal-conservative axis) and RWA + 8values (libertarian-authoritarian axis). Preliminary correlations had already indicated that some items were highly redundant (e.g., SDO9/SDO10, VAL29/VAL36) and this was later confirmed during CFA. After inspecting loadings and modification indices, one problematic item (MFQ15) was removed, leaving 43 items (19 and 24 per factor). To improve model fit, a few residual correlations were allowed between highly similar items, reducing misfit caused by shared variance not explained by the main factors. The final CFA model, estimated with maximum likelihood, showed moderate fit (CFI = 0.707, TLI = 0.688, RMSEA = 0.115, SRMR = 0.102), which is expected given the small sample size ( $n = 60$ ) and high item count. Fixed cutoff values for fit indices like RMSEA and SRMR may be misleading when applied universally[6]. Importantly, factor loadings were generally strong and aligned with theory, with most exceeding 0.5 and some particularly high (RWA14 = 0.875, VAL26 = 0.903). The two latent factors were moderately correlated ( $r = 0.57$ ), indicating that individuals with more conservative moral and social beliefs also tended to support stronger authority, while still supporting a two-dimensional structure of political ideology.

To evaluate the internal consistency of the shortened scales, we computed Cronbach's alpha for the final item subsets obtained from Random Forest selection. The liberal-conservative subset (19 items from MFQ and SDO) achieved  $\alpha = 0.91$ , with average inter-item correlations of 0.36 and item-total correlations ranging from 0.36 to 0.73, showing that all items contributed positively. The libertarian-authoritarian subset (24 items from RWA and 8values) achieved  $\alpha = 0.92$ , with average inter-item correlations of 0.44 and item-total correlations from 0.43 to 0.81, again indicating strong coherence. Both coefficients exceed the 0.80[4] threshold for high-stakes instruments, confirming that the reduced item sets remained reliable and internally consistent. Together with the CFA results, these findings suggest that the shortened scales form valid and dependable measures of ideological orientation on both dimensions.

Overall, the machine-learning approach successfully reduced the original 103 items to about 20-24 per axis without compromising validity or reliability, providing a concise two-dimensional questionnaire.

## 4. Discussion

This paper explored how political identity can be predicted and efficiently measured using a combination of scientific ideological questionnaires (MFQ, SDO, RWA) and the non-academic 8values test. Five research questions guided the study, ranging from predictive alignment to psychometric reduction.

### 4.1. RQ1: Alignment with self-identified ideology

The four questionnaires showed moderate to strong alignment with participants' political self-identification. MFQ and SDO were both correlated with the liberal-conservative axis, while RWA and 8values strongly captured the libertarian-authoritarian dimension. These findings confirm that traditional psychometric tools and even a popular online quiz can reflect how individuals perceive their own ideology.

### 4.2. RQ2: Capturing the two dimensions

Random Forest models demonstrated that MFQ and SDO together explained nearly half of the variance in liberal-conservative placement, while RWA and 8values provided similarly strong predictions for the libertarian-authoritarian axis. The results suggest that combining instruments adds complementary

information, but most explanatory power was already captured by the strongest single scale in each pair.

### 4.3. RQ3: Demographic influences

Gender differences emerged consistently. Women in the sample were more likely to identify as *Liberal* and *Libertarian* and their scores on MFQ, RWA and 8values reflected this orientation. Men leaned relatively more toward *Conservatism* and *Authoritarianism*. These effects should be interpreted cautiously, given the small and unbalanced sample, but they point to meaningful demographic influences.

### 4.4. RQ4: Constructing a concise questionnaire

Machine learning successfully reduced the original 103 items to 20-25 per axis without major losses in predictive power. Random Forest in particular outperformed Lasso, highlighting its advantage in modelling complex, non-linear patterns. The shortened scales captured most of the variance explained by the full instruments while significantly reducing item burden.

### 4.5. RQ5: Most predictive items

The strongest predictors mapped neatly onto theoretical expectations. On the liberal-conservative axis, moral foundations related to harm and fairness and SDO hierarchy items were most important. On the libertarian-authoritarian axis, RWA authority items and 8values statements about government control stood out. These results show that a small number of well-chosen items can represent the core of ideological measurement.

### 4.6. Overall interpretation

Taken together, the findings demonstrate that it is possible to build a shorter and reliable two-dimensional questionnaire for political identity. The proposed two-dimensional political identity questionnaire consists of 43 items: 19 capturing the liberal-conservative dimension (from MFQ and SDO) and 24 capturing the libertarian-authoritarian dimension (from RWA and 8values). Each item preserves its original phrasing, but the full set can be completed in about 7-10 minutes, compared to over 25 minutes for the original instruments. For example, liberal-conservative items focus on moral and social hierarchy statements such as "People should not do things that are disgusting, even if no one is harmed" or "It's probably a good thing that certain groups are at the top and other groups at the bottom," while libertarian-authoritarian items address attitudes toward governance and tradition, such as "Religion should play a role in government" or "Same-sex marriage should be legal."

The reduced item set provides a solid basis for a short-form political identity questionnaire. Such a tool would be useful in time-limited surveys, quick online quizzes or research contexts requiring efficient measurement. By focusing on the most predictive and non-redundant items, the final version preserves accuracy while improving user experience and data quality, turning lengthy instruments into practical tools.

## 5. Limitations and Future Work

Several limitations should be acknowledged. This study was limited by a small and relatively homogeneous sample ( $n=60$ ), which restricts generalisability and statistical power. Most participants were young and politically engaged, leading to similar responses across some demographic categories. Predictive performance was moderate ( $R^2$  up to 0.47), suggesting that while the models captured meaningful variance, they could not fully account for the complexity of ideological self-identification. The selected items were optimised on this dataset and may not generalise to other populations and potential biases such as socially desirable responding or anchoring effects cannot be ruled out. It is also worth mentioning that the original full-length questionnaires, when put together in our survey,

included a total of 103 items. This was time-consuming and may have contributed to participant fatigue, highlighting the value of creating shorter yet reliable scales. Future research should replicate the approach with larger and more diverse samples, test the stability of the reduced scales over time (e.g. test-retest reliability) and assess external validity in relation to real-world political behaviours. Exploring alternative feature-selection methods and additional psychometric checks could further strengthen the instrument.

## 6. Conclusion

This paper demonstrates that a concise, psychometrically valid political identity questionnaire can be constructed by combining existing instruments and applying machine learning alongside psychometric evaluation. A reduced set of around 20 items per axis was sufficient to capture self-identified positions on both the liberal-conservative and libertarian-authoritarian dimensions while maintaining strong reliability. By shortening lengthy scales without sacrificing validity, the proposed approach highlights a practical way to improve participant experience and research efficiency, paving the way for more accessible and data-driven tools for measuring political ideology.

## References

- [1] B. Altemeyer, *The Authoritarians*, Department of Psychology, University of Manitoba, Winnipeg, Canada, 2006.
- [2] J. Graham, J. Haidt, B. A. Nosek, Liberals and conservatives rely on different sets of moral foundations, *J. Pers. Soc. Psychol.* 96 (2009) 1029–1046. doi:10.1037/a0015141
- [3] F. Pratto, J. Sidanius, L. M. Stallworth, B. F. Malle, Social dominance orientation: A personality variable predicting social and political attitudes, *J. Pers. Soc. Psychol.* 67 (1994) 741–763. doi:10.1037//0022-3514.67.4.741
- [4] D. George, P. Mallery, *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*, 16th ed., Routledge, New York, NY, 2019. doi:10.4324/9780429056765.
- [5] P. K. Ozili, The Acceptable R-Square in Empirical Modelling for Social Science Research, Munich Personal RePEc Archive, Paper No. 115769 (2023), pp. 1–22. URL: <https://mp.ra.ub.uni-muenchen.de/115769/>
- [6] K. Groskurth, M. Bluemke, C. M. Lechner, Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions, *Behav. Res. Methods* 56 (2024) 3891–3914. doi:10.3758/s13428-023-02193-3
- [7] M. D. Finkelman, N. Smits, R. J. Kulich, K. L. Zacharoff, B. E. Magnuson, H. Chang, J. Dong, and S. F. Butler, Development of short-form versions of the Screener and Opioid Assessment for Patients with Pain–Revised (SOAPP-R): A proof-of-principle study, *Pain Medicine* 18 (2017) 1292–1302. doi:10.1093/pm/pnw210

## A. Full Item List of Ideological Questionnaires

### Moral Foundations Questionnaire (MFQ):

- MFQ1 Whether or not someone suffered emotionally.
- MFQ2 Whether or not some people were treated differently than others.
- MFQ3 Whether or not someone's action showed love for his or her country.
- MFQ4 Whether or not someone showed a lack of respect for authority.
- MFQ5 Whether or not someone violated standards of purity and decency.
- MFQ6 Whether or not someone cared for someone weak or vulnerable.
- MFQ7 Whether or not someone acted unfairly.
- MFQ8 Whether or not someone did something to betray his or her group.
- MFQ9 Whether or not someone conformed to the traditions of society.
- MFQ10 Whether or not someone did something disgusting.
- MFQ11 Whether or not someone was cruel.
- MFQ12 Whether or not someone was denied his or her rights.
- MFQ13 Whether or not someone showed a lack of loyalty.
- MFQ14 Whether or not an action caused chaos or disorder.
- MFQ15 Whether or not someone acted in a way that God would approve of.
- MFQ16 Compassion for those who are suffering is the most crucial virtue.
- MFQ17 When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.
- MFQ18 I am proud of my country's history.
- MFQ19 Respect for authority is something all children need to learn.
- MFQ20 People should not do things that are disgusting, even if no one is harmed.
- MFQ21 One of the worst things a person could do is hurt a defenseless animal.
- MFQ22 Justice is the most important requirement for a society.
- MFQ23 People should be loyal to their family members, even when they have done something wrong.
- MFQ24 Men and women each have different roles to play in society.
- MFQ25 I would call some acts wrong on the grounds that they are unnatural.
- MFQ26 It can never be right to kill a human being.
- MFQ27 I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.
- MFQ28 It is more important to be a team player than to express oneself.
- MFQ29 If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.
- MFQ30 Chastity is an important and valuable virtue.

### Social Dominance Orientation (SDO):

- SDO1 Some groups of people are just more worthy than others.
- SDO2 In getting what you want, it is sometimes necessary to use force against other groups.
- SDO3 It's OK if some groups have more of a chance in life than others.
- SDO4 To get ahead in life, it is sometimes necessary to step on other groups.
- SDO5 If certain groups stayed in their place, we would have fewer problems.
- SDO6 It's probably a good thing that certain groups are at the top and other groups are at the bottom.
- SDO7 Inferior groups should stay in their place.
- SDO8 Sometimes other groups must be kept in their place.
- SDO9 It would be good if groups could be equal.

- SDO10 Group equality should be our ideal.
- SDO11 All groups should be given an equal chance in life.
- SDO12 We should do what we can to equalize conditions for different groups.
- SDO13 Increased social equality is beneficial to society.
- SDO14 We would have fewer problems if we treated people more equally.
- SDO15 We should strive to make incomes as equal as possible.
- SDO16 No group should dominate in society.

Right-Wing Authoritarianism (RWA):

- RWA1 Our country desperately needs a mighty leader who will do what has to be done to destroy the radical new ways and sinfulness that are ruining us.
- RWA2 Gays and lesbians are just as healthy and moral as anybody else.
- RWA3 It is always better to trust the judgement of the proper authorities in government and religion than to listen to the noisy rabble-rousers in our society who are trying to create doubt in people's minds.
- RWA4 Atheists and others who have rebelled against the established religions are no doubt every bit as good and virtuous as those who attend church regularly.
- RWA5 The only way our country can get through the crisis ahead is to get back to our traditional values, put some tough leaders in power and silence the troublemakers spreading bad ideas.
- RWA6 There is absolutely nothing wrong with nudist camps.
- RWA7 Our country needs free thinkers who have the courage to defy traditional ways, even if this upsets many people.
- RWA8 Our country will be destroyed someday if we do not smash the perversions eating away at our moral fiber and traditional beliefs.
- RWA9 Everyone should have their own lifestyle, religious beliefs, and sexual preferences, even if it makes them different from everyone else.
- RWA10 The "old-fashioned ways" and the "old-fashioned values" still show the best way to live.
- RWA11 You have to admire those who challenged the law and the majority's view by protesting for women's abortion rights, for animal rights, or to abolish school prayer.
- RWA12 What our country really needs is a strong, determined leader who will crush evil, and take us back to our true path.
- RWA13 Some of the best people in our country are those who are challenging our government, criticizing religion, and ignoring the "normal way things are supposed to be done."
- RWA14 God's laws about abortion, pornography and marriage must be strictly followed before it is too late, and those who break them must be strongly punished.
- RWA15 There are many radical, immoral people in our country today, who are trying to ruin it for their own godless purposes, whom the authorities should put out of action.
- RWA16 A "woman's place" should be wherever she wants to be. The days when women are submissive to their husbands and social conventions belong strictly in the past.
- RWA17 Our country will be great if we honor the ways of our forefathers, do what the authorities tell us to do, and get rid of the "rotten apples" who are ruining everything.
- RWA18 There is no "one right way" to live life; everybody has to create their own way.
- RWA19 Homosexuals and feminists should be praised for being brave enough to defy "traditional family values."
- RWA20 This country would work a lot better if certain groups of troublemakers would just shut up and accept their group's traditional place in society.

8values - Government Axis Items:

- VAL1 Oppression by corporations is more of a concern than oppression by governments.

VAL2 Tariffs on international trade are important to encourage local production.

VAL3 The United Nations should be abolished.

VAL4 Military action by our nation is often necessary to protect it.

VAL5 I support regional unions, such as the European Union.

VAL6 It is important to maintain our national sovereignty.

VAL7 Wars do not need to be justified to other countries.

VAL8 Military spending is a waste of money.

VAL9 Governments should be accountable to the international community.

VAL10 Even when protesting an authoritarian government, violence is not acceptable.

VAL11 My religious values should be spread as much as possible.

VAL12 Our nation's values should be spread as much as possible.

VAL13 It is very important to maintain law and order.

VAL14 The general populace makes poor decisions.

VAL15 Physician-assisted suicide should be legal.

VAL16 The sacrifice of some civil liberties is necessary to protect us from acts of terrorism.

VAL17 Government surveillance is necessary in the modern world.

VAL18 The very existence of the state is a threat to our liberty.

VAL19 Regardless of political opinions, it is important to side with your country.

VAL20 All authority should be questioned.

VAL21 A hierarchical state is best.

VAL22 It is important that the government follows the majority opinion, even if it is wrong.

VAL23 The stronger the leadership, the better.

VAL24 Democracy is more than a decision-making process.

VAL25 Children should be educated in religious or traditional values.

VAL26 Religion should play a role in government.

VAL27 Drug use should be legalized or decriminalized.

VAL28 Same-sex marriage should be legal.

VAL29 No cultures are superior to others.

VAL30 Sex outside marriage is immoral.

VAL31 If we accept migrants at all, it is important that they assimilate into our culture.

VAL32 Abortion should be prohibited in most or all cases.

VAL33 Gun ownership should be prohibited for those without a valid reason.

VAL34 Prostitution should be illegal.

VAL35 We should open our borders to immigration.

VAL36 All people - regardless of factors like culture or sexuality - should be treated equally.

VAL37 It is important that we further my group's goals above all others.

# Optimizing Product Catalogue Design: A Comparative Study of Traditional Photography and 3D Modeling

Simon Kolmanič<sup>1,†</sup>, Jan Hrašar<sup>†</sup>, Štefan Horvat<sup>†</sup> and Domen Mongus<sup>†</sup>

*University of Maribor, Faculty of Electrical Engineering and Computer Science, Koroška cesta 46, 2000 Maribor, Slovenia*

## Abstract

This study examines the replacement of traditional studio photography with a 3D modelling and rendering workflow in product catalogue production. Using the case of Urnes, a manufacturer of wooden urns, we compared the two approaches in terms of time, cost, and visual fidelity. The 3D pipeline, combining Blender, Substance 3D Painter, and InDesign, reduced preparation time per variation from 179 minutes to 36 minutes—a 4.9-fold improvement. After an initial 50-hour setup, subsequent catalogue updates achieved time savings of 97.6%. An informal interview with 38 professional users confirmed that 3D renderings matched or exceeded photographs in colour and texture accuracy. Results demonstrate that 3D workflows significantly lower production costs, accelerate updates, and support product personalisation, positioning them as a sustainable alternative to photography in catalogue design.

## Keywords

Product Catalogue Design, Computer Generated Imagery, Visual Fidelity, Cost Reduction

## 1. Introduction

A literature review shows that 3D modelling plays a crucial role in the production process along with various phases of engineering and assembly [1]. The increase in computer power and the development of advanced rendering algorithms have led to a rise in the visual quality of computer-generated imagery (CGI), which has begun to replace traditional product photographs in catalogues. Several online sources suggest that since 2014, 75% of the images in the IKEA catalogue have been computer-generated rather than photographed. Although this claim has not been confirmed by IKEA, there are several potential advantages to using CGI instead of photography. These include faster turnaround times for creating marketing materials, lower costs compared to product photography, and the ability to implement design changes quickly.

Unfortunately, there is a lack of sources that would enable a more comprehensive and detailed quantification of those advantages. To close this gap, we conducted a study that included the creation of a printed catalogue for a niche Slovenian company, Urnes, producing wooden urns.

Each year, the company releases a new catalogue of its products, showcasing 93 different urn models. The urns had to be manufactured first and then colored. After that, a motif was added by engraving or UV printing. Once these steps were completed, the urns were photographed in a controlled environment to ensure consistent lighting conditions for the entire collection. However, it was often necessary to retake several photographs to achieve an identical colour tone for the urns within the same collection.

After the photo shoot, the photographs were transferred to a computer, where colours were manually corrected and images cropped. Finally, reflections were added to enhance the urns' appeal for customers.

---

<sup>\*</sup>Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia

<sup>1</sup>Corresponding author.

<sup>†</sup>These authors contributed equally.

✉ simon.kolmanic@um.si (S. Kolmanič); jan.hrasar@student.um.si (J. Hrašar); stefan.horvat@um.si (Š. Horvat)

 0000-0002-0776-1860 (S. Kolmanič); 0009-0004-9885-7252 (Š. Horvat); 0000-0002-2160-0529 (D. Mongus)

 © 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.12>

The problem arises when specific urns within the collection were altered or when customers requested personalisation of the urns. In such instances, it was necessary to repeat the entire photography procedure along with all subsequent tasks for the entire collection.

We replaced the previous procedure with a new 3D pipeline that includes Blender for urn creation and Substance 3D Painter for texture generation. For designing the catalogue, we used Adobe InDesign. We measured the time required for each stage of this process and compared it to the traditional catalogue production method. To evaluate the quality of catalogue production, we conducted interviews with 38 customers. This allowed us to effectively assess the benefits of replacing product photographs with CGI, which is the primary focus of this article. The article is organised as follows: Section 2 presents the theoretical background, followed by a description of the methodology. Section 4 outlines the results, which are then discussed, leading to our conclusions.

## **2. Theoretical Background**

### **Visual Communication and Product Presentation**

Visual communication relies on the deliberate arrangement of images, colour, typography, and space to direct attention and support message clarity. In product catalogues, these principles ensure that buyers can quickly recognise product features and variations. Research has shown that hierarchy, balance, and contrast reduce cognitive load and improve comprehension [2], [3], [4]. 3D visualisation provides consistent lighting, accurate proportions, and scalable variations, which strengthen visual fidelity and brand coherence [5].

### **Development of 3D Visualisation Technologies**

Over the last decade, advances in rendering engines and physically based materials have enabled photorealistic outputs that can be indistinguishable from photography [6]. Early applications of CGI in advertising date back to the 1980s (e.g., Coca-Cola's Brilliance campaign), but only in the 2010s did CGI become economically viable for large-scale catalogue production. Media outlets such as Fast Company and The Verge reported that by 2014, approximately 75% of IKEA's catalogue product images were computer-generated, based on interviews with IKEA's visualisation managers. While this figure is widely cited, it originates from media reports rather than an officially published statement. Nevertheless, IKEA's adoption illustrates the growing maturity of CGI workflows for mass-market communication.

### **Consumer Perception of CGI and Realism**

The effectiveness of CGI-based images depends not only on cost efficiency but also on consumer trust and perception. Studies have shown that realism, telepresence, and interactivity improve engagement and purchasing intent [7], [8]. Other research demonstrates that viewers increasingly struggle to distinguish CGI from photography, provided that lighting and material textures are handled correctly [6]. However, limitations remain in colour calibration between screen and print, and in the representation of highly specific surface finishes, which may impact perceived authenticity.

### **Cost Efficiency and Workflow Optimisation**

While academic studies on cost reduction remain limited, industry reports consistently highlight the economic benefits of CGI. Nfinite reports [9] that retailers adopting CGI experienced up to an 85% decrease in visual budgets, 40% lower image creation costs, and 63% faster production

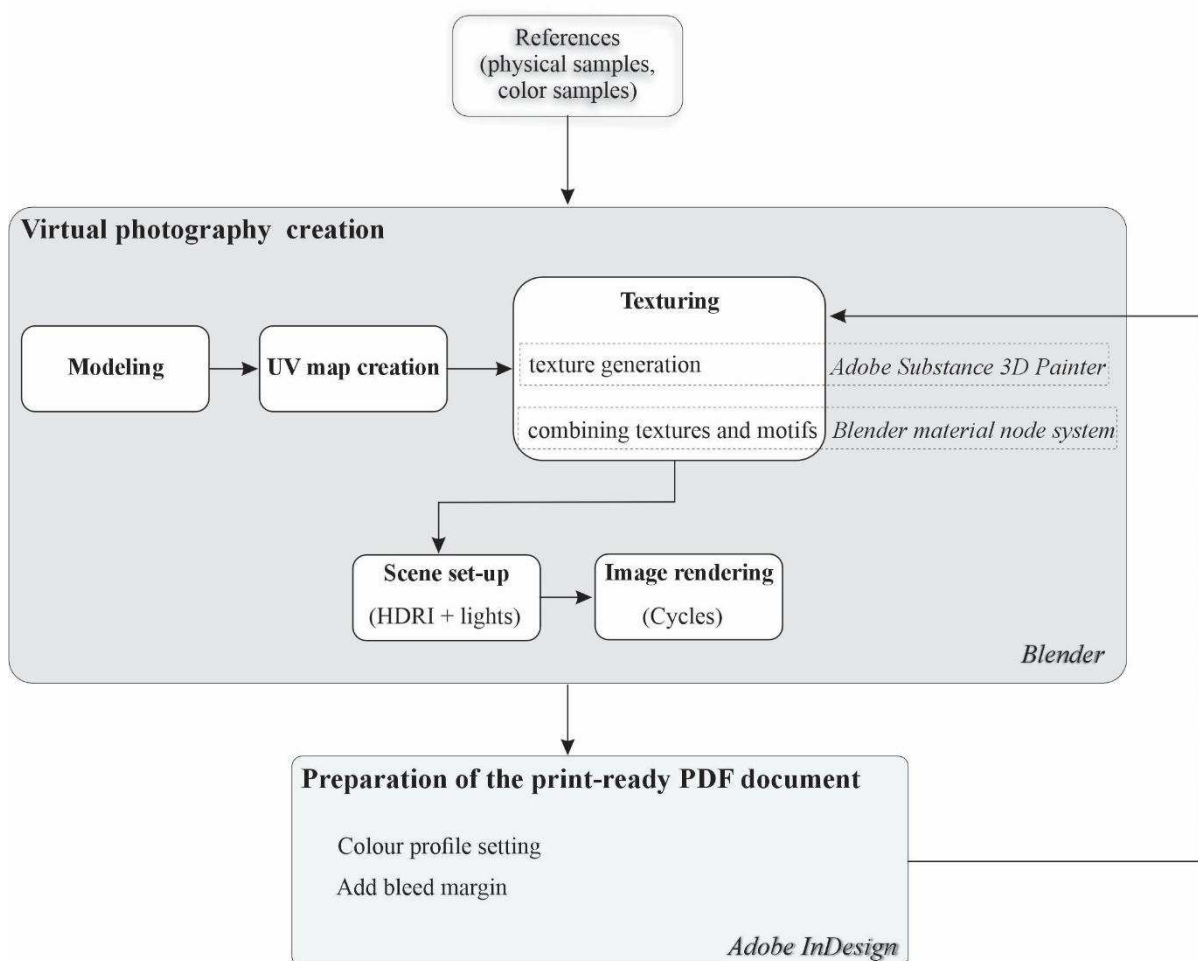
cycles. Similarly, Modelry/CGTrader [10] found that 3D virtual photography was up to six times cheaper than traditional photography for furniture products. These findings align with the concept of digital prototyping, where a single 3D asset can be reused across multiple variations, marketing channels, and even extended into AR/VR applications.

## Research Gap

Existing literature emphasises visual fidelity, consumer perception, and industrial case studies, but few peer-reviewed studies quantify cost and time savings in catalogue production using 3D workflows. This study addresses that gap by comparing traditional studio photography with a unified 3D pipeline in terms of preparation time, costs, and customer evaluations.

## 3. Methodology

The research was conducted as a case study of Urnes company catalogue. Urnes is a Slovenian company that produces wooden urns. Historically, the creation of these catalogues relied on studio photography, which necessitated physical prototypes, intricate lighting arrangements, and extensive post-production efforts. This study seeks to investigate the potential of employing a 3D visualisation workflow (see Figure 1) as a substitute for traditional methods. The primary objective is to decrease production time and costs while ensuring that visual fidelity is preserved or enhanced.



**Figure 1:** 3D visualisation workflow – from 3D model to printing.

The 3D pipeline incorporated several software tools to streamline the workflow. Blender was used to model the geometry of the urns and design the scene layouts. Adobe Substance 3D Painter was employed for material definition and texturing, allowing for a realistic simulation of wood surfaces and engraved motifs. Adobe InDesign was utilised for composing the catalogue, which included typography, page layout, and preparation for print.

This workflow enabled the reuse of base 3D assets across various product variations, making it highly adaptable for catalogue updates.

The effectiveness of the 3D workflow was assessed using three main metrics:

1. **Production time** – the duration required to prepare product variations and finalise catalogue pages, compared with traditional studio photography.
2. **Cost comparison** – an estimate of direct and indirect costs associated with both approaches, including labour, equipment, materials, and post-production.
3. **Customer evaluation** – assessment of perceived visual quality, realism, and usability of the catalogue by end users.

Data were collected using two complementary approaches:

- **Process logs** documented the steps and timing of catalogue preparation in both the photographic and 3D workflows, allowing for direct quantitative comparison.
- **Interviews** with 38 professional users, primarily funeral service providers who regularly interact with the catalogue, were conducted. Respondents were asked to evaluate the new 3D-based catalogue against previous photographic versions in terms of colour accuracy, texture fidelity, and response time to personalised orders.

## 4. Results

To evaluate the efficiency of the new 3D visualisation workflow, a direct comparison was made with the traditional photographic approach previously used in the Urnes product catalogue. The two methods were analysed along four dimensions: (1) process steps, (2) production time, (3) cost implications, and (4) visual and customer evaluation.

### 4.1. Process steps

The traditional catalogue production process at Urnes relied on studio photography. The workflow consisted of several sequential phases:

1. **Equipment setup** – A photographic studio was prepared with a camera, a tripod, and controlled lighting, as shown in Figure 2. This phase was performed once at the beginning of the session to ensure a consistent environment for capturing all urns.
2. **Photography** – Each urn variant was physically produced, placed in the studio, and photographed individually. Minor adjustments in lighting or positioning were often necessary to maintain uniformity across the catalog.
3. **Colour correction** – Captured images were imported to a computer and manually adjusted for colour balance, contrast, and tonal consistency. This step aimed to match the digital representation to the actual appearance of the physical product.

4. **Cropping** – Images were cropped to remove background elements and standardize framing.
5. **Addition of reflections** – Subtle reflections were manually created in 2D image editing software to simulate the impression of a product placed on a reflective surface. This was a labour-intensive step, often repeated for every product image.

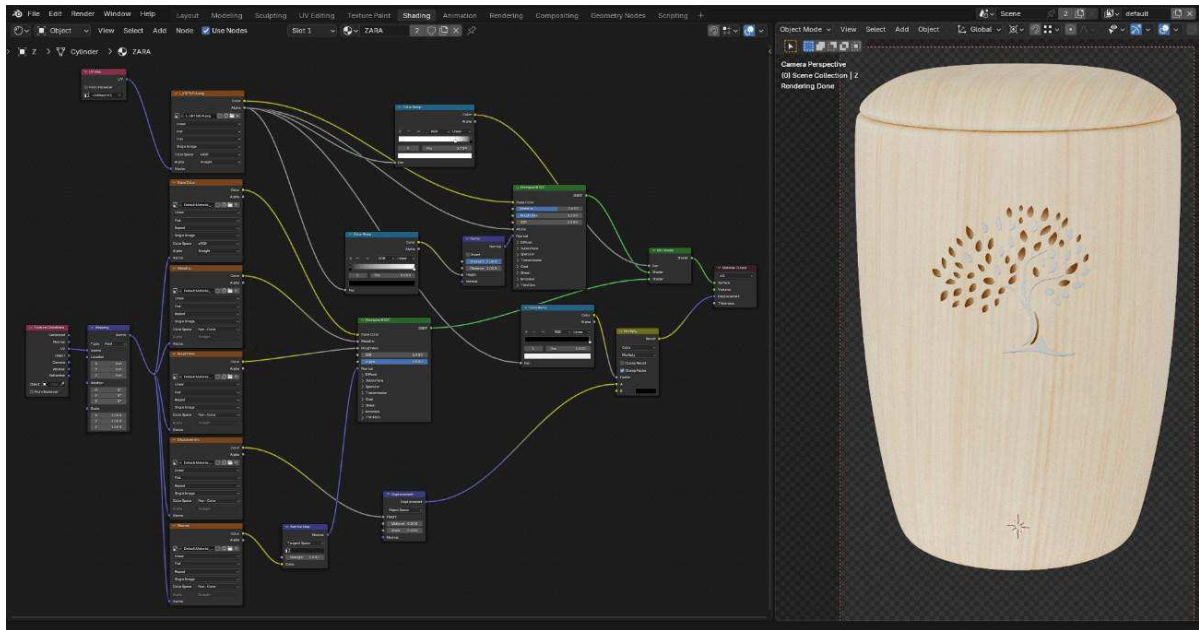


**Figure 2:** Capturing a photograph of a single urn as an example.

In the 3D visualisation pipeline, the following steps are involved, as illustrated in Figure 1:

1. **Modelling,**
2. **Texturing and adding motifs,**
3. **Scene setup,** and
4. **Image rendering.**

Most of the steps are carried out within the Blender environment, with step 2 being the most complex because it utilises the material node system, as illustrated in Figure 3. Additionally, a reflection had to be included. To achieve this, we needed to adapt Blender's built-in shadow catcher function to also capture the reflection of the urn while rendering. Both the rendered urn image and its reflection were then combined in a final image of an urn, which can be seen in Figure 4.



**Figure 3:** Incorporation of motifs through the material node system in Blender.



**Figure 4:** Example of the rendered image of an urn with a rose motif.

## 4.2. Production Time

Urnes offers six distinct series of urns in its catalogue, each featuring a different number of designs. The series includes Classic Natural, Classic Painted, Trend Natural, Trend Painted, Crystal, and Borders. The production times for each series vary within the traditional catalog production process, as shown in Table 1. The average production time is 174.74 minutes per urn.

**Table 1**

Production times for each series of urns included in the catalogue within the traditional production process

|                 | Production time<br>[min/unit] | Painting/oiling time<br>[min/unit] | Motif application time<br>[min/unit] | Number of urns in the<br>catalogue | Total production time<br>[min/series] |
|-----------------|-------------------------------|------------------------------------|--------------------------------------|------------------------------------|---------------------------------------|
| Classic Natural | 118                           | 12                                 | 135                                  | 6                                  | 810                                   |
| Classic Painted | 118                           | 54                                 | 177                                  | 24                                 | 4248                                  |
| Trend Natural   | 118                           | 12                                 | 133                                  | 6                                  | 798                                   |
| Trend Painted   | 118                           | 54                                 | 175                                  | 42                                 | 7350                                  |
| Crystal         | 118                           | 54                                 | 242                                  | 6                                  | 1452                                  |
| Borders         | 118                           | 54                                 | 177                                  | 9                                  | 1593                                  |

After completing the urns, they needed to be photographed. The total time required for equipment setup and studio preparation was 45 minutes. Taking photographs of all 93 urns took an additional 46.5 minutes, followed by 93 minutes dedicated to colour correction. Finally, 279 minutes were needed to add reflections and crop the images. In total, the average time required to prepare each urn image for the catalogue using this traditional method was approximately 179.72 minutes per urn.

In contrast, the implementation of a 3D visualization pipeline required about 50 hours. This time included modelling the urn, preparing basic textures and material nodes, and setting up the scene. On average, creating the final texture and rendering the image took approximately 4.2 minutes per urn. The entire process of creating the images needed for the catalogue took 56.51 hours, or 36.46 minutes per urn, making it 4.93 times faster than the traditional method. The time savings will be even more substantial when we create the next catalogue, as the existing 3D visualization pipeline can be reused, potentially reducing the time required by 97.6% (see Table 2).

**Table 2**

Time comparison between photographic and 3D workflows (n = 93 urns)

| Workflow                      | Time<br>breakdown<br>[h] | Total time<br>[h] | Savings vs.<br>photography | Improvement<br>factor |
|-------------------------------|--------------------------|-------------------|----------------------------|-----------------------|
| Photography<br>(baseline)     | $2.98 \cdot 93$          | 277,5             | -                          | -                     |
| 3D Initial<br>(with setup)    | $50 + 0.07 \cdot 93$     | 56.5              | 79.6%                      | 4.9                   |
| 3D Steady State<br>(no setup) | $0.07 \cdot 93$          | 6.5               | 97.6%                      | 42.6                  |

#### 4.3. Cost Implications

The photographic method incurred costs for physical prototypes, studio equipment, materials, and post-production labor. These costs scaled linearly with the number of product variations. By contrast, the 3D workflow required an upfront investment in asset creation but enabled the reuse of digital

models across all future variants. This significantly reduced operational costs, particularly in catalog updates and personalized orders.

4.4. Visual and Customer Evaluation

An interview with 38 catalog users, representatives of Urnes’ usual business partners, indicated that the 3D-based images were judged to match or surpass the photographic versions in terms of color fidelity and wood texture detail, see Figure 5. Furthermore, 61% of respondents reported faster response times when ordering personalized urns, suggesting that the 3D workflow improved not only efficiency but also customer experience.

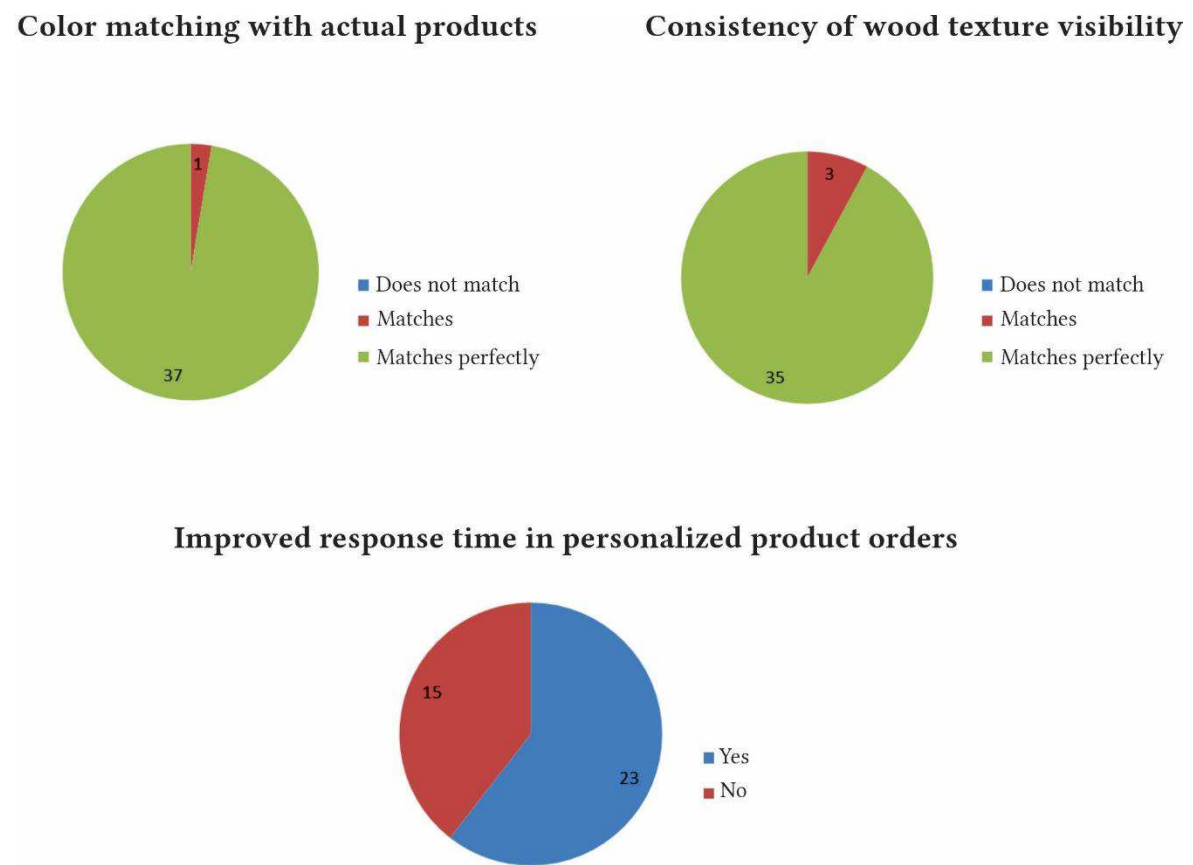


Figure 5: Results of the interviews with the catalog users.

5. Discussion

The results of this study demonstrate that a 3D visualization workflow can effectively replace traditional studio photography in the production of product catalogues. The most significant benefit lies in the drastic reduction of preparation time: once the digital pipeline was established, the average preparation time per urn variant decreased from 179 minutes to just 4.2 minutes, representing a 42.6-fold improvement and a 97.6% time saving. Even when the one-time setup investment of 50 hours is included, the 3D workflow proved more efficient after approximately 18 product variants. This finding highlights the scalability of 3D methods—initial costs are offset quickly in contexts where numerous variations or frequent catalogue updates are required.

These results align with industry reports indicating that CGI workflows reduce image creation costs by up to 85% [9] and are several times faster than photography [10]. Our findings extend this body of knowledge by providing a quantified, real-world case study that compares both methods

under controlled conditions. The break-even analysis provides a practical benchmark for other firms considering a transition to 3D visualization.

In addition to efficiency gains, the interview with 38 catalogue users confirmed that the 3D-generated images were perceived as highly accurate, particularly in terms of colour fidelity and wood grain representation. This supports previous research showing that CGI can achieve visual authenticity comparable to photography [6], [8]. 61% of respondents also reported faster response times when ordering personalized products, suggesting that the flexibility of digital assets contributes to improved customer experience and decision-making speed. This echoes findings from e-commerce research that highlight the role of visual fidelity and responsiveness in consumer satisfaction [7].

Nevertheless, the study also reveals certain limitations. The accuracy of colour reproduction across different print and digital media remains a challenge, particularly when dealing with subtle tones and specific surface finishes. This reflects a broader issue in digital visualization workflows, where calibration between devices and media can affect user perception. Future research should examine methods of standardized colour management to close this gap.

Finally, while this case focused on a single product line within a niche industry, the implications are broadly applicable. Sectors such as furniture, fashion, and consumer electronics, which involve numerous product variations and frequent updates, are likely to benefit most from the adoption of 3D workflows. Moreover, the reuse of 3D assets across catalogue design, online retail, and immersive platforms such as augmented and virtual reality could further amplify cost savings and customer engagement.

## 6. Conclusions

This study confirmed that a 3D visualization workflow can serve as a viable alternative to traditional studio photography in catalogue production. Beyond efficiency, the approach supports flexibility in personalization and offers a scalable foundation for future product communication strategies.

The findings encourage companies that manage frequent product updates or numerous variations to consider transitioning toward digital pipelines. Although some challenges remain, particularly in colour management, these are outweighed by the long-term advantages in adaptability and sustainability.

Future research should address technical refinements, such as standardized workflows for print fidelity, and explore the integration of 3D assets into emerging channels, including augmented and virtual reality.

## Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT-5 and Grammarly in order to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References

- [1] C. Kirpes, G. Hu, and D. Sly, 'The 3D product model research evolution and future trends: a systematic literature review', *Applied System Innovation*, vol. 5, no. 2, p. 29, 2022.
- [2] S. Covello, 'Visual Rhetoric', *Visual Communication*, 2019.

- [3] J. Krause, *Lessons in Typography: Must-know typographic principles presented through lessons, exercises, and examples*. New Riders, 2015.
- [4] W. Collins, A. Hass, K. Jeffery, A. Martin, R. Medeiros, and S. Tomljanovic, *Graphic design and print production fundamentals*. British Columbia, 2015.
- [5] D. J. Eck, 'Introduction to computer graphics', 2016.
- [6] L. Hausken, 'Photorealism versus photography. AI-generated depiction in the age of visual disinformation', *Journal of Aesthetics & Culture*, vol. 16, no. 1, p. 2340787, 2024.
- [7] H.-J. Kim, S. C. Jeong, and S.-H. Kim, 'Comparative Analysis of Product Information Provision Methods: Traditional E-Commerce vs. 3D VR Shopping', *Applied Sciences*, vol. 15, no. 4, p. 2089, 2025.
- [8] S. Debbabi, M. Daassi, and S. Baile, 'Effect of online 3D advertising on consumer responses: the mediating role of telepresence', *Journal of Marketing Management*, vol. 26, no. 9–10, pp. 967–992, 2010.
- [9] Nfinite, 'The shift is starting: Why retailers are moving to CGI and 3D imagery to improve business outcomes', 2022, [Online]. Available: [https://www.nfinite.app/blog/why-retailers-moving-to-cgi-and-3d-imagery?utm\\_source=chatgpt.com](https://www.nfinite.app/blog/why-retailers-moving-to-cgi-and-3d-imagery?utm_source=chatgpt.com)
- [10] Modelry, '3D virtual photography is 6x cheaper than traditional photography', 2021. [Online]. Available: [https://www.modelry.ai/3d-virtual-vs-traditional-photography-cost-comparison?utm\\_source=chatgpt.com](https://www.modelry.ai/3d-virtual-vs-traditional-photography-cost-comparison?utm_source=chatgpt.com)

# Integration of Hybrid Animation in a 360-degree Environment <sup>\*</sup>

Aleksandar Ilievski<sup>†</sup>, Suzana Žilič Fišer<sup>†</sup> and Simon Kolmanič<sup>†</sup>

<sup>†</sup> University of Maribor, FERl, Koroška Cesta 46, SI-2000 Maribor, Slovenia

## Abstract

This paper addresses the lack of an existing pipeline for integrating Hybrid Animation in a 360-degree environment. While works like Google Spotlight's *Pearl* showcase the expressive potential of cinematic VR (Virtual Reality) animation using stylized NPR (Non-Photorealistic Rendering) techniques, they lack the integration of 2D visual elements. We propose a pipeline that leverages accessible tools, including Blender's Grease Pencil for 2D animation, NPR shading for stylized 3D rigs, and a 360-degree render using a metadata injector. Our framework makes hybrid 360-degree animation more accessible and adaptable, enabling future developments and creative applications.

## Keywords

Virtual Reality, Hybrid Animation, 360-degree video, Blender, Non-photorealistic Rendering

## 1. Introduction

The evolution of animation and virtual reality has enabled new forms of immersive storytelling that combine artistic expression with technological sophistication. Hybrid animation, which merges 2D visual elements such as hand-drawn frames with 3D visual elements like computer-generated models, allows creators to achieve expressive, stylized visuals while leveraging modern rendering techniques [1]. Complementing this, non-photorealistic rendering (NPR) shaders enable 3D models to adopt 2D hand-drawn aesthetics, enhancing clarity, style, and narrative emphasis [2]. 360-degree cinematic VR provides immersive experiences through pre-rendered 3DOF content, often utilizing equirectangular projection mapping (EPM) to project spherical scenes onto a rectangular image for VR playback [3].

By exploring the intersection of hybrid animation, NPR shading, and cinematic VR, this paper examines how contemporary techniques can produce high-fidelity, stylized narratives that engage audiences beyond traditional media.

## 2. Theoretical Background

### 2.1. Hybrid Animation

Hybrid animation combines 2D visual elements, such as traditional hand-drawn animation, with 3D visual elements, including CGI models and environments, to create a visually rich and unique style [4]. This approach enables animators to retain the expressive qualities and fluidity of 2D hand-drawn techniques while leveraging the depth, perspective, and realism offered by 3D computer-generated elements to integrating these visual elements provides both creative flexibility and technical efficiency, enabling hybrid animation to achieve effects that are difficult to realize using only one medium [1]. Together, these methods demonstrate how the combination of 2D and

---

<sup>\*</sup>Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia

<sup>1\*</sup> Corresponding author.

<sup>†</sup> These authors contributed equally.

 aleksandr.ometov@tuni.fi (A. Ometov); t.princesales@utwente.nl (T. P. Sales); manfred.jeusfeld@acm.org (M. Jeusfeld)

 0000-0000-0000-0000 (A. Ometov); 0000-0000-0000-0000 (T. P. Sales); 0000-0000-0000-0000 (M. Jeusfeld)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.13>

3D visual elements enhances artistic storytelling and expands the possibilities of modern animation production.

## **2.2. Non-photorealistic rendering**

Non-photorealistic rendering (NPR) shaders are techniques used in 3D computer graphics to produce 2D visual effects such as stylized lines, cel shading, and painterly textures on 3D visual elements, giving them the appearance of traditional 2D art. By controlling color, lighting, and edges, NPR shaders allow 3D models to mimic hand-drawn or illustrative styles, bridging the gap between 2D hand-drawn aesthetics and 3D computer-generated elements[5]. This approach provides artists with creative flexibility, enabling the production of stylized animations, games, and simulations that emphasize artistic expression over photorealism. NPR shaders also enhance visual clarity and storytelling, particularly in hybrid animation, where 2D and 3D visual elements coexist.[6]

## **2.3. 360-degree Animation**

360-degree Animation, often used in cinematic VR, provides an immersive viewing experience by allowing users to explore a scene in all directions. This format is widely used in virtual tours, animated short films, and immersive storytelling, where audience presence in the environment enhances engagement [7]. To achieve seamless panoramic visuals, equirectangular projection mapping (EPM) is commonly employed, projecting the spherical environment onto a rectangular 2D image, which can then be wrapped onto a virtual sphere for VR playback [8]. This method ensures consistent image quality across the full 360-degree view and simplifies the rendering and compositing process. In many cinematic VR applications, content is pre-rendered rather than generated in real-time, allowing for high-fidelity graphics and complex visual effects that would be computationally expensive in interactive VR. Typically, this approach is paired with 3 Degrees of Freedom (3DOF) VR, enabling users to look around freely but not move through the scene, providing a controlled yet immersive narrative experience.

## **3. Case Study**

### **3.1. A case study of Google Spotlight Stories: Pearl**

Several immersive storytelling platforms have explored cinematic VR experiences. More specifically, Google Spotlight Stories allows users to experience 360-degree narratives with interaction and spatial audio. These experiences guide the viewer's attention within the environment, offering an engaging sense of presence.

*Pearl* is a VR short film created using non-photorealistic rendering (NPR) techniques and 360-degree cinematography [9]. It provides an immersive narrative experience within a fully 3D environment.

The production of *Pearl* involved several innovative methods, developed by the research team using a custom in-house real-time rendering engine. Entire cinematic shots were animated in a 3D environment, which allowed the team to generate three distinct versions of the film:

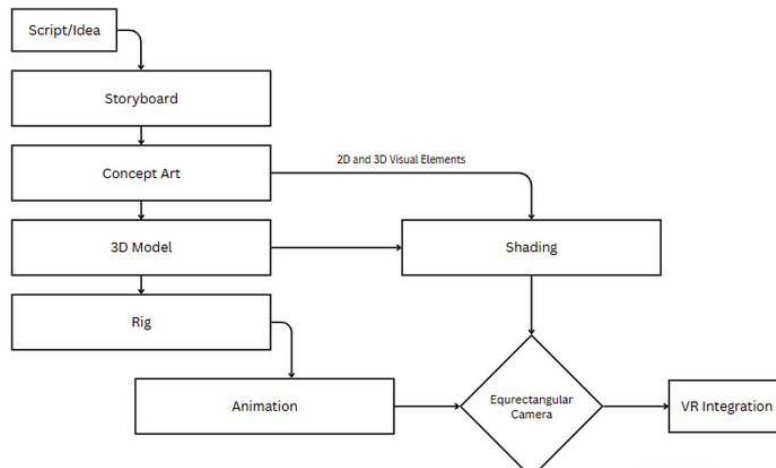
1. An interactive VR version that enables users to manipulate and explore scenes.
2. A theatrical version rendered in real-time for standard viewing.
3. Pre-rendered 360-degree spherical video designed for standalone playback.

## 4. Method

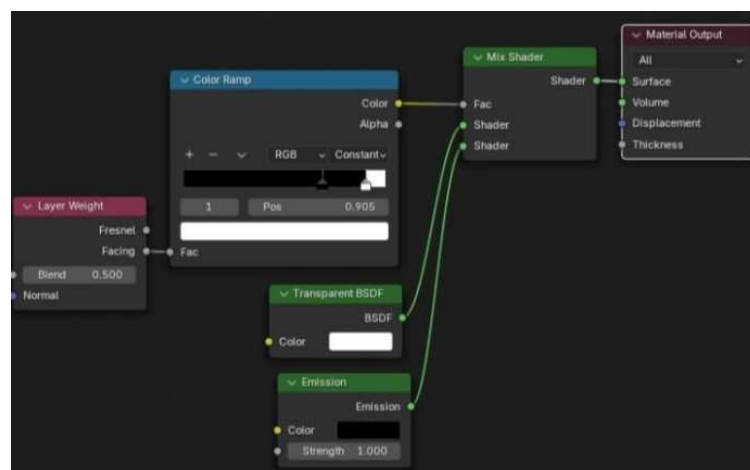
Our work focuses on developing a hybrid animation pipeline that integrates hand-drawn 2D visual elements, enabling immersive 360-degree VR storytelling. The entire production was carried out using open-source software, providing a standalone workflow from initial concept to final rendered 360° video. The pipeline was designed to maintain visual consistency across 2D and 3D components, ensuring a cohesive, illustrative aesthetic throughout the immersive environment.

### 4.1. Pipeline

The initial stage of the pipeline involved creating the visual style and look for the project, as shown in Figure 1. Once the conceptual sketches and visual development were completed, the focus shifted to shader development. A non-photorealistic rendering (NPR) shader was created using Blender's node editor as depicted in Figure 2. This step allowed for precise control over shading, outlines, and the overall visual tone of the scene. Through the combination of custom nodes, the shader replicated characteristics of traditional hand-drawn animation, while remaining compatible with real-time rendering requirements. Adjustments to lighting, color gradients, and edge detection were iteratively refined to achieve a stylized look that could be consistently applied to all 3D models in the environment.



**Figure 1:** Production Pipeline. Diagram created by the Author (2025).



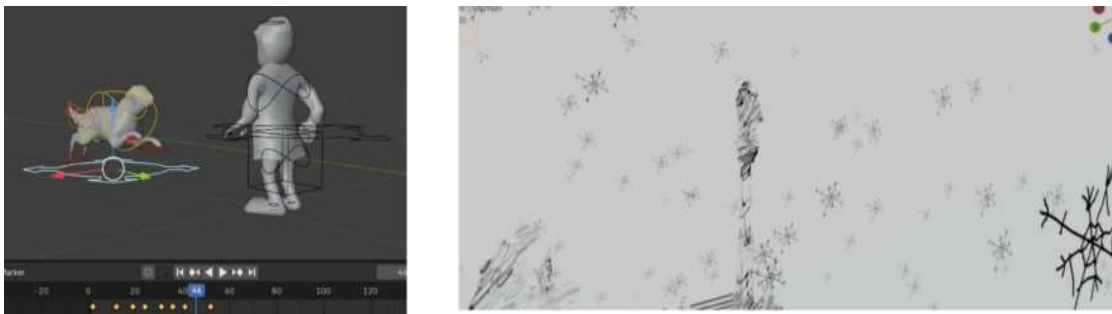
**Figure 2:** Node system. Created by the Author (2025).

## 5. Node Editor

Attention was directed toward matching the visual style between the 2D concept art and the 3D models. Concept planes were imported into the 3D scene to serve as reference, allowing for direct comparison and alignment. Textures, materials, and shading parameters were meticulously tuned to ensure that 3D objects maintained the illustrative qualities of the 2D designs. This stage was critical in creating a seamless visual integration, bridging the gap between hand-drawn aesthetics and three-dimensional depth.

### 5.1. 2D and 3D Matching

The style matching of the animation between the 3D models using the NPR shader, allowing for movement that visually resembled hand-drawn sequences. To further enhance the hybrid nature of the animation, Blender's Grease Pencil tool was employed to add frame-by-frame 2D animations and visual effects directly onto the 3D environment as demonstrated in Figure 3.

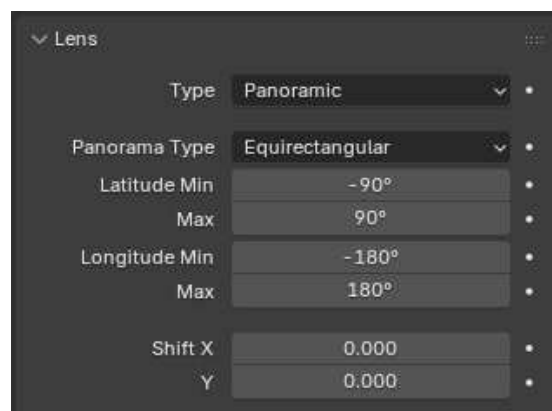


**Figure 3:** 2D and 3D Mixing. Created by the Author (2025)

This integration enabled the addition of secondary animations, such as dynamic effects of the snowflakes, while preserving the hand-drawn feel within a fully three-dimensional scene. The combination of rigged 3D animation and 2D frame-by-frame effects created a hybrid workflow, extending the expressive possibilities of the VR environment beyond traditional 3D animation.

### 5.2. Camera setup

The scene layout was then prepared for immersive 360-degree VR viewing. An equiangular camera setup was utilized to capture the entire environment, ensuring that the rendered video could be projected seamlessly onto a spherical surface as shown in Figure 4.



**Figure 4:** Equiangular: Camera Setup. Created by the Author (2025)

### 5.3. Equiangular Projection Mapping

To make the video compatible with 360-degree playback platforms, a metadata injector was employed. This tool encoded the equirectangular video with the necessary metadata for spherical projection, ensuring proper display in VR headsets and other immersive viewing devices. A demonstration of the hybrid animated scene into a fully immersive 360-degree VR experience can be seen in Figure 5.



**Figure 5:** Equirectangular Projection Mapping. Created by the Author (2025).

## 6. Conclusion

The methodology described above demonstrates how open-source tools can be leveraged to create hybrid animations that merge 2D hand-drawn artistry with 3D environments. The NPR shader development, 2D/3D style matching, Grease Pencil integration, equirectangular camera setup, and 360-degree metadata injection collectively establish a pipeline that supports both creative expression and technical precision. By combining these techniques, the project overcomes the limitations of traditional 3D-only immersive animation, providing a richer and more visually engaging VR storytelling experience.

The integration of hand-drawn animation within a 3D space not only enhances the visual expressiveness of the scene but also allows for greater flexibility in storytelling, enabling interactive and immersive narratives that retain the charm of traditional animation. The hybrid workflow presented here represents a practical solution for producing high-quality VR content with both aesthetic and interactive depth.

However, several limitations remain. Real-time rendering in VR still poses challenges, particularly in terms of maintaining high frame rates when combining complex NPR shaders with layered 2D animation.

Future work should address these limitations by exploring optimization strategies for real-time NPR shading, automation techniques created by a plugin for style transfer between 2D and 3D assets, and more efficient methods of integrating Grease Pencil animations into large-scale VR environments. Advancements in GPU processing and VR display technologies will further enable higher-quality hybrid content, paving the way for more interactive, adaptive, and visually rich immersive storytelling.

## Acknowledgements

The author(s) thanks the Blender and Open-Source Software Community.

## Declaration on Generative AI

The author(s) have employed Generative AI tools such as ChatGPT and Grammarly for minimal improvement by structuring this research paper and checking grammar mistakes.

## References

- [1] T. O’Hailey, *Hybrid Animation: Integrating 2D and 3D Assets*, 2nd ed., Taylor & Francis, 2014. Available: <https://www.taylorfrancis.com/books/mono/10.4324/9781315867885/hybrid-animation-tina-hailey>
- [2] B. Gooch, P. P. J. Sloan, A. Gooch, P. Shirley, and R. Riesenfeld, “Interactive technical illustration,” *Proceedings of the Symposium on Interactive 3D Graphics*, pp. 31–38, 1999, doi: 10.1145/300523.300526.
- [3] J. Jerald, *The VR Book: Human-Centered Design for Virtual Reality*, Morgan & Claypool Publishers, Oct. 2015, doi: 10.1145/2792790.
- [4] J. Kivistö, *Hybrid Animation: The Process and Methods of Implementing 2D Style in 3D Animation*, Tampere University of Applied Sciences, 2019. Available: <https://www.theseus.fi/handle/10024/265116>
- [5] C. J. Curtis, S. E. Anderson, J. E. Seims, K. W. Fleischer, and D. H. Salesin, “Computer-generated watercolor,” in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*, Los Angeles, CA, USA, Aug. 1997, pp. 421–430, doi: 10.1145/258734.258896.
- [6] B. Gooch and A. Gooch, *Non-Photorealistic Rendering*, A K Peters/CRC Press, 2001, doi: 10.1201/9781439864173.
- [7] M. Slater and M. V. Sanchez-Vives, “Enhancing our lives with immersive virtual reality,” *Frontiers in Robotics and AI*, vol. 3, Dec. 1, 2016, doi: 10.3389/frobt.2016.00074.
- [8] R. Szeliski and H.-Y. Shum, “Creating full view panoramic image mosaics and environment maps,” *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*, Los Angeles, CA, USA, Aug. 1997, pp. 251–258, doi: 10.1145/258734.258861.
- [9] P. Osborne, “Pearl: VR short film,” Nexus Studios, Oct. 19, 2025. Available: <https://nexusstudios.com/work/pearl/>

## Online Resources

[360-degree Hybrid Animation Test](#)

# NERVIS: An Interactive System for Graph-Based Exploration and Editing of Named Entities

Uroš Šmajdek<sup>†</sup>, Ciril Bohak<sup>\*,†</sup>

*University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, SI-1000 Ljubljana, Slovenia*

## Abstract

We present an interactive visualization system for exploring named entities and their relationships across document collections. The system is designed around a graph-based representation that integrates three types of nodes: documents, entity mentions, and entities. Connections capture two key relationship types: (i) identical entities across contexts, and (ii) co-locations of mentions within documents. Multiple coordinated views enable users to examine entity occurrences, discover clusters of related mentions, and explore higher-level entity group relationships. To support flexible and iterative exploration, the interface offers fuzzy views with approximate connections, as well as tools for interactively editing the graph by adding or removing links, entities, and mentions, as well as editing entity terms. Additional interaction features include filtering, mini-map navigation, and export options to JSON or image formats for downstream analysis and reporting. This approach contributes to human-centered exploration of entity-rich text data by combining graph visualization, interactive refinement, and adaptable perspectives on relationships.

## Keywords

Named Entity Visualization, Graph Exploration, Interactive Visualization

## 1. Introduction

Textual data continues to grow at an unprecedented scale across domains such as scientific publishing, journalism, and social media, but also with the digitization of historical text corpora. Extracting meaning from these large corpora requires more than keyword search or frequency analysis; it requires understanding the entities mentioned in the text and the relationships between them. named entity recognition (NER) and related information extraction methods provide a way to identify people, organizations, locations, and other key entities. However, the resulting collections of mentions and links are often complex, redundant, and difficult to navigate without appropriate visualization and interaction techniques.

Graph-based visualization offers a natural way to represent entities and their relationships. Nodes can represent documents, entity mentions, or consolidated entities, while edges capture relationships such as co-location of mentions or equivalence of entities across contexts. Yet, static graph layouts alone are insufficient for the exploration of large and ambiguous entity networks. What is needed are interactive approaches that allow users to flexibly explore different levels of granularity, filter and refine connections, and iteratively adjust the structure of the graph to reflect domain knowledge and emerging insights.

In this paper, we introduce an interactive visualization system for entity-centric text exploration. The system combines graph representations with multiple coordinated views, supporting fuzzy and precise connections, user-driven graph editing, and export capabilities for integration with downstream analysis pipelines. By emphasizing exploratory flexibility and human-centered control, our approach bridges automatic information extraction with interactive sensemaking, enabling users to uncover patterns and relationships in complex entity-rich text collections.

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

<sup>†</sup>These authors contributed equally.

✉ uros.smajdek@fri.uni-lj.si (U. Šmajdek); ciril.bohak@fri.uni-lj.si (C. Bohak)

🌐 <https://lgm.fri.uni-lj.si/ciril> (C. Bohak)

🆔 0000-0002-9015-2897 (C. Bohak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.14>

## 2. Related Work

A number of prior works provide foundations for designing interactive visualization systems that support exploration of large and complex text collections. Paul et al. [1] introduce *TexTonic*, a visual analytics system that enables exploration through hierarchical clustering and direct manipulation of terms. Their approach highlights how interactive interfaces can support users in dynamically engaging with data to discover patterns and relationships among entities and mentions. In a complementary direction, Lee et al. [2] expand design considerations for information visualization interactions beyond the traditional mouse and keyboard. Their work underscores the importance of multimodal interaction, which can broaden the accessibility and flexibility of visualization systems.

Scalability and performance are equally critical aspects of interactive visualization. Tao et al. [3] present *Kyrix*, a system designed for large-scale web-based visualization, demonstrating techniques for efficient pan and zoom interactions at scale. Their emphasis on performance optimization ensures responsiveness when managing very large datasets, an aspect directly relevant to maintaining smooth exploration of entity-rich corpora. Such techniques align well with our use of fuzzy views and approximate connections to facilitate exploratory analysis without sacrificing interactivity.

The need for robust navigation and filtering has also been well established. Al Nasar et al. [4] discuss principles of interactive personal information management systems, emphasizing personalized interfaces to help users manage large information collections. While their focus is on photographs and videos, the principles are transferable to text-based entity exploration. Similarly, Sedig and Parsons [5] propose a pattern-based framework for interaction design that supports complex cognitive activities, providing conceptual guidance for structuring interactions within entity exploration systems.

Editing and refinement capabilities form another important dimension of related work. Satyanarayan et al. [6] propose *Reactive Vega*, a declarative framework for specifying interactive visualizations. Their approach demonstrates how flexible editing, specification, and interaction can empower users to tailor visualizations to their analytical needs. This perspective informs our focus on enabling users to add, remove, and refine entities and connections within the graph representation.

Recent work has advanced domain-specific analytical systems enhanced with graph-based visualizations. Rahman et al. [7] present *DiaVis*, a dashboard for iterative exploration of diabetes-related data, highlighting the benefits of refinement cycles for sustaining engagement—an insight we adopt for entity-level analysis. Heberle et al. [8] propose a web-based system for biological networks that employs rule-based filtering and automated layout generation to support scalable exploration. Similarly, Kwak et al. [9] develop a visual analytics framework for stroke care networks, using multiple coordinated views to capture patterns across different levels of analysis, from national to community scale. In cybersecurity, a recent system by Razbelj et al. [10] introduces a graph-based approach to visualizing honeypot-captured attacks, enabling interactive exploration of large, heterogeneous datasets to detect recurring patterns and uncover hidden connections. Collectively, these systems demonstrate how interactive visualization, filtering, and scalable layouts can support domain-specific analysis.

Finally, Mongiovi and Gangemi [11] propose *GRAAL*, a graph-based retrieval system for collecting related passages across multiple documents. By emphasizing semantic interactions and subgraph representations, *GRAAL* directly addresses the challenge of identifying co-occurring entity mentions in large text collections. Their methodology provides valuable inspiration for incorporating semantic relationships into our proposed interactive visualization system.

Prior research has contributed foundational techniques in multimodal interaction, scalability, navigation, iterative exploration, and semantic graph analysis. Building on these works, our system integrates and extends these ideas into a unified framework for interactive exploration, refinement, and visualization of named entities across large text corpora.

### 3. Design Principles

In this section, we highlight key limitations of existing named entity visualization systems and draw from the broader visualization literature to derive five core design principles for our approach:

- **Integrated data exploration and visualization** – Graph-based visualizations of named entities often involve selecting a subset of entities and relations from a larger corpus, the scope and relevance of which may not be known in advance. This selection process can become labor-intensive if performed manually. Existing network visualization systems often assume that all entities and relations in a dataset are to be included in the visualization, overlooking the exploratory step required to identify subsets of interest. This can result in cluttered or unreadable graphs, as well as delays in achieving meaningful insights.
- **Cross-Document Entity Relations** – Beyond the classification of individual mentions, named entity linking (NEL) and coreference extraction are often employed to link multiple entity instances that refer to the same underlying entity. This task extends both within a document, where entities may appear under different surface forms, abbreviations, or titles, and across documents, where mentions must be linked despite variation in spelling, context, or language. Commonly used named entity visualizations, such as displaCy [12] (see fig. 1 (left), struggle to showcase such relations effectively as they are only depicted as properties of individual entity instances.
- **Retaining the Fuzziness of Entity Data** – Information extraction processes related to named entities, such as NER and NEL, rarely yield unambiguous results, especially for historical and heterogeneous corpora, where metadata is often incomplete [13]. Instead, they often produce multiple candidate entities and relations with associated confidence scores [14, 15]. Instead of only preserving the “best guess”, we aim to showcase multiple candidates and let the researcher explore the spectrum of possible interpretations.
- **Optional automation** – While automation is critical for making graph visualizations accessible, it should not eliminate authorial control. Meaningful network visualizations often require manual refinement, such as pruning peripheral nodes, re-weighting edges, or emphasizing particular subgraphs. Automation should therefore serve as an initial scaffold, enabling authors to iteratively adjust and enhance the visualization to highlight the most relevant aspects of the data.
- **Progressive visualization** – Interactive graph exploration depends heavily on timely and comprehensible visual feedback. Research indicates that user interactions should produce visual responses within 50–100 ms to maintain a fluid sense of control [16]. At the same time, sudden or large-scale layout changes can disorient users and obscure the perceived impact of their actions. Progressive visualization techniques [17] mitigate these issues by subdividing expensive computations, such as force-directed layouts or entity disambiguation steps, into incremental updates with smooth transitions [18]. This approach provides clarity and responsiveness while supporting iterative exploration of large or evolving networks.

### 4. System Design

NERVIS is an interactive web-based system for graph visualization and editing, implemented using HTML, CSS, TypeScript, and React framework<sup>1</sup>. The system is designed around a modular architecture that separates data management, visualization, and user interaction. At its core, NERVIS maintains structured data models that represent nodes, edges, and associated metadata, enabling efficient storage, retrieval, and manipulation of graph information. These models feed into a graph visualization pipeline,

---

<sup>1</sup><https://react.dev>

which computes layouts, applies styling and interaction rules, and renders the graph for exploration and analysis.

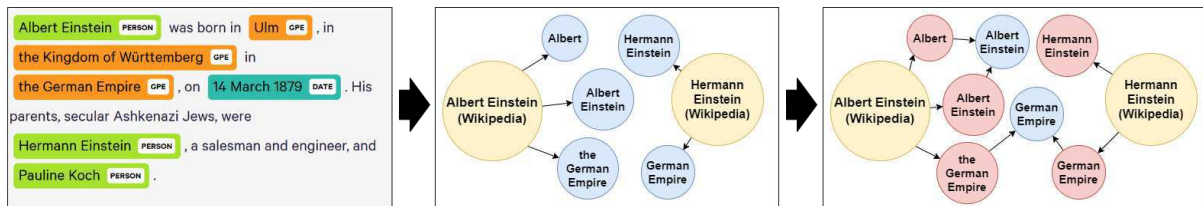
The user interface provides direct access to the system’s functionality, allowing users to inspect and modify graph elements, control visualization parameters, and apply filters. The interface is organized into distinct components that facilitate navigation, node-level editing, and rule-based filtering, providing a cohesive environment for interactive graph analysis.

#### 4.1. Data Model

In this section, we present the data model underlying NERVIS, designed according to the principles of *Cross-Document Entity Relations* and *Retaining the Fuzziness of Entity Data* (see section 3).

Typical text-based named entity visualizations represent entities as annotated spans over a sequence of tokens (see fig. 1, left). To extend this approach to a network graph spanning multiple documents, entities are first reformulated as *entity instance nodes*, while token sequences are abstracted into *document nodes* (fig. 1, middle). Although this representation captures basic entity-document relationships, it cannot effectively represent higher-level relations, such as links between mentions of the same entity across different documents. To address this, entity instance nodes are further separated into *mention nodes*, representing individual occurrences of an entity within a document, and *entity nodes*, representing the underlying entity itself (fig. 1, right).

Both *entity* and *mention nodes* include an entity class property derived from the CoNLL-2003 dataset [19], which classifies named entities as *Person*, *Organization*, *Location*, or *Miscellaneous*. This classification was chosen because it underpins many state-of-the-art NER methods [14]. To maintain the principle of data fuzziness, mentions are not strictly constrained to match the class of their connected entity, enabling the system to highlight potential inconsistencies while allowing users to correct them at their discretion.



**Figure 1:** Data model diagrams for text-based named entity visualizations from displaCy [12] (left), simple document-entity model (middle), and proposed document-mention-entity model (right).

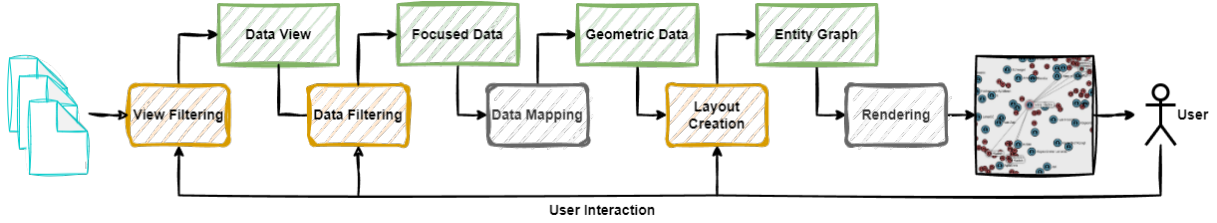
The data model also defines four types of connections i.e., graph edges. First, *mention-to-document* and *mention-to-entity* connections reflect relationships provided in the input data, and as per design principles, impose no limit on the number of entities a mention may be associated with. Second, *mention-collocation* connections capture which mentions occur within the same sentence. Finally, a virtual *entity-to-document* connection is introduced to facilitate observation of cross-document entity relations and to enable more efficient filtering and exploration of the graph.

#### 4.2. Visualization Pipeline

The NERVIS visualization pipeline is illustrated in fig. 2 and is modeled after the classical visualization pipeline framework [20]. It is designed to transform the structured data described in section 4.1 into an interactive graph representation, and is organized into five main stages:

1. **View Filtering** – The structural abstraction of the graph is adjusted to emphasize different levels of detail in the visualization.
2. **Data Filtering** – Input data is refined according to user-specified criteria, such as rule-based filters, user selection, or document subsets, to focus the visualization on relevant nodes and edges.

3. **Data Mapping** – Filtered data is mapped onto geometric structures, with selected elements optionally receiving additional visual emphasis based on user interactions.
4. **Layout Creation** – Node and edge positions are computed to produce a spatial arrangement that highlights relational structures and preserves graph readability.
5. **Rendering** - The graph is visually rendered on the display.

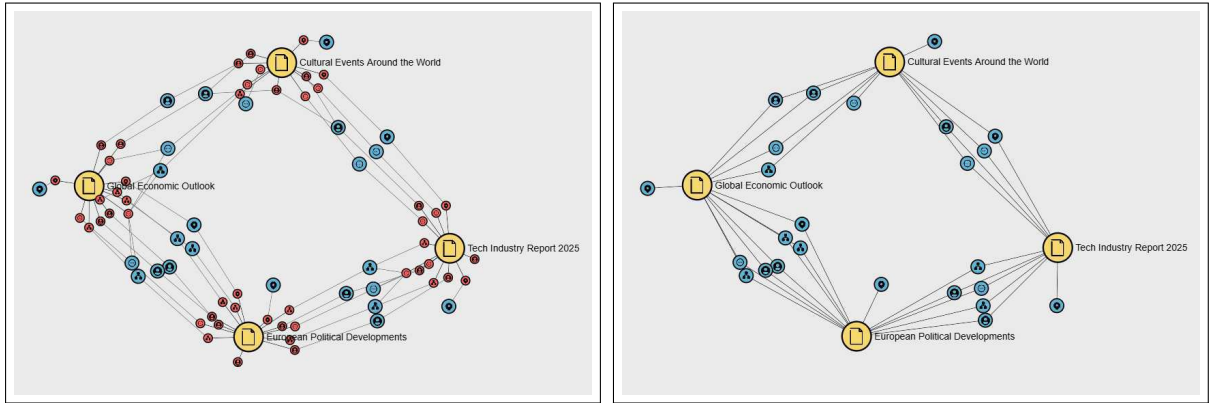


**Figure 2:** NERVIS visualization pipeline. Orange and grey nodes indicate intractable and uninterpretable stages respectively. Green nodes indicate intermediary data representations.

#### 4.2.1. View Filtering

The first stage of the pipeline allows users to toggle between two structural views of the graph: the default Document–Mention–Entity (D–M–E) view and a simplified Document–Entity (D–E) view. In the D–E view, mention nodes are abstracted away and replaced with virtual entity-to-document edges, producing a more concise visualization that emphasizes cross-document entity relations.

This choice directly shapes the set of nodes and edges made available to subsequent stages of the pipeline. By selecting the D–M–E view, users retain the full detail of entity mentions within documents, enabling fine-grained exploration of textual annotations. By contrast, the D–E view allows users to focus on higher-level relationships across documents without the visual clutter of mention nodes.



**Figure 3:** A comparison of Document–Mention–Entity view (left) and Document–Entity view (right).

#### 4.2.2. Data Filtering

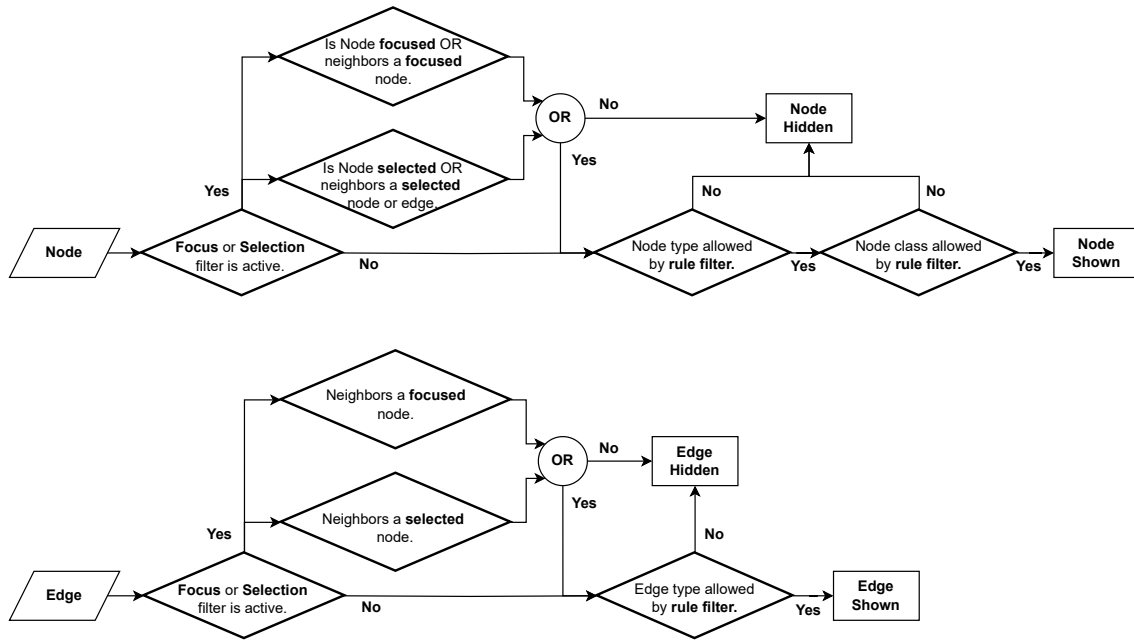
To satisfy the *Integrated Data Exploration and Visualization* design principle (see section 3), NERVIS supports interactive filtering of the input data. Inspired by other network visualization tools (e.g., PatientFlow [9] and CellNetVis [8]), we implement two filtering steps:

1. **Focus filter** – This filter allows the user to specify a node or edge of interest. Two focus filters can be applied concurrently: the *selection filter* and the *node focus filter*. The *selection filter* highlights a chosen node by removing all edges that do not connect it to its neighbors, reducing ambiguity in dense graphs. The *node focus filter* removes all nodes and edges not directly connected to a

specified node, enabling users to concentrate on the local neighborhood of interest. If applied concurrently, they are designed not to conflict with each other; the selection filter will not remove edges of the focused node and *vice versa*.

2. **Rule filter** – This filter allows users to control which types of nodes and edges are displayed based on their properties. Filtering can be performed by node type (e.g., *document*, *mention*, *entity*), by entity class (e.g., *Person*, *Location*), or by any combination of these criteria.

The complete filtering algorithm is illustrated in fig. 4.



**Figure 4:** Node (top) and edge (bottom) filtering algorithm flowchart.

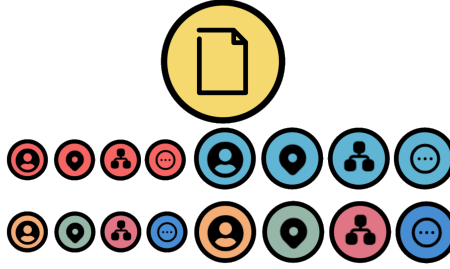
### 4.2.3. Data Mapping

In the data mapping stage, the filtered data is transformed into geometric structures that will be rendered in the final visualization. This process is subject to two key constraints. First, the number of nodes can reach several thousand, requiring the use of simple and recognizable shapes to maintain clarity in high-density graphs. Second, the layout algorithms, described in section 4.2.4, are similarly limited to simple geometric structures to ensure computational efficiency and maintain performance.

To satisfy these constraints, each node is represented as a circle, with size and color encoding its type. To support a variety of use cases, two color schemes are provided: one based on node type and another based on entity class. Additionally, to enhance visual distinction and improve user interpretation, a pictogram indicating the node's class and type is displayed. Examples of the resulting node structures are shown in fig. 5.

### 4.2.4. Layout Creation

In the layout creation step, the positions of nodes within the graph are determined. In line with the design principle of *optional automation* (see section 3), NERVIS allows users to either employ an automated layout algorithm or manually adjust node positions. For automated layout generation, we use the *ForceAtlas2* algorithm [21], which is efficient and supports iterative computation, aligning with our design principle of *progressive visualization*.



**Figure 5:** Visual structures used to depict different node types. The top row shows a *document* node. The middle row illustrates *mention* nodes (left) and *entity* nodes (right) across different entity classes. The bottom row depicts mention and entity nodes using an entity class-based color scheme. From left to right the individual node pictograms depict *person*, *location*, *organization* and *miscellaneous* entity classes.

We employ the implementation provided by the Graphology framework [22], which leverages the Web Workers API<sup>2</sup> to multithread the computation, ensuring that the user interface remains responsive during layout processing. An example of a layout produced by the ForceAtlas2 algorithm is shown in fig. 6.

#### 4.2.5. Rendering

Rendering constitutes the final stage of the visualization pipeline, responsible for producing the visual representation of the graph on the user interface. To ensure a smooth and responsive experience when visualizing high-density graphs, NERVIS integrates the *Sigma.js*<sup>3</sup> framework. Sigma.js employs an instance-based WebGL rendering pipeline, allowing computations to be offloaded to the GPU and fully leveraging modern graphics hardware for high-performance visualization.

One limitation of this rasterized rendering approach is that it does not produce vector-based outputs, which restricts the possibility of external post-processing or vector editing of the visualization. However, this is compensated by a comprehensive set of in-system editing tools, including node manipulation, focus and rule filtering, and layout adjustments, which allow users to modify the graph without requiring external editing.

#### 4.3. User Interface

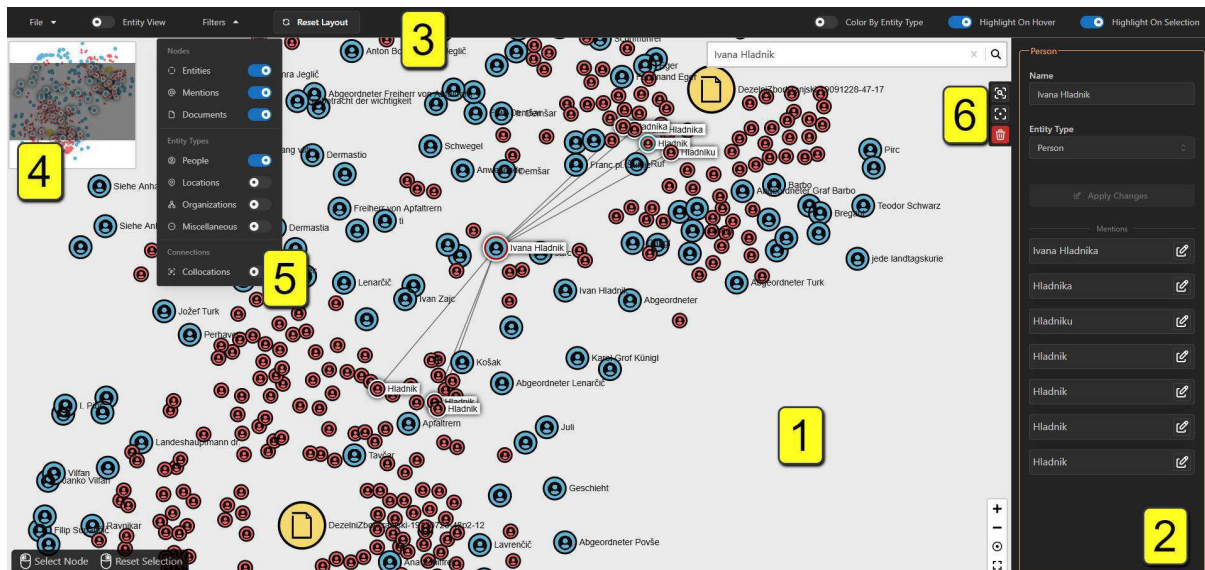
The NERVIS system provides a user interface for interactive graph exploration and editing (see fig. 6). The main display (1) presents the graph produced by the visualization pipeline described in section 4.2, enabling detailed inspection of nodes and edges. Additionally, the event-based interaction system allows direct selection, movement, and deletion of nodes and edges on the main display. The node editing widget (2) allows direct modification of node attributes and properties within the interface, and also supports immediate update of the visualized nodes.

The toolbar (3) offers core workflow functions, including graph import and export, view toggling, and activation of focus filters to highlight or hide specific portions of the graph. The toolbar also allows for efficient search and selection of nodes within the graph, utilizing MiniSearch<sup>4</sup>, an efficient full-text search engine for JavaScript. A minimap (4) provides an overview of the graph layout, assisting in navigation and orientation on larger graphs. The rule filter panel (5) permits selective filtering according to user-defined criteria, allowing attention to be directed toward relevant patterns and relationships. The node actions panel (6) allows access to three context-sensitive operations: contextual zoom, which sets the view to encompass the node and its visible neighbors, the activation of the node focus filter (see section 4.2.2), and node deletion.

<sup>2</sup><https://html.spec.whatwg.org/multipage/workers.html>

<sup>3</sup><https://github.com/jacomyl/sigma.js>

<sup>4</sup><https://github.com/lucaong/minisearch>



**Figure 6:** Overview of NERVIS graphical user interface. (1) Graph rendered using the visualization pipeline described in section 4.2. (2) Node editing widget (3) Toolbar. (4) Graph’s minimap. (5) Selection of rule filters. (6) Node-specific actions and filters.

## 5. Expert Evaluation

Because NERVIS builds on established technologies, such as the Sigma.js rendering framework for GPU-accelerated graph drawing, a dedicated performance evaluation was not conducted. Instead, we carried out an expert evaluation to assess the system as a whole, focusing on usability, interaction design, and the extent to which the implemented features support the intended workflows.

We discussed the resulting system with a Digital Humanities (DH) scholar studying historical print media, investigating relationships and representations of historical figures and geographic locations. Since systems for NER and NEL perform worse than humans, historians have learned to be very skeptical of purely computer-based approaches. In order to make historical research reliable, software for interacting with computer-produced results is a critical part of the DH research process.

After initial consultation set the stage and expectations, we performed two rounds of expert system evaluation with synthetic data. After preliminary evaluation, the feature requests were prioritized and the most important ones were implemented. After the final round, the expert judged the system to be mature enough for use in real-life setting. They also composed the list of possible future improvements, additions and extensions.

In the initial consultation we discussed software goals and previous experience when handling named-entity data with existing software. One of the key things pointed out during this consultation was the burden of human labor when dealing with computer-generated labels. For historians, a recall-focused approach is preferred to an accuracy-focused one. They prefer broad sets of result candidates, which they can browse through and edit manually as opposed to only seeing a single most likely result. This encouraged the design choice to develop a tool where the deletion of nodes and edges is the most common user operation, as opposed to creating new ones. A high recall approach inevitably generates relatively high amounts of noise, which we planned to tackle with a variety of content filters that allow researchers to focus on specific documents, entities, or types when cleaning the data. Since the users of DH tools often do not have an engineering background and often struggle with the installation process, we decided on the web platform, which also makes the system broadly available.

After the prototype was tested, we followed up with our second consultation. At this stage, the tool received generally positive feedback, with feature requests focusing on the most common user actions. As already mentioned, with high recall results, most user actions involve deleting nodes and edges, which was made additionally accessible with the implementation of hotkeys for quick node/edge

deletion. We discussed the noisiness of the visualization when there are many documents present, which exposed the need for users to focus on a single document when cleaning the data. While collocations are a common tool for studying text, in this particular task the question of which entities appear in the same sentence is of relatively limited importance. We decided to prioritize edges between documents, mentions, and entities while initially hiding the edges denoting collocations.

The final evaluation presented us with a new list of requirements and feature requests, which are the subject of future work. The need for undo/redo actions and action history was expressed alongside the wish to see specific mentions in the text context. Finally, the need to merge recognized entity nodes arose, due to errors in NEL process.

## 6. Conclusion

In this paper, we introduced an interactive web-based system for the visualization and editing of named entity graphs. The system builds on existing technologies such as Graphology [22] and Sigma.js, but contributes a tailored data model, a structured visualization pipeline, and interaction techniques specifically designed for large-scale named entity graphs.

Because the system leverages established rendering and layout technologies, we concentrated our evaluation on usability and interaction rather than raw performance. To this end, we conducted an expert evaluation with a digital humanities scholar working on historical print media. The evaluation provided valuable feedback on design decisions, particularly regarding the trade-off between recall and noise in named-entity data, the need for efficient editing operations, and the importance of content filters to reduce cognitive load when exploring large graphs. Iterative consultation shaped the prototype into a system the expert deemed ready for real-world use, while also identifying directions for further development, such as undo/redo functionality, contextual access to mentions, and support for merging entity nodes.

## Acknowledgment

We would like to thank Filip Dobranić for his expert feedback in evaluation of our system. This work was supported by the Slovenian Research and Innovation Agency research programme “Digital Humanities: resources, tools and methods” (2022–2027) [grant number P6-0436] and by the project “Large Language Models for Digital Humanities” (2024–2027) [grant number GC-0002].

## References

- [1] C. L. Paul, J. S. Chang, A. Endert, N. Cramer, D. Gillen, S. Hampton, R. Burtner, R. Perko, K. Cook, *TexTonic: Interactive Visualization for Exploration and Discovery of Very Large Text Collections*, Information Visualization (2018). doi:10.1177/1473871618785390.
- [2] B. Lee, P. Isenberg, N. H. Riche, S. Carpendale, *Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions*, IEEE Transactions on Visualization and Computer Graphics (2012). doi:10.1109/tvcg.2012.204.
- [3] W. Tao, X. Liu, Y. Wang, L. Battle, c. Demiralp, R. Chang, M. Stonebraker, *Kyrix: Interactive Pan/Zoom Visualizations at Scale*, Computer Graphics Forum (2019). doi:10.1111/cgf.13708.
- [4] M. R. Al Nasar, M. Mohd, N. M. Ali, *A Conceptual Framework for an Interactive Personal Information Management System* (2011). doi:10.1109/iuser.2011.6150545.
- [5] K. Sedig, P. Parsons, *Interaction Design for Complex Cognitive Activities With Visual Representations: A Pattern-Based Approach*, AIS Transactions on Human-Computer Interaction (2013). doi:10.17705/1thci.00055.
- [6] A. Satyanarayan, R. P. Russell, J. Hoffswell, J. Heer, *Reactive Vega: A Streaming Dataflow Architecture for Declarative Interactive Visualization*, IEEE Transactions on Visualization and Computer Graphics (2016). doi:10.1109/tvcg.2015.2467091.

- [7] M. F. Rahman, M. R. Islam, S. Akter, S. Akter, L. Islam, G. Xu, DiaVis: Exploration and Analysis of Diabetes Through Visual Interactive System, *Human-Centric Intelligent Systems* (2021). doi:10.2991/hcis.k.211025.001.
- [8] H. Heberle, M. F. Carazzolle, G. P. Telles, G. V. Meirelles, R. Minghim, Cellnetvis: a web tool for visualization of biological networks using force-directed layout constrained by cellular components, *BMC Bioinformatics* 18 (2017) 395. doi:10.1186/s12859-017-1787-5.
- [9] K. Kwak, J. Park, H. Song, A visual analytics framework for inter-hospital transfer network of stroke patients, *Applied Sciences* 13 (2023). doi:10.3390/app13095241.
- [10] M. Rabzelj, C. Bohak, L. v. Južnič, A. Kos, U. Sedlar, Cyberattack Graph Modeling for Visual Analytics, *IEEE Access* 11 (2023) 86910–86944. doi:10.1109/ACCESS.2023.3304640.
- [11] M. Mongiovì, A. Gangemi, GRAAL: Graph-Based Retrieval for Collecting Related Passages Across Multiple Documents, *Information* (2024). doi:10.3390/info15060318.
- [12] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python (2020). doi:10.5281/zenodo.1212303.
- [13] M. Ehrmann, A. Hamdi, E. L. Pontes, M. Romanello, A. Doucet, Named entity recognition and classification in historical documents: A survey, *ACM Comput. Surv.* 56 (2023). doi:10.1145/3604931.
- [14] Z. Hu, W. Hou, X. Liu, Deep learning for named entity recognition: a survey, *Neural Computing and Applications* 36 (2024) 8995–9022. doi:10.1007/s00521-024-09646-6.
- [15] Z. Zhang, Y. Zhao, H. Gao, M. Hu, LinkNER: Linking Local Named Entity Recognition Models to Large Language Models using Uncertainty, in: *Proceedings of the ACM Web Conference 2024, WWW '24*, Association for Computing Machinery, New York, NY, USA, 2024, p. 4047–4058. doi:10.1145/3589334.3645414.
- [16] Z. Liu, J. Heer, The Effects of Interactive Latency on Exploratory Visual Analysis, *IEEE TVCG* 20 (2014) 2122–2131. doi:10.1109/TVCG.2014.2346452.
- [17] M. Angelini, G. Santucci, H. Schumann, H.-J. Schulz, A Review and Characterization of Progressive Visual Analytics, *Informatics* 5 (2018). doi:10.3390/informatics5030031.
- [18] J. Heer, G. Robertson, Animated Transitions in Statistical Data Graphics, *IEEE TVCG* 13 (2007) 1240–1247. doi:10.1109/TVCG.2007.70539.
- [19] E. F. Tjong Kim Sang, F. De Meulder, Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition, in: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, 2003, pp. 142–147. URL: <https://www.aclweb.org/anthology/W03-0419>.
- [20] S. K. Card, J. D. Mackinlay, B. Shneiderman (Eds.), *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [21] M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software, *PLOS ONE* 9 (2014) 1–12. doi:10.1371/journal.pone.0098679.
- [22] G. Plique, Graphology, a robust and multipurpose Graph object for JavaScript., 2025. doi:10.5281/zenodo.14835805.

# Transparent Persona Generation With LLMs: An Evidence-based and Traceable Method for User-centred Design

Bojan Blažica<sup>1,\*†</sup>, Manca Topole<sup>1,†</sup> and Marko Debeljak<sup>1,†</sup>

<sup>1</sup>*Jožef Stefan Institute, Department of Knowledge Technologies, Jamova cesta 39, Ljubljana, Slovenia*

## Abstract

Personas are a cornerstone of user-centred design, but traditional methods for developing them are difficult to validate, prone to bias and labour-intensive. Data-driven approaches have improved scalability, but often lack the narrative richness and empathy that make personas effective. We present a methodology that uses large language models (LLMs) to accelerate the creation of personas while underpinning and constraining the results with contextual and empirical data. Our approach emphasises transparency and traceability: each generated persona attribute can be linked to its source material, including project documentation, workshop transcripts, survey results or other contextual corpora. By combining the narrative strengths of LLMs with the rigour of an evidence-based foundation, the method generates personas that are both descriptive and verifiable. We present a five-step workflow methodology: (1) generation of persona candidates from contextual data using LLMs, (2) iterative refinement to ensure representativeness of personas, (3) selection of the most relevant profiles through expert evaluation, (4) design of detailed persona profiles, and (5) enrichment with empirical evidence to ensure traceability and validation. The methodology is illustrated with a case study from the field of soil health, but can also be applied to other design contexts where alignment between different stakeholders is crucial. We argue that this approach positions LLMs not as a substitute for human expertise, but as an accelerator of persona work that improves accountability, reduces bias and facilitates communication in collaborative design processes.

## Keywords

personas, large language model, traceability, user-centered design, decision support systems

## 1. Introduction

Personas are an established method in user-centred product design to capture archetypal user needs, their motivations and challenges in a lively, narrative form [1]. They help designers maintain empathy with users and facilitate communication within design teams, but have been criticised for being difficult to validate and for reflecting the biases of the experts who create them, while also being time-consuming to produce [2]. These limitations have motivated interest in computational and data-driven persona approaches that use analytics and algorithms to speed up the creation of personas and ensure that profiles are based on real user data [3]. While quantitative data offers efficiency and representativeness, it alone cannot capture the attitudes and motivations that give personas depth. Recent work suggests that large language models (LLMs) can complement such approaches by analysing qualitative inputs at scale and enriching data-driven personas with contextual nuance and empathy-inducing qualities [4].

The rapid emergence of LLMs has opened up new possibilities for persona generation. LLMs can create plausible and detailed personas from prompts that are either completely synthetic or based on empirical data [5]. Compared to previous algorithmic methods, LLMs are attractive because of their ability to generate coherent narratives, integrate different attributes and adapt to different design contexts. Evaluations suggest that LLM-generated personas can match or even outperform human-generated personas in terms of perceived consistency, informativeness and credibility [6, 7].

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

†These authors contributed equally.

✉ bojan.blazica@ijs.si (B. Blažica); manca.topole@ijs.si (M. Topole); marko.debeljak@ijs.si (M. Debeljak)

ORCID 0000-0003-4597-5947 (B. Blažica); 0009-0004-5404-7549 (M. Topole); 0000-0003-4334-4295 (M. Debeljak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.15>

However, risks remain: LLM-generated personas are prone to stereotyping and may encode demographic or cultural biases inherited from training data [8, 9]. They often lack provenance and traceability, as links between persona attributes and underlying user data are rarely made explicit [3, 4]. They are also vulnerable to hallucinations, generating attributes not grounded in evidence [8, 9], and often produce oversimplified, generic or stereotypical profiles that miss contextual nuance and lived complexity [9, 2]. As a result, scholars emphasise the importance of combining LLMs with human expertise, using them as accelerators rather than replacements in persona workflows [4].

New research has begun to assess the use of LLMs in persona development. Salminen and colleagues [10] show that most published persona prompts produce concise, structured results rather than richer, empathetic narratives, highlighting the lack of established best practices for prompt design. De Paoli [11] and Schuller et al. [6] further explore prompting strategies such as role-play, one-shot examples and gradual refinement and show how these decisions influence the quality and credibility of the generated personas. This work suggests that synthetic personas can support rapid prototyping, simulation and pilot studies, but also indicates that human expertise and participatory input remain important to ensure their validity, transparency and relevance in HCI [4, 5].

The literature suggests that LLM-based persona generation is a promising but unsettled field. Compared to traditional and data-driven personas, LLMs offer unprecedented efficiency and flexibility, but their reliability and inclusivity remain contested [6, 2]. This evolving debate emphasises the potential of LLMs to reshape persona-based design [11, 3] while highlighting the need for critical reflection on how such tools are embedded in design practices and decision-making processes [5, 12]. A sentiment echoed in relation to LLM based tools in general; [13] propose a roadmap with four common approaches that the community has taken to achieve transparency: model reporting, publishing evaluation results, providing explanations, and communicating uncertainty. Similarly, recent work on multi-layered human-centered explainability frameworks [14] introduces an architecture that combines (1) foundational models with built-in explainability mechanisms, (2) a human-centered explanation layer tailored to users' cognitive load and expertise, and (3) a dynamic feedback loop that refines explanations through real-time interaction. Together, these perspectives highlight that transparency extends beyond technical implementation as it must also account for how people interpret and use explanations.

Among the unresolved challenges highlighted in previous work, our study focuses on traceability - the ability to link persona attributes to concrete sources. We argue that without traceability, LLM-generated personas run the risk of being dismissed as arbitrary, especially in collaborative contexts where they serve as shared artifacts among different partners. To address this problem, we propose a transparent method that combines the narrative capabilities of LLMs with contextual data, enabling the creation of personas that are both descriptive and verifiable. Our aim is to demonstrate this approach using a case from the field of soil health, although the method is also transferable to other fields.

## **2. Material and Methods**

This section describes the process used to develop LLM-accelerated and traceable personas. It is divided into three parts. Section 2.1 outlines the overall methodology, from the initial generation of a broad pool of candidate personas to the refinement of the final, evidence-based profiles. Section 2.2 presents the prompt design and the models used. Section 2.3 describes the data sources used to ground and calibrate the personas.

Together, these steps form a reproducible process in which LLMs are used to accelerate the creation of personas, while contextual and empirical data ensure that the resulting personas remain relevant and credible.

### **2.1. Methodology**

The methodology consists of five main steps that structure the process of creating, selecting and refining personas with LLM support.

### **2.1.1. Step 1: Creation of candidate personas**

- Objective: Create an initial pool of personas based on contextual data.
- Description: An LLM was used to generate a wide range of candidate personas driven by contextual data to provide meaningful starting examples. In our case, this contextual data was project documentation (described in section 2.3.1). However, the method is not limited to documents. Other data sources, such as user-generated content, relevant literature (scientific or non-scientific) or other domain materials could equally serve to guide the generation of initial candidates. - Extend the methodology section to include discussion of traceability not only for workshop data but also for document-based persona generation. At this stage, traceability is not explicitly applied, as the contextual documents serve only to provide general thematic direction and domain framing to the LLM. The goal is to obtain diverse but plausible first drafts rather than to attribute specific persona statements to individual sources; detailed evidence linking and traceability are introduced in later steps.

### **2.1.2. Step 2: Iterative refinement**

- Objective: Ensure representativeness.
- Description: The candidate pool was iteratively expanded to ensure representativeness for the domain and the task at hand. In our case, the development of a soil health dashboard meant ensuring coverage of all user types (land steward, private actor, public actor, knowledge actor, citizen), land uses (agriculture, forestry, urban) and scales (local, landscape, regional, European). This was achieved through manual iteration in chat sessions with the LLM, for example, by asking for “more public actor persona examples,” or through domain-specific prompts such as “what about NGOs and activists as examples of citizen personas”.

### **2.1.3. Step 3: Internal selection**

- Objective: Identification of the most relevant personas for the design task.
- Description: Domain experts and project team members were asked to rank the personas created according to their relevance to the project objectives, using a scale from 1 (not at all relevant) to 5 (highly relevant). This ranking helped to focus attention on the most important personas. The top 10% of personas, as determined by ranking were selected as the narrowed pool to be further developed in the next step.

### **2.1.4. Step 4: Drafting persona profiles**

- Objective: Create a set of persona drafts that reflect key contextual dimensions (scale, land use, stakeholder role).
- Description: Persona drafts were developed from the narrowed pool obtained in Step 3 (Section 2.1.3). Additional personas were introduced where stakeholder types were missing based on editorial decision and what is needed by the project. We’ve structured persona drafts around the Soil Health Dashboard’s main functionalities (monitoring, assessing, and managing soil health) and included stakeholder-relevant sections (e.g., context, primary goals and tasks, pain points). We’ve further elaborated drafts with the help of iterative LLM chat sessions, which resulted in additional relational elements such as hobbies, motivational quotes, and persona images. These elements were included to support relatability in design and use contexts, while remaining secondary to the functional sections.

### **2.1.5. Step 5: Evidence-based refinement**

- Objective: Refine the persona profiles using empirical data from workshops to ensure that they reflect the actual needs, challenges, goals and priorities of soil health stakeholders.

- **Description:** The persona drafts created in the previous step (Section 2.1.4) were systematically enriched with empirical data from stakeholder workshops. Integration was carried out using LLM, with each refinement cycle combining a detailed and structured prompt with the draft persona, five Excel files with filtered stakeholder- and land-use-specific data (as described in 2.3.2) and a sample persona file to ensure consistent formatting and structure. The resulting persona profiles were then manually reviewed for internal consistency, clarity and logical alignment. This process ensured that each persona could be traced back to empirical evidence while maintaining coherence.

## 2.2. LLM Use and Prompts

At the time of developing this methodology, it was not possible to feed proprietary data directly into the OpenAI interface; therefore, we used an internal deployment of Onyx (formerly Danswer) for Step 1 (Section 2.1.1) and Step 2 (Section 2.1.2). Onyx is a knowledge-augmented chat tool that combines contextual retrieval with LLM chat integration, allowing documents to be indexed locally and used in conversation with an LLM. The project documentation was entered into Onyx and connected to GPT-4 via an API. In all steps, LLMs were used to generate text and, in the final step, also to generate images for persona profiles. All interactions were carried out either via a chat interface, although automation via API would have been possible.

In Step 1 (Section 2.1.1) and Step 2 (Section 2.1.2), GPT-4 was prompted through Onyx to generate candidate personas in open dialogue sessions. Each request asked for the following information: user, user type, land use, scale, and persona description. In our case, the contextual data came from the project documentation (see Section 2.3.1). A single researcher conducted these sessions, iteratively expanding the pool to improve the coverage of user types, land uses and scales.

Persona drafts were developed through iterative chat sessions with GPT-4o, performed directly in the OpenAI interface, which by then allowed direct grounding with user-supplied data. Prompts were applied in a structured yet conversational manner to elaborate persona drafts while maintaining alignment with the predefined framework (candidate persona pool) established in earlier steps and the project documentation.

In Step 5 (Section 2.1.5), persona refinement was carried out with GPT-5 (auto setting). Here, the process of prompt development was more elaborate: several cycles of testing and comparison were used to arrive at a stable version that provided consistent, evidence-based results. The final prompt included explicit instructions on how to map empirical workshop data to persona sections, how much data to include, and how to cite individual stakeholder statements to ensure traceability. Additional instructions refined during testing specified selection boundaries, narrative style, and formatting rules to keep the profiles consistent and concise. This final version was applied to all completed personas, and the full prompt text is included in Appendix A.

Default parameters for seed and temperature parameters were used as it is not possible to set them in the web-based interface. Comparing various LLMs for persona generation was not in the scope of this work.

## 2.3. Data

To mitigate the limitations of LLM-based personas mentioned in the **Introduction**, we grounded our persona development in contextual data. This grounding occurred in two stages: first, a light, broad contextual seeding in the initial persona generation; then later, a deeper grounding for the final profiles using domain-specific empirical data.

### 2.3.1. Preliminary grounding via contextual documents

In the initial phase, we engaged in conversational prompting with the LLM while supplying project documentation as context. In our case, the project is the EU funded research project BENCHMARKS (<https://soilhealthbenchmarks.eu/>), which aims to develop soil health indicators across different land

uses, stakeholders, and spatial scales. The intended personas are to be used in user-centered design and development of a Soil Health Dashboard: a tool that integrates local, landscape, regional, and European data (e.g., statistical indicators, remote sensing, soil sampling) to help users choose indicators suited to their objectives, land use, and context. By feeding the model the project's DoA (description of the action, also description of work - DoW or Annex 1) document (which covers objectives, scientific rationale, work packages, partner roles, and more), we enabled the LLM to anchor persona suggestions in the project's domain, goals, and structure.

### **2.3.2. Final profile calibration data**

The project workshops involved over 500 stakeholders across 11 European countries and 9 languages. Each of the 24 local workshops combined plenary and group sessions, where participants worked with sticky notes and posters to record inputs. Activities were designed to capture stakeholder perspectives on soil health. This empirical input provided stakeholder-grounded information essential for developing a Soil Health Dashboard that is both relevant and useful for its intended users. While the workshops were not originally conducted for persona development, their results proved highly suitable for this purpose, as integrating them into this process further ensured that the final tool addresses stakeholder needs and priorities.

The workshops comprised different activities in which stakeholders explored soil health challenges, their objectives and management practices, needs, and prioritized soil functions. Outputs from these activities were translated into English by local facilitators and transcribed into standardized Excel sheets. We extracted data subsets corresponding to each persona's stakeholder and land-use type. This resulted in four datasets linked to personas, namely their needs, objectives, practices, and prioritized soil functions. Challenges were also extracted but could not be linked to specific stakeholder types due to the structure of the workshops. Each dataset was stored in a dedicated Excel file, forming the empirical evidence base for Step 5 (Section 2.1.5) of our persona development methodology.

## **3. Results**

The proposed methodology yielded intermediate and final results. Intermediate outputs show the evolution from a large pool of seed personas to a smaller set of prioritized candidates. The final output consists of fully elaborated and traceable persona profiles, calibrated with contextual and empirical data.

### **3.1. Seed personas**

After the initial generation and iterative refinement (Sections 2.1.1 and 2.1.2), we obtained a list of 100 seed personas (Table 1). The list was representative of our domain and balanced across land use, user type, and scale (Table 2).

### **3.2. Ranked personas**

In the internal selection (Section 2.1.3), eight project participants rated each persona on a scale from 1 (not relevant) to 5 (very relevant). Based on the average scores, the top 10% were selected as candidates for further development (Table 3). The average score was 3.50 with a standard deviation of 0.92. After this step, some of the user types (private actors, citizens) and scales (landscape) were not represented, so an editorial decision had to be made as to whether or not they should be included in the final selection for profile development.

### **3.3. Final personas**

The final stage produced a set of 9 calibrated personas representative of the project. Each profile was generated in two steps: first, a draft including general personal data (Section 2.1.4), and second,

**Table 1**

Excerpt of 5 rows from the full set of 100 seed personas showing the simplified version of personas used in the first step of persona creation.

| n  | User Type       | User                 | Land Use    | Scale     | Persona Description                                                                                                |
|----|-----------------|----------------------|-------------|-----------|--------------------------------------------------------------------------------------------------------------------|
| 1  | Land Steward    | Farmer               | Agriculture | Local     | John Doe; 45, Male; Goal: Improve crop yield; Hobbies: Gardening, Reading                                          |
| 18 | Private Actor   | Agribusiness Manager | Agriculture | Landscape | Emily Brown; 42; Female; Goal: Integrate soil health into business practices; Hobbies: Traveling, Cooking          |
| 37 | Public Actor    | Regional Policymaker | Agriculture | Regional  | Anna Black; 40; Female; Goal: Implement soil health policies; Hobbies: Volunteering, Yoga                          |
| 59 | Knowledge Actor | Soil Scientist       | Agriculture | Europe    | Dr. Alan Brown; 60; Male; Goal: Research soil health indicators; Hobbies: Writing, Lecturing                       |
| 82 | Citizen         | Citizen Scientist    | Agriculture | Local     | Alex Grey; 30; Non-binary; Goal: Contribute to soil health data collection; Hobbies: Community gardening, Blogging |

**Table 2**

Summarised distribution of personas across land use, user type and scale, showing balanced coverage

| User Type       | Land Use       | Scale        |
|-----------------|----------------|--------------|
| Land Steward    | 17 Agriculture | 24 Local     |
| Private Actor   | 19 Forestry    | 20 Landscape |
| Public Actor    | 22 Urban       | 23 Regional  |
| Knowledge Actor | 23 All         | 33 Europe    |
| Citizen         | 19             |              |

refinement and calibration with empirical data (Section 2.1.5). This process resulted in a persona profile containing two distinct categories of information:

- **General persona data** (personal details that set the broad context and help readers empathize with the persona): name, age & nationality, primary scale & land use, role & production system, farm & climate context, tech proficiency & access, decision horizon & authority, hobbies and motivational quote.
- **Empirical, domain-specific data** (attributes grounded in empirical evidence from stakeholder workshops): primary goals, pain points, needs, dashboard expectations and scenario of use.

To illustrate, *Ana Martínez Sánchez* (Figure 1), an organic arable farmer from Castilla-La Mancha from Spain, was developed into a complete profile. The following examples illustrate how empirical

**Table 3**

The short list of personas with their average vote, standard deviation and overall ranking. Ranking resulted in missing stakeholder types in the selection calling for an editorial decision to complete the list.

| n  | User Type       | Land Use    | Scale    | Min | Max | Avg  | Std  | Rank |
|----|-----------------|-------------|----------|-----|-----|------|------|------|
| 3  | Land Steward    | Agriculture | Local    | 4   | 5   | 4,75 | 0,43 | 4    |
| 8  | Land Steward    | Agriculture | Local    | 4   | 5   | 4,88 | 0,33 | 2    |
| 9  | Land Steward    | Forestry    | Local    | 4   | 5   | 4,75 | 0,43 | 4    |
| 37 | Public Actor    | Agriculture | Regional | 3   | 5   | 4,50 | 0,71 | 10   |
| 38 | Public Actor    | Agriculture | Regional | 4   | 5   | 4,75 | 0,43 | 4    |
| 39 | Public Actor    | Agriculture | Europe   | 4   | 5   | 4,88 | 0,33 | 2    |
| 41 | Public Actor    | All         | Europe   | 3   | 5   | 4,50 | 0,71 | 10   |
| 52 | Public Actor    | Forestry    | Regional | 4   | 5   | 4,50 | 0,50 | 9    |
| 54 | Public Actor    | Urban       | Europe   | 3   | 5   | 4,63 | 0,70 | 8    |
| 59 | Knowledge Actor | Agriculture | Europe   | 5   | 5   | 5,00 | 0,00 | 1    |
| 60 | Knowledge Actor | Agriculture | Europe   | 4   | 5   | 4,75 | 0,43 | 4    |
| 64 | Knowledge Actor | All         | Europe   | 3   | 5   | 4,50 | 0,71 | 10   |

evidence shaped her persona:

- **Primary goals** – safeguard soil fertility and resilience under dryland organic farming, grounded in workshop objectives on soil fertility and nutrient cycling.
- **Pain points** – recurrent drought limiting yields, linked to challenges identified by workshop participants.
- **Needs** – validated protocols to measure soil improvements, based on stakeholder needs for objective, standardized assessment methods.
- **Dashboard expectations** – scenario simulations comparing crop rotations and cover crops, connected to workshop objectives and practices on nutrient cycling and water management.

A full version of Ana’s persona, including both general and empirical elements with explicit source references, is provided in Appendix B.

## 4. Discussion

The aim of this study was to address several problems in the development of personas: lack of tractability, bias, oversimplification, limited empirical foundation, and the challenge of balancing narrative richness and methodological rigour. Our approach addresses these issues by combining the generative power of LLMs with structured contextual and empirical data and iterative expert supervision. By anchoring persona characteristics in data, including project documentation (contextual data) and data from stakeholder workshops (empirical data), we enhanced transparency and credibility while reducing the risk of arbitrary or hallucinated results. Each persona characteristic could be linked to a verifiable source, strengthening accountability in collaborative design processes.

A second important innovation of the presented research work lies in the structured, multi-step methodology that integrates generation, refinement, validation and evidence-based calibration into a single workflow. Unlike traditional persona methodologies, which are often static and subjective, our approach accelerates the creation of personas while maintaining contextual relevance, stakeholder alignment and motivational depth. Compared to existing data-driven approaches that focus on scalability at the expense of empathy and tone, our method creates personas that are both representative and meaningful for user-centred design. Recent research has begun to tackle aspects of this challenge. For example, De Paoli [11] proposed iterative workflows combining LLM-based analysis and persona drafting, Shin et al. [4] studied how to distribute tasks between humans and AI in persona generation, and Jung et al. [3] introduced structured data-to-persona pipelines to improve transparency. These efforts, however, focus on isolated stages, whereas our approach integrates generation, refinement, validation, and calibration into a single workflow. This structured methodology is not only novel in itself but also lays the foundation for practical integration into decision support systems (DSSs).

The integration of personas into the development of DSSs is expected to bring several practical benefits. We anticipate that personas can act as a bridge between technical development and user expectations by helping to recognise inconsistencies early on and guiding adjustments to the design of the user interface and data presentation. They are also expected to facilitate communication between the various stakeholders and support the negotiation of trade-offs between complexity and usability. Most importantly, we see the potential for personas to serve as recurring reference points throughout the project, influencing design priorities and evaluation criteria. In future work, we plan to evaluate whether personas indeed evolve from static artifacts into dynamic tools that support decision-making and increase user confidence during the continued development of the DSS.

However, our approach is also associated with some limitations. The quality and representativeness of the contextual data had a strong influence on the results. Biases or gaps in the source material were sometimes reproduced, even if they were more transparent. Human oversight also remained crucial. Careful, timely development, iterative testing and expert validation were required to maintain contextual accuracy, which increased resource requirements. Nevertheless, each validation cycle improved the quality of the persona, but also increased the complexity, and the human judgment required inevitably led to a degree of subjectivity. This reflects the fact that incorporating human contexts into design artifacts requires interpretive involvement in addition to automatic generation.

These considerations highlight both the potential and the current limitations of LLM-based persona generation and point to future methodological refinements as briefly outlined in the conclusions.

## 5. Conclusions and future work

This study presents a transparent, evidence-based methodology for creating user personas that combines the narrative capabilities of LLMs with contextual and empirical data. The approach addresses key



Figure 1: Persona illustration of Ana Martínez Sánchez, an organic arable farmer from Castilla-La Mancha, Spain. The image was generated with an LLM and is intended to support empathy and visualization rather than depict a real individual.

limitations of existing methods by improving tractability, reducing bias, and preserving narrative richness while accelerating the development of DSSs. Applied to the soil health domain, the method shows potential to strengthen user-centred development of DSS by improving stakeholder alignment and system relevance. We expect that, beyond their direct design function, personas can also serve to connect project partners by providing a shared understanding of stakeholder groups and by linking different project objectives to the same shared personas.

Future work will evaluate these expected benefits in practice and will aim to automate provenance tracking, integrate dynamic data, and include additional stakeholders to improve the scalability and impact of persona-driven design. Replicating the workflow across multiple LLM families and diverse application domains will allow us to validate its consistency, broaden its relevance, and strengthen its contribution to transparent, evidence-based persona generation.

In addition, drawing on established persona evaluation methodologies such as the Persona Perception Scale [15], future studies should compare evidence-grounded LLM–human personas with human-only and LLM-only baselines to assess perceived realism, usefulness, and empathy.

In line with earlier work comparing algorithmically generated and human-crafted personas [? ], future studies will also conduct comparative evaluations contrasting fully synthetic LLM personas, the proposed evidence-grounded LLM + human workflow, and human-only baselines to understand how personas are perceived by persona users and how likely they will adopt them. Finally, since transparency is central to our contribution, further research should also explore more objective and automated methods for verifying footnote-to-source correctness, as the current manual review process, while effective, remains labour-intensive.

## Acknowledgments

This study was supported by the Slovenian Research Agency (grant P2-0103) and by the European Union’s Horizon Europe research and Innovation program, under Grant Agreement: 101091010, Project BENCHMARKS. We gratefully acknowledge the partners of the BENCHMARKS project and all participants of the stakeholder workshops. Special thanks to Fabio Volkmann, who led the workshops where the data were collected.

## Declaration on Generative AI

During the preparation of this work, the author(s) used OpenAI GPT-5 in order to: assist with literature review, text drafting, rephrasing, and grammar correction. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the publication’s content.

## References

- [1] T. Miaskiewicz, K. A. Kozar, Personas and user-centered design: How can personas benefit product design processes?, *Design studies* 32 (2011) 417–430.
- [2] C. Lazik, C. Katins, C. Kauter, J. Jakob, C. Jay, L. Grunske, T. Kosch, The impostor is among us: Can large language models capture the complexity of human personas?, *arXiv preprint arXiv:2501.04543* (2025).
- [3] S.-G. Jung, J. Salminen, K. K. Aldous, B. J. Jansen, Personacraft: Leveraging language models for data-driven persona development, *International Journal of Human-Computer Studies* 197 (2025) 103445.
- [4] J. Shin, M. A. Hedderich, B. J. Rey, A. Lucero, A. Oulasvirta, Understanding human-ai workflows for generating personas, in: *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, 2024, pp. 757–781.

- [5] M. Prpa, G. Troiano, B. Yao, T. J.-J. Li, D. Wang, H. Gu, Challenges and opportunities of llm-based synthetic personae and data in hci, in: Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, 2024, pp. 716–719.
- [6] A. Schuller, D. Janssen, J. Blumenröther, T. M. Probst, M. Schmidt, C. Kumar, Generating personas using llms and assessing their viability, in: Extended abstracts of the CHI conference on human factors in computing systems, 2024, pp. 1–7.
- [7] J. Salminen, C. Liu, W. Pian, J. Chi, E. Häyhänen, B. J. Jansen, Deus ex machina and personas from large language models: Investigating the composition of ai-generated persona descriptions, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–20.
- [8] H. A. Haxvig, Concerns on bias in large language models when creating synthetic personae, arXiv preprint arXiv:2405.05080 (2024).
- [9] L. Sun, T. Qin, A. Hu, J. Zhang, S. Lin, J. Chen, M. Ali, M. Prpa, Persona-l has entered the chat: Leveraging llms and ability-based framework for personas of people with complex needs, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–31.
- [10] J. Salminen, D. Amin, B. Jansen, Using ai for user representation: An analysis of 83 persona prompts, arXiv preprint arXiv:2508.13047 (2025).
- [11] S. De Paoli, Improved prompting and process for writing user personas with llms, using qualitative interviews: Capturing behaviour and personality traits of users, arXiv preprint arXiv:2310.06391 (2023).
- [12] R. Y. Pang, H. Schroeder, K. S. Smith, S. Barocas, Z. Xiao, E. Tseng, D. Bragg, Understanding the llm-ification of chi: Unpacking the impact of llms at chi through a systematic literature review, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–20.
- [13] Q. V. Liao, J. W. Vaughan, Ai transparency in the age of llms: A human-centered research roadmap, arXiv preprint arXiv:2306.01941 (2023).
- [14] C. De Silva, T. Halloluwa, D. Vyas, A multi-layered research framework for human-centered ai: Defining the path to explainability and trust, arXiv preprint arXiv:2504.13926 (2025).
- [15] J. Salminen, J. M. Santos, H. Kwak, J. An, S.-g. Jung, B. J. Jansen, Persona perception scale: development and exploratory validation of an instrument for evaluating individuals’ perceptions of personas, International Journal of Human-Computer Studies 141 (2020) 102437.

## Appendices

### A. Prompt used for final persona calibration

0. Canonical inputs (will be provided)  
 Prompt document: the canonical instructions. If anything is left unclear, consult me before making a decision when amending the persona profile.  
 Profile template document (Word, finalized persona; e.g., Ana): copy fonts, sizes, spacing, section sequence, bullets, headings from this template.  
 Persona draft document (Word) for the new persona (includes persona type which specifies personas stakeholder type and land use type + includes information section (from the Benchmarks milestone reports) which sets the framework for the persona profile)  
 Five Excel workshop files filtered to the personas stakeholder type land use:
  - i. Challenges (land-use only)
  - ii. Needs (stakeholder land use)
  - iii. Objectives (stakeholder land use)
  - iv. Practices (stakeholder land use)
  - vi. Priority soil-function votes (stakeholder land use)
1. File-to-section mapping (How to integrate workshop data to user persona profile)

- Challenges Pain points  
 Needs Needs (general requirements); also inform Dashboard expectations  
 Objectives Primary goal, Dashboard expectations  
 Practices Dashboard expectations, Scenario of use  
 Soil-function votes Weave naturally into Primary goal / Dashboard expectations / Scenario of use (not a separate section)
2. Pre-processing & checks
    - i. Open the Persona draft and identify sections:  
 Keep (in this order) unless contradicted: title (name and a few main details), persona type, Inspiration, Age & nationality, Primary scale & land use, Context (title of this section differs from persona to persona, depending on persona type), Tech proficiency & access, Decision horizon & authority, Hobbies, Motivational quote  
 Replace (or blend per step 6): Primary goals, Pain points, Needs, Dashboard expectations, Scenario of use
    - ii. Open Excels and confirm ID columns:  
 challenge\_id, need\_id, objective\_id, practice\_id.  
 In Soil-function document use rank\_in\_slice column and soil\_function\_name; treat rank as the canonical identifier.
    - iii. Multi-item rows: if a cell contains multiple statements (newlines/semicolons), treat each as a separate statement unless clearly one phrase. You will cite only the fragment used later.
    - iv. Slice correctness: only use rows matching the personas stakeholder type and land use (for Challenges: land use only).
  3. Selection heuristics (what to keep)
 

Objectives: choose the 2 to 4 most relevant to the personas ambition and the product (dashboard).

Pain points: 3 to 4 max, concrete and role-true (systemic barriers for public/knowledge actors; on-farm constraints for land stewards).

Needs: focus on dashboard-relevant needs (analytics, indicators, comparability, reporting, scenarios, data/monitoring, evidence flows), possibly use others as well for broader context

Practices: pick 1 to 3 that the dashboard can model/compare/monitor.

Soil functions: take the top 3 by rank\_in\_slice for this personas slice.
  4. Integration style (how to write)
 

Bold any sentence/phrase grounded in workshop data.

Blend naturally: do not say workshops say. Let the narrative sound like the persona; the footnotes carry the traceability.

Soil functions in text: refer to them naturally (e.g., habitat provision, carbon & climate regulation, water regulation), without explicit (1st/2nd/3rd priority) unless the request explicitly asks for showing ranks.

Footnote(s) will still point to Soil function (priority X) in the Evidence Table.

Bullets only for Pain points; all other sections are short narrative paragraphs.
  5. Superscript footnotes (no brackets in text)
 

Footnotes are in the form of superscript numbers

For new workshop-grounded statements you add, insert a new superscript number.

Footnotes must be numbered sequentially across the whole profile
  6. Section-by-section procedure
    - i. Primary goal (blend or replace):
      - If the draft goal fits the Inspiration, blend workshop objectives/soil-function priorities naturally into 2 to 3 sentences.
      - If misaligned, replace with a short narrative merging: 2 to 3 Objectives + (naturally referenced) soil functions + (optionally) 1 dashboard-relevant Need.
      - Keep to ~70 to 120 words max.
    - ii. Pain points:
      - 3 to 4 bullets from Challenges, concise and role-true.
      - Each bullet grounded in a challenge row (footnote).

- iii. Needs:
    - 1 short paragraph (2 to 4 sentences), focus on dashboard-relevant needs (3 core ideas).
    - Examples: indicators/thresholds, scenario analysis, monitoring/benchmarking, data comparability, evidence export/sharing.
  - iv. Dashboard expectations:
    - 1 short paragraph (3 to 4 ideas), combining Objectives + Practices + (naturally referenced) soil functions.
    - Speak in terms of what the dashboard should enable (compare, simulate, monitor, export).
  - v. Scenario of use:
    - 3 to 5 sentences. A simple story of how the persona interacts with the dashboard, weaving 1 to 2 practices and at least one soil-function aspect (naturally).
    - Avoid feature laundry lists.
  - vi. Other draft sections: keep as is unless contradicted by data.
7. Evidence building (footnotes table)
 

Create an Evidence Table on a new page after the profile text.

Columns: Footnote | Source | Workshop quotation.

For each superscript number, map to its source:

    - Objective N quote objective\_text.
    - Need N quote need\_text.
    - Challenge N quote challenge\_text.
    - Practice N quote practice\_text.
    - Soil function (priority X) quote soil\_function\_name for that rank from the personas slice.

If a cell had multiple statements: include only the fragment used and append (excerpt).

If text was used verbatim: show plain (no label).

Maintain the same numbering order as they appear in the profile.
  8. Formatting (must mirror template)
 

Use the Profile template document for:

    - Fonts, sizes, line spacing, margins, headings, bullet style.
    - Section sequence exactly as in the template.
    - Superscripts for footnote markers (not inline digits).

No duplicate sections: remove any draft versions of the five updated sections.
  9. Length control (1-page profile)
 

The profile text (from Primary goal through Scenario of use) should fit within ~1 page in the templates formatting.

If over limit, trim in this order:

    1. Dashboard expectations (compress to 3 ideas)
    2. Needs (down to 3 tight items)
    3. Primary goal (keep to 2 sentences)
    4. Pain points (max 3 bullets)

Do not trim the Evidence Table for length; it belongs after the page break.
  10. Quality & consistency checks (before exporting)
 

No contradictions across sections.

All bolded claims have a footnote.

Footnote numbers are sequential and each appears in the Evidence Table.

Soil functions appear naturally in text; priority is represented in footnotes/table, not parenthetical numbers in narrative (unless explicitly requested).

Needs are dashboard-relevant (funding/bureaucracy excluded by default; minor policy context allowed for public actors).

Fits ~1 page profile text in template formatting.
  11. When to ask
 

Ask the user before proceeding if:

The prompt doc and this instruction conflict and its unclear which rule to apply.

The personas slice (stakeholder land use) is ambiguous or missing in an Excel. An objective/need/practice cannot be matched to a credible row. Fitting to 1 page would require removing substantive content.

## 12. Output

Word document (.docx) that:

- Mirrors the template formatting and section order,
- Contains the amended profile with bold workshop-grounded statements and superscript footnotes,
- Ends with an Evidence Table mapping each footnote to the exact workshop quotation (or (excerpt)).

## B. Final persona profile with traceable data

|                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Name</b>                           | Ana Martínez Sánchez                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Age and nationality</b>            | 44, Spanish (Castilla-La Mancha)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Primary scale and land use</b>     | Agriculture, Local (20 ha certified organic arable land)                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <b>Farm and climate context</b>       | Mediterranean climate — hot, dry summers; mild, wetter winters; occasional intense autumn rains typical of Spain's central plateau                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>Role and production system</b>     | Owner-manager of a 20 ha rainfed (dryland) organic farm. Three-year rotation — year 1: barley for grain; year 2: chickpeas for export; year 3: vetch/cover crops grown as green manure and grazed by a neighbour's ewes. Production relies solely on seasonal rainfall (no irrigation), shallow non-inversion tillage, and composted sheep manure. Her aim is to conserve soil moisture, build organic matter, and keep weeds in check with mechanical tools rather than herbicides.                                       |
| <b>Tech proficiency and access</b>    | Basic desktop PC in her home office; smartphone in the field; intermittent 4G — prefers offline ready exports.                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Decision horizon and authority</b> | Seasonal; full decision power over field operations and budget.                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Hobbies</b>                        | Runs a stall at the village cooperative market; experiments with heirloom tomato varieties in her kitchen garden; makes preserves with her two teenage children; and occasionally hikes with friends in the nearby sierras.                                                                                                                                                                                                                                                                                                |
| <b>Motivational quote</b>             | "Take care of what takes care of you — soil, people, and community."                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Primary goals</b>                  | Ana's main ambition is to safeguard soil fertility and resilience under dryland organic farming, while staying economically viable and compliant with organic certification <sup>1</sup> . She wants to compare how practices such as cover crops, shallow tillage, and compost affect yields, soil organic matter, and erosion control <sup>2</sup> . Maintaining soil functions like nutrient cycling <sup>3</sup> and water regulation and purification <sup>4</sup> is central to her vision of long-term stewardship. |
| <b>Pain points</b>                    | Recurrent drought limiting yields <sup>5</sup> ; Erosion susceptibility on some fields <sup>6</sup> ; Administrative burden pulling time from field management <sup>7</sup>                                                                                                                                                                                                                                                                                                                                                |
| <b>Needs</b>                          | She values hands-on, validated guidance she can trust. She needs objective, validated measurement protocols to assess soil improvements <sup>8</sup> , plus stronger connections between farmers and researchers for practical problem-solving <sup>9</sup> . Peer learning matters: exchange of good practices and materials with nearby farmers <sup>10</sup> , and active spaces to talk through needs and experience <sup>11</sup> help turn ideas into confident action.                                              |

| Footnote | Source ID                  | Workshop transcript quote                                                                                                                                                                                         |
|----------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1        | Objective 4                | Primary production, profitability, crisis preparedness, intergenerational (excerpt)                                                                                                                               |
| 1        | Objective 48               | Enrich the soil by improving the soil content of NPK, CaCO <sub>3</sub> , SO <sub>3</sub> and trace elements, organic matter (excerpt)                                                                            |
| 2        | Objective 2                | In the nutrient cycling, the goal is as closed cycles as possible (excerpt)                                                                                                                                       |
| 2        | Objective 3                | From land drainage to water management (excerpt)                                                                                                                                                                  |
| 3        | Soil function (priority 1) | Nutrient cycle                                                                                                                                                                                                    |
| 4        | Soil function (priority 2) | Water regulation and purification                                                                                                                                                                                 |
| 5        | Challenges 5, 39           | Drought                                                                                                                                                                                                           |
| 6        | Challenge 31               | Acidity and erosion susceptibility (soil properties)                                                                                                                                                              |
| 7        | Challenge 56               | The burden of bureaucracy and administration takes away resources and resources from managing the country                                                                                                         |
| 8        | Need 59                    | Objective protocols are validated for measuring soil health, Present and disseminate the results obtained on experimental sites for the use of organic waste products, accompany visits and tests on soil quality |
| 9        | Need 52                    | Interconnection offer between farmers and research to share experiences on agricultural techniques and carry out activities with these audiences (excerpt)                                                        |
| 10       | Need 47                    | Promotion of a reception area on the Farm for inter-stakeholder meetings (excerpt)                                                                                                                                |
| 11       | Need 10                    | Active interaction, talk about needs and experiences (excerpt)                                                                                                                                                    |
| 12       | Objective 2                | In the nutrient cycling, the goal is as closed cycles as possible (excerpt)                                                                                                                                       |
| 12       | Objective 3                | From land drainage to water management (excerpt)                                                                                                                                                                  |
| 12       | Objective 4                | Primary production, profitability, crisis preparedness, intergenerational (excerpt)                                                                                                                               |
| 13       | Objective 48               | Enrich the soil by improving the soil content of NPK, CaCO <sub>3</sub> , SO <sub>3</sub> and trace elements, organic matter (excerpt)                                                                            |
| 14       | Practice 4                 | Include grasses/clovers in the crop rotation (excerpt)                                                                                                                                                            |
| 14       | Practice 5                 | Diversifying crop rotations (excerpt)                                                                                                                                                                             |
| 15       | Practice 4                 | Include grasses/clovers in the crop rotation (excerpt)                                                                                                                                                            |
| 16       | Practice 1                 | Diverse crop rotation depending on the situation (excerpt)                                                                                                                                                        |

### Dashboard expectations

She expects the dashboard to integrate field data with clear, easy-to-interpret soil health indicators (e.g., organic matter, erosion risk)<sup>8</sup>, and to simulate scenarios comparing different crop rotations, cover crop mixes, or tillage intensities<sup>12</sup>. It should highlight the contribution of practices to water retention and nutrient availability<sup>13</sup>, and allow her to download concise, certification-ready reports<sup>14</sup>.

### Scenario of use

After harvest, Ana logs her tillage operations and uploads soil organic carbon results. The dashboard visualises the organic matter trend and alerts her to compaction and erosion risk<sup>15</sup>. Curious about alternatives, she runs a scenario comparing barley–chickpea with a more diverse rotation including oats and lupins<sup>16</sup>, seeing how soil functions like water holding capacity respond. She exports a PDF summary to share with her farm advisor before deciding on next season's rotation.

# Inferring a Mobile User's Valence and Arousal through On-Screen Text Analysis

Edita Džubur<sup>1</sup>, Veljko Pejović<sup>1</sup>

<sup>1</sup>University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia

## Abstract

Understanding a user's emotional state is critical for building adaptive and intelligent mobile applications. In this paper we investigate the feasibility of inferring valence and arousal from the text displayed on smartphone screens. We developed AV-Sense, a mobile application that combines the Experience Sampling Method, a technique that prompts users to report their feelings in the moment, with passive screentext logging. In a two-week study with 12 participants, we collected 787 ESM responses and over 650,000 screentext entries. Data analysis revealed meaningful temporal and individual patterns in reported affect. We then explored the use of large language models to predict valence and arousal from screentext, but results indicated limited predictive power in this setting. Our findings highlight both the potential and current challenges of screentext-based affect inference, laying the groundwork for future research on emotion-aware applications and naturalistic psychological studies.

## Keywords

Text analysis, experience sampling method, screentext sensing, valence, arousal, large language models

## 1. Introduction

Smartphones have become the primary medium for communication, information access, and digital services. Despite their ubiquity, these devices have little understanding of the emotional state of their users. Applications typically adapt based on context such as location or activity, but they rarely consider affective states such as valence and arousal.

Valence and arousal are the two fundamental dimensions of affect used in psychological models of emotion. Valence describes how pleasant or unpleasant an experience feels, whereas arousal reflects its intensity or level of activation, ranging from very calming to highly exciting [1]. This two-dimensional representation is widely used in affective science and has been established through decades of research.

Recognizing these states could enable more adaptive applications such as adjusting notifications, recommending suitable activities, or supporting well-being interventions. It could also provide valuable insights for psychological research in naturalistic settings [2].

Unlike physical activity or location, emotional state cannot be directly sensed with smartphone hardware. However, there is evidence that the content users engage with, particularly on-screen text, reflects their affective state [3, 4]. Simultaneously, recent advances in natural language processing, especially through large language models (LLMs), provide an attractive avenue for seamless analysis of texts in search for potential link between the content and the affect of a user consuming the content.

In this study, we developed *AV-Sense*, a mobile application based on the AWARE-Light [5] framework that integrates the Experience Sampling Method (ESM) with passive screentext logging. The application periodically prompted users to self-report their affective state on a two-dimensional valence–arousal grid, while continuously recording the textual content displayed on the smartphone screen. We conducted a two-week study with 12 participants, during which the system collected 787 ESM responses paired with approximately 650,000 screentext entries. This dataset enabled us to analyze temporal and individual patterns in reported affect, assess the availability of screentext data as contextual information, and explore the feasibility of applying large language models to predict valence and arousal from naturalistic

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

✉ veljko.pejovic@fri.uni-lj.si (V. Pejović)

ORCID 0000-0002-9009-0024 (V. Pejović)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.16>

screen text. While the predictive performance of LLMs was limited, the study provides a first step towards understanding the potential and boundaries of screentext-based affect inference.

## 2. Related Work

### 2.1. Mobile sensing and the Experience Sampling Method

Mobile sensing has become an important way to study human behavior, since smartphones are always with users and carry a variety of built-in sensors [2]. When combined with ESM, this makes it possible to collect data about people’s daily lives in a more ecologically valid way [6, 3]. ESM asks participants to provide in-the-moment self-reports, which helps reduce recall bias and capture affective states in their natural environment. Previous studies have shown that smartphones can measure mood and mental states by combining signals such as location, activity, and communication patterns [7, 8]. These works confirm that frequent affect measurement is feasible, but they also point out challenges around privacy, technical stability, and participant burden.

### 2.2. Inferring affect from text and screentext sensing

Language is one of the clearest ways people express emotions. Research has shown that linguistic markers are often linked to psychological and affective states [3]. This insight led to the development of sentiment analysis methods and affective lexicons, which are now widely used in psychology and computer science [9]. With the recent progress in machine learning, large language models have also been tested for emotion detection from text, offering better contextual understanding than earlier lexicon-based tools [10].

More recently, researchers have looked at the text people actually see on their smartphone screens, known as *screentext*. Teng et al. introduced a screentext sensor for Android as part of the AWARE-Light framework, which enables continuous and privacy-conscious logging of on-screen text [11]. In later work, they showed that screentext data can be used to predict affective states, with LLM-based prompting approaches performing better than simple classifiers [4]. These studies suggest that screentext could serve as a useful, non-intrusive signal of user affect, while also highlighting the difficulty of modeling subtle and context-dependent emotions.

Our work continues in this direction by combining screentext sensing with real-time affect reporting on a two-dimensional valence–arousal grid. Where earlier studies often used weekly or retrospective questionnaires, our approach links screentext with immediate self-reports, giving us a closer look at how affect can be inferred in everyday settings.

## 3. Data collection

### 3.1. AV-Sense application

AV-Sense was built on top of the AWARE-Light framework, which is an open-source platform for mobile sensing. The main goal of the app is to combine active self-reporting with passive data collection. For self-reports, we introduce a new type of an ESM question in the form of a two-dimensional grid as shown in Figure 1. On this grid, participants rate their current valence, ranging from negative to positive on a scale from -3 to 3, and their arousal, ranging from calm to excited on the same scale. The grid also displays examples of emotions placed on the coordinate system to illustrate the combinations of the two dimensions. When a prompt appeared, participants saw the instruction: *"Please rate how you feel right now based on your current mood."* They were encouraged to report their general emotional state at that moment, not necessarily emotions related to the phone content. This clarification ensured consistency across self-reports and reduced ambiguity about whether responses should reflect momentary feelings or reactions to specific on-screen material.

How do you feel right now?

Move the dot to represent your current emotional state.

Valence: How positive or negative the emotion feels.  
Arousal: How intense or relaxed the emotion feels.

AROUSAL

high

|            |         |           |       |
|------------|---------|-----------|-------|
| Infuriated | Fear    | Excited   | Happy |
| Angry      | Alarmed | Delighted |       |
| Frustrated | Amused  | Glad      |       |
|            | Annoyed | Curious   |       |
| Miserable  | Bored   | Content   |       |
| Depressed  | Gloomy  | Satisfied |       |
| Sad        | Tired   | Relaxed   |       |
|            |         | Calm      |       |

low

negative

positive

VALENCE

CANCEL

SUBMIT

**Figure 1:** Valence-Arousal Questionnaire

The grid prompt is triggered after five minutes of continuous phone usage, based on screen-on and screen-off events. If the notification was not answered within five minutes, it disappeared. After each answered prompt, a timeout of two hours is applied to avoid overloading participants.

Alongside the ESM, the app also includes a screentext sensor. This sensor uses Android’s accessibility service to access the view hierarchy of the operating system and extract the textual content displayed on the screen, without actually taking screenshots. To avoid collecting potentially highly-sensitive data, communication and banking apps are excluded from screentext sensing by default, with participants being able to add any other apps to the blacklist. Only the five minute window of screentext before an answered ESM prompt is saved into the database, while all other screentext is discarded.

### 3.2. User study

We conducted an AV-Sense field study with 12 participants, mostly university students between 21 and 30 years old. The study lasted for two weeks, during which participants had the app installed and running on their personal phones. The ESM prompts appeared several times per day, asking the users to report their current emotional state on the valence–arousal grid. Over the course of the study, we collected 787 ESM responses together with around 650,000 screentext entries. This provided us with a dataset where every self-reported emotional state could be linked to the text that was visible on the phone at the immediately preceding time.

The collected data makes it possible to explore both how users’ valence and arousal changes over time, as well as whether the screentext could serve as a signal for predicting the valence/arousal change.

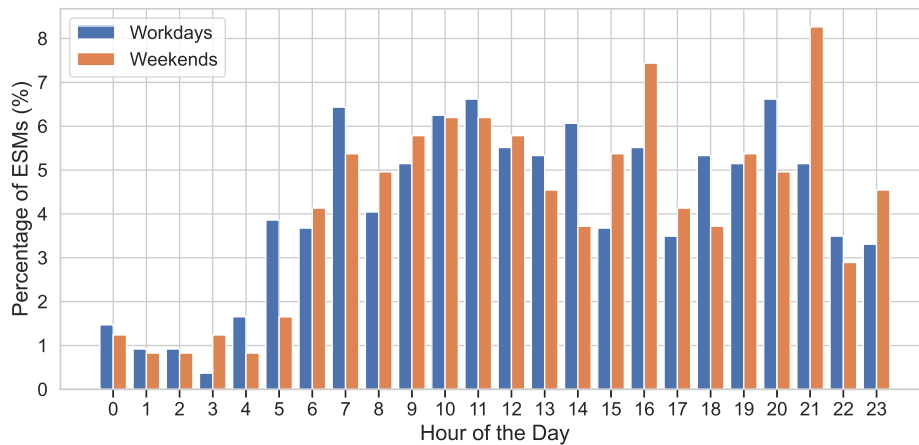
## 4. Data overview

The ESM data are stored in form of individual records, each containing the notification time, the response time, and the selected valence and arousal values. The screentext data are stored separately and linked to ESM records by timestamps. For every answered prompt we only kept screentext entries from the five minutes before the response. This way each self-report can be directly connected to the text that was visible on the screen in the time leading up to it.

In our study, the number of answers varied among participants. Because of this, all distributions and averages in our analysis are calculated as weighted values, so that each participant contributed proportionally to their number of responses.

### 4.1. Temporal distribution

In Figure 2 we plot the temporal distribution of responses to ESM questionnaires. We observe a clear daily pattern already reported in similar previous studies [12]. Participants responded less during the night and early morning, with activity increasing throughout the day. We also compared workdays and weekends, with the distributions differing slightly between the two, but in both cases the majority of responses come during the active parts of the day.



**Figure 2:** Average valence and arousal during workdays and weekends.

We also analyzed the response times to ESM prompts using a cumulative distribution function. The results showed that participants usually answered quickly. The median response time was about 13 seconds, and more than 90% of all responses were submitted within the first minute after the notification appeared. This means that most self-reports can be considered reliable in-the-moment reflections of the user’s current state, since long delays that could introduce recall bias were rare [13]. The relatively fast response times also suggest that participants engaged with the study consistently and did not find the notifications overly disruptive.

### 4.2. Descriptive statistics of valence and arousal

Table 1 summarizes the mean and standard deviation of valence and arousal for each participant. Most valence averages are close to zero, with a slight shift toward positive values, while a few participants report mildly negative valence on the average. This suggests that neutral and slightly positive moods are most common, with extreme states appearing less often, which is typical for ESM data [6].

Arousal means were often slightly below zero, placing many responses in calmer or moderately active states. This is consistent with our notification scheme, which triggered prompts after five minutes of continuous phone use, usually during routine interaction rather than moments of high activation.

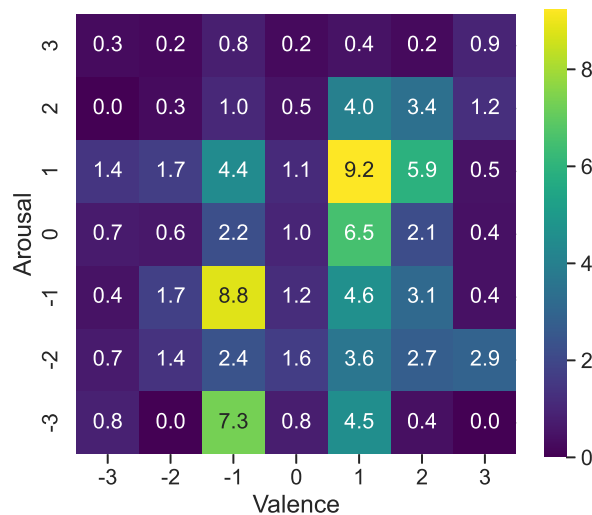
Standard deviations show clear individual differences, with some participants displaying more variability in arousal than valence. This aligns with the idea that arousal tends to fluctuate more with context, while valence often remains closer to neutral. Although group-level summaries are informative, they also hide strong individual differences, underlining the value of examining both perspectives.

**Table 1**  
Descriptive statistics of valence and arousal by participant

| Participant | Valence $\mu$ | Valence $\sigma$ | Arousal $\mu$ | Arousal $\sigma$ |
|-------------|---------------|------------------|---------------|------------------|
| 1           | 0.35          | 1.45             | 0.02          | 2.16             |
| 2           | 0.43          | 1.50             | -0.97         | 1.73             |
| 3           | 1.35          | 1.14             | -1.32         | 1.72             |
| 4           | 0.47          | 1.92             | -0.13         | 1.32             |
| 5           | 0.61          | 1.06             | 0.11          | 1.08             |
| 6*          | 0.12          | 0.71             | 0.00          | 0.71             |
| 7           | 0.72          | 1.88             | -0.80         | 1.85             |
| 8           | -0.08         | 1.82             | -0.12         | 1.66             |
| 9           | 0.02          | 1.13             | -0.74         | 1.67             |
| 10          | -0.20         | 1.34             | 0.16          | 1.34             |
| 11          | -0.07         | 1.45             | -0.14         | 1.12             |
| 12          | 0.35          | 1.36             | -0.24         | 1.89             |

\* Participant 6 reported near-constant values (zeros on the grid); retained here for completeness, but excluded from the heatmap in Fig. 3.

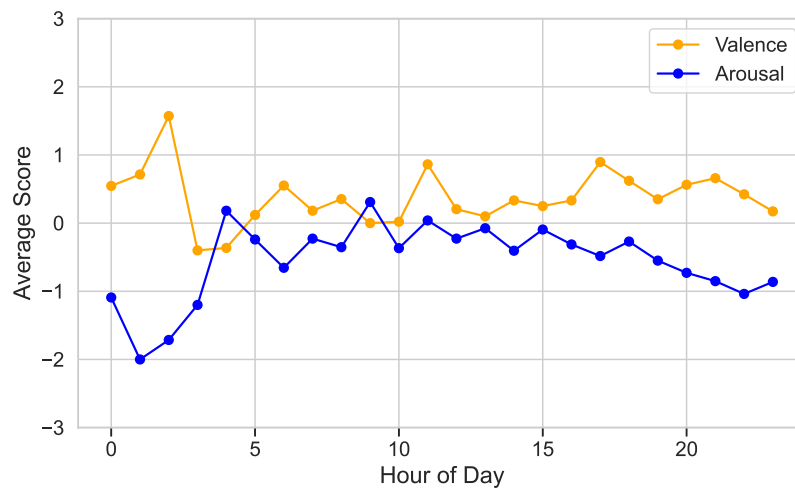
We also look at the overall distribution of valence and arousal answers across participants. One participant (User 6) is removed from the analysis because almost all of their answers are zeros, possibly indicating quick, meaningless answering. Figure 3 shows a heatmap of the normalized distribution for the remaining participants. The majority of answers are located around neutral to slightly positive valence and low to moderate arousal. Extremes on the edges of the grid are rare, which fits with the idea that everyday ESM data tends to capture mostly neutral and mild states [6]. There is also a weak positive relationship between valence and arousal, with higher valence sometimes paired with higher arousal, and lower valence with lower arousal, although the spread is wide.



**Figure 3:** Heatmap of the normalized distribution of ESM answers (without User 6).

### 4.3. Daily pattern of valence and arousal

In our data we do not see a clear daily pattern for valence. As shown in Figure 4, the hourly averages move without a strong trend across the day, which differs from reports that often find higher positive valence in the morning and a decline toward evening [14]. Arousal shows a more typical daily rhythm. After waking up, subjective energy rises and reaches a peak in the first half of the day, followed by a noticeable afternoon dip and lower levels in the evening and at night [14, 15]. Hourly averages are computed with the weights described above so that participants contribute in proportion to their number of responses.



**Figure 4:** Hourly averages of valence and arousal across the study.

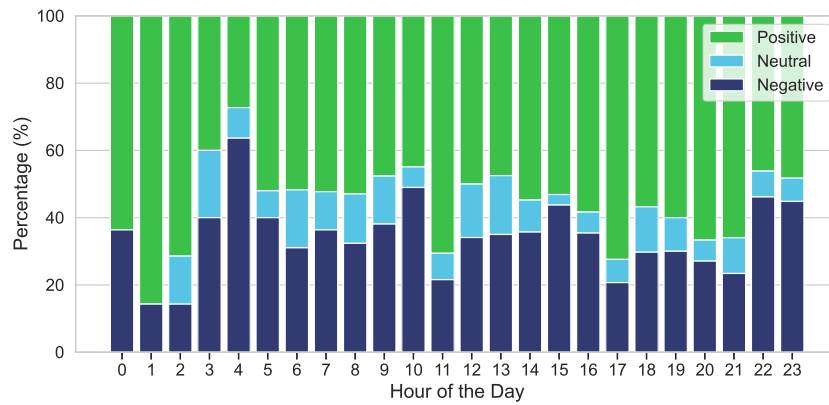
Average values alone can sometimes hide the variety of affective experiences during the day. To capture this diversity we grouped the answers into categories of valence (negative, neutral, positive) and arousal (low, neutral, high). Figures 5 and 6 show how the shares of these categories change across hours of the day. This view reveals aspects that averages miss. For example, a few very positive ratings can push the average upward even if there are many slightly negative ones, creating a misleading impression.

In our data, average valence stays slightly positive through most of the day, but the categorical view shows that negative moods became more common in the evening. For arousal we saw a similar pattern. Although the average drops in the afternoon, the distributions make it clear that this is driven by a higher share of low-energy states such as tiredness or calmness, together with a decrease in high-energy states. These categorical distributions therefore complement the averages and provide a more detailed look at daily affective dynamics.

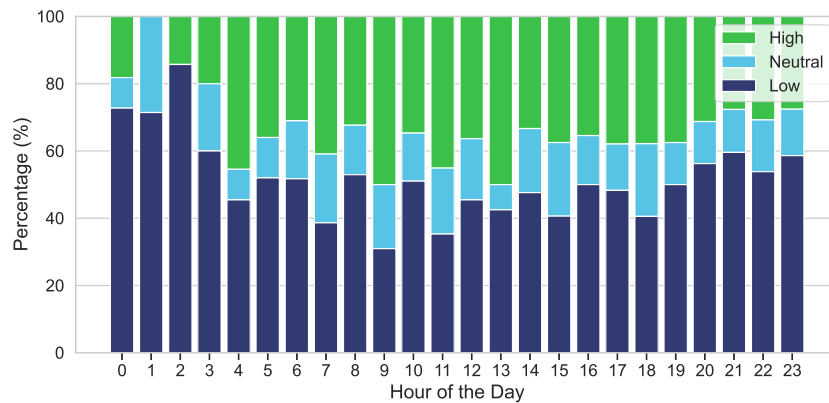
### 4.4. Availability of screentext data

Figure 7 shows, for each participant, the share of ESM answers where screentext data are available in the five minutes before the response. For most participants, more than half of their ESM answers are linked with screentext, and in many cases the coverage is above 90 %. A few participants have very low availability, with one showing no screentext data at all and another only about 22 %. These cases were most likely caused by missing permissions or technical issues that prevented the sensor from working, or by frequent use of apps that were blocked from screentext capture. Such answers were excluded from further analyses, since models in the next section require at least some text context. For the remaining participants, the high availability rates mean that ESM answers could usually be interpreted together with surrounding screentext, giving us a rich basis for later analysis.

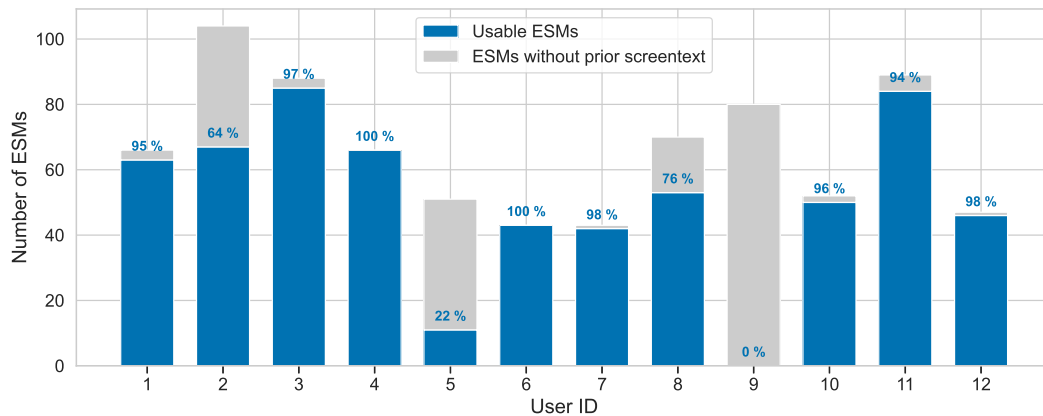
To better understand the context of the collected screentext, we also looked at the categories and



**Figure 5:** Share of answers by valence categories (negative, neutral, positive) across hours of the day.



**Figure 6:** Share of answers by arousal categories (low, neutral, high) across hours of the day.



**Figure 7:** Share of ESM answers per participant with screentext available in the five minutes before the response.

specific apps that appeared most often before ESM answers. Social, messaging, and video apps were by far the most common, while other categories like browsing, reading, games, or productivity were much less frequent. At the app level, Instagram, YouTube, and Reddit were the top sources of screentext, followed by Discord, Chrome, and TikTok. This confirms that most of the captured screentext came from social and media-related contexts, which is important for interpreting the results of our predictive models.

## 5. LLM inference

After collecting ESM responses and screentext, we explored whether LLMs could predict valence and arousal from the text captured on participants’ devices. We tested two prompting strategies and compared them to a simple baseline. For each participant, we divided their examples into a training group and a test set (80/20 split). The first approach used few-shot prompting, where a subset of training examples (we included  $k = 3\text{--}5$  per participant) was inserted into the prompt together with their associated arousal–valence scores. The model was then asked to predict the score for a new example from the test set. The second approach used rubric-based prompting, where instead of concrete examples, the model was given a set of short written instructions. These instructions described how to assign arousal and valence values based on emotional tone and intensity (e.g., positive language as higher valence, negative language as lower valence, urgent text as higher arousal, calm text as lower arousal). The baseline model simply predicted the average score of the training group for each participant. We evaluated all approaches using Mean Absolute Error (MAE).

We tested these strategies with GPT-5-nano accessed via API.

**Table 2**

MAE of predicted valence and arousal per each participant with different prediction methods. Highlighted is the best performing approach for each user and separately for valence and arousal.

| Participant | Valence (MAE)  |                   |                 | Arousal (MAE)  |                   |                 |
|-------------|----------------|-------------------|-----------------|----------------|-------------------|-----------------|
|             | Basic approach | Few-shot approach | Rubric approach | Basic approach | Few-shot approach | Rubric approach |
| 1           | 1.385          | 1.692             | 1.154           | 1.154          | 1.615             | 1.385           |
| 2           | 0.692          | 1.538             | 1.538           | 1.615          | 1.846             | 1.846           |
| 3           | 1.000          | 1.235             | 1.176           | 1.235          | 1.529             | 1.529           |
| 4           | 1.154          | 1.231             | 1.308           | 1.692          | 1.846             | 1.923           |
| 6           | 0.375          | 0.500             | 0.500           | 0.250          | 0.375             | 0.375           |
| 7           | 1.500          | 1.625             | 1.250           | 1.000          | 1.250             | 1.125           |
| 8           | 2.000          | 1.818             | 1.455           | 1.455          | 1.545             | 1.455           |
| 10          | 1.100          | 1.100             | 1.200           | 1.400          | 1.500             | 1.000           |
| 11          | 1.188          | 1.438             | 1.375           | 1.813          | 1.875             | 1.938           |
| 12          | 0.556          | 1.333             | 1.222           | 0.889          | 0.778             | 0.889           |

Table 2 summarizes the results. Prediction performance varied substantially among participants and prompting strategies. For some participants, the rubric-based prompt gave the lowest MAE, while for others, the few-shot prompt worked slightly better. However, in most cases, the baseline model outperformed both prompting strategies. This demonstrates that predicting valence and arousal from screentext is highly challenging and that simple prompting setups with LLMs do not yield consistent improvements over a trivial average-based predictor.

Participant-level variability also shaped the results. Participant 6 had uniformly low errors for all approaches because their ESM responses were highly repetitive and clustered around neutral values. This made predictions easier but at the same time less informative. In contrast, participants with more varied responses introduced more complexity for the models, and errors rose accordingly. A broader limitation is that the models had difficulty adapting to different personal styles of smartphone use, from routine messaging to more expressive content. Personalization may therefore be important for making these models useful in practice.

### 5.1. Correlation of sentiment with valence and arousal

The premise behind the analysis above was that the text a user sees on the screen reflects the user’s valence and arousal. This link was further reinforced in the few-shot prompting setup, where texts were explicitly paired with the reported valence and arousal. Nevertheless, the LLM may sometimes

infer only the sentiment of the text itself rather than the user’s emotional state. In such cases, it is useful to know whether the sentiment of the text, on average, correlates with the expressed sentiment of the user.

We therefore conducted an additional experiment in which we classified screentext into coarse sentiment categories of polarity (negative, neutral, positive) and intensity (none, low, medium, high). These categories were then mapped to numerical values to enable comparison with the self-reported ESM scores. Polarity was mapped to  $-1, 0, 1$  and intensity to the interval  $[0, 1]$ . We then calculated Pearson and Spearman correlations with the valence and arousal dimensions.

The results showed a weak but statistically significant correlation between sentiment polarity and valence ( $n = 591$ , Pearson  $r = 0.143$ ,  $p = 0.00048$ ; Spearman  $\rho = 0.132$ ,  $p = 0.0013$ ). This indicates that more positive sentiment in the screentext slightly increased the likelihood of a higher reported valence, although the effect was small. For intensity and arousal, however, no meaningful relationship was found ( $n = 591$ , Pearson  $r = 0.004$ ,  $p = 0.919$ ; Spearman  $\rho = 0.002$ ,  $p = 0.959$ ).

These findings highlight a key problem. If the models rely heavily on sentiment cues in the text, they only recognize whether text is broadly positive or negative. This overlaps slightly with reported valence, but it does not capture differences in arousal. As a result, calm and excited states with the same polarity may be confused by the model. This gap reflects a well-known limitation of sentiment analysis approaches, which emphasize polarity while overlooking the activation dimension. In naturalistic screentext, this limitation becomes even more pronounced because text fragments are short, noisy, and lack broader context.

## 6. Discussion and Conclusions

Our results showed that combining ESM with screentext collection was feasible in everyday settings. Most ESM answers were linked with screentext, which provided useful context for analysis. The data was often noisy and sometimes missing due to blocked apps or technical limits, highlighting the trade-off between privacy and coverage in real-world deployments.

In terms of affect patterns, arousal showed a clearer daily rhythm than valence. Category distributions revealed changes that averages alone would have hidden, such as more negative moods in the evening and more low-energy states in the afternoon. These patterns are consistent with known affect dynamics but were also shaped by the way prompts were triggered.

LLM inference proved more challenging. Performance varied across participants and prompting strategies, with neither few-shot nor rubric-based prompting showing a consistent advantage over the baseline. This suggests that models were sensitive to individual differences and that simple prompting was not sufficient to capture subtle affective variation. The sentiment-based experiment confirmed this limitation. The weak correlation with valence and almost no correlation with arousal showed that models relied heavily on sentiment cues, which explains why predictions often missed the difference between calm and excited states of the same polarity. The gap between sentiment detection and full valence–arousal inference therefore remains large, especially in naturalistic data.

The study has several limitations. The sample was small, with only twelve participants, and the duration was short, which limits generalizability. Participants were also relatively homogeneous in age and background. The data itself was restricted to screentext, without additional modalities such as audio, images, or sensor data that might capture arousal more effectively. Screentext was often noisy, containing UI elements and symbols rather than only meaningful content. Privacy rules also reduced coverage, since communication and banking apps were blocked by default and participants could blacklist additional apps. Finally, the prompting scheme, which triggered questions after five minutes of continuous screen use followed by a two-hour timeout, shaped the contexts in which responses were collected and likely emphasized routine phone use.

Even with these limits, the main takeaway is clear. ESM plus screentext is feasible on personal phones and yields a dataset that links self-reports with the text people actually see. Text alone is not enough for robust affect inference, especially for arousal. Looking forward, there is potential to combine screentext

with other lightweight signals, such as app usage context, time of day, or sensor data, to capture a fuller picture of affect. With larger and more diverse samples, and with careful attention to privacy, these approaches could support the development of more reliable affect-aware mobile applications. Possible applications include digital well-being tools that adapt notifications based on the user's state, chat or messaging apps that detect emotionally charged interactions, or learning and productivity apps that adjust difficulty and pacing depending on energy and mood. They could also support psychological research by providing ecologically valid affect data without requiring intrusive sensing.

## References

- [1] E. A. Kensinger, D. L. Schacter, Processing emotional pictures and words: Effects of valence and arousal, *Cognitive, Affective, & Behavioral Neuroscience* 6 (2006) 110–126.
- [2] G. Miller, The smartphone psychology manifesto, *Perspectives on Psychological Science* 7 (2012) 221–237.
- [3] Y. R. Tausczik, J. W. Pennebaker, The psychological meaning of words: Liwc and computerized text analysis methods, *Journal of language and social psychology* 29 (2010) 24–54.
- [4] S. Teng, T. Zhang, S. D'Alfonso, V. Kostakos, Predicting affective states from screen text sentiment, in: *Companion of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Melbourne, Australia, 2024.
- [5] N. van Berkel, S. D'Alfonso, R. K. Susanto, J. Goncalves, V. Kostakos, Aware-light: a smartphone tool for experience sampling and digital phenotyping, *Personal and Ubiquitous Computing* 27 (2023) 435–445.
- [6] S. Shiffman, A. A. Stone, M. R. Hufford, Ecological momentary assessment, *Annual Review of Clinical Psychology* 4 (2008) 1–32.
- [7] R. LiKamWa, Y. Liu, N. D. Lane, L. Zhong, Moodscope: Building a mood sensor from smartphone usage patterns, in: *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, 2013.
- [8] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, A. Aucinas, Emotion-sense: a mobile phones based adaptive platform for experimental social psychology research, in: *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010.
- [9] S. M. Mohammad, Sentiment analysis: Detecting valence, emotions, and other affectual states from text, in: *Emotion measurement*, Elsevier, 2016.
- [10] S. M. Mohammad, P. D. Turney, Crowdsourcing a word–emotion association lexicon, *Computational intelligence* 29 (2013) 436–465.
- [11] S. Teng, S. D'Alfonso, V. Kostakos, A tool for capturing smartphone screen text, in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, 2024.
- [12] V. Pejovic, M. Musolesi, Interruptme: Designing intelligent prompting mechanisms for pervasive applications, in: *ACM UbiComp*, Seattle, WA, USA, 2014.
- [13] A. A. Stone, S. Schneider, J. M. Smyth, Evaluation of pressing issues in ecological momentary assessment, *Annual Review of Clinical Psychology* 19 (2023) 107–131.
- [14] F. Bu, J. K. Bone, D. Fancourt, Will things feel better in the morning? a time-of-day analysis of mental health and wellbeing from nearly 1 million observations, *BMJ Ment Health* 28 (2025).
- [15] A. A. Stone, J. M. Smyth, T. Pickering, J. Schwartz, Daily mood variability: Form of diurnal patterns and determinants of diurnal patterns, *Journal of Applied Social Psychology* 26 (1996) 1286–1305.

# Exploring the Effects of Multimodal User Interfaces in Autonomous Vehicles

Kristina Stojmenova Pečičnik<sup>1\*</sup>, Timotej Gruden<sup>1</sup>, Grega Jakus<sup>1</sup>, Sašo Tomažič<sup>1</sup>, and Jaka Sodnik<sup>1</sup>

<sup>1</sup> University of Ljubljana, Faculty of Electrical Engineering, Tržaška cesta 25, Ljubljana, 1000, Slovenia

## Abstract

This short paper presents a study design created to explore how multimodal user interfaces (UIs) influence comfort, user experience, and overall well-being in fully autonomous vehicles. The study uses a motion-based driving simulator equipped with real vehicle components to provide realistic driving conditions. Participants are asked to complete two trials: a baseline trial and one with either a simple or extended multimodal UI combining auditory, tactile, and speech-based feedback. Driving scenarios include lane changes, emergency braking, V2X and V2V communication, and road surface variations. Data collection integrates physiological measures (EGG, EDA, PPG), self-assessment questionnaires on motion sickness (MSAQ) and user experience (UEQ), and performance on a reading comprehension task simulating non-driving activities. The paper provides some initial data about the participants' demographics and their user experience with the UI, and gives an insight into the planned next steps in terms of analysis and results application. The findings from this study will provide an insight into the potential of multimodal HMI design to enhance safety, comfort, and user acceptance in autonomous vehicles.

## Keywords

User interface, autonomous vehicle, multimodal interaction, non-driving related task

## 1. Introduction

This short paper presents a user study design focusing on investigating how the integration of a multimodal user interface (UI) in a fully automated vehicle (i. e. autonomous vehicle (AV)) affects the AV user's comfort, user experience and overall well-being when using these technologies. By focusing on human-vehicle interaction, the study explores the role of and potential of human-machine interaction (HMI) in enhancing user satisfaction and mitigation of potential discomfort and motion sickness. Additionally, it examines how drivers engage in non-driving related tasks when using autonomous vehicles and how these activities influence their overall well-being and interaction with the vehicle.

This study is conducted as part of Ljubljana Pilot, one of four Pilots in the HorizonEurope project Federated cyber-physical infrastructure for operational design domain continuity (FRODDO) [1].

## 2. Methodology

The study is conducted in a motion-based driving simulator, with the visuals presented on a triple 49-inch screen configuration. Although an AV is used for the study, the simulator is equipped with real car parts including a steering wheel, pedals, gear box and a dashboard, to increase the immersiveness and realistic representation of a driving environment.

---

*Human-Computer Interaction Slovenia 2025, October 13, 2025, Koper, Slovenia*

\*Corresponding author.

✉ kristina.stojmenova@fe.uni-lj.si (K. S. Pečičnik); timotej.gruden@fe.uni-lj.si (T. Gruden); grega.jakus@fe.uni-lj.si (G. Jakus); saso.tomazic@fe.uni-lj.si (S. Tomažič); jaka.sodnik@fe.uni-lj.si (J. Sodnik)

ORCID 0000-0001-6584-7147 (K. S. Pečičnik); 0000-0002-8099-9832 (T. Gruden); 0000-0001-9373-7885 (G. Jakus); 0000-0002-2968-8879 (S. Tomažič); 0000-0002-8915-9493 (J. Sodnik)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<https://doi.org/10.26493/978-961-293-559-7.17>

## 2.1. Driving scenarios

Two versions of one driving scenario (please see **Figure 1**) were developed for the study, differing in the starting and ending points, and consequently route used by the AV during the trial. This enables very similar driving conditions that allow for comparison of results, while making them different enough to avoid anticipation effects with subsequent trials for the same user. The scenario included driving on a highway, city, and rural roads and lasted for about 13 minutes.



**Figure 1:** Driving scenario map used in the study.

The driving scenarios were developed in SCANeR Studio [2]. To explore the effects of the UI on the AV user's comfort, five different types of driving situations were developed:

- 1) **Lateral movement** (e.g., lane change, turning at intersection),
- 2) **Longitudinal movement** (e.g., emergency braking, front vehicle stop, see **Figure 2**),
- 1) **Preventive aggressive maneuvering of the vehicle to prevent an accident caused by other traffic participants** (e.g., street cams detecting dangerous situation),
- 2) **Unexpected acceleration and avoiding the obstacle** (e.g., next vehicle will move out of the way, so we can speed up), and
- 3) **Road deformation** (e.g., street bumper, don't spill your coffee).

Each trial featured 12 driving situations.

## 2.2. User interface

To communicate the vehicle's actions in each of the five groups of driving situations, a multimodal user interface was developed, which featured:

- simple auditory sounds (beeps, chimes, and turning signal sounds),
- synthesized speech (description of the vehicle's action and/or the event in the environment that provoked the vehicle's action), and
- tactile information (in the form of vibrations, placed in the driving seat).

Different combinations of modalities were then used for different driving situations, based on the nature of the event and its potential impact on the driver's safety. To further access the effect of abstract from representational messages, two versions of the multimodal interface were explored:

- Simple UI: featuring only simple auditory sounds, and for critical events also tactile information
- Extended UI: all of the elements of the simple UI with the addition of synthesized speech.



**Figure 2:** An example of a situation where drivers could benefit from appropriate UI design: Emergency braking due to congestion on a highway road.

### 2.3. Study design

The study design was approved by the University of Ljubljana's Ethical Committee, no. 049-2025. It was developed following guidelines as seen in Research Methods in Human-Computer Interaction [3].

This study has a mixed design. The effects of using an UI in AVs are observed within-subject. For this, each participant completes two trials: a baseline trial without any UI, and one trial with one of the UIs. The effects of the format of the messages (abstract beeps vs. representational speech) are observed between-subjects, comparing only the results from trials each participant performs with one of the UIs. The trial order and version of driving scenario used was randomized to avoid confound effects (such as increased comfort due to increased familiarity with the equipment in the second trial), that would emerge if the same order was used. This was made possible by using 2 versions of the driving scenario and three versions of the UI assistance (baseline or no UI, simple UI and extended UI). Table X presents the possible 8 combinations designed following the rules of a Latin square. This was then repeated until the predefined number of participants was reached.

**Table 1**

Trial order randomization using Latin square design.

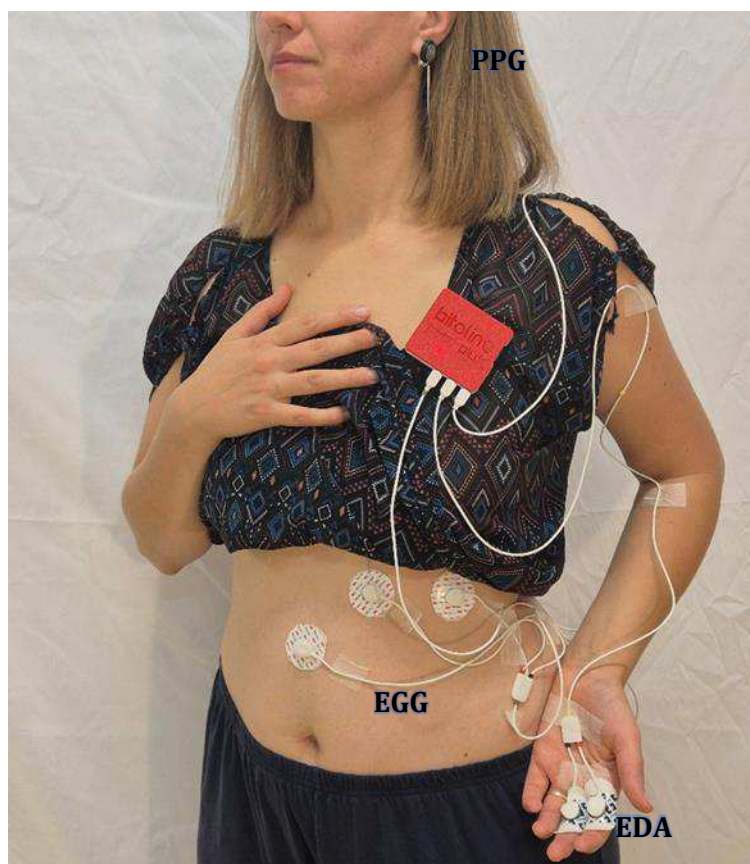
| Trial order combination | Driving scenario - version 1 | Driving scenario - version 2 |
|-------------------------|------------------------------|------------------------------|
| 1                       | 1B                           | 2S                           |
| 2                       | 1B                           | 2E                           |
| 3                       | 2S                           | 1B                           |
| 4                       | 2E                           | 1B                           |
| 5                       | 2B                           | 1S                           |
| 6                       | 2B                           | 1E                           |
| 7                       | 1S                           | 2B                           |
| 8                       | 1E                           | 2B                           |

\*B = Baseline trial; S = Trial with Simple UI; E = Trial with Extended UI.

Prior to the start of the study, each participant is provided with a written study description, explaining the goals of the study and tasks of the participant. This is followed by an informed consent, where the participants are further informed about the potential risks and benefits of participation in the study. The informed consent was prepared based on the Informed template for conducting studies with human participants at University of Ljubljana. Each participant is provided with a 10€ gift card, as a compensation for their time and effort for participating in the study.

The effects are observed through three groups of dependent variables: physiological data, self-assessment data and performance data of a non-driving related task the participants are asked to complete during each trial.

Physiological data is collected with Bitalino [4] with 100 Hz sampling frequency, using an electrogastrogram [EGG] sensor, an electrodermal activity [EDA] sensor, and a photoplethysmogram [PPG] sensor as shown in **Figure 3**. In addition, gaze and other pupilometry variables are collected with Tobii Pro Glasses 3 wearable eye tracker with 100 Hz sampling frequency [5].



**Figure 3:** Sensors used for collections of physiological data.

The self-assessment data focuses on gathering information about motion sickness and user experience with the UIs. Motion sickness is observed using the Motion Sickness Questionnaire (MSQ) [6]. With this questionnaire, the participant is asked to rate their current physical status prior the start of the study, and after each trial. The results are then compared to explore any effects of the UI on motion sickness. The user experience is collected with the User Experience Questionnaire [7]. Each participant completes the UEQ immediately after completing the trial with the UI. The questionnaire consists of 26 questions, which can be answered using a 7-point Likert scale (1 – Completely agree, 7 – Completely disagree). The answers are used to provide scores on six aspects of user experience – Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty.

Lastly, participants are asked to complete a non-driving task (NDRT) during each trial. The introduction of a NDRT is simulation of how AV users are expected to spend their time when using

an AV, i.e. engage in tasks such as reading, writing or interacting with electronic devices. In this study, a reading comprehension task is introduced, which requires the participant to read a text and then complete a short quiz asking them about information from the text. The reading comprehension task is an adaptation from an exam aimed at 8th grade pupils in Slovenia. The reasoning behind using this task is two-fold: a) it is engaging and cognitively demanding enough to require the participants to focus on the task and not on driving, and b) it is relatively simple enough for all participants to complete it, so that it does not discourage or evoke any unnecessary performance pressure. Two versions were used, to provide different texts in each of the trials. Lastly, the study collected information about the participant's demographics, such as age, gender, driving and simulation technology experience.

### 3. Participants demographics and experience

24 participants (10 male and 14 female) aged between 21 and 64 ( $M = 40.21$ ,  $SD = 13.71$ ) took part in the study until September 30<sup>th</sup>. All of the participants had a valid driving license ( $M = 20.79$ ,  $SD = 13.45$ ), with 66.67% of them reporting to drive daily, 20.83% few times a week, 8.33% few times a month, and 4.17% only few times a year.

Participants were further asked about their past experiences with driving simulation, playing mobile and video games, and use of advanced driving assistive systems (ADAS), to better understand their levels of familiarity with the topic and lab environment used for conducting the study. The results are presented in **Table 2**.

**Table 2**

Participant's experience with driving simulation, video and mobile games and use of ADAS.

| Criteria                 | Never   | Once                | Few times        | Multiple times |
|--------------------------|---------|---------------------|------------------|----------------|
| Used a driving simulator | 29.17 % | 25 %                | 33.33%           | 12.5 %         |
|                          | Never   | Once a week or less | Few times a week | Every day      |
| Plays games              | 54.16 % | 29.17 %             | 12.5%            | 4.17 %         |
|                          | Never   | Once                | Few times        | Multiple times |
| Uses ADAS                | 37.5 %  | 8.33 %              | 12.5 %           | 41.67 %        |

### 4. Next steps

The intermediate next step is completing the data collection. A total of 30 participants is foreseen to take part in this study. This will be followed by data analysis of all three groups of dependent variables.

The physiological EGG, EDA and PPG signals will be used to assess the effect of the UI on the participants' comfort, stress, arousal and overall wellbeing, by comparing the results from the baseline trial with the trials with the UI. A mixed model will be applied, to further explore whether the information modality and message format (abstract vs. representational) has any effect on these physiological states. The gaze and pupilometry data will be used to explore the effects of the UI on trust, with the gaze data considered as a direct indicator of trust. The results from the NDRT performance will be used as an indirect indicator about the participant's trust in the AV and the impact of the UI on it, with positive correlation indicating the participants trust and vice versa, negative correlations low trust and urge for monitoring the driving rather than focusing on the NDRT.

The self-reported motion sickness will be also used to assess the effects of the absence/presence of an UI – and the version of it - on the participant's motion sickness. The self-reported user experience will be used to obtain an understanding of the impact of the UI on the participants' experience, and by using the benchmark values featured in the UEQ Data Analysis

Tool (UEQ, Germany), obtain an insight on how the evaluated UI compares to other products (business software, web pages, web shops, social networks) featured in the benchmark.

Lastly, the demographic data will be used to investigate if there are any age, gender or experience correlations with the user experience, comfort and trust when using an AV and any potential impacts the presence/absence of an UI has on them.

After completing the analysis, we hope the findings from this study will provide an insight into the potential of multimodal HMI design to enhance safety, comfort, and user acceptance in autonomous vehicles.

## Acknowledgements

The authors would like to thank all of the participants for their time and effort in participating in the study. The authors would like to especially thank Simon Pestotnik for conducting the study and developing the simulation scenarios.

The work presented in this paper was financially supported by the Slovenian Research and Innovation Agency within the program ICT4QL, grant no. P2-0246 and by the European Union's HorizonEurope research and innovation program for the project FRODDO, grant agreement no. 101147819.

## References

- [1] ERTICO. Federated cyber-physical infrastructure for ODD continuity (FRODDO). Available at: <https://froddo-project.eu/web/>
- [2] AVSimulation. SCANer studio. Available at: <https://www.avsimulation.com/scanerstudio/>.
- [3] Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.
- [4] PLUX Biosignals. Bitalino. Available at: <https://www.pluxbiosignals.com/collections/bitalino>
- [5] Tobii Pro Glasses 3 wearable eye tracker. Availabe at: <https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3>
- [6] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3), 203-220.
- [7] Laugwitz, B., Held, T., & Schrepp, M. (2008, November). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group (pp. 63-76). Springer, Berlin, Heidelberg.
- [8] UEQ online. Data Analysis Tools. Available at <https://www.ueq-online.org/>

## Seznam avtorjev/List of Authors

Kristina Stojmenova Pečečnik: University of Ljubljana, FE, Slovenia  
Grega Jakus: University of Ljubljana, FE, Slovenia  
Matevž Pesek: University of Ljubljana, FRI, Slovenia  
Helena Jeretina: University of Ljubljana, FRI, Slovenia  
Emir Hodžić: University of Ljubljana, FRI, Slovenia  
Ciril Bohak: University of Ljubljana, FRI, Slovenia  
Jurij Anžič: University of Ljubljana, FRI, Slovenia  
Ilija Gavrilović: University of Ljubljana, FRI, Slovenia  
Jaka Kužner: University of Ljubljana, FRI, Slovenia  
Matija Marolt: University of Ljubljana, FRI, Slovenia  
Gašper Leskovec: University of Ljubljana, FE, Slovenia  
Eva Gaberšček: University of Ljubljana, FE, Slovenia  
Jaka Sodnik: University of Ljubljana, FE, Slovenia  
Janez Koprivec: University of Ljubljana, FRI, Slovenia  
Lea Pajnič: University of Primorska, FAMNIT, Slovenia  
Klen Čopič Pucihar: University of Primorska, FAMNIT, Slovenia  
Matjaž Kljun: University of Primorska, FAMNIT, Slovenia  
Maheshya Weerasinghe: University of Primorska, FAMNIT, Slovenia  
Ana Nikolić: University of Primorska, FAMNIT, Slovenia  
Marko Tkalcic: University of Primorska, FAMNIT, Slovenia  
Uroš Sergaš: University of Primorska, FAMNIT, Slovenia  
Simon Kolmanič: University of Maribor, FERI, Slovenia  
Jan Hrašar: University of Maribor, FERI, Slovenia  
Štefan Horvat: University of Maribor, FERI, Slovenia  
Domen Mongus: University of Maribor, FERI, Slovenia  
Aleksandar Ilievski: University of Maribor, FERI, Slovenia  
Suzana Žilič Fišer: University of Maribor, FERI, Slovenia  
Uroš Šmajdek: University of Ljubljana, FRI, Slovenia  
Bojan Blažica: Jožef Stefan Institute, Slovenia  
Manca Topole: Jožef Stefan Institute, Slovenia  
Marko Debeljak: Jožef Stefan Institute, Slovenia  
Edita Džubur: University of Ljubljana, FRI, Slovenia  
Veljko Pejović: University of Ljubljana, FRI, Slovenia  
Timotej Gruden: University of Ljubljana, FE, Slovenia  
Sašo Tomažič: University of Ljubljana, FE, Slovenia

