The International Series on Information Systems and Management in Creative eMedia is advancing the knowledge of the use of informaiton systems and management in the wider field of creative eMedia industries. The series covers a wide range of media , such as television, publishing, digital games, radio, ubiquitous/ambient media, advertising, social media, motion pictures, online video, eHealth, eLearning, and other eMedia industries.

Printed and published by the International Ambient
Media Association (iAMEA)
Ihanakatu 7-9/A1
FIN-33100 Tampere
FINLAND
www.ambientmediaassociation.org
www.ambientmediaassociation.org/Journal

ISBN 978-952-7023-17-4 (PDF)

ISSN 2341-5584 (Paperback)
ISSN 2341-5576 (PDF)
ISSN 2341-6165 (CD-ROM)

International Series on Information Systems and Management in Creative eMedia

Nr. 2017/2

INTERNATIONAL SERIES ON INFORMATION SYSTEMS AND MANAGEMENT IN CREATIVE eMEDIA (CreMedia)

NUMBER 2017/2

CreMedia

**Artur Lugmayr, Kening Zhu, Xiaojuan Ma (eds.)**

# Proc. of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017)

## ARTIFICIAL INTELLIGENCE MEETS VIRTUAL AND AUGMENTED WORLDS

**in conjunction with SIGGRAPH ASIA**

**Bangkok, Thailand, Nov. (2017)**

**Artificial Intelligence meets VR and AR**


IS&M in Creative eMedia

**International Series on Information Systems and Management in Creative eMedia (CreMedia)**

The International Series on Information Systems and Management in Creative eMedia (CreMedia) is advancing the knowledge of the use of information systems and management in the wider field of creative eMedia industries. The series covers a wide range of media, such as television, publishing, digital games, radio, ubiquitous/ambient media, advertising, social media, motion pictures, online video, eHealth, eLearning, and other eMedia industries. More information about this series at: http://www.ambientmediaassociation.org/Journal.

CONTRIBUTING DOMAINS

- computer science, in particular entertainment computation, media technology and multimedia;
- human-computer-interaction, and user-experience;
- ubiquitous, pervasive, ambient, and semantic intelligent technologies;
- media management, business, economics, information systems research in media industries;
- media art, content production, content systems, and services;
- tools, software/hardware architectures, and their solutions;
- methods, algorithms, and paradigms.

FOUNDERS AND MANAGING DIRECTORS

- A/Prof. Dr. Artur Lugmayr, AUSTRALIA
- Bjoern Stockleben, HFF, GERMANY
- Emilija Stojmenova, Univ. of Ljubljana, SLOVENIA
- Estefania Serral Asensio, KU-Leuven, BELGIUM

PAST SERIES TITLES

- Managing and Leading Creative Universities – Foundations of Successful Science Management, A Hands-On Guide for (Future) Academics 2017/1 (2017)
- Proceedings of the 9th Workshop on Semantic Ambient Media Experiences (SAME 2016b): Visualisation, Emerging Media, UX, No 2016/2 (2016)
- Proceedings of the 7th Workshop on Semantic Ambient Media Experiences (SAME 2014), No 2014/1 (2014)
- Proceedings of the 8th Workshop on Semantic Ambient Media Experiences (SAME 2016): Smart Cities for Better Living with HCI and UX - SEACHI Extended Papers, No 2016/1 (2016)
- 2nd Workshop on (Re)Creating Lively Cities Through Ambient Techn. in Arts, Culture, and Gastronomic Experiences, No 2013/3 (2013)
- Proc. of the 1st Workshop on Defining a European Research Agenda on Inform. Systems and Management in eMedia Industry, No 2013/1 (2013)
- Proceedings of the 6th International Workshop on Semantic Ambient Media Experience (SAME), No 2013/2 (2013)
- Proceedings of the 5th International Workshop on Semantic Ambient Media Experience (SAME), No 2012/1 (2012)
- Proceedings of the 4th International Workshop on Semantic Ambient Media Experiences, in conjunction with C&T, No 2011/1 (2011)
- Proceedings of the 3rd Semantic Ambient Media Experience (SAME) Workshop in Conjunction with AmI 2010, No 2010/1 (2010)

**INTERNATIONAL SERIES ON INFORMATION SYSTEMS AND MANAGEMENT IN CREATIVE eMEDIA (CreMedia)**

Artur Lugmayr, Kening Zhu, Xiaojuan Ma (eds.)

# Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017):

# ARTIFICIAL INTELLIGENCE MEETS VIRTUAL AND AUGMENTED WORLDS (AIVR)

Number 2017/2
Bangkok, Thailand, 27th November 2017

Printed and published by the International Ambient Media Association (iAMEA)
Tampere, Finland 2017

Published and printed by International Ambient Media Association (iAMEA),
Ihanakatu 7-9/A1
FIN-33100 Tampere
FINLAND
ISBN 978-952-7023-17-4 (PDF)

ISSN 2341-5584 (Paperback)
ISSN 2341-5576 (PDF): www.ambientmediaassociation.org/Journal
ISSN 2341-6165 (CD-ROM)

## Preface

The objective of this workshop was to invite scholars and practitioners to discuss synergies between virtual and augmented reality (VR & AR) and artificial intelligence (AI) and machine learning (ML). The aim was to gather a cross-disciplinary team of experts with a background in computer science, Human-Computer Interaction (HCI), psychology/cognitive sciences, culture/communication studies, design and art to develop this fascinating intersection. The aspects to discuss range from user-experience, technologies applications, methods, cultural implications, communication theories, to artistic approaches. The detailed workshop description can be found in [1] and on the workshop website on http://aivr.ambientmediaassociation.org. Please subscribe to our email list for further information: http://emaillists.artur-lugmayr.com/.

The 10th Semantic Ambient Media Experience (SAME) proceedings where based on the academic contributions to a two-day workshop that was held in conjunct with SIGGRAPH Asia, Bangkok, Thailand. The papers of this workshop are freely available through http://www.ambientmediaassociation.org/Journal under open access as provided by the International Ambient Media Association (iAMEA) Ry. iAMEA is hosting the international open access journal/series entitled "International Journal/Series on Information Systems and Management in Creative eMedia". For any further information, please visit the website of the Association: http://www.ambientmediaassociation.org.

The International Ambient Media Association (AMEA) Ry organizes the Semantic Ambient Media (SAME) workshop series, which took place in 2008 in conjunction with ACM Multimedia 2008 in Vancouver, Canada; in 2009 in conjunction with AmI 2009 in Salzburg, Austria; in 2010 in conjunction with AmI 2010 in Malaga, Spain; in 2011 in conjunction with Communities and Technologies 2011 in Brisbane, Australia; in 2012 in conjunction with Pervasive 2012 in Newcastle, UK; and in 2013 in conjunction with C&T 2013 in Munich, Germany; and in 2014 in conjunction with NordCHI 2014 in Helsinki, Finland. In 2015 we had no workshop, but have been collaborating with the SEACHI Workshop Smart Cities for Better Living with HCI and UX, which has been organized by UX Indonesia and held in conjunction with Computers and Human-Computer Interaction (CHI) in San Jose, CA, USA in 2016. In 2016 we organised a second workshop (VisEmEx), which has been held at Curtin Univ., Perth, WA, Australia.

The series is double-blind peer reviewed; indexed in Scopus, CORR/arXiv, DBPL, OpenAIRE, WorldCat, Google Scholar; ranked on JUFI 1 in the Finnish national publication ranking system; and archived by Zenodo, and the National Library of Finland.

The workshop organizers present you a fascinating crossover of latest cutting-edge views on the topic of ambient media, and hope you will be enjoying the reading. We also would like to thank all the contributors, as only with their enthusiasm the workshop can become a success. At least we would like to thank the lovely organizing team of SIGGRAPH Asia for the help in the organizational aspects of the workshop.

The Editors — Bangkok, Thailand, 2017

[1] A. Lugmayr, Z. Kening, and X. Ma, "Artificial Intelligence MEETS Virtual and Augmented Realities," *SIGGRAPH Asia 2017 Workshops*, Bangkok, Thailand: ACM, 2017, pp. 2a:1–2a:1.

## Contents

**Preface** .................................... v
**Contents** .................................... vii
**List of Contributors** .................................... ix
**Workshop Programme and Description** .................................... xi

**Virtual Reality Techniques for Eliciting Empathy and Cultural Awareness: Affective Human-Virtual World Interaction** .................................... 1
*Ivonne Chirino-Klevans*, International School of Management, Paris, France

**Deep Learning for Classification of Peak Emotions within Virtual Reality Systems** .................................... 6
*Denise Quesnel*, School of Int. Arts&Techn., Simon Fraser University, CA
*Steve DiPaola*, School of Int. Arts&Techn., Simon Fraser University, CA
*Bernhard E Riecke*, School of Int. Arts&Techn., Simon Fraser University, CA

**Adaptive Tutoring on a Virtual Reality Driving Simulator** .................................... 12
*Sandro Ropelato*, ETH Zurich, CH
*Fabio Zuend*, ETH Zurich, CH
*Stephane Magnenat*, ETH Zurich, CH
*Marino Menozzi*, ETH Zurich, CH
*Robert W. Sumner*, ETH Zurich & Disney Research Zurich, CH

**Combining Intelligent Recommendation and Mixed Reality in Itineraries for Urban Exploration** .................................... 18
*Giulio Jacucci*, University of Helsinki, FI
*Salvatore Andolina*, Aalto University, FI
*Denis Kalkhofen*, Graz University of Technology, AT
*Dieter Schmalstieg*, Graz University of Technology, AT
*Antti Nurminen*, Aalto University, FI
*Anna Spagnolli*, University of Padova, IT
*Luciano Gamberini*, University of Padova, FI
*Tuukka Ruotsalo*, University of Helsinki, FI

**Interacting with Intelligent Characters in AR** .................................... 24
*Gokçen Çimen*, ETH Zürich, CH
*Ye Yuan*, Carnegie Mellon, US
*Robert W. Sumner*, Disney Research, CH
*Stelian Coros*, ETH Zürich, CH
*Martin Guay*, Disney Research, CH

**Chalktalk VR/AR** .................................... 30
*Ken Perlin*, New York University, US
*Zhenyi He*, New York University, US
*Fengyuan Zhu*, New York University, US

**Visualization Methods of Hierarchical Biological Data: A Survey and Review** .................................... 32
*Irina Kuznetsova*, Medical Univ. Graz, AT
*Artur Lugmayr*, Curtin University, AU
*Andreas Holzinger*, Medical Univ. Graz, AT

**Children Road SafetyTraining with Augmented Reality (AR) [Demo]** ........ 40
*Artur Lugmayr*, Curtin University, AU

**Data-Driven Approach to Human-Engaged Computing** ........ 43
*Xiaojuan Ma*, Hong Kong Univ. of Science & Technology, HK

**DupRobo: An Interactive Robotic Platform for Physical Block-Based Autocompletion** ........ 48
*Taizhou Chen*, School of Creative Media, City Univ. of Hong Kong, HK
*Yi-Shiun Wu*, School of Creative Media, City Univ. of Hong Kong, HK
*Feng Han*, School of Creative Media, City Univ. of Hong Kong, HK
*Baochuan Yue*, School of Creative Media, City Univ. of Hong Kong, HK
*Arshad Nasser*, School of Creative Media, City Univ. of Hong Kong, HK
*Kenning Zhu*, School of Creative Media, City Univ. of Hong Kong, HK

**Towards Emergent Play in Mixed Reality** ........ 51
*Patrick Misteli*, ETH Zurich, CH
*Steven Poulakos*, Disney Research, CH
*Mubbasir Kapadia*, Rutgers University, USA
*Robert W. Summer,* Disney Research/ETH Zurich, CH

**AI, You're Fired! Artwork** ........ 57
*Aleksandra Vasovic*, Independent Artist, RS

# List of Contributors

Aleksandra Vasovic
Andreas Holzinger
Anna Spagnolli
Antti Nurminen
Arshad Nasser
Artur Lugmayr
Baochuan Yue
Bernhard E Riecke
Denis Kalkhofen
Denise Quesnel
Dieter Schmalstieg
Fabio Zünd
Feng Han
Fengyuan Zhu
Giulio Jacucci
Gokçen Çimen
Irina Kuznetsova
Ivonne Chirino-Klevans
Ken Perlin
Kening Zhu
Li Mei
Luciano Gamberini
Marino Menozzi
Martin Guay
Mubbasir Kapadia
Patrick Misteli
Robert W. Sumner
Salvatore Andolina
Sandro Ropelato
Stelian Coros
Stéphane Magnenat
Steve DiPaola
Steven Poulakos
Taizhou Chen
Tuukka Ruotsalo
Xiaojuan Ma
Ye Yuan
Yi-Shiun Wu
Zhenyi He

## Workshop Programme and Description

| | |
|---|---|
| 09:00-09:30 | **VIDEO, DEMO, POSTER SHOWCASE and MORNING COFFEE** |
| 09:30-09:45 | **OPENING**<br>Artur Lugmayr, Kenning Zhu, Xiaojuan Ma |
| 09:45-10:45 | **KEYNOTE**<br>Prof. Yoshifumi Kitamura, Tohoku University, Japan<br>*Interactive Content Design* |
| 10:45-11:00 | **BREAK** |
| 11:00-12:30 | **1st PAPER SESSION**<br>• ***Adaptive Tutoring on a Virtual Reality Driving Simulator***<br>*Sandro Ropelato (presenter)*, ETH Zurich, CH<br>• ***Virtual Reality Techniques for Eliciting Empathy and Cultural Awareness: Affective Human-Virtual World Interaction***<br>*Ivonne Chirino-Klevans (presenter)*, International School of Management, FR<br>• ***Interacting with Intelligent Characters in AR***<br>*Gokçen Çimen (presenter)*, ETH Zürich, CH<br>• ***Chalktalk VR/AR***<br>*Ken Perlin*, New York University, US<br>• ***Child Safety Training Through Augmented Reality***<br>*Artur Lugmayr (presenter)*, Curtin University, AU<br><br>**12:00-12:30 Group Discussions and Paper Re-Cap** |
| 12:30-14:00 | **LUNCH BREAK** |
| 14:00-15:30 | **2nd PAPER SESSION**<br>• ***Combining Intelligent Recommendation and Mixed Reality in Itineraries for Urban Exploration***<br>*Giulio Jacucci (presenter)*, University of Helsinki, FI<br>• ***Deep Learning for Classification of Peak Emotions within Virtual Reality Systems***<br>*Denise Quesnel (presenter)*, School of Interactive Arts & Technology, Simon Fraser University, CA<br>• ***DupRobo: An Interactive Robotic Platform for Physical Block-based Autocompletion***<br>*Taizhou Chen*, School of Creative Media, City University of Hong Kong, HK<br>• ***Data Visualisation Methods of Hierarchical Biological Data***<br>*Irina Kuznetsova (presenter)*, *Andreas Holzinger*, Holzinger Group, Medical University Graz, AT<br>• ***Toward Emergent Play in Mixed Reality***<br>*Patrick Misteli*, ETH Zurich, CH<br><br>**15:00-15:30 Group Discussions and Paper Re-Cap** |
| 15:30-16:00 | **BREAK** |
| 16:00-17:00 | **GROUP DISCUSSIONS** |
| 17:00-17:45 | **DEBRIEFING & GROUP RESULT PANEL** |
| 17:45-18:00 | **CLOSING & WRAP UP** |
| 18:00-LATE | **DINNER & DRINKS** |
| DEMOS & OTHER WORKS | ***AI, You're Fired! Artwork***<br>*Aleksandra Vasovic*, Independent Artist, RS |

# Virtual Reality Techniques for Eliciting Empathy and Cultural Awareness: Affective Human-virtual world interaction

Ivonne Chirino-Klevans, PhD
International School of Management
Boulevard Raspail. Paris, France
ivonne.chirino-klevans@faculty.ism.edu

**Figure 1:** *Recording of the virtual reality simulation.*

## ABSTRACT

On the average human beings have about 50,000 thoughts every day. If we consider that thoughts influence how we feel there is little doubt that the way we perceive reality will strongly correlate with how we act upon that reality. Let's contextualize this thinking process within the realm of global business where interacting with individuals from other cultural backgrounds is the norm. Our own perceptions and stereotypes towards those cultural groups will strongly influence how we interact with them in business situations. The problem is that stereotypes, being cognitive shortcuts, not necessarily accurately represent intentions. Stereotypes provide us with a false sense of security enabling us to believe that we "understand" the reasons behind certain actions and reactions. This false sense of security often results in conflict in global business situations. That is one of the reasons why becoming globally competent without falling into stereotyping will provide us with the tools to increase success in cross-cultural business interactions.

This paper describes an approach to design a virtual reality (VR) scenarios aimed at developing abilities to work across cultures using the principles of empathy and perspective taking. The approach we are taking in this design innovation paper moves away from only using the understanding of cultural dimensions in cultural competence skills development as research shows that focusing on "preconceived" differences in cultures can enhance stereotyping. Instead our approach provides users with the opportunity of exploring the thought process as a character in first person whose cultural background is different from that of the user. This scenarios provide opportunities for perspective taking which is conducive to empathy across cultures.

## CONCEPTS

**Virtual reality, empathy, cultural competence, leadership, global competence**

## 1 INTRODUCTION

Educators everywhere are discovering that Virtual reality, an emergent technology, can potentially provide opportunities for learning to happen (Freina, & Ott, 2015). In the field of business education we can find multiple approaches to skill development through the use of case studies, role plays, technology-mediated simulations, to name a few. Virtual reality (VR) is starting to have traction in the area of skill development. VR can be used to recreate "real-life" situations where interactions can occur in a "safe" environment (Grant, 2008). One area in particular that benefits from the use of VR is the development of Global Leadership skills. These virtual scenarios will become powerful tools in the process of becoming globally competent only when the instructional design is robust enough and is able to address how each component of being globally competent can be broken down and represented in different virtual scenarios.

### 1.1 Development of global skills: empathy vs cultural dimensions

A traditional approach to develop global skills development (cross cultural competence) has been the understanding of cultural dimensions that may distinguish one culture from another. Cultural dimensions (Hofstede, 2001) are a set of behaviors and attitudes



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

that are shaped by cultural values. It is not uncommon to see many executive education programs and simulations aimed at understanding and identifying cultural dimensions in global business interactions. Although there is value in the ability of recognizing cultural dimensions, mere intellectual understanding of these differences do not necessarily result in effective cross-cultural negotiations. These can become intellectual exercises that may promote cultural stereotypes. On the other hand, helping executives/managers develop skills such as empathy will help in the process of understanding other cultures, taking perspective and eventually move to action by producing behaviors that address these cultural differences. It is only important to have a virtual reality scenario that resembles real life but it is also very important to create content that is actually aligned with the principles of the teaching of the skill it is aiming to develop. And because VR is starting to gain attention from different stakeholders, trainers and professionals in the field of business education it becomes more relevant that the content that is created using VR actually develops the skills it is saying it develops

## 1.2 Virtual Reality for empathy development across cultures

Virtual Reality Simulations (VRS) have the potential of being powerful tools for skill development. These VRS have the ability of recreating "real-life" scenarios where the user can be immersed giving him/her the experience of actually being "there". For example some authors have described how "altered media" (digital environments embedded in real life scenarios) can be used to provide unique opportunities of interaction in Augmented Realities (Lugmayr, Serral, Scherp, Pogorelc, Mustaquim (2014).

Virtual scenarios allow for the practice of skills in "life-like" situations, such as a virtual business meeting between a Chinese and a Canadian executive before facing the real situation. During these virtual encounters executives can learn best practices for different potential business negotiations across cultures increasing the sense of efficacy as well as providing tools to conduct effective cross-cultural negotiations. Mentoring relationships are traditionally used in organizations to help high potentials connect with more experienced executive and get guidance derived from these experienced business people. These mentorship relationships often require formal structures and availability of both parts in order to be effective. VR simulations not only are not time bound but the offer as well element of emotion in the contextualization of the VR experience. These virtual scenarios, when they involve emotions portrayed by the "actors", allow for a more realistic approach requiring the use of specific skills in order to successfully complete a multicultural business encounter (Sternberg, 2007). Many virtual simulations are starting to intentionally integrate emotions enacted by virtual agents and the use of the virtual space to elicit a response in the user (Zhu, Ma, Chen, & Liang, M. (2016).

Research in the area of cultural competence indicate that this ability requires different skills. Many authors have attempted to define these skills (Byram, Nichols and Stevens, 2001). Daniel Goleman (2006), an expert in emotional intelligence identifies 3 elements involved in empathy: cognitive empathy, emotional empathy and compassionate empathy. In order to better understand the role of empathy in intercultural interactions we will provide a brief description of each one of them. Cognitive empathy refers to the ability of perspective taking. This implies that individuals are able to see reality from behind "someone else's eyes", and are able to act upon this understanding of the world. On the other hand, emotional empathy refers to the ability of "feeling" what the other person is feeling because we have created a strong understanding of their perspective of a specific reality. This connection is related to social intelligence which addresses others' feelings. It will be the task of the viewer to calibrate his/her responses as a result of emotional empathy. The third kind of empathy is "compassionate empathy" which is the type of empathy that moves the individual to act. After reading these definitions it becomes evident that in the process of becoming cultural competent executives working in global businesses should not stop at understanding someone else's perspective (cognitive empathy) or creating a connection, but the outcome of this process should be moving to action after having accurately (or as accurate as possible) understood their counterpart's perspective. Research on empathy in global business has shown that executives who are effective at demonstrating empathy are perceived to have more productive inter-cultural negotiations (House & Javidan, 20004). Finally, some studies have revealed that empathy in global business is more important in high-power cultures, that is, cultures where power is perceived to be held by higher levels of an organization) and where power seems to be the source of stability and social order. Therefore empathic approaches in these situations reveal a climate of support and protection that enhance positive job performance (Yan & Hunt, 2005). The VR scenario needs to provide situations in which the participant is faced with the task of demonstrating empathy at the three different levels: cognitive, emotional and compassionate. It will be the task of the content expert to identify specific tasks that result in the development of these three levels.

The next question is how to measure effectiveness. Even when there are measures of effectiveness the question remains, why are these VR experiences effective?

## 1.3 What makes a VR simulation engaging?

Keeping students engaged into what they are learning is a question that many educators have tried to answer throughout time in many ways. Integrating technology into education and skill development can potentially be a source of increased engagement in the learning of the content. But the novelty of a new approach can quickly wear out. That is why educators and instructional designers should be asking themselves, what attributes of technology in the context of VR contribute to the learning that is obtained? In a meta-analysis study Howard (2017) concluded that three were the main reasons for a VR simulation to work: excitement, physical fidelity, and cognitive fidelity. This author concludes that a novel situation when presented in an engaging way, such as VR, encourages participants to try new behaviors because of the excitement that novelty brings within. Virtual environments are not widely known yet and create a sense of inquiry and curiosity. Having access to an immersive environment can in itself create excitement and get users engaged into the VR experience, more so if the environment includes exploration and an opportunity of playing (Bowman, Kruijff, LaViola, & Poupyrev, 2004), even when the required behaviors are not that complex.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

The second variable that research has identified associated with the effectiveness of VR is "physical fidelity". This means providing the user with an opportunity to practice a behavior in a similar way (or as close to it) as he/she would in real life. This element has proven to be very relevant in situations that involve physical rehabilitation where it is very important that patients are able to "translate" movements that are observed in a video into movements that are similar to those they engage in real life (Lucca, 2009).

The third variable discussed in the literature refers to increased "cognitive fidelity", defined as the extent in which a program prompts similar psychological processes as the real environment would. That is, when an individual is rehearsing a new behavior (mental or physical) there are cognitive processes that are happening simultaneously (attention, focus, concentration, etc.). Cognitive fidelity refers as well as the ability to elicit an emotion in the user and is the focus of this project.

The purpose of this document is to present a state of the art approach to Leadership skills development through the use of a simulation mediated by Virtual Reality, specifically empathy. In the following paragraphs I will describe the work involved in the design of the VR mediated simulation aimed at helping in the development of perspective taking as a pre-cursor of empathy. This simulation was created to address the gap in action learning practices for skill development.

# 2. VR SIMULATION DETAILS

## 2.1 General Description of the Simulation

The VR simulation provides users with an opportunity of immersing themselves in a virtual reality scenario where they take the role of an **observer or an actor** in a global business meeting happening in Asia. The simulation is divided in two parts: the first one where the participant is an observer and the second one where the participant takes on a specific character's role within the simulation. The first part of the simulation is set in an office environment. In the meeting there are several executives who reside in different countries (and come from different cultures) and who have chosen the China office as a meeting point. Their goal is to discuss and define the team's strategy since the team is charged to work on a global project. The scene represents several interactions among these executives who engage into different conversations that lead to conflict. The second part of the virtual experience allows the user to take the role of **one of the executives** involved in these interactions and be immersed into their thoughts. These thoughts reveal cultural assumptions that represent different cultural dimensions that are involved in communication. The ability to be immersed in someone else's thoughts provides an opportunity for perspective taking strongly associated with the development of empathy. The virtual reality scene must be considered as a component of a more comprehensive approach to developing global competence as we understand that global competence goes beyond recognizing differences in cultures but has to include decisions based on this new knowledge. This virtual reality exercise is aimed at developing awareness of difference and empathy through action learning which are key components of cultural competence, but not enough nor sufficient. By no means have we assumed that participants will be culturally competent after this experience. This is just the awareness component of a greater set of skills that being culturally competent in global business setting requires.

## 2.2 The setting and characters' psyche

Most digital simulations will face the challenge of creating virtual agents (virtual humans) that can accurately represent human behavior. Simulations that use digital worlds must be able to meet the following criteria: create environments that act in a 3D world, virtual agents need to be able to engage in one-on-one "dialogues" with the user and with other virtual humans and with real humans (users of the simulation), and those agents need to exhibit human-like behaviors and emotions. None of the above is an easy task. Because we believe that the recreation of human emotion and behavior needs to be as accurate as possible in order to trigger cognitive accuracy we approached this simulation integrating videos of real people as characters in a business setting. These characters need to demonstrate cultural dimensions through their dialogues, thoughts, and behaviors which will be accompanied or influenced by emotions that are triggered during different intercultural interactions. What makes this simulation unique and different from other business simulations is the intentionality in the design of the psyche of each character. This intentionality means that each character will think and react according to pre-defined cultural dimensions. We emphasize the importance of looking at this simulation as a training exercise in global competence and it should not be taken as an exercise in stereotyping. Instructors should be very careful at emphasizing the fact that behavioral and attitudinal manifestations of culture can vary across individuals and that this simulation represents one perspective of how cultural dimensions may be manifested in intercultural business interactions. The design of the dialogues and thoughts within the simulation reflect cultural dimensions and were coded as metadata that included the emotion as well. Emotions give characters energy and momentum to take action to solve problems during the actual recording of the simulation. The following three cultural dimensions were integrated to build each character's psyche reflected through their dialogues and thoughts: power distance, context, and task orientation. Here is a definition of each one of these dimensions. We define **power distance** (Hofstede's, 2001) as the "extent to which the less powerful members of organizations and institutions (including family) accept and expect that power is distributed unequally." A higher power distance reveals that inequality and power is perceived by the followers. It also indicates that hierarchy is clearly established and executed without reason nor resistance. A lower degree of the power distance means that people will question authority and will look for an equal distribution of power.

The second dimension that was included in the characters psyche was **context**. When we talk about context as it relates to culture we assume polarities. That is, there are cultures that represent high-context while others will represent low-context (Forsyth, 2010). In **high-context** cultures meanings are the result of a shared understanding which belongs to a particular group. There are several unspoken rules and symbols that would be difficult to understand to members outside of the group or to the untrained eye. In high-context cultures a great deal of communication happens through symbols and shared meanings. In **low-context** cultures "what you see is what you get" (Forsyth, 2010). Communication is more direct and relies on how the sender uses explicit communication. Messages are generally conveyed through conversation and, generally speaking, engage in business relationships rather faster than high-context cultures. The third dimension integrated into the dialogues was **task orientation** (Forsyth, 2010). Cultures with task orientation tend to approach business interactions focusing first on the task and then on the relationship. In these types of cultures achievement and meeting



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

goals are more important than forming a relationship. Task and deadlines are at the core of any business transaction. **Relationship oriented** cultures will consider important to create trust before any business transaction can successfully take place. Leaders that use a relationship oriented approach will focus on motivating, developing and meeting the needs of their teams, encouraging team work and collaboration. Then results will flow. Both approaches can be effective. The challenge is having a keen eye, understanding differences in approaches, and finding effective ways to address and act upon each preference when needed.

### 2.3 Setting empathy-evoking scenarios in VR

As described above Daniel Goleman, in his extensive work related to emotional intelligence, has consistently discussed the importance of empathy in effective leadership (Goleman, 2006). For the purpose of this paper Empathy is defined as the ability to identify and understand someone else's perspective (cognitive empathy), followed by the ability to "feel" what the other person is feeling and finally experience compassionate empathy which is what moves an individual to act. This VR simulation focused first on creating cognitive empathy. This ability starts with "perspective taking" which implies an ability to see the world from someone else's eyes. It is well known that leaders who are able to take perspective are perceived as someone who cares about his/her follower's development (Bass, 1989). With this in mind this simulation allows for users to be immersed in any of the character's thoughts. These characters are Chinese, Indian. Singaporean, and American. Each of the characters' thoughts were carefully designed considering the three dimensions of cultural competence previously identified (context, power distance, task-relationship orientation). Conveying emotions through dialogues is difficult enough. Adding cultural dimensions to these dialogues reveals a second layer of complexity in the creation of the psyche through dialogues and thoughts. This is one of the reasons why the creation of dialogues and thoughts for each character in the simulation requires an iterative process where writers are knowledgeable not only about how to express emotion but also what that expression would look like in different cultures. These dialogues and thoughts designed for each character required validation by different individuals who had the ability to describe and understand emotions across cultures through dialogue and contextual cues. We consulted with executives and natives from the countries represented in the simulations. Their observations were integrated and dialogues and thoughts were modified as needed. Being able to observe and make sense of these cues is part of the skills required to develop empathy in global competence. The simulation provides an opportunity to experience the world from a specific culture's point of view through being immersed in someone else's thoughts, allowing the user to take perspective and open their eyes to better understanding of the potential sources of conflict that arise from interpreting the world from our own cultural perspective. The second component of empathy, emotional empathy, requires a more complex process of design since it is aimed at helping the observer experience and "feel" the way the other is feeling. When users are able to take perspective by taking the role of "the other" and when they are able to listen how "the other" is interpreting the world from their own cultural values, the user will be able to experience emotional empathy. The final element of empathy- compassionate empathy- was not considered for this project but will be integrated into a more advanced simulation.

### 2.4 Video recording

After the dialogues had been created we hired actors that took on the role of the different "executives" in the VR simulation. These actors represented executives from China, India, Singapore, and the USA. A 360 degree camera was used to record all interactions and other cameras were used as well on each character to provide the user with the ability of taking the role of each actor during the actual simulation. This way the user, when taking the role of the actor, would be able to see the business setting as if he/she was sitting at the table where the meeting was taking place and from a different cultural perspective. The recording was done in such a way that the user could "see" their own hands and legs if he decided to look down recreating as much as possible the sensation of "being there".

## 3 METHOD

### 3.1 Conducting the simulation

Throughout this paper we have described the relevance of using simulations for leadership development in global contexts. VR provides a perfect means to bring this learning to life through the integration of action learning. In the following paragraphs we will describe how the VR simulation was used in 3 pilot workshops. The users were given 360 Head Mounting Displays (HMD) where they can see the simulation that has been downloaded to a mobile phone. The user observes the global business interaction that is taking place in an office in China. The user sees this interaction as a "fly on the wall" perspective (third person). The user is asked to remove the HDM and the facilitator prompts the group to identify what has happened in the scene. Discussions are around cultural dimensions impacting global business practices. Participants are asked to put on the goggles again and to choose one character from the business interaction. When the user clicks on the character in the VR world the user becomes that character and observes the interaction from their own cultural perspective. The user is also able to "listen" to his/her thoughts that reveal cultural dimensions as the character is giving meaning to the interaction that is going on where he/she is immersed. After observing this new scene participants are asked to take the HMD off and discussions are facilitated around cultural dimensions that impacted the scene they were being part of and are given the opportunity to analyze how different views (revealed by thoughts and dialogues) represented cultural dimensions, and more importantly, resulted in conflicts and misunderstandings. The user

## 4 FUTURE WORK

This simulation should be contextualized as the beginning of a more comprehensive approach to empathy development to work across cultures. We understand that perspective taking is only tapping on the tip of the iceberg of what being globally competent means. We see this first simulation as the beginning of a more comprehensive training that will later involve other skills that promote interactivity and decision making on situations that require negotiation, motivation, decision making, and conflict resolution among many other skills related to global leadership development integrating elements of Artificial Intelligence that will address responses to cross cultural conflict. Our final aim is to provide future leaders safe hands-on experiences mediated through the use of virtual reality technologies. We want to provide users interactive virtual experiences that require the use of different skills to handle a variety of interpersonal and intercultural business situations such as motivation, negotiation, conflict resolution, and strategic



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

decision making in global environments. We also understand that being global competent is a life-long process that requires an ability to be open to experience, nimble and respectful while avoiding stereotyping.

## ACKNOWLEDGMENTS

This work was possible thanks to a DELTA grant from NCSU.

## REFERENCES

Bass, B. M. (1989). *Leadership and performance beyond expectations.* New York, NY: Free Press.

Bowman, D. A., Kruijff, E, LaViola, J. J., &, Poupyrev, I. (2004). *3D user interfaces: Theory and practice*. New York, NY: Addison-Wesley.

Byram, M., Nichols, A., Stevens, D. (2001). Developing Intercultural Competence in Practice. New Yor, NY: Cromwell Press.

Freina, L., & Ott, M. (2015). A literature review of immersive virtual reality in education. State of the Art and perspectives, ELSE Conference, Bucharest, 2015. Bucharest, Romania: Academia.

Goldstein, A., & Sorcher, M. (1974). *Changing supervisory behavior*. New York, NY: Pergamon.Goleman, D. (2006). *Working with emotional intelligence*. New York, NY: Bantam Dell.

Goleman, D. (2006). Social Intelligence. New York, NY: Bantam Dell.

Grant, R. (2008). UALR's virtual reality center aiding businesses. *Arkansas Business, 25*(33), 20. Retrieved from http://proxying.lib.ncsu.edu/index.php?url=http://search.proquest.com.prox.lib.ncsu.edu/docview/220417323?accountid=12725

Heiden, E., & Lajoie, Y. (2010). Games-based biofeedback training and the attentional demands of balance in older adults. *Aging Clinical and Experimental Research*, *22*, 367–373.

Hofstede, G. (2001). *Culture's Consequences: comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

House, R. J., & Javidan, M. (2004). Overview of GLOBE. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), Culture, leadership, and organizations: The GLOBE study of 62 societies (pp. 513–563). Thousand Oaks, CA: Sage Publications.

Howard, M. (2017). A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior*, *70*, 317–27.

Kesselman, M. (2016). Current CITE-ings from the popular and trade computing literature: Google Cardboard – virtual reality for everyone. *Library Hi Tech News*, 33(4), 15-16.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development.* Thousand Oaks, CA: SAGE Publications.

Lucca, L. F. (2009). Virtual reality and motor rehabilitation of the upper limb after stroke: A generation of progress? *Journal of Rehabilitation Medicine*, *41*, 1003–1006.

A. Lugmayr, E. Serral, A. Scherp, B. Pogorelc, and M. Mustaquim, "Ambient media today and tomorrow," Multimedia Tools and Applications, vol. 71, 2014, pp. 7–37.

McCauley, C. D., & McCall, M. W. (2014). *Virtual reality and leadership development*. San Francisco, CA.: John Wiley & Sons.

Ott, J. S. (1989). *The organizational culture perspectives*. Homewood, IL: The Dorsey Press.

Shema, S. R.; Brozgol, M., Dorfman, M, Maidan, I., Sharaby-Yeshayahu, l., & Malik

Kozuch, H. (2014). Clinical experience using a 5-week treadmill training program with virtual reality to enhance gait in an ambulatory physical therapy service. *Physical Therapy, 94*, 1319–1326.

Sternberg, R. (2007). A systems model of leadership: WICS. *American Psychologist*, *62*(1), 34–42.

Swartout, W., Hill, R., & Gratch, J.(2001). Toward the Holodeck: Integrating graphic

Yan, J., & Hunt, J. G. (2005). A cross cultural perspective on perceived leadership effectiveness. International Journal of Cross Cultural Management, 5, 49–66.

Zhu, K., Ma, X., Chen, H., and Liang, M. (2016). Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures under the Influence of Operation, Handheld Posture, and Interaction Space. International Journal of Human-Computer Interaction.

# Deep Learning for Classification of Peak Emotions within Virtual Reality Systems

Denise Quesnel
School of Interactive Arts & Technology
Simon Fraser University
Canada
dquesnel@sfu.ca

Steve DiPaola
School of Interactive Arts & Technology
Simon Fraser University
Canada
sdipaola@sfu.ca

Bernhard E. Riecke
School of Interactive Arts & Technology
Simon Fraser University
Canada
ber1@sfu.ca

## ABSTRACT

Research has demonstrated well-being benefits from positive, 'peak' emotions such as awe and wonder, prompting the HCI community to utilize affective computing and AI modelling for elicitation and measurement of those target emotional states. The immersive nature of virtual reality (VR) content and systems can lead to feelings of awe and wonder, especially with a responsive, personalized environment based on biosignals. However, an accurate model is required to differentiate between emotional states that have similar biosignal input, such as awe and fear. Deep learning may provide a solution since the subtleties of these emotional states and affect may be recognized, with biosignal data viewed in a time series so that researchers and designers can understand which features of the system may have influenced target emotions. The proposed deep learning fusion system in this paper will use data collected from a corpus, created through collection of physiological biosignals and ranked qualitative data, and will classify these multimodal signals into target outputs of affect. This model will be real-time for the evaluation of VR system features which influence awe/wonder, using a bio-responsive environment. Since biosignal data will be collected through wireless, wearable sensor technology, and modelled through the same computer powering the VR system, it can be used in field research and studios.

## CCS CONCEPTS

• Human-centered computing ~ Virtual reality • Computing methodologies ~ Artificial intelligence

## ADDITIONAL KEYWORDS AND PHRASES

Virtual reality; Deep Learning; Affect; Emotion; Biosignals

## 1 INTRODUCTION

Awe experiences can lead to shifts in perspective, changes to how people see their relationship with the world. Furthermore, awe tends to provide therapeutic and educational benefits [1-5].

Additionally, awe is correlated with increased willingness to volunteer, and increased life satisfaction [6]. Being awestruck is also good for physical health: of six positive emotions, awe was found to be the strongest predictor of reduction in inflammatory cytokines [7], which are responsible for the initiation and persistence of pain. There are more examples of how being awestruck is good for the environment, social connectivity, and health, and many individuals have their own stories of how natural wonders, spaceflight, transformative life events, and artistic artifacts have elicited awe.

Awe-inspiring events in our natural world are quite rare but when they occur, it is with an immediate and intense manner that can be felt as a sensation of 'chills' or with the elicitation of goose bumps on the skin [8], see **Fig. 1**. This can be measured via physiological biosignals via skin surface video recording, which delivers an objective, unobtrusive means of recording and analyzing triggers of these states. In the study of profound transformative experiences, it is important to be able to objectively monitor the phenomenon for validity, but also to remove the burden of self-report from the participant. Reliance on self-reports of feelings can have major drawbacks, including the need for users to continuously self-monitor which can have the unwanted effect of removing them from an emotional state, and this data typically is retrospective. To address this, Benedek & Kaernbach [9] suggest physiological monitoring of these peak emotional events with many real-time physiological devices now that can collect data. While researchers now understand many of the psychophysiological mechanics that correlate with powerful affective states like awe and fear, there is currently few ways to measure the magnitude of this correlation, and therefore classify the target states accurately.

**Figure 1**: *Appearance of goose bumps on skin.*

### 1.1 Motivation

The consequence of successfully prompting an awe-inspiring experience with immersive virtual environments could be profound. Strong experiences of affect like awe are rare phenomena, and suffer from lower intensities in a lab environment due to possible issues with ecological validity [10]. For example, it is difficult to simulate an experience of great natural beauty and vastness to the same level of aesthetic detail and embodiment as in real life, thus missing a critical wellness experience for the participant, especially those with mobility or travel logistic issues. To address this, immersive, virtual environments can be personalized, with stimuli presented in a controlled, private,



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

physically accommodating manner superior to similar immersive delivery methods (large format screens like IMAX theatres, motion-tracked CAVE automatic virtual environments). Chirico [11] posits that these very personalized features may assist in eliciting higher-intensity awe in a controlled setting. Except for our pilot study of 16 participants in VR [12], this is largely unexplored especially with interactivity features. The reasons for this are largely two-fold: 1) awe-inspiring events are rare in controlled-laboratory environments; and 2) there is a lack of common verbal expression for profound, emotional experiences. Phenomenological methodology, self-reports, and interview data collection methods provide some insight into the experience of awe [8-10, 13-15], but the aforementioned drawbacks exist, with data being retrospective, subjective, with possible misinterpretations between participant and facilitator; and requiring a laboratory environment and extensive time to collect data. Using these methods alone, it is challenging to collect data, or suggest that findings in the lab may transfer into the field.

At the same time, researchers need an affordable tool for measuring and delivering experiences of awe for further understanding of the phenomenon and the events that trigger it. With our proposed study, we aim to utilize a deep learning multimodal fusion model that can extract salient features from multiple biosensors and annotated qualitative events in performing classification of affective states. This system will be part of a wearable tool comprising of a wireless webcam for video of goose bumps and skin conductance sensors (Fig. 2) that can record the user's continuous biosignal data without the need to be in a lab environment. Our aim is that by learning salient features of peak positive emotional experiences like awe, we can identify features in the VR system that prompt these feelings, which in turn will allow us to create more effective VR experiences for profound emotion elicitation.

## 1.2 Data corpus for emotion and affect in VR

It has been reliably demonstrated that changes with the body's physiological signals represents a shift in emotional state. These changes reflect the way the autonomic nervous system (ANS) and central nervous system (CNS) process internal events within the body and mind. Many physiological sensors can validate emotional states and affective moments, with electrocardiograms (ECG), heart rate (HR), galvanized skin response (GSR), skin conductance (SC), electrodermal activity (EDA), electroencephalogram (EEG), and electromyogram (EMG) all reliably tested [16-17].

More recently, profound shifts in emotional states of awe and wonder have been detected through visible goose bumps on the skin. Kelter & Haidt [1] demonstrated that goose bumps are a distinct central autonomic nervous system marker of awe associated with sympathetic activity. The correlation of chills and goose bumps to awe inspiring, affective experiences are well established [8, 13, 18]. Sumpf, Jentschke, & Koelsch [19] proposed heart rate variability (HRV), and Galvanic Skin Response (GSR) as additional measures to goose bumps [20].

For human computer interaction (HCI) and UX, physiological biosensing are becoming popular for the exploration of user affective experience. Findings have revealed significant correlations between psychophysiological arousal via HR and EDA with self-reported emotion in gameplay [21], with biosensing technology now readily available with the surge of wearable biosensing technology like FitBit for health monitoring, and biosensing interfaces for entertainment (gaming) to replace traditional game controllers.

**Figure 2**: *Interior view of the wireless wearable research tool, consisting of a HD video camera for macro 1 x 1 inch coverage of participant's skin surface for goose bump recording.*

# 2 DESIGNING AFFECTIVE VR THROUGH AI

## 2.1 Use of the time series

A time series (TS) is a collection of values obtained from sequentially ordered measurements or values, often visually represented as a database or graph. Measurements are uniformly spaced (time instants) at a given sampling rate. Represented as *T*, a time series is as follows:

$$T = (t_1,...,t_n),\ t_i \in \mathbf{R}. \quad (1)$$

While clusters are frequently used in data mining where labels are often unknown in the dataset, classification differs in that classes are known in advance, and the algorithm is trained on this example dataset [22]. First, we must understand what the features are, as they will belong to different classes. This way, when an unlabeled dataset is fed into the system, it can assign the appropriate class. For this study, TS classification will be the aim rather than clustering, as we can appropriately label intense feelings of positive emotion into categories. We need our system to identify the features as learned through the labeled training data set and classify new datasets. This can be done effectively with Convolutional Neural Networks (CNNs), which are suitable for TS classification through the use of window slicing for classification at the slice level [23], and are robust to scaling issues, i.e: different time series may contain differing time scales, common with biosignal data. As a deep learning method, the features extracted by CNNs are not handcrafted (manually added) as they are with non-deep learning models.

## 2.2 Pattern recognition and multimodal fusion

Pattern recognition consists of recognizing an object based on its unique attributes and features. While TS classification and pattern recognition is complex enough within a single biosignal TS, it is important to note that a comprehensive, accurate classification



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

may not be possible without a multimodal approach. It should be noted that humans use multiple modalities to feel and express a state [24]; humans do not emote through one channel alone. An advantage to using deep learning (DL) in CNNs for multimodal biosensing recognition and classification is that the sensors can be collected from an unconstrained environment, meaning no baseline, and no significant data preprocessing or handcrafted feature selection is required [25]; learning can occur through ordered feature representation from raw data. Additionally, CNN based classifiers have been more accurate than k-Nearest Neighbor algorithm and Support Vector Machine models for event classification of multi-senor environments [26-27]. It is for these compelling reasons and because of the accuracy of DL CNN models of emotion and affect that we propose to use a similar method utilizing multi-sensor physiological data. Deep multimodal fusion models demonstrate filter-pooling methods provide the most effective fusion for accurate predictions [28]. User events such as annotated comments made by the participant (discrete signals) are recorded in parallel with physiological (continuous) biosignals, and used in a CNN with a Single Layer Perceptron (SLP). Through fusing aspects of the system such as user events with physiological data, we can better evaluate the VR system for target affect and emotion by understanding the elicitor of the target affect and emotional state, and then further facilitate these targets in VR.

# 3 METHODS

## 3.1 Emotional Inducement in VR

For this study, we are interested in feelings of peak positive affect while interacting with immersive VR. Our stimulus is a visualization of the Earth, in which many participants experience a sense of awe and wonder through visiting places on earth [12]. We call this stimulus the 'Earthgazing' content, and is presented in VR, as seen in Fig. 3. To understand how affect-inducing the Earthgazing content is, we need to compare it to content that was created for dissimilar purposes. We choose an educational game of the same length, designed to be informative in nature and not designed to induce positive affect. Both sets of content are created by the researchers in Unity3D, and use the same navigational interface (HTC Vive hand controllers), and head-mounted display (HTC Vive).

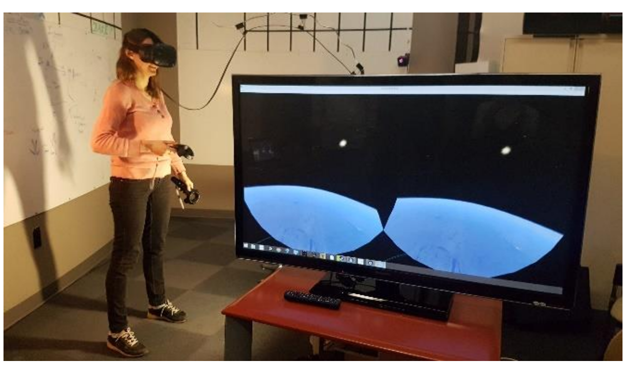


**Figure 3**: *experimental setup with participant 'earthgazing'.*

**DEFINITIONS**

***Cued recall debrief:*** a video situated recall methodology applied in naturalistic research, with the aim of re-immersing the participant and gaining insight into their thoughts and feelings of the system being evaluated. (Bentley et al., 2005)

***Affect****:* a short term, discrete, and conscious subjective feeling that may have an influence upon a person's overall emotion (Bentley et al., 2005, p. 3). Affective meaning is the degree at which a situation or object changes an individual's reaction to the environment (Duncan & Barrett, 2007).

## 3.2 Data collection and labelling

For the training data, it is important to accurately label the samples accordingly to their affective state. This will be conducted through a process of cued recall debriefing, which re-immerses the participant with the VR system content and gains insight into their thoughts and feelings [29]. This is a trusted method of data collection and analytics for UX evaluation, because it can isolate features of the system design that may be responsible an affective event. During cued recall, the facilitator can view the biosignals on TS and annotate a comment made by the participant about their affective state (see **Fig. 4**). Such an affective statement may be "Whoa!" indicating surprise or wonder, or "I felt I was part of a greater collective", indicating a potential awe-inspiring experience. Participants may also directly categorize their affective states, such as statements like "entering this scene made me feel afraid right there", or "I felt awe when I looked at the Earth". Such statements can be coded into their appropriate affective states, through thematic analysis. Regardless of how they are coded, the comments made by the participant will be annotated on the time series, but coded after the session with the participant is concluded.

These discrete affective annotations can be transformed into binary continuous signals and when a window is generated around them, features from biosignals may be seen within the window. As each annotated event will be within its own window, the window as a sample will be pairwise ranked for feelings of awe, excitement, motion sickness, and frustration based on the interpretation from the cued recall debriefing (coded thematic analysis).

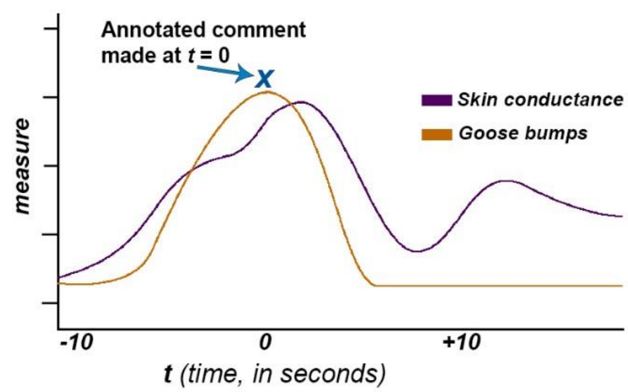



**Figure 4**: *affective comment annotated within a joint display of physiological data.*

A final step after the cued-recall debriefing is concluded, is that the participant will be asked to rank the entire experience of the Earthgazing content and educational content. The facilitator will



3
In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

ask the participant to rank the more awe-inspiring/exciting/motion sickness inducing/frustrating pair: A or B; both are equally so; neither are so. The purpose of ranking the entire experience and annotating the events themselves is so that we have rankings for the preference learning of the model across the experience, and within windows centered on events. With this method, we aim for experimental validity and to utilize the framework set by Yannakakis & Martínez [30] that demonstrates rank-based questions yield more reliable models on the constructs of emotion and preference than rating-based questions.

# 4 MODELING, TRAINING, & CLASSIFICATION

Each neuron defines one local feature which is extracted at every position of the input signal, with the output creating a new signal called a feature map (lower time resolution output). Reduced dimensionality of the convolutional layer is created through filtered feature maps consisting of the fusion of the biosignals and discrete data events (annotated comments from cued recall debriefing) that are transformed to continuous binary signals with a decay rate; see **Fig. 5**. Windows are centered on events, displaying the newly created signal of patterns.

The classification is the output layer of the neural network. In this case, the classification/output is a representation of the affective states. The output will be determined from the outcome of the initial corpus data collection, where researchers will be assigning annotated events from participants into categories of awe, excitement, motion sickness, and frustration. From these categories, classifications will be generated. Automation of feature extraction is possible through Preference Deep Learning (PDL), which was first applied to psychophysiology by Martínez et al., [31] with pairwise preference events of affect across two biosignals. Preference learning handles pairs of data samples (xP, xN), as a model that "outputs higher values for the samples preferred on each pair and lower for the non-preferred" sample [28, p. 4], xP is preferred over/greater than xN. Since we know some of the main features and labels that we'd like to use in our training data but still wish to discover new and unknown features, this is an appropriate technique. With PDL, the SLP is trained to predict the affective state of a user via their biosignals through backpropagation, like Rank Margin error function.

If accurate, this will allow for the VR system to become bio-reactive and adjust the participant's environment to personalize their experience, IE: if they are predicted to be experiencing fear, the system integrates calming audio-visual stimuli; if the participant is possibly experiencing awe, the system may respond with a crescendo of music and aesthetic beauty.

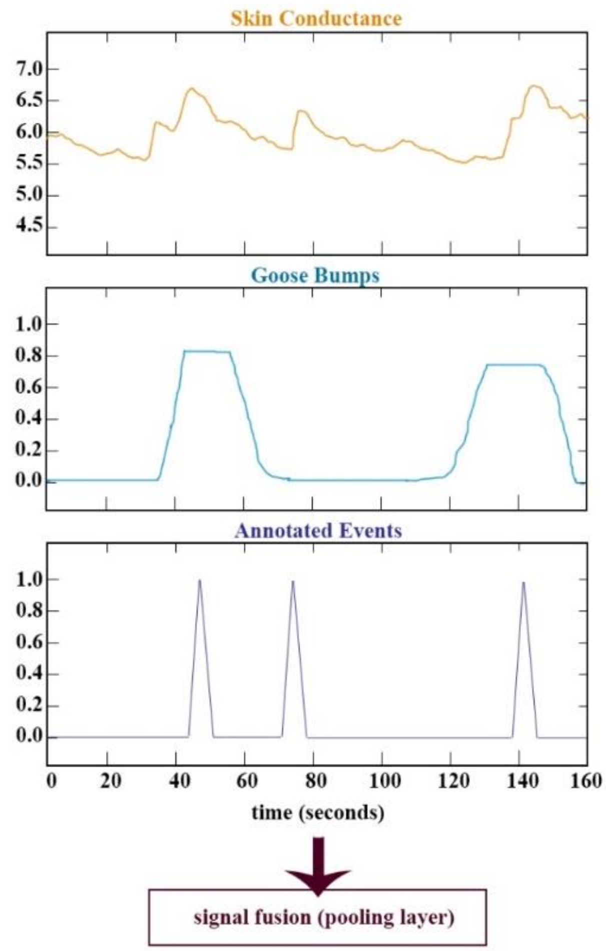



**Figure 5**: *Fusion of continuous biosignals and events (binary continuous).*

# 5 FUTURE WORK

Through our proposed study, we aim to augment existing rank-based PDL systems [28, 31] with a validation technique of qualitative cued recall data and categorization of participant affective states for the labelling of our corpus datasets for positive, profound emotion classification. In fusing multimodal physiological signals with reported events of affect like awe and wonder, our aim is to utilize a CNN for accurate classification of output states, demonstrating that complex time series data can be integrated in a model of affect. This work involves the collection of skin conductance and goose bump physiological data alongside the cued recall annotated events from participants for training data (the corpus), and once collected will be made open-sourced for public use. We will test the model and make iterative adjustments to the model parameters, including number of layers, the model will be evaluated for accuracy, and reported with descriptive and inferential statistics. If the system accurately classifies the target output affect states, we will observe whether new salient features have been discovered.

Such a tool as this can be regarded a type of "emotive media", specifically fitting into the characteristic of bio-feedback sensors creating a pathway for interactive applications to discover emotional states [32]. Because emotional VR has been seen to elicit



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

specific verbal and physical responses in our studies, such as a sharp intake of breath, outbursts of 'whoa!', or physically relaxing in posture [33], it may be valuable to collect basic social signals. Such signals are used in human-robot interaction, with speech, posture, and body orientation data collected and modelled into behavior generation modules for state-sensitive behaviours [34]. Such data could be collected during the creation of the data corpus, and coded based on the cues noted in video recordings. This data could be particularly useful in helping to differentiate between profound affective states that may have similar biosignal features in at least one sensor modality, for example between a fear-inducing moment, or a profound awe-inducing moment, both of which may induce physiological goosebumps- yet verbal or posture data may clearly indicate which state is more probable.

In utilizing a deep learning approach to a wearable biosensing research tool for study of positive peak emotion, we hope the work may be utilized by researchers, designers, and artists for evaluation and creation of bio-responsive, personalized VR environments.

## ACKNOWLEDGMENTS

This work is partially supported through a Simon Fraser University graduate student merit-based scholarship award, KEY Big Data Initiative. The authors thank Simon Fraser University for its support.

## REFERENCES

[1] Keltner, D. and Haidt, J. 2003. Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and Emotion*. 17, 2 (Jan. 2003), 297–314. DOI:https://doi.org/10.1080/02699930302297.

[2] Piff, P.K. et al. 2015. Awe, the small self, and prosocial behavior. *Journal of Personality and Social Psychology*. 108, 6 (2015), 883–899. DOI:https://doi.org/10.1037/pspi0000018.

[3] Riva, G. et al. 2016. Transforming Experience: The Potential of Augmented Reality and Virtual Reality for Enhancing Personal and Clinical Change. *Frontiers in Psychiatry*. 7, (2016), 164. DOI:https://doi.org/10.3389/fpsyt.2016.00164.

[4] Shiota, M.N. et al. 2011. Feeling good: autonomic nervous system responding in five positive emotions. *Emotion (Washington, D.C.)*. 11, 6 (Dec. 2011), 1368–1378. DOI:https://doi.org/10.1037/a0024278.

[5] Shiota, M.N. et al. 2007. The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and Emotion*. 21, 5 (Aug. 2007), 944–963. DOI:https://doi.org/10.1080/02699930600923668.

[6] Rudd, M. et al. 2011. Awe Expands People's Perception of Time, Alters Decision Making, and Enhances Well-Being. *NA - Advances in Consumer Research Volume 39*. (2011). DOI:https://doi.org/10.1177/0956797612438731.

[7] Stellar, J.E. et al. 2015. Positive affect and markers of inflammation: discrete positive emotions predict lower levels of inflammatory cytokines. *Emotion (Washington, D.C.)*. 15, 2 (Apr. 2015), 129–133. DOI:https://doi.org/10.1037/emo0000033.

[8] Grewe, O. et al. 2009. The Chill Parameter: Goose Bumps and Shivers as Promising Measures in Emotion Research. *Music Perception: An Interdisciplinary Journal*. 27, 1 (Sep. 2009), 61–74. DOI:https://doi.org/10.1525/mp.2009.27.1.61.

[9] Benedek, M. and Kaernbach, C. 2011. Physiological correlates and emotional specificity of human piloerection. *Biological Psychology*. 86, 3 (Mar. 2011), 320–329. DOI:https://doi.org/10.1016/j.biopsycho.2010.12.012.

[10] Silvia, P.J. et al. 2015. Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*. 9, 4 (2015), 376–384. DOI:https://doi.org/10.1037/aca0000028.

[11] Chirico, A. et al. 2016. The Potential of Virtual Reality for the Investigation of Awe. *Human-Media Interaction*. (2016), 1766. DOI:https://doi.org/10.3389/fpsyg.2016.01766.

[12] Quesnel, D., & Riecke, B. E. (2017). Awestruck: Natural interaction with virtual reality on eliciting awe (pp. 205–206). IEEE. https://doi.org/10.1109/3DUI.2017.7893343

[13] Benedek, M. et al. 2010. Objective and continuous measurement of piloerection. *Psychophysiology*. 47, 5 (Sep. 2010), 989–993. DOI:https://doi.org/10.1111/j.1469-8986.2010.01003.x.

[14] Gallagher, S. et al. 2015. *A Neurophenomenology of Awe and Wonder*. Palgrave Macmillan UK.

[15] Grewe, O. et al. 2011. Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*. 39, 2 (Apr. 2011), 220–239. DOI:https://doi.org/10.1177/0305735610362950.

[16] Li, L. and Chen, J. h 2006. Emotion Recognition Using Physiological Signals from Multiple Subjects. *2006 International Conference on Intelligent Information Hiding and Multimedia* (Dec. 2006), 355–358.

[17] Szwoch, W. 2015. Emotion Recognition Using Physiological Signals. *Proceedings of the Mulitimedia, Interaction, Design and Innnovation* (New York, NY, USA, 2015), 15:1–15:8.

[18] Schurtz, D.R. et al. 2011. Exploring the social aspects of goose bumps and their role in awe and envy. *Motivation and Emotion*. 36, 2 (Sep. 2011), 205–217. DOI:https://doi.org/10.1007/s11031-011-9243-8.

[19] Sumpf, M. et al. 2015. Effects of Aesthetic Chills on a Cardiac Signature of Emotionality. *PLOS ONE*. 10, 6 (Jun. 2015), e0130117. DOI:https://doi.org/10.1371/journal.pone.0130117.

[20] Colver, M.C. and El-Alayli, A. 2016. Getting aesthetic chills from music: The connection between openness to experience and frisson. *Psychology of Music*. 44, 3 (May 2016), 413–427. DOI:https://doi.org/10.1177/0305735615572358.

[21] Drachen, A. et al. 2010. Correlation Between Heart Rate, Electrodermal Activity and Player Experience in First-person Shooter Games. *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games* (New York, NY, USA, 2010), 49–54.

[22] Esling, P. and Agon, C. 2012. Time-series Data Mining. *ACM Comput. Surv.* 45, 1 (Dec. 2012), 12:1–12:34. DOI:https://doi.org/10.1145/2379776.2379788.

[23] Guennec, A.L. et al. 2016. Data Augmentation for Time Series Classification using Convolutional Neural Networks. (Sep. 2016).



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

[24] Chen, L.S. et al. 1998. Multimodal human emotion/expression recognition. (1998), 366–371.

[25] Sarkar, S. et al. 2016. Wearable EEG-based Activity Recognition in PHM-related Service Environment via Deep Learning | PHM Society. *INTERNATIONAL JOURNAL OF PROGNOSTICS AND HEALTH MANAGEMENT*. 7, Special Issue Big Data and Analytics (Sep. 2016), 10.

[26] Tong, C. et al. 2017. A convolutional neural network based method for event classification in event-driven multi-sensor network. *Computers & Electrical Engineering*. (Jan. 2017). DOI:https://doi.org/10.1016/j.compeleceng.2017.01.005.

[27] Zhang, Y. et al. 2013. Multi-metric Learning for Multi-sensor Fusion Based Classification. *Inf. Fusion*. 14, 4 (Oct. 2013), 431–440. DOI:https://doi.org/10.1016/j.inffus.2012.05.002.

[28] Martínez, H.P. and Yannakakis, G.N. 2014. Deep Multimodal Fusion: Combining Discrete Events and Continuous Signals. *Proceedings of the 16th International Conference on Multimodal Interaction* (New York, NY, USA, 2014), 34–41.

[29] Bentley, T. et al. 2005. Evaluation Using Cued-recall Debrief to Elicit Information About a User's Affective Experiences. Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future (Narrabundah, Australia, Australia, 2005), 1–10.

[30] Yannakakis, G.N. and Martínez, H.P. 2015. Ratings are Overrated! *Frontiers in ICT*. 2, (2015). DOI:https://doi.org/10.3389/fict.2015.00013.

[31] Martinez, H.P. et al. 2013. Learning deep physiological models of affect. *IEEE Computational Intelligence Magazine*. 8, 2 (May 2013), 20–33. DOI:https://doi.org/10.1109/MCI.2013.2247823.

[32] Lugmayr, A. 2016. Emotive media: a review of emotional interfaces and media in human-computer-interaction. Proceedings of the 28th Australian Conference on Computer-Human Interaction (OzCHI '16 ) (2016), 338–342. DOI: 10.1145/3010915.3010982

[33] Quesnel, D. and Riecke, B. 2017. Connected Through Awe: Can Interactive Virtual Reality Elicit Awe for Improved Well-Being? Poster Presented at the 3rd Annual Innovations in Psychiatry and Behavioral Health: Virtual Reality and Behavior Change (Stanford University, CA, Oct. 2017). DOI: 10.13140/RG.2.2.22177.10088

[34] Sun, M., Zhao, Z. and Ma, X. 2017. Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17 ) (2017), 556–567. DOI: 10.1145/3025453.3025469



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# Adaptive Tutoring on a Virtual Reality Driving Simulator

Sandro Ropelato
ETH Zurich
sandro.ropelato@gmail.com

Fabio Zünd
ETH Zurich
fabio.zund@inf.ethz.ch

Stéphane Magnenat
ETH Zurich
stephane@magnenat.net

Marino Menozzi
ETH Zurich
mmenozzi@ethz.ch

Robert W. Sumner
ETH Zurich / Disney Research Zurich
robert.sumner@inf.ethz.ch

## ABSTRACT

We propose a system for a Virtual Reality (VR) driving simulator including an Intelligent Tutoring System (ITS) to train the user's driving skills. The VR driving simulator comprises a detailed model of a city, Artificial Intelligence (AI) traffic, and a physical driving engine, interacting with the driver. In a physical mockup of a car cockpit, the driver operates the vehicle through the virtual environment by controlling a steering wheel, pedals, and a gear lever. Using a Head-Mounted Display (HMD), the driver observes the scene from within the car. The realism of the simulation is enhanced by a 6 Degrees of Freedom (DOF) motion platform, capable of simulating forces experienced when accelerating, braking, or turning the car. Based on a pre-defined list of driving-related activities, the ITS permanently assesses the quality of driving during the simulation and suggests an optimal path through the city to the driver in order to improve the driving skills. A user study revealed that most drivers experience presence in the virtual world and are proficient in operating the car.

## KEYWORDS

virtual reality, driving simulation, intelligent tutoring system

**ACM Reference Format:**
Sandro Ropelato, Fabio Zünd, Stéphane Magnenat, Marino Menozzi, and Robert W. Sumner. 2017. Adaptive Tutoring on a Virtual Reality Driving Simulator. In *Proceedings of 1st Workshop on Artificial Intelligence Meets Virtual and Augmented Worlds (AIVRAR), BITEC, Bangkok, Thailand, November 2017 (AIVRAR'17),* 6 pages.

## 1 INTRODUCTION

Learning how to drive a car involves many hours of training. Combinations of complex activities require the driver's full attention, and even experienced motorists make mistakes or show wrong reactions when faced with unexpected events. Having a simulated setup to improve car driving skills can be useful to both novice and experienced drivers, as a variety of scenarios that might occur on real roads can be exercised in a safe environment. In order to imitate real-life driving scenarios, an immersive Virtual Reality (VR) environment is required.

Training softwares running in a workplace-like environment, with a single screen and a keyboard, can be used to simulate an interactive car ride. However, presence and immersion in such setups are strongly limited. A keyboard does not resemble the instruments used to control a car, and a regular monitor fails to offer a sufficiently large field of view to experience movement. In addition, there is no physical feedback of acceleration and no intuitive way of looking around in the virtual environment.

Computer programs have been used to assist in training skills and have shown improvement in the learning progress when adapting to the individual learner. We combine a VR headset with a 6 Degrees of Freedom (DOF) motion system to improve the presence and immersion of the driving simulation, and include an Intelligent Tutoring System (ITS) to adapt the training to the individual user. In our proposed system, the ITS suggests an optimal sequence of exercises, such as stable driving, turning, and reaction, to the user. These exercises are spatially distributed in the virtual city. As such, the ITS is well-suited to be integrated into the driving environment as a Satellite Navigation System (satnav), which leads the user from one exercise to the next. Hence, the ITS does not interfere with the driving immersion and the user can follow the instructions provided by the satnav without being distracted from driving.

## 2 RELATED WORK

**Virtual Reality Simulations.** The Railway Technical Research Institute (RTRI) has developed a VR safety simulation system where users can train how to respond to critical situations [9]. In a virtual environment, various types of problems can be simulated. Users of the system are required to cooperate with each other and resolve problems in order to prevent further complications and restore services as soon as possible. The goal of this system is to offer a safe environment where unpredictable or dangerous incidents can be handled by public transport staff. The knowledge and experience gained in training situations in the virtual environment can be projected onto real-life scenarios and improve people's performance in problem solving.

In a different simulation, Augusto et al. [1] show how VR can be used to train security staff in securing and protecting nuclear facilities. Based on construction plans of a nuclear power plant, a virtual environment is created. In a game-like training mode, security staff watch the facilities while the system simulates an infiltration attempt where intruders try to access restricted areas. The authors propose a way to improve physical security of nuclear facilities in two ways, by offering a method to analyze the facility's infrastructure and by enabling security personnel to actively practice operations in VR.

Both of the aforementioned projects aim at improving real-life performance in difficult scenarios by providing training in virtual environments, from which users learn how to handle real-life situations. In both examples, however, the exercises in the simulated scenarios are manually created and not adapted to the specific skills of the user. Carefully matching the tasks and their order to the needs of the user requires manual input.

**Learning Progress.** Previous work [3] has shown that the order in which exercises are solved can have a major influence on the learning progress. The optimal sequence depends on the subject



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

solving the exercise and varies between individuals. The authors proposed a method to estimate the learning progress in each exercise and generate a sequence tailored to each user. When testing their algorithm, Zone of Proximal Development and Empirical Success (ZPDES), on primary school pupils solving basic math exercises, they showed that the system-generated order of activities yields a better overall learning progress than one defined by experts.

In our work, we combine the use of a VR training environment with the approach of dynamically adapting the sequence of trained activities using the ZPDES algorithm.

## 3 HARDWARE AND SOFTWARE ENVIRONMENT

Applying a tutoring system to car driving requires an environment where different skills can be trained. For this purpose, we created a VR driving platform to simulate an interactive car ride through a city. We used Unity, a 3D game engine, to combine visuals, a physics simulation, interaction with input devices, and a motion system. The following section presents an overview of all components used in the simulation, and describes how they interact with each other.

**Figure 1: Cockpit mockup on the motion platform.**

### 3.1 Hardware

**Motion System.** To exert physical forces on the driver, we employ a Thruxim Pro 6 DOF motion simulator by CKAS Mechatronics Pty Ltd. It supports linear displacement of the driving platform along the X, Y and Z axes, as well as rotation in all three directions. This allows simulating linear acceleration by moving and tilting the platform in the corresponding direction. For example, when accelerating in a real car, the driver is being pushed back into the seat. On the simulator, this can be imitated by rotating the platform around the horizontal axis. Lateral forces that occur while turning a car can be simulated by tilting the platform around the longitudinal axis. When the simulated acceleration is constant or changes very slowly, this creates an illusion of linear acceleration without an actual linear movement. In car driving, however, there are strong changes in acceleration. For instance, when driving at a constant speed and then immediately braking, the acceleration along the forward axis changes in almost no time from 0 G to as much as 1 G [8]. When the rotation of the platform changes too quickly, the motion is perceived as a rotation around the center of the platform rather than a change in velocity, thus destroying the illusion of linear acceleration. To avoid this, the tilting is supported by an actual linear movement along the corresponding axis. Especially at higher acceleration change rates, this can reduce the perception of a rotational movement [2].

**Cockpit Mockup.** The platform is equipped with a driving seat taken from an old Ford Ka and a wooden frame for mounting the steering wheel, the pedals, and three screens. Three 27-inch monitors have been arranged to offer a feld of view of up to 120 degees, depending on the driver's position. A Thrustmaster T500RS steering wheel and pedal set along with an 8-gear shifter imitate input devices present in real cars. Force feedback can be applied to the steering wheel. Fig. 1 shows the cockpit mockup mounted on the motion platform.

**Head-Mounted Display.** The virtual environment is either presented on the three screens or through an HTC Vive Head-Mounted Display (HMD). The HMD is tracked using two base stations installed on the ceiling above the simulation platform. When moving or turning the head, the camera is moved accordingly in the virtual scene so that the driver is able to examine objects in the cockpit from different angles. Movement of the motion platform is subtracted from the HMD's position and rotation so that the driver's perspective remains relative to the cockpit when linear acceleration is simulated. The Vive's display has a resolution of 2160×1200 pixels (1080 × 1200 per eye) and its optics offer a field of view of up to 110 degrees. The display refresh rate is 90 Hz which allows for low-latency updates when moving and rotating the head [10]. The tracking system records the HMD's position with a maximum tolerance of 2 mm [6]. For our application, this is precise enough so that the user does not detect any jitter. We considered the total weight of 550 g [7] to be acceptable in that, even after test runs on the simulator above 30 minutes, no driver reported discomfort from the headset's weight.

**Computer.** A gaming computer with a 4 GHz Intel Core i7-6700K processor, 32 GB memory, and two Nvidia GeForce 1080 graphics cards, with 8 GB graphics memory each, drive the simulation.

### 3.2 Physics Simulation

The way a car behaves is influenced by various parameters such as mass, engine power, tire friction, and suspension. Some of these can be easily modeled with Unity's built-in physics engine. For others, we had to build new models based on specific characteristics of the car.

**Suspension and Tires.** On an abstract level, a vehicle such as a car consists of a rigid body with a number of wheels attached. In this case, there are four wheel, two of which are driven by an engine. In Unity, each wheel is configured to push the body of the car upwards. As on a real car, the wheels are not directly connected to the body but use a simplified suspension, simulating a spring and a damper. The amount of force applied by the wheel colliders depends on the configuration of the suspension model.

Along with the suspension properties, the tire friction is defined for each wheel. Unity uses a two-spline curve to specify the force exerted on the contact point between the wheel and the road as a function of tire slip, the magnitude of the motion vector between a tire's contact point and the road. The tire slip is zero when the wheel has full traction and increases when the tire slides on the road,



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

e.g. in an emergency brake. The wheel friction spline is defined by two points, (expremum slip, extremum force) and (asymptote slip, asymptote force). Given a value for tire slip, the force on the wheel is taken from evaluating the spline. Since this is a vague approximation of a tire's behavior, the values do not correspond to any specification but have been evaluated by testing the slipping behavior when accelerating, braking, or turning the car at high speeds.

**Engine and Transmission.** The simulated car has a combustion engine, which means that the torque produced is a non-linear function of the engine speed. It is usually specified in revolutions per minute (RPM). In other words, when pushing down the accelerator pedal, the force that is being output by the engine depends on how fast the engine is already going. For this simulation, we used the specification of a Fiat 500's engine [12]. The engine keeps its speed at a predefined RPM value when the car is stationary or driving very slowly. This is done by gradually increasing the throttle until the idling speed is reached. The engine stops when its speed drops below a minimum RPM value.

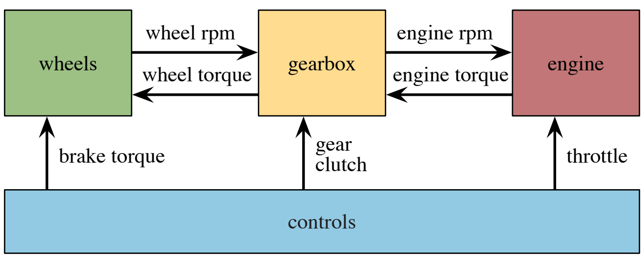



**Figure 2: Components of the transmission model. The driver's input controls how fast the engine should accelerate, how much brake force is applied, how far the clutch is engaged, and which gear is selected.**

The wheels are not directly driven by the engine. They are connected to the gearbox, translating the engine's speed to a different output speed defined by the gear's transmission ratio. As shown in Fig. 2, engine, gearbox, and wheels are connected to each other and are controlled by the driver's input.

### 3.3 3D Content Generation

Creating an appealing visual design substantially contributes to the realism of a virtual reality application. 3D models of cars, a detailed model of the car cockpit, and a set of traffic signals have been manually created to add to a life-like car driving experience. A city generator has been used to generate 3D models of buildings and a street layout upon which roads are dynamically constructed in Unity.

**Car and Cockpit Model.** We created a 3D model of a Fiat 500 and integrated it with Unity's built-in shaders and support for performance-saving Level of Detail (LOD) rendering. A detailed model of the car's interior has been designed to resemble the real cockpit. It contains a speedometer, a tachometer, mirrors, control LEDs for the indicators, and a satnav, which displays directions provided by the ITS. The mirrors correctly display the scene behind the car. This has been realized by placing three cameras in front of the mirrors, each rendering the virtual environment as seen through the respective mirror. The camera's rendered output is stored in a render texture and displayed on the mirror. Fig. 3 shows the view presented to the driver when sitting inside the car's cockpit.


**Figure 3: Interior view of the car. The mirrors reflect the environment behind the car. The satnav displays directions and distance to the next junction.**

**City Generation.** Manually creating a large-scale virtual environment, such as a city with road junctions, buildings, and traffic signals, is a time-consuming task. For the generation of the simulated city we used CityEngine, a city generation software developed by Esri. It features rule-based geometry generation and offers highly-customizable models of buildings and roads. In a first step, we defined a road network graph containing nodes that are connected by road segments. Each road segment holds information about the number of lanes, the lane width. as well as the sidewalk width. Additional user-defined properties are added to define the maximum allowed speed on each lane and a flag indicating which lane goes in which direction. Once the road network has been defined, CityEngine subdivides areas enclosed by streets into footprints of buildings. It generates the building models and road geometry, including sidewalks. All generated models can be exported into a variety of 3D formats, including FBX, which are then imported into Unity.

For the driving simulator, we required more control over the exact shape and textures of the roads, especially at junctions. We therefore decided to not export the street geometry but to dynamically generate roads in Unity. A Python interface provides access to the coordinates and other attributes of all nodes and segments defined in the road network. With an export script, we write all information that is required to construct the roads into an XML file. An import script in Unity reads the exported file. In a first step, the road graph is created from the information contained in the node and segment tags. Then, the road geometry is created along the shape of the road segment and overlaid with asphalt textures containing road markings. Where two road segments join, the geometry is aligned so that there is no gap. In addition to the road geometry, sidewalks are generated along both borders of the road. As on real streets, the sidewalks are rounded on corners to enable proper turning at junctions. Fig. 4 shows a junction generated based on a CityEngine export. The complete city is shown in Fig. 5.

**Performance.** The city we created contains nearly 9 kilometers of roads, 519 buildings, and 40 cars that drive around the streets. Using the following optimization techniques, we achieved a framrate constantly above 60 Frames per Second (FPS). The number of draw



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

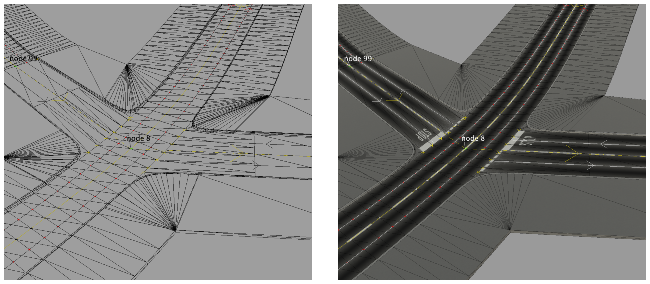


**Figure 4: Road geometry with and without textures applied. The yellow lines mark the center of the road segments connecting the nodes. The lanes are symbolized with a white line and the direction of each lane is indicated by the arrows.**

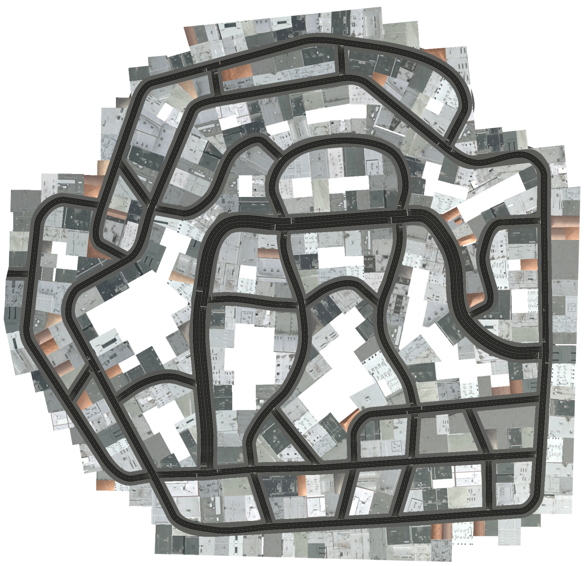


**Figure 5: Imported city with 8896 m of roads, sidewalks, and 519 buildings.**

calls could be drastically reduced with Unity's built-in occlusion culling. The physics calculation for other cars is only activated when they are closer than 60 meters to the driver's car. Communicating with the motion platform and calculating the shortest path to the next target are handled in separate threads to not block the main thread.
The 90 FPS refresh rate of the HTC Vive caps the frame rate of the simulation.

## 4 AI AND ADAPTIVE LEARNING

With a working environment of a driving scenario in place, we extended the simulation with a traffic simulation of Artificial Intelligence (AI)-controlled cars and implemented an adaptive learning system to train the driver.

### 4.1 AI Cars

In order to simulate lifelike behavior of other road users, the system must know where, when and how fast the computer generated cars drive. In this simulation, other cars are required to be able to follow a lane, automatically accelerate and brake, respect each other's right of way, and indicate where they go using their turn signal. With precise information about the position of each lane, it is easy to have other cars just follow the road. In order to allow AI cars to correctly handle turning at junctions, we extended the lane information by junction segments, connecting incoming and outgoing lanes at each intersection. Each junction segment is then assigned a unique priority, specifying which car can go first. While driving along the lanes and turning at junctions, all computer controlled cars obey the following rules:

**Do not exceed the allowed speed.** When driving, the cars accelerate up to a speed of 50 km/h, the allowed speed in the city.

**Keep enough space to the car ahead.** The gap between the cars is always big enough to safely come to a stop when the driver in front suddenly brakes. It is calculated from the car's current velocity and the configured braking acceleration.

**Respect right of way.** Before crossing a junction, yield to other road users that have the right of way.

**Complete stop.** Respect the same rules as on natural intersections, but come to a complete stop before driving onto the junction.

**Stop at red lights.** When on a red light, stop behind the signalization. After the light turns green, drive onto the junction but respect other vehicles who have the right of way (e.g. when turning left, yield to oncoming traffic).

### 4.2 Intelligent Tutoring System

The ability to drive a car involves skills in a set of activities. Clement et al. [3] show that the order in which activities are trained has an impact on the overall learning progress. Their proposed algorithm, ZPDES, optimizes the activity sequence based on continuous evaluation of a driver's skills. Activities are organized by exercise type and difficulty level. The Zone of Proximal Development (ZPD) defines a subset of activities that are expected to improve the user's skills when being trained. ZPDES updates this set based on the evaluation of each performed activity and selects the next activity with the highest expected learning progress.

In this section, we will show how we created an ITS by adapting the the ZPDES algorithm to the task of car driving, how a driver's skill in various activities is continuously tested, and how the activities chosen by the ITS are presented.

**Activities.** During a discussion with a professional driving instructor, we assembled a list of abilities that define a good driver. While we agreed that automatically deciding whether a person is proficient in car driving is not possible in an artificial environment, we were able to define a subset of these abilities that make sense to be trained on a driving simulator as they can be tested under controlled conditions and evaluated by the system:

**Stable driving (on straight roads).** The driver maintains a stable track with only little variance in the distance to the center of a straight lane.
*Evaluation:* When on a lane segment, the distance between the car's position and the closest position on the lane segment is recorded every meter. The recording starts a certain distance $d$ after the start node of the segment and ends $d$ before the end node to ignore deviation from the lane caused by turning. After enough samples have been recorded, a score between 0 and 1



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

is given based on the variance of the samples (lower variance yielding a higher score).

**Stable driving (on curved roads).** This activity has the same objective and uses the same technique of evaluation as the *Stable driving (on straight roads)* activity but is tested on curved roads for increased difficulty.

**Turning.** Execute all steps required to properly turn the car at a junction (check mirrors, look over shoulder, set indicator). *Evaluation:* When approaching a turn, check head rotation and indicator state. A score of 1 is rewarded when all steps have been executed. Failing to set the indicator reduces the score by 0.5, not checking the mirror by 0.3, and a missing shoulder check by another 0.2.

**Complete stop.** Bring the vehicle to a complete stop at a stop sign. *Evaluation:* When passing a junction from a road signaled with a stop, the vehicle must be completely stationary within a certain distance before the stop line. This activity is graded in a binary way, yielding either 1 or 0 points.

**Constant speed (without elevation).** The driver maintains constant speed throughout a lane segment on an even road. *Evaluation:* In a fixed time interval $t$, the car's velocity is recorded. If the speed needs to be reduced due to traffic driving slower, the recorded sample is discarded. Similar to the stable driving activity, the variance is calculated to give a score between 0 and 1. If the end of the lane segment is reached without collecting enough samples, the activity is aborted and not scored in order to prevent incorrect evaluation.

**Constant speed (on up or downhill roads).** This activity has the same objective and uses the same technique of evaluation as the *Constant speed (without elevation)* activity but is tested on either ascending or descending roads.

**Reaction.** React to a vehicle unexpectadly crossing the driver's path. *Evaluation:* A computer controlled car is positioned at a junction crossing the way of the driver's vehicle. The car then pulls out to provoke a collision if the driver does not react quickly. The score is calculated from the time between the other car starts moving and the moment the driver hits the brakes.

**Adaptation of the ZPDES Algorithm.** In a first step, the activities are structured into the activities graph, as depicted in Fig. 6, which orders them by exercise type and difficulty level. The two stable driving activities, as well as the constant speed activities, are connected as they are activities of the same type. The second version of each is considered more difficult, which is why they are positioned at a higher difficulty level in the graph. Amongst all other activities, the difficulty level is the same. When the simulation starts, all activities of the lowest difficulty level are included in the ZPD while the more difficult ones are excluded.

**Selecting the Next Activity.** Based on the result of previously solved activities, ZPDES suggests the next activity to be chosen. Each activity can be tested at various locations on the map. In order to complete as many activities as possible within a given time, the closest instance of the activity should be found. Given an activity and the current position of the driver, the distance of the shortest path to each activity instance is determined using Dijkstra's shortest path algorithm [4] and the one with the minimal distance is set as the target on the virtual satnav. Once the path to the chosen activity is known, all activity instances on the way can be tested, gaining additional information about the driver's performance.

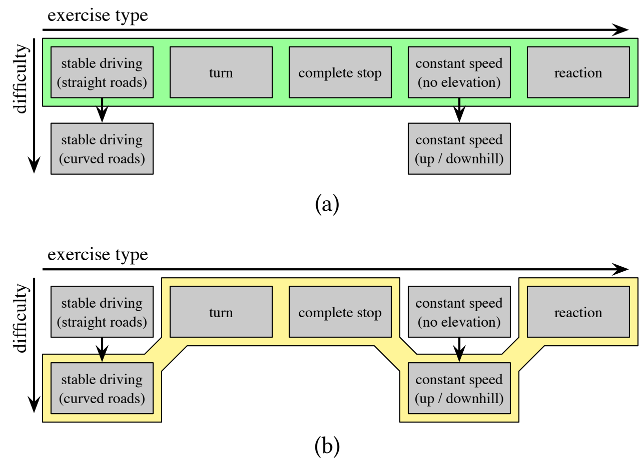



**Figure 6: Activities graph with initial ZPD (a) and updated ZPD where the more complex activities have been activated (b).**

# 5 EVALUATION

In order to evaluate the driving simulator, we conducted a user study. The goal was to determine if the overall quality is high enough, to validate if the simulator can be used for future experiments, and to collect user feedback for possible improvements.

**Experiment Design.** We invited a total of 17 participants from various departments, as well as people not related to our facility. All of them had a valid driving license and were familiar with the Swiss traffic rules. 5 (29.4%) of the participants were female and the average age was 29.5 years (SD: 8.3 years). After being instructed how to operate the vehicle, the participants were asked to wear the HMD and drive through the virtual city for 15 minutes, following the directions given by the satnav. A simulator sickness questionnaire [5] was filled in before and after the driving session. A presence questionnaire [11] was answered after the test run. The drivers were told to immediately abort the experiment as soon as they experience any kind of discomfort.

## 5.1 Results

4 participants (23.5%) aborted the run due to symptoms of simulator sickness. The rest managed to complete the 15 minutes run without experiencing major discomfort.

**Presence.** In a total of 22 questions, the participants indicated how strongly they experienced presence in the simulator by giving values between 1 (not at all) and 7 (completely). These questions were then evaluated and converted into a scoring scheme in 7 subscales as shown in Fig. 7. The overall presence is on a satisfactory level. Of the measured criteria, the quality of the *sound* effects yields the lowest score, which we figured we can improve in a future version. A relatively high score on the *possibility to act* subscale suggests that the simulated behavior of the car resembles a real-life car to a level that enables users to control the vehicle through the virtual scene.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

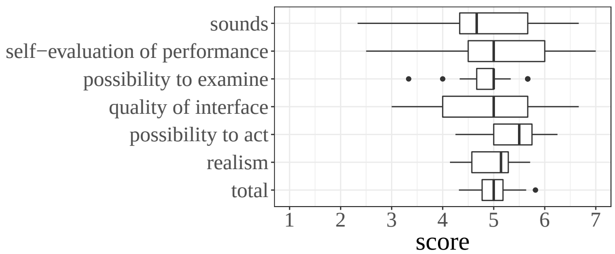



**Figure 7: Evaluation of the presence questionnaire (N = 17). A score between 1 and 7 is calculated for each of the subscales. The median is marked with a bold black line. The boxes show the 25th and 75th percentile. The whiskers are limited at 1.5 IQR and outliers above or below are symbolized by black dots.**

**Simulator Sickness.** The participants indicated how strong they experience discomfort by assigning values (0: none, 1: slight, 2: moderate, 3: severe) to 16 symptoms. The same set of questions was asked before and after the test run. Evaluating the simulator sickness questionnaire summarizes symptoms on three subscales: Nausea, oculomotor disturbance, disorientation, and a total score. Fig. 8 shows the results of the questionnaire before and after the test subjects participated in the driving activity. The most considerable increase in discomfort is reflected on the *nausea* subscale. This can be explained by the fact that, with our hardware setup, real acceleration cannot be experienced. The acceleration imitated by tilting the platform pretends that the user is moving but a discrepancy between the visual representation and the physically perceived movement remains, which for some people leads to feeling nauseous.

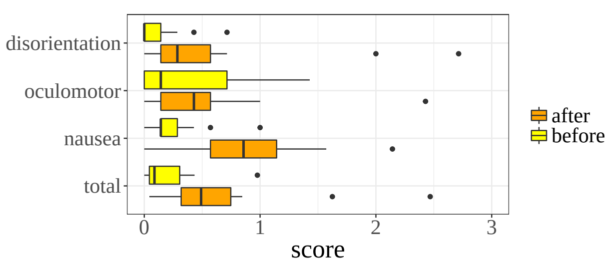



**Figure 8: Evaluation of the simulator sickness questionnaire (N = 17). A score between 0 and 3 is calculated for each of the subscales before (yellow) and after driving (orange). The median is marked with a bold black line. The boxes show the 25th and 75th percentile. The whiskers are limited at 1.5 IQR and outliers above or below are symbolized by black dots.**

**Lessons Learned.** We interpret the results of the presence questionnaire as an indication that the behavior of our simulation resembles a real-life car to a level that enables users to control the vehicle through the virtual scene without major difficulties. The simulator sickness questionnaire revealed that driving through the virtual environment leads to a slight increase in discomfort for test runs below 15 minutes. Future experiments should therefore be designed to last only for short periods of time. Applied to the objective of improving car driving skills in a virtual reality environment, this suggests a setup with a series of many shorter training sessions rather than few long ones.

## 6 CONCLUSION

We have shown how VR technologies can be applied to create an immersive car driving experience. We explained how physical properties influencing the driving characteristics of a car can be simulated, and presented in a way to model the behavior of a car's engine and transmission system. Our software connects to a 6 DOF motion system to simulate acceleration while driving, and queries input devices in the cockpit mockup, controlling the virtual car. AI cars drive through the city, follow the existing traffic rules, and interact with each other, as well as with the user's car. We presented five different types of driving-related activities that can be trained and automatically evaluated through an ITS. By adapting the ZPDES algorithm for car driving, we have shown how a personalized teaching sequence can be generated. Challenges involving limited computational performance, integration of motion hardware, and efficiently simulating city-wide traffic could be addressed. A user study revealed that most users experience a good level of presence in the virtual world and are proficient in operating the car on the VR driving simulator.

**Future Work.** The information gathered from the user study provides useful information when setting up further experiments. With a future user study, we aim to evaluate how strong the ITS' effect is on the learning progress. The content generation framework we provided can be used to create driving environments tailored to specific requirements posed by future research in VR car driving.

## REFERENCES


[1] S. C. Augusto, A. C. A. Mol, P. C. Mol, and D. S. Sales. 2009. Using Virtual Reality in the Training of Security Staff and Evaluation of Physical Protection Barriers in Nuclear Facilities. *International Nuclear Atlantic Conference* (2009).

[2] D. R. Berger, J. Schulte-Pelkum, and H. H. Bülthoff. 2007. *Simulating believable forward accelerations on a Stewart motion platform.* Technical Report 159. Max Planck Institute for Biological Cybernetics.

[3] B. Clement, D. Roy, P.-Y. Oudeyer, and M. Lopes. 2015. Multi-Armed Bandits for Intelligent Tutoring Systems. *Journal of Educational Data Mining (JEDM)* 7, 2 (2015).

[4] E. W. Dijkstra. 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1, 1 (1959), 269–271.

[5] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220.

[6] O. Kreylos. 2016. Analysis of Valve's 'Lighthouse' Tracking System Reveals Accuracy. Available online at https://www.roadtovr.com/analysis-of-valves-lighthouse-tracking-system-reveals-accuracy/. (2016).

[7] B. Lang. 2017. New HTC Vives Weigh 15Launch. Available online at https://www.roadtovr.com/htc-vive-weight-15-percent-lighter-than-original-headset-vs-oculus-rift-comparison/. (2017).

[8] K. Reif. 2014. *Fundamentals of Automotive and Engine Technology.* Springer Fachmedien Wiesbaden. 15–21 pages.

[9] T. Shibata and H. Fujihara. 2006. Development of Railway VR Safety Simulation System. *Quarterly Report of RTRI* 43, 2 (2006), 87–89.

[10] Digital Trends Staff. 2017. Spec Comparison: Does the Rift's Touch Update Make it a True Vive Competitor? Available online at https://www.digitaltrends.com/virtual-reality/oculus-rift-vs-htc-vive/. (2017).

[11] B. G. Witmer and M. J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *U.S. Army Research Institute for the Behavioral and Social Sciences* 7, 3 (1998), 225–240.

[12] P. Zal. 2015. 2015 Fiat 500 1.2 engine Horsepower / Torque Curve. Available online at http://www.automobile-catalog.com/curve/2015/2182295/fiat_500_1_2.html. (2015).




In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# Combining Intelligent Recommendation and Mixed Reality in Itineraries for Urban Exploration

Giulio Jacucci[1,3], Salvatore Andolina[1,2], Denis Kalkhofen[5], Dieter Schmalstieg[5], Antti Nurminen[2], Anna Spagnolli[4], Luciano Gamberini[4], Tuukka Ruotsalo[1,3]

[1]Helsinki Institute for Information Technology HIIT

[2]Department of Computer Science, Aalto University, Finland
name.surnanme@aalto.fi

[3]Department of Computer Science, University of Helsinki, Finland
name.surnanme@helsinki.fi

[4]Human Inspired Technologies Research Centre, University of Padova, Italy
name.surnanme@unipd.it

[5]Institute of Computer Graphics and Vision (ICG) Graz University of Technology, Austria
name.surname@icg.tugraz.at

## ABSTRACT

Exploration of points of interest (POI) in urban environments is challenging for the large amount of items near or reachable by the user and for the modality hindrances due to reduced manual flexibility and competing visual attention. We propose to combine different modalities, VR, AR, haptics-audio interfaces, with intelligent recommendation based on a computational method combining different data graph overlays: social, personal and search-time user input. We integrate such features in flexible itineraries that aid different phases and aspects of exploration.

## CCS CONCEPTS

**• CCS → Human-centered computing → Human computer interaction (HCI) → HCI theory, concepts and models**

## KEYWORDS

Intelligent recommendation, Augmented Reality, Urban Exploration , Multimodal Interaction

## 1 INTRODUCTION

Urban environments offer copious points of interest (POI) comprising sites, services, or cultural artifacts which are distributed in space and can also be encountered as dense assemblies in particular areas. This offers great opportunities for personalised exploration but also the challenges. Firstly considering the choice of interaction modality aggravated by limited manual flexibility and competing visual attention for safety. Moreover, the large amount of POI near the user or potentially reachable, suggests personalized recommendations which can make use of the extensive social data available on the web, e.g., tags, ratings and reviews. We propose to combine multimodal AR and intelligent recommendations for a new generation of urban exploration systems (Figure 1). The approach has been prototyped in an actual system integrating different modalities, such as virtual reality (VR), augmented reality (AR), and haptics-audio interfaces, as well as advanced features. Intelligent recommendations use a computational method combining different data graph overlays: social, personal and search-time user input. The integration of multimodality, MR and intelligent recommendations is synthesized in flexible itineraries that aid planning, serendipitous discovery and wayfinding.

## 2 REQUIREMENTS FOR THE URBAN EXPLORER

### 2.1 Facilitate serendipity

Most touristic applications are based on finding the route from visitor position to a particular Point of Interest (POI), e.g., a museum, restaurant, or church. Such an approach usually relies on geographical visualization of the environment (such as a 2D map) complemented with a route and POI list. This information is suitable for wayfinding, but less for the broader issues of exploration, like getting an overview or choosing a worthwhile destination. Facilitating serendipity comes a significant issue in designing for exploration rather than wayfinding. Users should be allowed to have significant experiences by chance. One possible way to facilitate the process is to provide users with an extended awareness of the surroundings by delivering cues, either proactively or by responding to explicit user requests.

### 2.2 Connect the user to the surroundings



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

While exploring a place, the user's focus of attention should be on the surroundings. Virtual information (e.g., 2D maps or POI lists) redirect the user's attention to the device screen and spare little cognitive resource in exploring the surroundings. To avoid safety concerns and a possible negative impact on the quality of experience, technology for urban exploration needs to redirect the attention of the visitor from the device to the surroundings. Multimodal displays provide the necessary redundancy so that the user can remain connected to the surroundings, while receiving visual, haptic, or aural cues from the system.

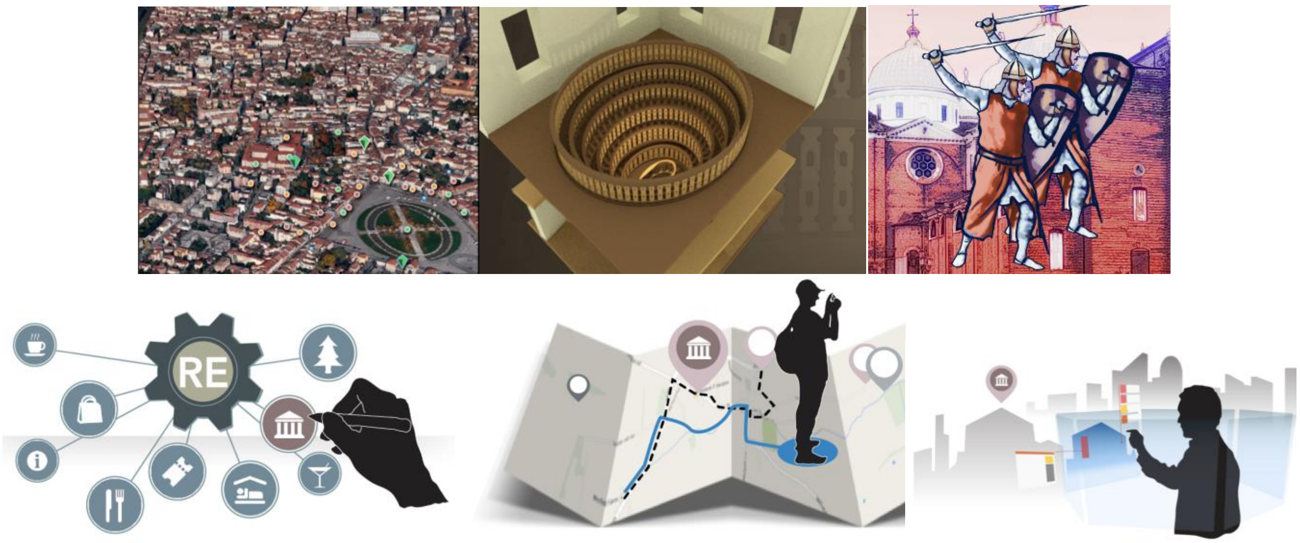



Figure 1: **Above example of cultural content in Padova, Italy, including POIs of different categories copiously distributed around the city with several dense areas, models, narratives of interiors or exterior artefacts such as the famous anatomical theater, or the history rich Abbey of Santa Giustina. Below: core support components responding to requirements of the explorer of urban environments with dense POIs: personalized recommendation, itinerary support, multimodal interaction responding to mobile and situational hindrances.**

## 2.3 Personalize Recommendations and Search

A rich urban environment can contain a large number of POIs. Yet, identifying relevant information is challenging with common search tools, resulting in an impoverished experience. For example, the same top-rated POI would be pushed to all nearby users. It is crucial to combine search engines with personalized recommendations based on the user profile and contextual information. A technology for urban exploration needs to retrieve information that matches user interests and external factors (e.g., opening hours, weather forecast, etc..) with the POI available in a particular place. This requires a model of users and their context, select the most appropriate content, and deliver it in the most suitable way (Ardissono et al., 2012). We also see a need to acknowledge variability in the visitors' profile. Users should be allowed to transfer from one profile to another one, rather than being confined to a static (or not easily changeable) profile.

## 2.4 Respect perceptual requirements

Multimodal interaction methods aim at improving human-computer communications by utilizing all available modalities of human input and output, in a natural manner (Turk 2014). Multimodality can free cognitive resources from the device toward the environment. This is especially important in a mobile setting (Lemmelä et al. 2008). Depending on the situation, users can choose the modality providing the most suitable level of safety, obtrusiveness, social acceptance, level of detail, and so on. For example, the visual modality may provide more detailed information, but requires more cognitive load. Multimodal access also makes information more accessible for disabled people. However, a user augmented with an interface supplying information through multiple modalities may easily be overwhelmed by information overload, or may fail to acknowledge information which is presented in a way that does not meet perceptual requirements. Occlusions on the screen, or clutter in general (including auditive and haptic jamming) defeat the objective of augmentation. Perception management must include strategies that avoid such situations. For visual displays, view management is required to ensure visibility of both real objects to relocate POIs to inconspicuous regions of the display (Bell at al., 2001). Multimodal interfaces need to generalize across multiple display dimensions (visual, audio, haptics) and decide where and how information items should best be displayed. These presentation management tasks require continuous optimization and can be computationally demanding.

# 3 COMPONENTS FOR URBAN EXPLORATION



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

## 3.1 Personalized Social Discovery


**Figure 3: Itinerary building. Recommended POIs can be selected to create a custom path.**

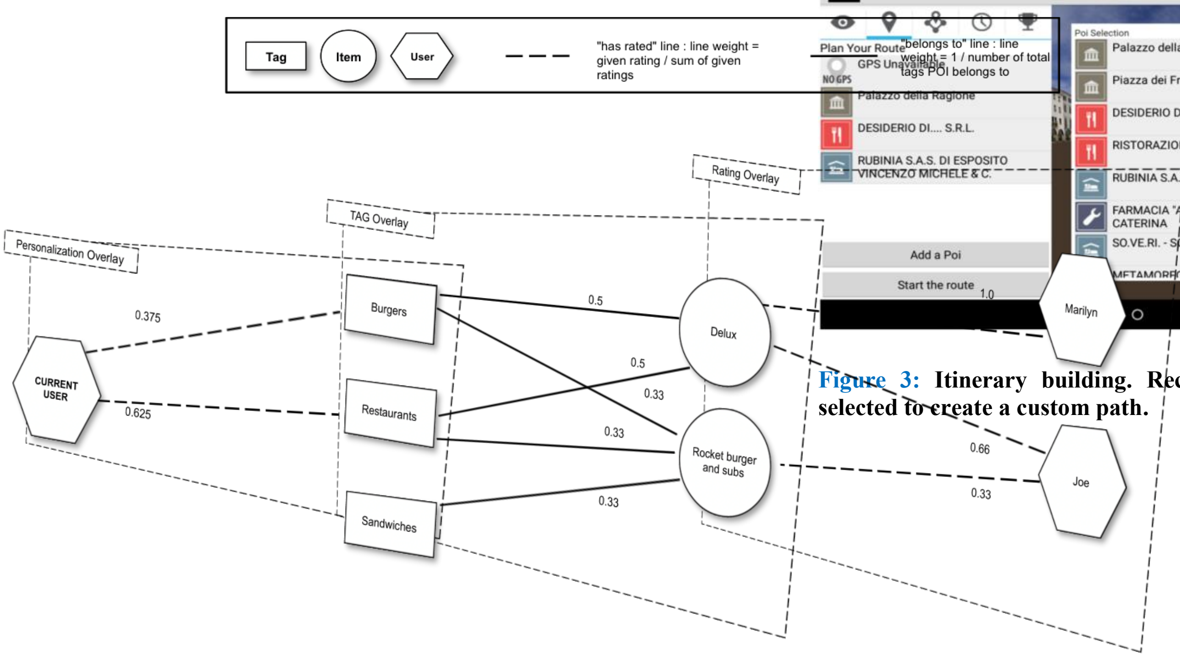



**Figure 2: Social web data are modeled as a set of overlaid graphs composed of social tags, social ratings, users' personal preferences, and search-time preferences observed from the interaction with the system. Each of these data are connected to the retrievable items on a graph overlay.**

We used two criteria for generating recommendations, one based on social rating and the other based on the visitors' profile. Social rating, as used in Tripadvisor or Booking.com, is a Web 2.0 technology, which analyzes the experience of other visitors to single out highly rated POI. Its credibility relies on two aspects: First, the information is neutral, since the ratings come from peer visitors. Second, reliability increases with the number of collected ratings, because unfair ratings are diluted with increasing sample size. Alas, the reason for certain destinations becoming highly popular are obscure and may be owed to reasons that do not concern the inquiring visitor, especially, if visitor deviates from the evaluation group.

Therefore, recommendations in our system are always considering the user's profile. Weighting by both crowd-sourced ratings and by personal preferences better facilitates serendipity, since recommendations can be based on either. Users may also turn off either ratings or profiles (or both), allowing unfiltered access to all POI information.

Our technique for personalized search and recommendations is based on three main parts (Orso et al. 2017): a data model that defines the representation of multiple data sources (content, social, and personal) as a set of overlaid graphs (see Figure 2), the relevance-estimation model that performs random walks with restarts on the graph overlays and computes a relevance score for information items, and the user interface (Figure 2).

## 3.2 Flexible Itineraries

Recommended POIs as delivered by the component in the previous section can be flexibly managed in planning and during visits providing support for situational plan changes and wayfinding (see Figure 3 for the UI where the user can choose recommended POI for an itinerary). In order to do some planning and optimize the visit, the tourist plans a path connecting the different POIs that are retrieved by the recommendation engine, creating a customized itinerary. The itinerary is built with the help of a routing service (e.g., Nokia Here). However, regular way-finding is replaced by a joint space+time optimization concerning the recommended POIs: Which set of POIs could be visited this morning? Which area of the city should be visited today? Alternatively, itineraries curated by professional tour guides can be retrieved, which highlight a topic of the urban environment (e.g., religious places, art museums, famous nightlife places or restaurants) or even tell a story (e.g., life of Galilei). Both types of itineraries do not force the visitor to follow the suggested path. Instead, the itinerary is merely an initial suggestions, from which the visitor can divert and return to afterwards. Visual or haptic cues regarding the location of the next POI to be visited can be provided as requested by the visitor.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

## 3.3 Wearable and haptic guidance

Deploying aural and tactile channels as a complement to a visual channel into the system would benefit the users in terms of safety, while enhancing awareness of the surroundings. Touch based visual interfaces have presented high responsiveness and enabled integrated input and output concentration on the same surface. However, this comes at the cost of drawing much visual attention of users onto the interface and would result in ignoring surrounding environment, which could raise concerns in the context of designing for urban exploration. Eyes-free interaction could be an alternative to visual interfaces, releasing the visual attention back to the environment and increasing awareness.

The advance of wearable technologies enables increasing awareness of the surrounding as well as a more engaged experience through interaction design. For example, a vibrotactile vest could offer a hands and eyes free navigation, leveraging both location on the body and distinctive vibration patterns to indicate cues such as direction, degree of turn, speed or user error. This literally enables embodied interaction with the environment. The wearers are steered towards their chosen POIs, prompted in a ‘natural-enough’, pleasant and easily understood manner. Similarly, a headset capable of 3D audio could not only provide essential textual information, but also spatial information of the environment, such as distance, increasing or decreasing proximity and direction of a POI as well as category of POI (e.g., cultural, shop or service). Moreover, a sensor and actuator equipped glove provides hand gesture recognition for direct interaction with the real environment, such as selection via pointing. While a POI is distant and untouchable, it could be represented through the vibration on the hand. The non-visual multimodal interaction is complemented visually by the real scene seen by the user. Data from a comparative study with a context-aware mobile app has shown that, while experiencing similar performance of different evaluation metrics, smartphone users spent in average 70% of time looking at the screen when exploring the urban area, while users wearing our haptic glove (Figure 4) were able to have a good exploration experience while leaving their visual attention on the surroundings (Jylhä et al., 2015).

## 3.4 Augmented Reality of Urban Artifacts

Mobile AR browsers have recently become very popular (Grubert et al., 2011). They commonly augment an urban POI using GPS in combination with the orientation sensor of modern smartphones. Although this approach works well for distant POI locations surrounding the user, the applications often suffer from an imprecise user localization for artifacts in the close proximity of the user.

Therefore, we have built an image-based localization pipeline, which is able to precisely estimate the camera pose of the user’s AR device. Our implementation includes a localization system based on an analysis of visual features and an incremental tracker. The goal of our system is to localize a query image (the current video frame) within a known 3D world. We represent the world using a 3D point cloud and we accurately identify the 6-degrees-of-freedom pose (i.e. position and orientation) of the

**Figure 4: Left the haptic glove; and right the audio-haptic exploration, in which a user points at a POI with the haptic glove and listens the associated audio description.**

acquired camera image with respect to the 3D world model by searching for corresponding points within the 3D model and the 2D camera image.

On top of the localization system, we have developed a fast incremental 3-degrees-of-freedom tracker in order to update the user’s pose after its initial localization. For AR in urban environments, typical scenarios use rotational movement only (i.e., panning the device around while staying in place). Combining the absolute 6-degrees-of-freedom camera pose (which we retrieved from the localizer) and relative 3-degrees-of-freedom camera rotation yields a full 6-degrees-of-freedom pose for every camera frame.

We also reliably suppress clutter by restricting the maximum annotation density on the display. If the POI list from the recommendation engine is too long and does not comfortably fit on the screen, conventional view management works by just omitting low-ranking annotations, making serendipitous discovery impossible. Adaptive view management (Tatzgern et al., 2016) folds similar annotations of low importance into a single “group” POI, which initially takes less space, but can be unfolded by the user on demand (Figure 5).

The AR visualization points the user to relevant objects which he or she can select for detailed information. A green icon indicates that animated content is associated with the astronomical clock. Next to the center, an abstract visualization indicates a cluster of search results, including the overall amount and the type encoded in color. The reticle in the center of the screen can be used to aim at POI icons by moving the tablet. When the reticle matches with a POI, users can tap with the thumb on the target button on the edge of the screen to confirm selection. In the case of Figure 4 the user can activate an animation on the . astronomical clock (Figure 6).



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal


Figure 5: **Mixed reality cues and affordances through computer-vision-based AR provide visual augmentations of a POI.**

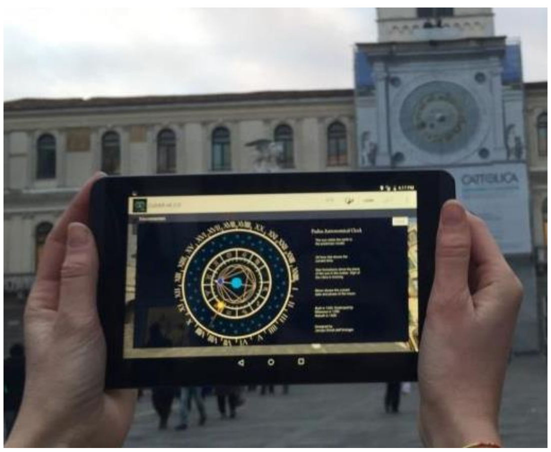


Figure 6: **The animated explanation of the astronomical clock activated from the AR interface in Figure 4.**

## 3.5 3D Engine and Virtual Reality.

As an alternative to AR we provide a virtual reality view of the environment. The AR view has the fundamental property of being egocentric. On one hand, it provides intuitivity due to the short cognitive distance between the real view and the view on the screen—they are always aligned (see section Augmented Reality of Urban Artifacts for AR Content Registration). However, it also poses limitations. Augmented content is overlaid on top of video feed with no capability to resolve real world occlusions, such as people walking by. A label that is supposed to be on a building facade will hang in front of everything. This contradicts human depth perception, where occlusions provide the strongest depth cue. Similarly, content that belongs to the far side of a building would be also visible. Indeed, incorrect depth interpretation is the most common perceptual problem in AR applications. In our system, the second case is handled by an online visibility test hosted in the client device, utilizing our 3D city model as a virtual occluder. For the first case, we offer an alternative.

AR's egocentric view does not allow easy virtual exploration or visual navigation planning. Our realistic 3D city model compensates for this, allowing free motion of the viewpoint in the full 3D environment (Figure 7). Furthermore, the default interaction mode for the 3D map follows the AR pointing metaphor: the initial view position and orientation are aligned with the position of the user and the orientation of the device. Users can choose either AR or 3D representation, with a smooth transition. The 3D map implementation is based on the m-LOMA mobile 3D city map engine, utilizing visibility preprocessing, level-of-detail management, and temporal coherence for optimal resource usage (Nurminen, 2008).

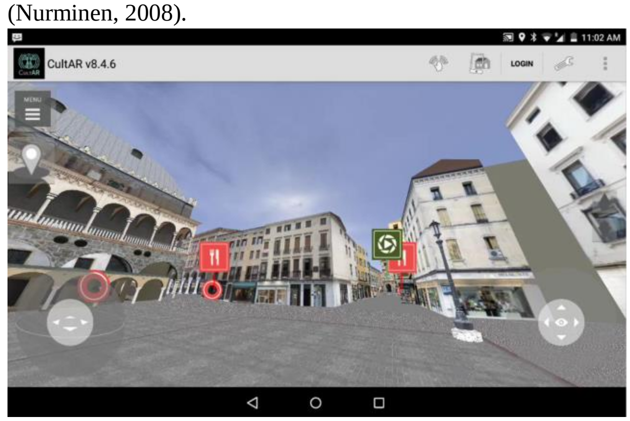



Figure 7: **Virtual Exploration.**

## 4. CONCLUSIONS AND WORKSHOP INPUT

In the approach presented we demonstrated the combination of machine learning and Mixed Reality in exploring urban environments. In particular we show how a recent machine learning technique utilizing social media data can support personalization for recommendation and search in mobile VR and AR in urban settings. The approach has been evaluated only in partial experimentation with separate components , however it demonstrates important opportunities in benefiting from big data such as social media reviews, ratings and tags for application in MR. Recently important investment have been made in big data including ambitious promises of benefits. However only recently approaches have started to (e.g. cognitive big data Lugmayr et al. 2017).

According to estimates, our digital universe is growing at increasing speed. From 2013 to 2020, it is expected to grow from 4.4 trillion gigabytes to 44 trillion by a factor of 10 (EMC Digital Universe with Research & Analysis by IDC). Not only does it seem to more than double every two years, but the proportion of analyzed and used data, which is merely 20% now, will double to 40% by 2020.

Employing machine learning and making use of Big data includes a variety of challenges and opportunities for human-computer interaction. In particular it is interesting to investigate what interaction techniques are particularly interesting or should be researched in mining, analyzing, searching and exploring data? These might include, for example, implicit (Eugster et al 2013, Barral et al 2017, subliminal (Zhu et al 2016, Aranyi et al 2016) or



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

gestural and multimodal interaction (Zhu et al. 2016, Jylhä et al 2015).

The approach taken has focused on using machine learning to personalize recommendations using social media thanks to the integrative role flexible itineraries accessible in VR and AR interfaces. We believe this is however only one possible instance of utilizing machine learning and big data in MR for urban exploration.

Other large cultural data sets including pictures, videos, 3D models and texts (Lugmayr et al. 2016) can be exploited to enhance cultural experiences through machine learning techniques.

## ACKNOWLEDGMENTS

This work was partially funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No 601139 (CultAR).

## REFERENCES

[1] Aranyi, G., Kouider, S., Lindsay, A., Prins, H., Ahmed, I., Jacucci, G., ... & Cavazza, M. (2014). Subliminal cueing of selection behavior in a virtual environment. PRESENCE: Teleoperators and Virtual Environments, 23(1), 33-50.

[2] L. Ardissono, T. Kuflik, and D. Petrelli, “Personalization in cultural heritage: the road travelled and the one ahead,” User Modeling and User-Adapted Interaction, Vol. 22, no. 1, 2012, pp. 73-99.

[3] Barral, O., Kosunen, I., Ruotsalo, T., Spapé, M. M., Eugster, M. J., Ravaja, N., ... & Jacucci, G. (2016). Extracting relevance and affect information from physiological text annotation. User Modeling and User-Adapted Interaction, 26(5), 493-520.

[4] B. Bell, S. Feiner, and T. Höllerer, “View management for virtual and augmented reality,” Symposium on User Interface Software and Technology, 2001, pp. 101–110.

[5] Eugster, M. J., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., ... & Kaski, S. Predicting Term-Relevance from Brain Signals.Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14). ACM, New York, NY, USA, 425-434.

[6] J. Grubert, T. Langlotz, R. Grasset (2011). Augmented reality browser survey, Technical Report 1101, Institute for Computer Graphics and Vision, University of Technology Graz, 2011

[7] Jylhä, A., Hsieh, Y. T., Orso, V., Andolina, S., Gamberini, L., & Jacucci, G. (2015, November). A Wearable Multimodal Interface for Exploring Urban Points of Interest. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 175-182). ACM.

[8] A. Lugmayr, B. Stockleben, C. Scheib, and M. Mailaparampil, “Cognitive Big Data. Survey and Review on Big Data Research and its Implications: What is Really ‘New’? Cognitive Big Data!,” Journal of Knowledge Management (JMM), vol. 21, no. 1, 2017.: http://www.emeraldgrouppublishing.com /products/journals/ call_for_papers.htm ?id=5855kÝ, Journal of Knowledge Management/Emerald

[9] A. Lugmayr, A. Greenfeld, A. Woods, and P. Joseph, “Cultural Visualisation of a Cultural Photographic Collection in 3D Environments – Development of ‘PAV 3D’ (Photographic Archive Visualisation),” in Entertainment Computing - ICEC 2016: 15th IFIP TC 14 International Conference, Vienna, Austria, September 28-30, 2016, Proceedings, G. Wallner, S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, and H.-S. Yang, Eds. Cham: Springer International Publishing, 2016, pp. 272–277

[10] Nurminen, A. 2008. Mobile 3D City Maps. IEEE Computer Graphics and Applications 28, 4 (2008), 20–31.

[11] Orso, V., Ruotsalo, T., Leino, J., Gamberini, L., & Jacucci, G. (2017). Overlaying social information: The effects on users’ search and information-selection behavior. Information Processing & Management, 53(6), 1269-1286.

[12] S. Lemmelä, A. Vetek, K. Mäkelä, and D. Trendafilov, “Designing and evaluating multimodal interaction for mobile contexts,” Proc. 10th international conference on Multimodal interfaces, (ICMI '08), 2008, pp. 265-272.

[13] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg, “Adaptive Information Density for Augmented Reality Displays,” In Proc. IEEE Virtual Reality, November 2016.

[14] M. Turk, “Multimodal interaction: A review,” Pattern Recognition Letters, Vol. 36, 2014, pp. 189-195.

[15] Zhu, K., Ma, X., Chen, H., and Liang, M. (2016). Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures under the Influence of Operation, Handheld Posture, and Interaction Space. International Journal of Human-Computer Interaction.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# Interacting with Intelligent Characters in AR

Gokcen Cimen[*] ETH Zurich
Ye Yuan Carnegie Mellon University
Robert W. Sumner Disney Research
Stelian Coros ETH Zurich
Martin Guay Disney Research

**Figure 1:** *Our intelligent virtual characters can navigate real world environments (right) and react to objects that collide with them (left).*

## Abstract

In this paper, we explore interacting with virtual characters in AR along real-world environments. Our vision is that virtual characters will be able to understand the real-world environment and interact in an intelligent and realistic manner with it. For example, a character can walk around un-even stairs and slopes, or be pushed away by collisions with real-world objects like a ball. We describe how to automatically animate a new character, and imbue it's motion with adaption to environments and reactions to perturbations from the real world.

**Keywords:** Character animation, locomotion tracking, procedural character animation.

**Concepts:** •**Computing methodologies** → **Animation;** *Graphics systems and interfaces;*

## 1 Introduction

Augmenting our environment with intelligent virtual characters that can walk around and interact with our environment is an exciting and promising vision. However, achieving this idea represents several technical challenges. It remains a challenge to model the motion of a character, have it understand its environment and navigate the world in a natural way.

In this work, we take a first step in the direction of making a character intelligent, and able to interact in AR. We separate the problem into three main components. First is the modeling of the character's motion and its ability to move around. Given a character's skeleton, how should the joints move in order to go from point A to point B, including on un-even terrain. We describe a parametric model of quadruped locomotion, which we use to fill a blendtree that outputs motions conforming to control directions. The second problem is how to adapt the characters motion to un-even terrains, as well as collisions with objects (such as a ball). We full-fill this by layering on top of the blendtree, an inverse kinematics solver for terrain adaptation, and a physically simulated ragdoll for character-object collisions. The last problem is the character's ability to understand the environment and naviguate it. Online consumer-level scanning of 3D worlds remains inaccurate, and we describe our solution which cleverly combines pre-defined objects with off-the-shelf scanning solutions to provide high-resolution 3D reconstructions of the environment.

Given our intelligent character that can understand the real world and move around, we describe the types of AR interactions we support with real world objects, as well as the experiments we conducted using these interactions.

## 2 Related Work

**AR:** An early example of AR technology is the MagicBook [Billinghurst et al. 2001]. Large markers are integrated into a book's pages, which enable viewing virtual content through VR glasses (and later through mixed reality [Grasset et al. 2008]), based on which page of the book is open. They add various visual and auditory effects to an existing illustrated book to enhance the reader's immersion. The Haunted Book [Scherrer et al. 2008] is a prime example of well-integrated AR content. The camera is mounted on a lamp on the table and the augmented book is viewed through a computer screen. Their focus lies on interaction between the virtual content and the physical book.

Seeing a virtual character walking on different terrains is ordinary in a game environment, but to see it walking on your desk and interacting with the physical objects on the table is not as common. A few games coming with the Hololens ( [Hololens 2017]), have a character navigating the real world, using the device's built 3D scanning technology. However, the accuracy of the interaction is quite low as the scans are of low resolution. From the idea of mixing physical objects with virtual 2D character, Kim et al. developed an interactive interface where a virtual character jumps over the physical blocks which the user can change their position in real-time [Kim et al. 2014]. It doesn't require any special equipment, such as wearable glasses, but a depth sensor and a projector.

**Character Motion:** The current practice for real-time interactive characters is to manually craft motion clips by key-framing in software such as Maya. The motions are then fed into a blendtree that blends the motion [Kovar et al. 2002; Arikan et al. 2003]. The controllers are then given a layer of inverse kinematics to adapt to different terrains, and use short-lived ragdoll dynamics for the character to react to perturbations and body collisions. We follow a similar path, which we describe in detail, but are different in one regard: when a new character comes in, the animator has to craft new motion clips which is a time consuming process that hinders the ability to scale the applications to many different characters. Hence in our paper we describe an automatic way of synthesizing the motion clips, similar to [Kokkevis et al. 1995], [Torkos and van de Panne 1998], [Megaro et al. 2015], [Bhatti et al. 2015].

[*]goekcen.cimen@inf.ethz.ch



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# 3 Character Motion

Our animation model provides a virtual quadruped character the ability to navigate real world environments, and react to objects in it in real-time—given only the character's skeleton as input. The motion model is composed of motion clips (walk straight, left, right, backward, etc) that are blended in real-time. Then an inverse kinematics and short-lived ragdoll retargeting method are layered on top to adapt the motion to terrains and perturbations. We start by describing how we generate motion clips from a skeleton.

## 3.1 Parametric Locomotion Model

We use mechanical simulation, together with characterizations of quadruped motion, to generate locomotion for characters of different shapes and sizes. Internally, our parameterized motion generation system is based on constrained multi-objective optimization. The parameters are what we call the motion plan: a gait pattern (which foot falls at which time), foot height, center-of-mass velocity and rotational velocity. We optimize to match these values together with various regularizers ensuring smooth transitions between clips. Some constraints are implicit, or by construction. To support a wide range of characters, we constrain the skeleton to a known simplified template (see Fig. 2) that has only hinge joints and pre-defined masses. The final stage consists in upscaling the motion from the simplified template to the higher-resolution template (see Figure 2)

### 3.1.1 Parameterization

We use a parametric model of quadruped that are composed of articulated chain like structures, in particular, of serially connected and actuated links. The design parameters $\mathbf{s}$ is used to specify the quadruped morphology, which is given by

$$\mathbf{s} = (l_1, \ldots, l_g, \mathbf{a}_1, \ldots, \mathbf{a}_n, b_w, b_l)\ , \qquad (1)$$

where $g$ is the number of links, $l_i \in \mathbb{R}$ is the length of each link, $n$ is the number of actuators, and $\mathbf{a}_i \in \mathbb{R}^3$ is the actuator parameters. For linear actuators, $\mathbf{a}_i$ defines the 3D attachment points, while for rotary actuators, it corresponds to orientation of axis of rotation. Apart from these parameters that represent the kinematic tree morphology of the quadruped, we use two additional parameters $b_w$ and $b_l$ to represent the physical dimensions of the quadruped body (width and length respectively).

Likewise, the motion parameters $\mathbf{m} = (\mathbf{P}_1, \ldots, \mathbf{P}_T)$ are defined by a time-indexed sequence of vectors $\mathbf{P}_i$, where $T$ denotes the time for each motion cycle. $\mathbf{P}_i$ is defined as:

$$\mathbf{P}_i = \left(\mathbf{q}_i, \mathbf{x}_i, \mathbf{e}_i^1, \ldots, \mathbf{e}_i^k, \mathbf{f}_i^1, \ldots, \mathbf{f}_i^k, c_i^1, \ldots, c_i^k,\right)\ , \qquad (2)$$

where $\mathbf{q}_i$ defines the pose of the quadruped, i.e., the position, and orientation of the root as well as joint information such as angle values, $\mathbf{x}_i \in \mathbb{R}^3$ is the position of the quadruped's center of mass (COM), and $k$ is the number of end-effectors. For each end-effector $j$, we use $\mathbf{e}_i^j \in \mathbb{R}^3$ to represent its position and $\mathbf{f}_i^j \in \mathbb{R}^3$ to denote the ground reaction force acting on it. We also use a contact flag $c_i^j$ to indicate whether it should be grounded ($c_i^j = 1$) or not ($c_i^j = 0$).

### 3.1.2 Motion Optimization

The purpose of motion optimization is to take a quadruped design $\mathbf{s}$ and optmize its motion for user specified task while satisfying certain constraints. We used a cost function $F(\mathbf{s}, \mathbf{m})$ to encode the task specifications. We now describe how $F(\mathbf{s}, \mathbf{m})$ is constructed. To this end, we use a set of objectives that capture users' requirements, and constraints that ensure task feasibility.

**Objectives** We allow the users to define various high-level goals to be achieved by their quadruped designs such as moving in desired direction with specific speeds, different motion styles, etc. To capture the desired direction and speed of motion, we define the following objectives:

$$E_{\text{Travel}} = \frac{1}{2}||\mathbf{x}_T - \mathbf{x}_1 - \mathbf{d}^D||^2\ ,$$
$$E_{\text{Turn}} = \frac{1}{2}||\tau(\mathbf{q}_T) - \tau(\mathbf{q}_1) - \tau^D||^2\ , \qquad (3)$$

where $\mathbf{x_i}$ is the quadruped's COM as defined in eq. 2, $\tau(\mathbf{q_i})$ is the turning angle computed from pose $\mathbf{q_i}$, while $\mathbf{d}^D$ and $\tau^D$ are desired travel distance and turning angles respectively. $E_{\text{Travel}}$ ensures that the quadruped travels a specific distance in desired time, while $E_{\text{Turn}}$ can be used to make a quadruped move on arbitrary shaped paths.

Motion style is highly effected by gait or foot-fall patterns that define the order and timings of individual limbs of a quadruped. We internally define various foot-fall patterns for different motion styles such as trotting, pacing, and galloping. When users select a specific motion style, our system automatically loads the necessary foot-fall patterns, thereby allowing novice users to create many expressive quadruped motions. Motion style is also effected by quadruped poses. For expert users, we allow the capability to specify and achieve desired poses, if needed, using the following objectives:

$$E_{\text{StyleCOM}} = \frac{1}{2}\sum_{i}^{T}||\mathbf{x}_i - \mathbf{x}_i^D||^2\ ,$$
$$E_{\text{StyleEE}} = \frac{1}{2}\sum_{i}^{T}\sum_{j}^{k}||\mathbf{e}_i^j - {\mathbf{e}_i^j}^D||^2\ , \qquad (4)$$

where $k$ is the number of end-effectors, $\mathbf{x}_i^D$ and $\mathbf{e}_i^D$ represent desired quadruped COM, and end-effector positions respectively. Apart from these, motion smoothness is often desired by the users, which is encoded by the following objective:

$$E_{\text{Smooth}} = \frac{1}{2}\sum_{i=2}^{T-1}||\mathbf{q}_{i-1} - 2\mathbf{q}_i + \mathbf{q}_{i+1}||^2\ . \qquad (5)$$

**Constraints** We next define various constraints to ensure that the generated motion is stable.

**Kinematic constraints:** The first set of constraints ask the position of COM, and end-effectors to match with the pose of the quadruped. For every time step $i$, and end-effector $j$:

$$\varphi_{COM}(\mathbf{q}_i) - \mathbf{x}_i = 0\ ,$$
$$\varphi_{EE}(\mathbf{q}_i)^j - \mathbf{e}_i^j = 0\ , \quad \forall j, \qquad (6)$$

where $\varphi_{COM}$ and $\varphi_{EE}$ are forward kinematics functions outputting the position of COM and end-effectors respectively.

We also have a set of constraints that relate the net force and torque to the acceleration and angular acceleration of the quadruped's COM:

$$\sum_{j=1}^{k} c_i^j \mathbf{f}_i^j = M\ddot{\mathbf{x}}_i\ ,$$
$$\sum_{j=1}^{k} c_i^j(\mathbf{e}_i^j - \mathbf{x}_i^j) \times \mathbf{f}_i^j = \mathbf{I}\ddot{\mathbf{o}}_i\ , \qquad (7)$$



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

where $M$ is the total mass of the quadruped, and $\mathbf{I}$ is the moment of inertia tensor. The acceleration $\ddot{\mathbf{x}}_i$ can be evaluated using finite differences: $\ddot{\mathbf{x}}_i = (\mathbf{x}_{i-1} - 2\mathbf{x}_i + \mathbf{x}_{i+1})/h^2$, where $h$ is the time step. Similarly, the angular acceleration $\ddot{\mathbf{o}}_i$ can be expressed as $\ddot{\mathbf{o}}_i = (\mathbf{o}_{i-1} - 2\mathbf{o}_i + \mathbf{o}_{i+1})/h^2$. We note that the orientation of the root $\mathbf{o}_i$ is part of the pose $\mathbf{q_i}$, and it uses axis-angle representation.

**Friction constraints:** To avoid foot-slipping, we also have the following constraints for each end-effector $j$:

$$c_i^j(\mathbf{e}_{i-1}^j - \mathbf{e}_i^j) = 0,\ c_i^j(\mathbf{e}_i^j - \mathbf{e}_{i+1}^j) = 0\,, \tag{8}$$

for all $2 \leq i \leq T-1$, which implies that the end-effectors are only allowed to move when they are not in contact with the ground. Further, to account for different ground surfaces, we enforce the friction cone constraints:

$$f_{i\,\parallel}^j \leq \mu f_{i\,\perp}^j\,, \tag{9}$$

where $f_{i\,\parallel}^j$ and $f_{i\,\perp}^j$ denote the tangential and normal component of $\mathbf{f}_i^j$ respectively, and $\mu$ is the coefficient of friction of the ground surface.

**Limb collisions:** For physical feasibility, we propose a collision constraint that ensures a safe minimum distance between the limb segments of quadruped over the entire duration of the motion.

$$d(\mathbf{V}_i^{k_1}, \mathbf{V}_i^{k_2}) \leq \delta\,, \tag{10}$$

where $\mathbf{V}_i^k$ represents a 3D segment representing the position and orientation of $k^{th}$ limb, $d(\cdot)$ computes the distance between $k_1$ and $k_2$ limbs, and $\delta$ is the threshold distance beyond which collisions may happen.

**Motion periodicity:** If the users prefer a periodic motion, we can add an additional constraint that relates the start pose $\mathbf{q}_1$ and the end pose $\mathbf{q}_T$ of the quadruped:

$$\mathbf{J}(\mathbf{q}_T) - \mathbf{J}(\mathbf{q}_1) = 0\,, \tag{11}$$

where $\mathbf{J}(\mathbf{q_i})$ extract the orientation of the root and joint parameters from pose $\mathbf{q_i}$.

#### 3.1.3 High-Resolution Motion

The motion planning algorithm described above mainly cares about the root and the end-effectors of the skeleton, and it does not optimize for the motion style of the limbs. Thus, for motion planning, we choose to use a reduced version of the modeled skeleton which only has two joints for each limb (Figure 2(a)). After the motion is generated, we do IK post-processing to match all joints (except intermediate limb joints) and end-effectors between the original high-resolution skeleton and the reduced one (Figure 2(c)).

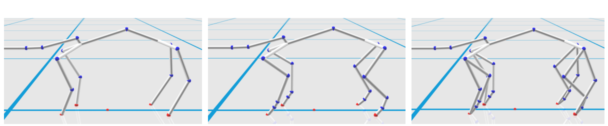



**Figure 2:** *(a) Low-resolution skeleton. (b) High-resolution skeleton. (c) Joint correspondence between low-resolution and high-resolution skeletons.*

**IK post-processing** We will just use front limb for the discussion. Since motion planning only produces the positions of the shoulder and the end-effector for each limb and we have two additional joints, *i.e.* wrist and finger, we need to add two parameters to constrain the limb and provide stylizing interface for the user. As illustrated in Figure 3(a), $L$ is the distance from the elbow to the end-effector, which can help determine the elbow's position. $\theta$ is the angle between the finger and the upright direction, which infers the positions of the finger and the wrist. Additionally, Figure 3(b-c) tells us that there are two solutions for the elbow, and similarly for the wrist. Thus, we need two binary parameters choose which way we want the elbow and the wrist to bend. If we inspect the motion of real animals, we will find that their joint angles keep changing during a motion cycle. To mimic such behavior, we use a different set of $L$ and $\theta$ when the limb is in full swing, and linearly interpolate these two sets of parameters for other motion phases.

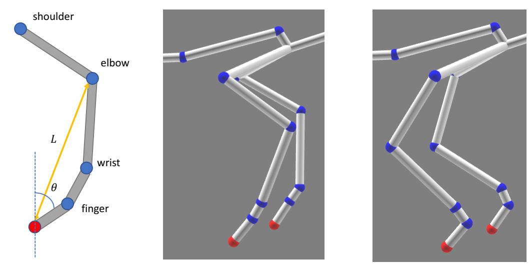



**Figure 3:** *(a) Illustration of the front limb. Two parameters $L$ and $\theta$ are used to constrain the joints and stylize the limb motion. (b-c) Since their are two analytical solutions for the elbow position, a binary parameter is used to select one of them.*

### 3.2 Kinematic Controller and Adaptation

The real-time motions is produced with an animation controller which transitions and blends between motion clips based on two input parameters: the *speed* and *direction* of the character's root. The controller is a state machine holding idle, walk forward, and walk backward states. The walking states (forward and backward) each blend between 3 motion clips: left, straight and right. The parameters for blending between clips or transitioning between states are detailed in Figure 5, and are automatically created from the generated motion clips, which is described in the previous section.

This motion controller performs only motions over flat terrain, and cannot react naturally to pushes and perturbations. To walk over different terrains, we adapt the current frame of the animation using inverse kinematics (IK), based on the terrain height. The ground height is computed by raycasting from the ground foot position, as shown in Figure 4.

Finally, to have the character react realistically to physical perturbations, such as being pushed or hit, we added a simulated character (ragdoll) layer on top. For this, we used PuppetMaster ( [PupperMaster ]) which is a character's physics tool for automatically generating ragdolls for bipeds. It enables creating active ragdolls that can follow kinematic reference motions. We extended its ragdoll layer for quadruped characters, and used it for simulating reactions.

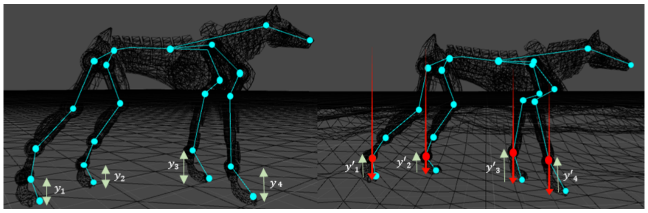


**Figure 4:** *Terrain adaptation is maintained by the estimation of the ground height at the position of the each feet by casting a ray and adding the feet offsets at the current animation frame.*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Speed | [-1.0, -0.1] | Backward Locomotion | Speed | [-1.0, -0.8] | Walk | Direction | [-1.0, -0.1] | Back. Walk Left |
| | | | | | | | [-0.1, 0.1] | Back. Walk Straight |
| | | | | | | | [ 0.1, 1.0] | Back. Walk Right |
| | | | | [-0.8, -0.1] | Trot | | [-1.0, -0.1] | Back. Trot Left |
| | | | | | | | [-0.1, 0.1] | Back. Trot Straight |
| | | | | | | | [ 0.1, 1.0] | Back. Trot Right |
| | [-0.1, 0.1] | Standing | Direction | [-1.0, -0.1] | Turn Left | | | |
| | | | | [-0.1, 0.1] | Idle | | | |
| | | | | [ 0.1, 1.0] | Turn Right | | | |
| | [ 0.1, 1.0] | Forward Locomotion | Speed | [ 0.1, 0.8] | Walk | Direction | [-1.0, -0.1] | Forw. Walk Left |
| | | | | | | | [-0.1, 0.1] | Forw. Walk Straight |
| | | | | | | | [ 0.1, 1.0] | Forw. Walk Right |
| | | | | [ 0.8, 1.0] | Trot | Direction | [-1.0, -0.1] | Forw. Trot Left |
| | | | | | | | [-0.1, 0.1] | Forw. Trot Straight |
| | | | | | | | [ 0.1, 1.0] | Forw. Trot Right |

**Figure 5:** *A blending diagram is automatically created from the generated motion clips that controls the motion transition using parametric inputs- speed and direction (of root).*

## 4 3D reconstruction

We describe our approach to understanding the environment for AR purposes. Because current consumer level hardware devices such as the Hololens only offer coarse reconstructions online, we cannot use them for having characters walk over as they appear to be floating in air.

Hence, we developed pre-defined objects that are recognized and localized in space using feature-based technology (Vuforia Engine [Vuforia 2017]). For each real world object, we define a corresponding 3D digital geomtric counter-part that matches in shape and size. Then, we scan the real world object from all directions using an RGB camera to obtain a data-base of image-based features and transformation pairs. At runtime, Vuforia searches for matching features and returns the id of the object, together with its transformation that we apply to the 3D object in the scene.

We encountered a few issues recognizing objects with Vuforia. One problem is when the objects are transparent, or have plain textures. In this case, the lack of features causes the recognition to fail. Similarly to object recognition, objects with shiny and reflective properties do not give successful image recognition and tracking. Hence for some objects, we add a rich texture on top to make them distinguishable, as shown in our figures bellow.

## 5 Interactions

We take our animated character together with its ability to navigate the real world environment, and design AR interactions in 3D. In particular, we provide ways for the user to specify where the character should go, ways to have 3D virtual and real objects collide with the character, as well as ways to configure different terrains.

**Specifying trajectories via touch on the screen.** The quadruped character can be directed through any arbitrary paths in the physical environment, over different terrains. The paths are generated by *projecting* onto the environment, the user's fingertip when drawing on the touchscreen, as shown in Figure 6.

**Walking over different real-world slopes.** The virtual character's behavior depends on the purpose of the interaction. Therefore we label objects either as terrain or non-terrain in order for the motion model to behave in the correct manner. We label the terrain automatically by defining objects as terrain if their height is bellow a certain threshold, that corresponds to the maximum height the character can climb.

Different slopes and platforms can be formed with different arrangements of the objects as shown in Figure 7. While the character's motion model will only employ the inverse kinematics for adapting to objects labeled as terrain, it should react differently for the other objects, as described next.

**Pushing characters with real-world objects.** Non-terrain objects can be used for interactions like colliding with or pushing the quadruped, as shown in Figure 8. For animating the reactions, a ragdoll simulation (un-controlled passive dynamics) is activated for a short period of time, letting the character react to the perturbation. After the short period of time, the state of the simulated (ragdoll) character is blended back into the animation state over another small window of time. Completely switching to a ragdoll simulation causes to the character to fall. Hence, above a certain force threshold, we do not blend back to the animation and simply let the character flow.

**Interacting with virtual objects.** We also experimented the interactions between the character and virtual objects. We designed a simple platformer game (which is shown in our accompanying video), where the user can use various *props* to carry the virtual character from the beginning to the end of the platform puzzle, while trying to prevent him from falling. The character only moves forward, and its moving direction can only be changed if it hits a *wall prop*. When the character meets an *elevator platform* which goes up and down, the user needs to use a *fan prop* (shown in Figure 10) to stop the character such that it can wait. For an increased challenge, we added an *enemy cannon* which shoots 3D balls at the character, possibly causing it to fall from the platform, as shown in Figure 9.

## 6 Discussion and Conclusion

While we did a first step in the direction of having intelligent character we can interact with in AR, characters that can understand their environments and navigate them, our system has a few limitations. First we can only interact slowly with objects, as the tracking is remains at low frequency. We use pre-defined 3D objects instead of scanning the world. We believe that both of these issues will improve with the evolution of hardware.

In the future, we plan on investigating chameleon technology with in-painting, allowing real-world characters to "come-to-life", by replacing their background and animating their virtual counter-parts. We also plan on integrating additional components to our character's intelligence, such the ability to talk and reason about the objects in the world.

## References

ARIKAN, O., FORSYTH, D. A., AND O'BRIEN, J. F. 2003. Motion synthesis from annotations. *ACM Transasctions on Graphics 22*, 3, 402–408.

BHATTI, Z., SHAH, A., WAQAS, A., AND KARBASI, M. 2015. Automated animation of quadrupeds using procedural program-



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

**Figure 6:** *Path drawing with touch is used to direct the character in the physical environment.*


**Figure 7:** *Different arrangements of predefined physical objects creates different slopes for the character to walk on. For the details of the 3D object reconstruction, we kindly refer to Section 4.*

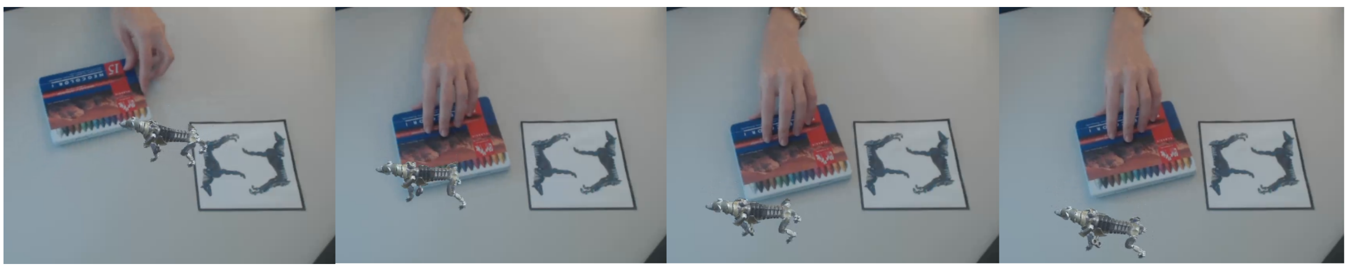


**Figure 8:** *The character reacts to the pushes by real objects.*

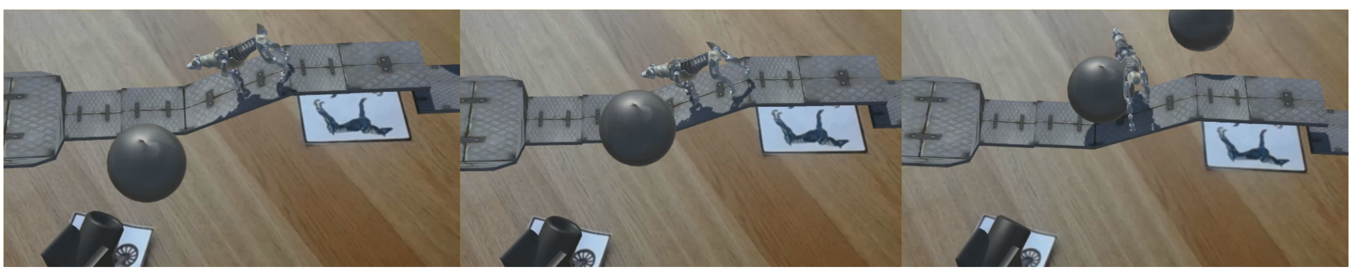


**Figure 9:** *During the platformer game, the character can be hit by a virtual cannon ball and fall down.*

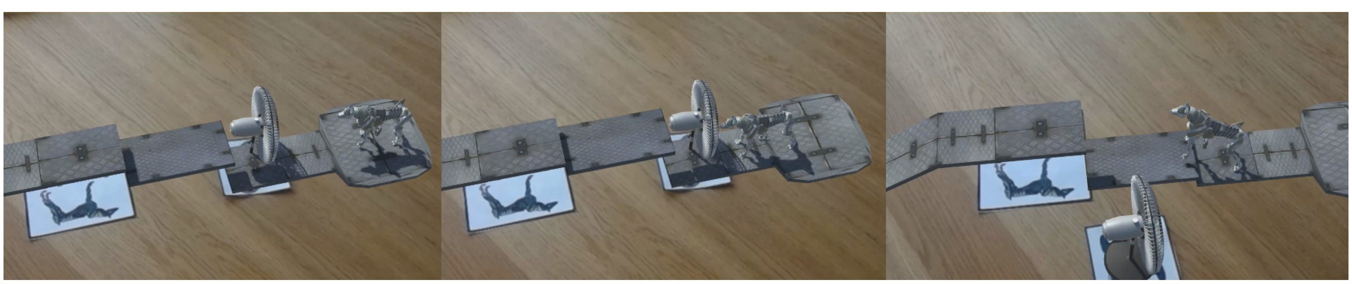


**Figure 10:** *A virtual fan can be used to stop the character which creates virtual forces.*

ming technique. In *Asian Journal of Scientific Research*, vol. 8, 165–181.

BILLINGHURST, M., KATO, H., AND POUPYREV, I. 2001. The magicbook - moving seamlessly between reality and virtuality. *Computer Graphics and Applications, IEEE 21*, 3 (May), 6–8.

GRASSET, R., DÜNSER, A., AND BILLINGHURST, M. 2008. Edutainment with a mixed reality book: A visually augmented illustrative childrens' book. In *Proceedings of the 2008 International*

*Conference on Advances in Computer Entertainment Technology*, ACM, New York, NY, USA, ACE '08, 292–295.

HOLOLENS, 2017. Microsoft, https://www.microsoft.com/en-us/hololens/.

KIM, H., TAKAHASHI, I., YAMAMOTO, H., MAEKAWA, S., AND NAEMURA, T. 2014. Mario: Mid-air augmented reality interaction with objects. *Entertainment Computing 5*, 4, 233 – 241.

KOKKEVIS, E., METAXAS, D., AND BADIER, N. I. 1995. Autonomous animation, control of four-legged animals. In *Proceedings of Graphics Interface '95*, GI '95, 10–17.

KOVAR, L., GLEICHER, M., AND PIGHIN, F. 2002. Motion graphs. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, USA, SIGGRAPH '02, 473–482.

MEGARO, V., THOMASZEWSKI, B., NITTI, M., HILLIGES, O., GROSS, M., AND COROS, S. 2015. Interactive design of 3d-printable robotic creatures. *ACM Trans. Graph. 34*, 6 (Oct.), 216:1–216:9.

PUPPERMASTER. Rootmotion, http://root-motion.com/.

SCHERRER, C., PILET, J., FUA, P., AND LEPETIT, V. 2008. The haunted book. In *ISMAR*, IEEE Computer Society, 163–164.

TORKOS, N., AND VAN DE PANNE, M. 1998. Footprint–based quadruped motion synthesis. In *Proceedings of the Graphics Interface 1998 Conference, June 18-20, 1998, Vancouver, BC, Canada*, 151–160.

VUFORIA, 2017. Qualcomm, http://www.qualcomm.com/vuforia.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# Chalktalk VR/AR

**Ken Perlin**
New York University
ken.perlin@gmail.com

**Zhenyi He**
New York University
zh719@nyu.edu

**Fengyuan Zhu**
New York University
zhufyaxel@gmail.com

## ABSTRACT

When people want to brainstorm ideas, currently they often draw their ideas on paper or on a whiteboard. But the result of those drawings is a static visual representation. Alternately, people often use various tools to prepare animations and simulations to express their ideas. But those animations and simulations must be created beforehand, and therefore cannot be easily modified dynamically in the course of the brainstorming process.

Chalktalk VR/AR is a paradigm for creating drawings in the context of a face to face brainstorming session that is happening with the support of VR or AR. Participants draw their ideas in the form of simple sketched simulation elements, which can appear to be floating in the air between participants. Those elements are then recognized by a simple AI recognition system, and can be interactively incorporated by participants into an emerging simulation that builds more complex simulations by linking together these simulation elements in the course of the discussion.

.CCS CONCEPTS
**•Human-centered computing → Mixed / augmented reality**
**KEYWORDS**
Virtual Reality, Augmented Reality

**ACM Reference format:**
Ken Perlin, Zhenyi He, and Fengyuan Zhu. 2017. Chalktalk VR/AR. In *Proceedings of SIGGRAPH 2017 ASIA, BITEC, Bangkok, Thailand, November 2017 (SIGGRAPH ASIA 2017)*, 2 pages.
DOI: 10.1145/1234

### 1 INTRODUCTION

There has been much work in recent years in supporting Virtual Reality (VR) and Augmented Reality (AR) display and interaction, but the bulk of this work has focused on individual users, or on interaction between people who are remotely located. Chalktalk VR/AR [1] [2]is a system for supporting face to face brainstorming between people in the context of VR/AR, who can be physically co-located.

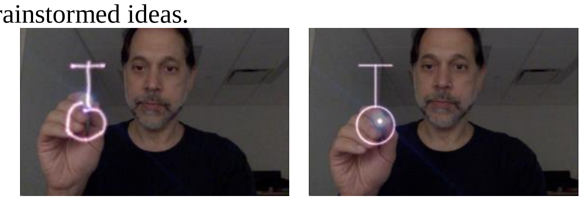
Traditionally, when people want to brainstorm ideas, currently they often draw their ideas on paper or on a whiteboard. But the result of those drawings is a static visual representation. Alternately, people often use various tools to prepare animations and simulations to express their ideas. But those animations and simulations must be created beforehand, and therefore cannot be easily modified dynamically in the course of the brainstorming process.

Chalktalk VR/AR allows participants to draw their ideas in the form of simple sketched simulation elements, which can appear to be floating in the air between participants. Those elements are chosen from a large vocabulary of available simulation element types. Each individual sketch element drawn by a participant is recognized at run-time by a simple AI recognition system, and converted to the corresponding simulation element, which can be interactively incorporated by participants into an emerging simulation, in the spirit of [3]. In this way, participants can collaborate to build complex simulations by linking together individual simulation elements in the course of their discussion.

### 2 CHALKTALK VR/AR DESIGN

Chalktalk VR/AR is an approach to shared virtual and augmented reality that begins with the concept that people communicating with each other in a shared VR or AR experience can talk to each other face-to-face while using a paradigm of creating "smart drawings" in the air.

The shapes that are drawn by participants are interpreted to find the closest match within a dictionary of known glyphs, see Figure 1. Those drawings can "come to life" and become active simulations. Those simulation objects can be joined together to create composite simulations to support complex expressions of collaborative thoughts and brainstormed ideas.



**(a) original drawing** **(b) result after drawing**

**Figure 1: Pendulum example.**
**(a) shows the original hand drawing in mid-air and (b) shows what it will evolve to after recognition.**

Once a drawing has been recognized, the specific geometric variations in how the participant draw the shape can be used to customize its resulting simulation object in ways that influence how that simulation will behave. For example if a participant draws a pendulum, then the length of the pendulum shaft and the size of the bob are reflected in the simulated pendulum, see Figure 2, in such a way that the



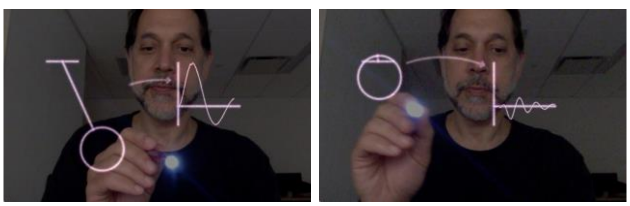
In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

physics of the swinging pendulum properly reflects the shape that was drawn, see Figure 3a and Figure 3b.

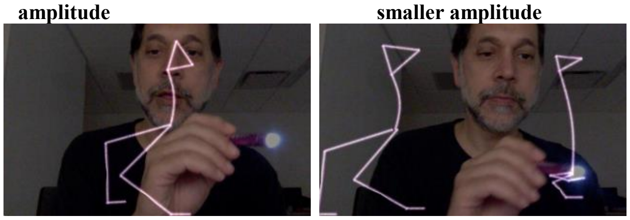
**(a) pendulums** **(b) shoes** **(c) dinosaurs**

**Figure 2: Geometric variation in drawing multiple glyphs**
**(a) The left glyph shows a pendulum with longer shaft and smaller bob, meanwhile the right glyph indicates a pendulum with shorter shaft and larger bob, which will affect their physical behavior.**
**(b) The upper glyph shows a shorted fish with larger head and the lower one shows a fish with longer body and smaller head. The length of the body will lead to different swimming speeds.**
**(c) The dinosaur in the left has longer legs than the right one. Therefore the left one will walk faster than the right now.**

Similarly if a participant draws a walking creature, than the proportions of the different parts of the creature, such as its torso length the size of its legs, are recognized by Chalktalk and are used to modify the way the creature walks and otherwise behaves, see Figure 3c and Figure 3d.



**(a)Longer shaft leads to larger amplitude** **(b) Shorter shaft leads to smaller amplitude**

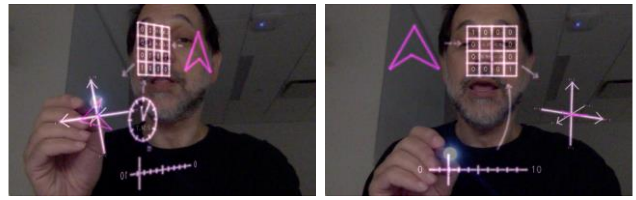


**(c) Longer legs lead to larger movement** **(d) Shorter legs lead to slower pace**

**Figure 3: How physics of the swinging pendulum properly reflects the shape that was drawn**

A fundamental decision was made that all drawing by participants be performed in a virtual plane that floats at a fixed distance in front of that participant, so as to best build upon our human ability to draw accurately in two dimensions. In contrast, the resulting recognized simulated objects exist in three dimensions, see Figure 4.

For example, a participant can make a drawing of what looks like an image of a hypercube. Then once that drawing has recognized by Chalktalk, the resulting interactive simulation object behaves just like a mathematical projection onto a three dimensional space of a four dimensional hypercube, see Figure 5.



**(a) The polygon in coordinate system is three dimensional** **(b)The whole AR scene is three dimensional**

**Figure 4: See the objects in three dimensions**

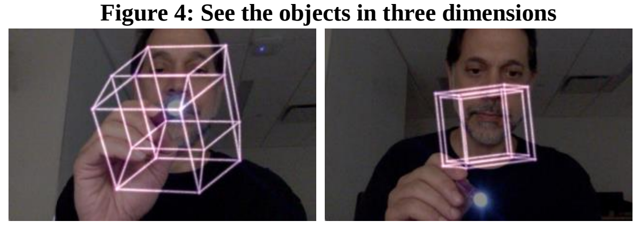


**Figure 5: Looking into a hypercube in three dimensions**

## 3 CONCLUSION AND FUTURE WORK

We have introduced Chalktalk VR/AR, a dynamic sketch-based simulation tool for face to face brainstorming in VR and AR. This approach preserves the immediacy of face to face collaboration. It also allows sophisticated simulation elements to be invoked and included in an emerging larger simulation. This approach of using freehand sketching to invoke simulation elements is well suited to the brainstorming process, as compared with less socially oriented approaches such as typing or choosing from menus. Also, the use of freehand sketching as an input modality allows participants to customize the properties of simulation elements during the course of their conversation, without needing to interrupt the flow of conversation to do so.

One potential drawback of Chalktalk VR/AR, as opposed to menu based approaches is that it requires participants to have a certain level of shared expert knowledge of the sketch language. However, we maintain that as VR and AR become more ubiquitous parts of our everyday communication, this general “language based” approach to visual support for face to face communication will become increasingly the norm, rather than the exception, in much the same way that natural language itself is a powerful and ubiquitous form of communication that relies on shared expert knowledge of the language being spoken.

## REFERENCES


[1] K. Perlin, "The coming age of computer graphics and the evolution of language," in *Proceedings of the 2nd ACM symposium on SUI*, 2014.

[2] K. Perlin, "Future Reality: How emerging technologies will change language itself}," in *IEEE computer graphics and applications*, 2016.

[3] W. R. Sutherland, "The on-line graphical specification of computer procedures," 1966.




In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# Visualization Methods of Hierarchical Biological Data: A Survey and Review

Irina Kuznetsova
Holzinger Group, HCI-KDD,
Institute for Medical Informatics/Statistics
Medical University, Graz, Austria
irina.kuznetsova@student.tugraz.at

Artur Lugmayr
Visualisation and Interactive Media, Lab.
(VisLab) Curtin University, Perth, AU
& Aalto Univ., Helsinki, Finland
lartur@acm.org

Andreas Holzinger
Holzinger Group, HCI-KDD,
Institute for Medical Informatics/Statistics
Medical University, Graz, Austria
a.holzinger@hci-kdd.org

## ABSTRACT

The sheer amount of high dimensional biomedical data requires machine learning, and advanced data visualization techniques to make the data understandable for human experts. Most biomedical data today is in arbitrary high dimensional spaces, and is not directly accessible to the human expert for a visual and interactive analysis process. To cope with this challenge, the application of machine learning and knowledge extraction methods is indispensable throughout the entire data analysis workflow. Nevertheless, human experts need to understand and interpret the data and experimental results. Appropriate understanding is typically supported by visualizing the results adequately, which is not a simple task. Consequently, data visualization is one of the most crucial steps in conveying biomedical results. It can and should be considered as a critical part of the analysis pipeline. Still as of today, 2D representations dominate, and human perception is limited to this lower dimension to understand the data. This makes the visualization of the results in an understandable and comprehensive manner a grand challenge.

This paper reviews the current state of visualization methods in a biomedical context. It focuses on hierarchical biological data as a source for visualization, and gives a comprehensive survey of visualization techniques for this particular type of data.

## CCS CONCEPTS

• Human-centered computing~Visualization techniques

## KEYWORDS

Visualization, hierarchical data, computer graphics, information visualization, big data, bioinformatics.

## 1 INTRODUCTION

The research domain of information visualization is broad, and involves a wide range of research fields, such as computer graphics (e.g. 2D and 3D graphics), information design to increase communication and sense making [24], creative aspects (e.g. design, layouts, colour use) [72], and methods from human-computer interaction. Making data understandable from a cognitive and machine learning point of view has emerged recently in the literature [40, 41], including the idea to render data throughout smart environments [37–39][49].

The biomedical domain is a complex field of biological processes. In addition, the advancements of biological technologies has led to a dramatic increase in data volumes [26], which has presented new challenges in knowledge extraction. Working with high-volume data requires the application of data mining that draws upon machine learning techniques. These methods help in extracting knowledge patterns and narrowing data to the smaller volumes. These include e.g. *Support Vector Machines (SVM*); *Artificial Neural Networks (ANN)*; clustering [46][4]; statistical techniques (e.g. Bayesian statistics [51]; *Hidden Markov Models (HMMs)* [10]; *Principle Component Analysis (PCA)*; classification methods [28][26][1].

The interpretation of extracted knowledge and the result of analysed data through visualisation is an essential step in the analysis pipeline, and becoming an important tool in bioinformatics. These not only includes simple visualizations (i.e. bar plots, pie charts, flow charts), but also advanced visualization techniques for representing final results in the biomedical domain (e.g. 3D). As also valid for other domains, visualisations should follow information design principles, which are defined in [60]:

(1) providing an overview of the data;
(2) zoom in/out options;
(3) filtering of unnecessary information;
(4) detailization of region of interests;
(5) relation between data points of interest;
(6) history of actions; and
(7) a possibility of extracting required parameters.

Another example for applying visualisation in biology is the application of new technologies such as the *Next Generation Sequencing (NGS),* which delivers enormous volumes of genomic data in a digital format. To visualise the data, genome browser applications are utilized to allow a real-time visualisation and exploration of genomic sequences; of any region of interest; in any required scale within a genome [53][31][56].

Within the scope of this paper, we firstly give an overview of existing visualization techniques, followed by description of the characteristics of hierarchical data. We focus mainly on traditional visualization techniques, as i.e. the visualization of hierarchically organized data in a 2D space at first place. We illustrate our approach based on a typical biological analysis workflow as previously discussed in [35]. The workflow narrows genetical information into a meaningful smaller subset called differentially expressed genes. These represent the active genes in the overall genetic information, and can be utilized to obtain *Gene Ontologies*



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

*(GO)*, which are categorizing the functions of various genes. Visualisation supports the understanding of results, as well as the obtained ontologies.

## 2 RELATED WORKS

We would like to point to the following works in information visualisation and design for further reading: [60], [42], and the excellent introductory guide [64]. Visualization plays a key role in the biomedical domain. Various techniques aim to deliver the correct representation of results in visual format, and follow visualization design principles described in [60]. To state an example, the visualization of a protein structure in 3D space enables researchers to have an overview of the studied protein; to rotate a protein image in different dimensions; to see protein-protein interactions; to measure an atomic distance; and to zoom into the region of interests. Several visualisation tools have been developed and support the analysis process. Remaining in the domain of protein structures Web3DMol, UCSF Chimera and POLYVIEW-3D are just a few examples for visual protein structure investigation [59][48][50].

**Table 1:** *Examples of visualization software*

| Biological Task | Visualization Task | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Overview** | **Zoom** | **Filter** | **Detailization** | **Relate** | **History** | **Extract** | **2D** | **3D** |
| **1. A Protein Structure Visualization** | | | | | | | | | |
| Web3DMol see [59] | + | + | + | + | + | + | + | | + |
| UCSF Chimera see [48] | + | + | + | + | + | + | + | | + |
| POLYVIEW-3D see [50] | + | + | + | + | + | + | + | | + |
| **2. NGS Data Visualization** | | | | | | | | | |
| IGV see [53] | | + | + | + | + | + | + | + | |
| UCS see [31] | | + | + | + | + | + | + | + | |
| ZEMBU see [56] | | + | + | + | + | + | + | + | |
| **3. Hierarchical Data Visualization (Phylogeny)** | | | | | | | | | |
| ETE Toolkit see [21] | + | | + | + | + | + | + | + | |
| PhyD3 see [33] | + | + | + | + | + | + | + | + | |
| EvolView see [33, 78] | + | + | + | + | + | + | + | + | |

Another set of tools allowing real-time visualisation and exploration of genomic sequences are the IGV, UCS, or ZEMBU genome browsers, which have been mentioned in the introduction section of the publication [53][31][56]. These, and many others allow the exploration of various samples of sequenced genomic data.

Another example of information visualization in biology is phylogeny, where hierarchical data structure is considered for image development. The intuitive way of representing hierarchies is as a tree diagram. The ETE Toolkit, PhyD3, EvolView and other visualization programs enable phylogenetic trees to be studied in more detail [21][33][33, 78].

Several tools mentioned within this section are summarized in Table 1, which classifies them according data type, and biological task. Visualisation is a very important research tool, which enables researchers to explore and study biological structures, investigate biological data in various digital formats, and understand molecular processes in an intuitive and comprehensive way.

## 3 CHARACTERISTICS AND PROCESSING OF HIERARCHICAL DATA

The hierarchical pattern is observed in numerous aspects of our life and various biological fields are not an exception: phylogenetics, GO, microarray analysis, differential expression analysis (dendrograms), and protein similarities represent data as hierarchies. Hierarchically organized data facilitates results comprehension and interpretation and provides a global overview of the data. Hierarchical clustering belongs to an unsupervised machine learning technique used for building hierarchies. Data is usually presented as a parent-child relation, where a parent can have zero or more related children (see Fig. 1 A-B).

### 3.1 Processing Hierarchical Data Structures

To process hierarchical structures, numerous computational methods have been developed such as neighbour-joining [54], UPGMA [63], maximum parsimony [11][13], and maximum likelihood [57]. These utilize distance-based, character-based, or statistical method of approaching hierarchies respectively.

Distance-based clustering is the traditional method for hierarchical clustering. The input data is a matrix, where rows characterize a unique object, and columns show the object's features. The distance matrix, also called a proximity matrix, is calculated with a linkage method (see Table 2) which enables the estimation of dissimilarity/similarity between objects. There are two types of algorithms for hierarchical clustering: (1) agglomerative, and (2) divisive [1]. The agglomerative or bottom-up is one of the popular hierarchical clustering algorithms that starts from grouping the two closest data points of the distance matrix into a cluster, updating the distance matrix for the just-generated cluster and the original matrix based on the selected linkage method, and continuing this process until only a single cluster remains [25]. In other words, it starts from grouping the closest data points of the input data ("bottom"), and ends when each data point is assigned to its related cluster ("up"). In contrast, the divisive method or top-down follows an opposite way of grouping data values. It considers an input data as one whole cluster and splits the data into smaller clusters by moving from the "top" – (one cluster) to the "down"- (many clusters) [1].



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

In addition to existing clustering algorithms, new techniques have been developed with the aim of addressing the emerging issues associated with large data volumes produced by new technologies. Loewenstein and team proposed a memory-constrained UPGMA (MC-UPGMA) algorithm that enables clustering of the large data sets that was implemented in C++ [36]. Kannan and Wheeler extended the parsimony score to phylogenetic networks; the algorithm was implemented in OCAML [30].

**Table 2:** *Hierarchical clustering - types of linkage methods* [26].

| Linkage methods | Formula |
|---|---|
| Single-linkage | $D(Ci,Cj) = \min_{x_p \in Ci,\ x_q \in Cj} d(x_p, x_q)$ |
| Complete-linkage | $D(Ci,Cj) = \max_{x_p \in Ci,\ x_q \in Cj} d(x_p, x_q)$ |
| Average-linkage, WPGMA | $D(Ci,Cj) = \frac{D(Ci,Cm) + D(Cj,Cn)}{2}$ |
| Average-linkage, UPGMA | $D(Ci,Cj) = \frac{D(Ci,Cm)\ \lvert Cm \rvert + D(Cj,Cn)\ \lvert Cn \rvert}{2}$ |
| Centroid-linkage | $D(Ci,Cj) = d(c_i, c_j)$ where $c_i = \frac{1}{\lvert c_i \rvert} \sum_{x_p \in Ci} x_p$ |
| Median-linkage, WPGMC | $D(Ci,Cj) = d(w_i, w_j)$ where $w_j = \frac{1}{2}(w_m + w_n)$ |
| Ward's linkage | $ESS = \sum_{x_n \in C} \lVert x_n - \bar{x} \rVert_2$ |

## 3.2 Data Representation, Storage and Queries

Other essential aspects in dealing with hierarchical data is data representation, and data encoding for e.g. storage in a database or as in-memory representations for applying algorithms. Theoretical considerations can be found in database theory, formal languages, and query languages. An overview of these techniques can be found in an interesting online article [82], are listed in Table 3, and more details about the theoretical aspects can be found in [81][8][52]. Several techniques offer different ways of accessing the required information.

**Table 3:** *Methods for storing hierarchical data* [82].

| Technique | Description | Links |
|---|---|---|
| *Adjacency List* | Recursive method. Each node of the tree has a pointer to a parent node. Intuitive and simple for implementation, but slow in performing queries. | [45] |
| *Path Enumeration* | Each entry is stored as a full path to the root. | [70] |
| *Nested Set* | Applies traversal method of numbering nodes. Each node is visited twice where each time the number of the visit is assigned (has two pointers) and stored. Fast for retrieving required information, but becomes slow for updating a tree. | [6] |
| *Nested Intervals* | Similar to nested set techniques, however the numbering can apply real/float/decimal numbering. | [69, 70] |
| *Flat Table* | Similar to adjacency method with addition of rank and a level information. | [14] |
| *Closure Table* | Transitive way of representing hierarchies. Applied if database does not support iterative query. | [80] |
| *Multiple Lineage Columns* | | [82] |

Representation of hierarchical data as adjacency list model (see Fig. 1-A), where each element of the table has a pointer to its parent, or it can be visualized as a tree diagram (see Fig. 1-B).

A

| ID | Name | Parent |
|---|---|---|
| 1 | World | NULL |
| 2 | Oceania | 1 |
| 3 | Micronesia | 2 |
| 4 | Polynesia | 2 |
| 5 | Australia & NZ | 2 |
| 6 | Melanesia | 2 |
| 7 | Australia | 5 |
| 8 | New Zealand | 5 |
| 9 | Latin America | 1 |
| 10 | Caribbean | 9 |
| 11 | South America | 9 |
| 12 | Central America | 9 |

B

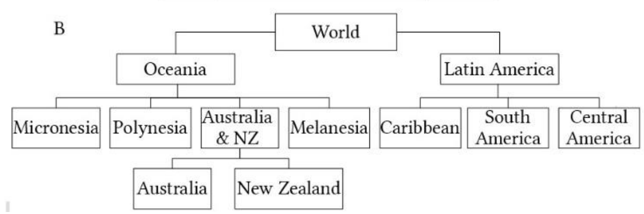



**Figure 1:** *Hierarchical data representation as: A- the adjacency list model; B- a tree diagram.*

## 3.3 Challenges in Hierarchical Data Processing

**3.3.1 Data Accessibility.** The problem of modern health sciences is that generated data may not be accessible to a health science researcher directly [18], because certain patterns ("knowledge") are hidden in arbitrarily high-dimensional spaces. Examples range from longitudinal rheumatology data sets, in which cohorts of patients are attributed with vectors in $R^{100}$ [62], to the uncertainties of RNA sequence base pairing variants with a potentially arbitrary number of dimensions [27]. The results gained by machine learning and knowledge extraction techniques need to be mapped down into the lower dimensions to make them accessible to a human expert [73]. This calls for a closer cooperation between machine learning and visualization experts [74]. A crucial factor in clustering techniques is the curse of dimensionality [32]. With increasing



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

dimensionality, the volume of the space increases so quickly that the available data becomes sparse, hence becoming extremely difficult to find reliable clusters. A further significant problem is that distances become imprecise as the number of dimensions grows, since the distance between any two points in a given data set converges; moreover, different clusters might be found in totally different sub spaces. Consequently, a global filtering of attributes on its own is not sufficient.

**3.3.2 Subspace Clustering.** The subspace clustering problem is difficult, as very different characteristics for grouping can be used: this can be highly subjective and context-specific and requires an expert-in-the-loop [19][20]. What is recognized as comfort for end-users of individual systems? It is interesting to note that human experts are quite capable in determining similarities and dissimilarities, which has been described by nonlinear *Multidimensional Scaling (MDS)* [58][68].

We can represent similarity relations between entities as a geometric model consisting of a set of points within a metric space. The output of an MDS routine is a geometric model of the data, with each object of the data set represented as a point in n-dimensional space. Consequently, there is urgent need to map very high-dimensional data into a small number of relevant dimensions to make it accessible for human expert analysis. For example, the similarity between patients may change by considering different combinations of relevant dimensions [22]. This is called subspace analysis and is a very interesting and relevant field of current research [12]. For example, with the goal of finding a k-dimensional subspace of $R^d$ in a way that the expected squared distance between instance vectors and the subspace is a minimum. This so-called subspace learning can also be used as a dimensionality reduction technique [15]. Common tools include the stationary subspace analysis toolbox [44], SubVIS [23] and Morpheus [43] – just to mention three.

# 4 VISUALISATION OF HIERACHICAL DATA

Tree-like structured graphs are a common way of representing hierarchical data. In general, a tree-structured graph is defined as a root node, which is connected through links or edges to the parent and children nodes [77]. The traditional tree view is visualized in upside-down way, where the root is on the top and a parent-child relation is shown towards the bottom. However, a tree graph can be also represented as a left-to-right diagram [76].

According to [55], visualization of hierarchical organized data can be represented as (see Table 4) [55]:

(1) explicit vs implicit; or

(2) axes-oriented vs radial (see Table 4).

The implicit method belongs to the space-filling technique that fits provided data into a defined space for example, rectangular, triangle, circular etc. The explicit method utilizes a traditional tree-like structure.

**Table 4:** *Types of visualization methods of hierarchical data.*

| **Type** | **Description** | **Axes-Oriented Layout** | **Radial Layout** |
|---|---|---|---|
| *Explicit* | Visualization method representing hierarchy as a node-link diagram [55] | Dendrogram, Intended tree (see Fig. 2, Fig. 4) | Circular tree (see Fig. 3) |
| *Implicit* | Visualization method representing hierarchy in a space-filling way [55] | Tree-maps (see Fig. 5) | Sunburst (see Fig. 6) |

There are a range of visualization graphs that enables hierarchies to be shown in 2D format. Dendrograms and intended layouts (see Fig. 2, Fig. 4) are examples of the explicit method in an axes-oriented layout, whereas a circular tree (see Fig. 3) is an explicit method in a radial layout. Space-filling techniques can also use axes-oriented layouts, such as tree-maps (see Fig. 5); or in a radial layout such as Sunburst (see Fig. 6).

## 4.1 Explicit Visualization

This part of the paper describes 2D visualization techniques of hierarchical data on GO data. The example subset data is taken from the REVIGO Web server at (http://revigo.irb.hr/), which applies a neighbor-joining hierarchical clustering algorithm to achieve hierarchies [66].

The output data of REVIGIO tool may be used as input data for tree-maps, or Sunburst visualization methods, whereas a distance-based clustering used for a tree structured diagrams. The second and the third columns of Table 5 named "representative" and "description" show the parent-child relationships respectively. For example, "nucleoside triphosphate metabolism" is a parent node of five related children annotations such as "nucleoside triphosphate metabolic process", "alanine biosynthetic process", "inositol biosynthetic process", "isocitrate metabolic process" and "regulation of translation". The "response to herbicide" and "ion transmembrane transport" have two and four related children correspondingly, whereas "protein refolding" has no children at all.

Although a tree diagram is a traditional way of representing hierarchies, REVIGO can visualize hierarchies as scatter plots, interactive graphs, tree-maps, tag clouds and intended trees [66].

**4.1.1 Dendrograms.** A dendrogram, also called a binary tree (see Fig. 2), is a visualization technique commonly used in representing groups of similarities (clusters) in the data produced by the hierarchical clustering method [16][25].

It has a traditional tree-like structure, where leaves are placed at the same level. The y-axis (height) shows the distance at which a cluster is formed. The labels across the x-axis are equally distributed for readability purposes. The dotted line is an example of a selected distance cut-off that enables the reader to see the number of clusters that found within that distance. Figure 2 illustrates that four distinct clusters were identified (represented in red, purple, blue, and green colours) if the closeness of objects was defined as a distance of value two. In biology, a clustering approach



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

**Table 5:** *The hierarchical data as GO. The example data is taken from REVIGO* [66], *and modified for explanatory purposes.*

| Term_ID | Representative (parent) | Description (child) | Freq InDb | log10 pval | Unique-ness | Dispens-ability |
|---|---|---|---|---|---|---|
| GO:0009141 | nucleoside triphosphate metabolism | nucleoside triphosphate metabolic pr. | 1.61% | -300 | 0.689 | 0 |
| GO:0006523 | nucleoside triphosphate metabolism | alanine biosynthetic process | 0.05% | -26 | 0.791 | 0.273 |
| GO:0006021 | nucleoside triphosphate metabolism | inositol biosynthetic process | 0.02% | -10 | 0.817 | 0.331 |
| GO:0006102 | nucleoside triphosphate metabolism | isocitrate metabolic process | 0.02% | -10 | 0.839 | 0.276 |
| GO:0006417 | nucleoside triphosphate metabolism | regulation of translation | 0.69% | -119 | 0.786 | 0.55 |
| GO:0009635 | response to herbicide | response to herbicide | 0.00% | -31 | 0.892 | 0 |
| GO:0010447 | response to herbicide | response to acidic pH | 0.01% | -14 | 0.863 | 0.546 |
| GO:0034220 | ion transmembrane transport | ion transmembrane transport | 3.53% | -300 | 0.912 | 0 |
| GO:0055085 | ion transmembrane transport | transmembrane transport | 8.92% | -300 | 0.932 | 0.497 |
| GO:0006818 | ion transmembrane transport | hydrogen transport | 1.15% | -300 | 0.866 | 0.366 |
| GO:0015797 | ion transmembrane transport | mannitol transport | 0.01% | -11 | 0.903 | 0.377 |
| GO:0042026 | protein refolding | protein refolding | 0.07% | -204 | 0.957 | 0.029 |

is commonly used to find groups of genes that share similar features based on results from *Differential Expression (DE)* analysis. Heatmaps represent a matrix of values of genes expression in a color-coded way and are accompanied with dendrograms. Dendrograms are illustrated along the heatmap on the top and/or left sides. The left side dendrogram represents the similarity between genes, and the top dendrogram the similarity between samples. Dendrograms are also common structures in representing phylogenetic trees [7].

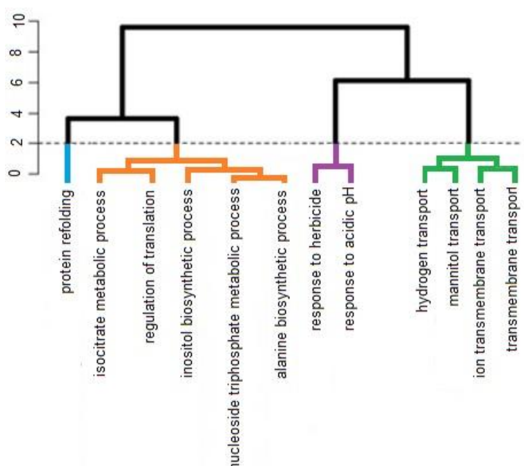



**Figure 2:** *Dendrogram, (visualized with R).*

**4.1.2 Circular Trees.** There are another two ways of visualizing trees as radial trees, and circular trees [76]. In the radial tree the hierarchical tree structure is visualized in an annulus wedge; the algorithm was proposed by P. Eades [9]. In the circular tree visualization, the root is placed at the central positions and leaf nodes are equally distributed around on the perimeter of a circle (see Fig. 3) [76]. The hierarchy in this case is shown with a tree-structure graph. Coloring and labelling are used to improve representation of tree graphs. The circular tree layout that is also explicit method used for representing phylogenetic trees [7].

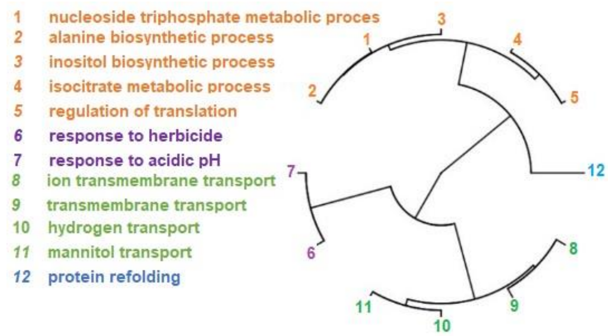
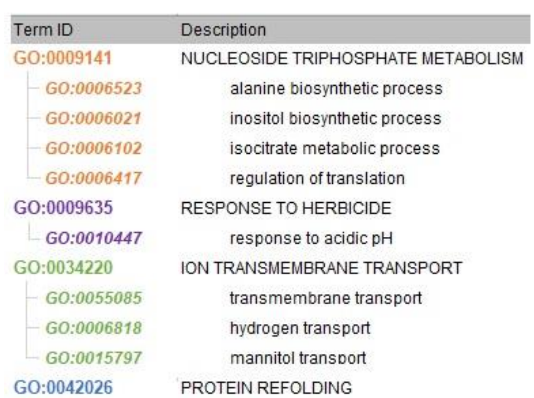



**Figure 3:** *Circular Tree, (visualized with R).*

**4.1.3 Intended Trees.** The intended layout is another way of representing hierarchies (Fig. 4). The data is plotted along the vertical axis and indentations are used in representing parent/children relationships. This type of visualization is commonly used for interface systems or online, as it allows easy access to required information by scrolling down. However, this technique has an unpublishable format and hence cannot be used as an effective overview of the data.



| Term ID | Description |
|---|---|
| GO:0009141 | NUCLEOSIDE TRIPHOSPHATE METABOLISM |
| GO:0006523 | alanine biosynthetic process |
| GO:0006021 | inositol biosynthetic process |
| GO:0006102 | isocitrate metabolic process |
| GO:0006417 | regulation of translation |
| GO:0009635 | RESPONSE TO HERBICIDE |
| GO:0010447 | response to acidic pH |
| GO:0034220 | ION TRANSMEMBRANE TRANSPORT |
| GO:0055085 | transmembrane transport |
| GO:0006818 | hydrogen transport |
| GO:0015797 | mannitol transport |
| GO:0042026 | PROTEIN REFOLDING |

**Figure 4:** *Intended Tree.*



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

### 4.2 Implicit Visualizations

**4.2.1 Tree-maps.** Tree-maps are a space-filling technique, also known as implicit, used to represent hierarchical structures, proposed in 1992 and are described in [55][67][61]. Tree-maps apply a recursive algorithm for visualizing nested rectangles. The tree-map uses outer rectangle as a tree's root and the inner space of this rectangle is filled with nested rectangles representing the parent/children relationship. This space is divided between parent nodes according to its assigned weight in the shape of rectangles, and each parent is subdivided into the amount of related children as further rectangles (see Fig. 5). Alternative algorithms have been proposed [5][79], as the original method suffers from the creation of narrow rectangles that impair the visualization's readability. The original tree-map layout was "slice and dice"; the idea has since been extended with the development of the web-based tree-map by Wattenberg [75],the strip or ordered tree-map algorithm [3] and the spiral layout algorithm that enables the reader to see changes in hierarchical data [66].

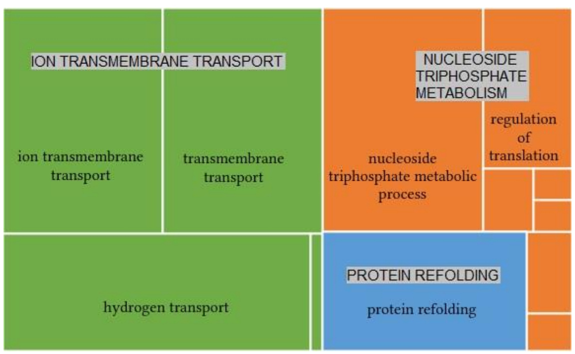



**Figure 5:** *Tree-map, R(treemaps).*

Tree-maps offer efficient usage of the available display space, and provides a good overview of the entire data hierarchy. The size of rectangles is relative to the size of the related data object, which simplifies data interpretation and evaluation. The color-coding helps to distinguish between different cluster groups and also helps show the relationship between children to parent nodes. The main graphical parameters for tree-map plotting are visualization area size, position and color-coding [71].

On the other hand, tree-map visualization becomes poor with the increase of input data size. While tree-map graphs still provide a data overview, supporting visualization objects such as labels cannot be drawn on small rectangles. Visualization of GO terms with tree-maps is an example of a using tree-map layout in biology [66].

**4.2.2 Sunbursts.** An alternative space-filling visualization is to represent data in a radial layout such as the Sunburst (see Fig.6) [65][29]. The hierarchy is represented from the center outwards from it. The inner circle is the root of the hierarchical data, and multiple layers of rings represent the parent-child relationships next to each other [67]. As the Sunburst is a circular space-filling technique, the edges of the provided display space are unused. The wedge size is relative to the cluster size. The interpretation of wedges sizes is relatively easy for the reader as each slice is represented in a familiar proportional way. However, in the case of narrow wedge sizes, the readability and evaluation of the visualization becomes poor. This leads to the similar problem of losing some graph labels, but can be addressed by using the empty space around the circular layout. As with tree-maps, the Sunburst uses colouring to improve readability of the visualization. Other space-filling visualization techniques available are the Voronoi diagram[2], Ellimaps [47], icicle plots [34] and Beamtree [17].

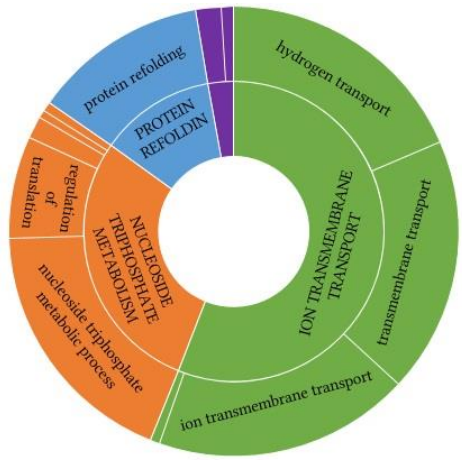



**Figure 6:** *Sunburst diagram.*

## 5 CONCLUSIONS

We reviewed the most common visualization techniques available for hierarchical structured data in 2D space within the scope of this paper. Within the conclusion section, we pinpoint to the 7 most relevant categories for classifying and characterizing biological visualizations to support the development of visualization taxonomies.

### 5.1 Visualization Technique

Visualization techniques for hierarchical structured data in 2D can be classified as explicit (dendrograms, circular tree, intended trees) and implicit (tree-maps, Sunburst) methods. All of them have advantages and disadvantages, and the choice of the most suitable visualization technique depends on the final representation goal. For example, space-filling techniques are the best for representing a global overview of final results. However, with the increasing size of the data, details such as labels are often omitted to avoid cluttering the final picture. Tree-maps utilize the complete display space, while the Sunburst uses only part of it. On the other hand, the Sunburst provides a more intuitive understanding of the relationship between data values due to proportional representation of relationships [29]; it is harder to see the size difference between rectangles in tree-maps.

### 5.2 Visualization Design

In addition to appropriate selection of the visualization method, it is important to apply suitable supporting visualization features:



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

(1) applying color-coding;
(2) providing legend information if necessary;
(3) ordering results appropriately;
(4) zooming into a region of interest;
(5) displaying additional supporting parameters (e.g. numerical proportions of pie slices); and
(6) using the same font for labeling, and others.

The appropriate utilization of such features makes visualization more intuitive to comprehend.

## 5.3 Interactive Multimedia Features

Modern technologies provide various techniques for exploring big data in real-time, such as interactive methods and web-based visualizations. However, representation of final results for big datasets in 2D space remains a challenge. In this review, we focus on hierarchical structured data specifically visualization methods. Taking into consideration the best features of existing techniques and applying them into development of new visualization methods may help in representing big data results in a clear, informative way.

## 5.4 Primary Visualization Tasks

Visualization aims to facilitate perception, comprehension of the data. The primary visualization tasks are presented in Table 1, which enable the exploration of data, and decision making during the analysis process of biological data.

## 5.5 Algorithms and Data Processing

Algorithms and pipelines involved in the analysis process as e.g. clustering (see Section 3).

## 5.6 Data Representation, Storage, and Query

Another issue to be addressed, is data representation, especially regarding the particularities of the organization of hierarchical data as tree-like structures. Table 3 lists these relevant features.

## 5.7 Software Tools and Analysis Pipelines

Utilized software tools as e.g. Web3DMol and analysis pipelines as listed in Table 1.

## 5.8 Future Work

In future works, we will focus on the development of a taxonomy for reviewing and classifying visual techniques of hierarchical biological data. We will utilize the criteria for the taxonomy, listed within the conclusion section.

# REFERENCES


[1] Aggarwal, C.C. and Reddy, C.K. 2013. *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC.

[2] Balzer, M., Deussen, O. and Lewerentz, C. 2005. Voronoi Treemaps for the Visualization of Software Metrics. *Proceedings of the 2005 ACM Symposium on Software Visualization* (St. Louis, Missouri, 2005), 165–172.

[3] Bederson, B.B., Shneiderman, B. and Wattenberg, M. 2002. Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies. *ACM Trans. Graph.* 21, 4 (Oct. 2002), 833–854.

[4] Boser, B.E., Guyon, I.M. and Vapnik, V.N. 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (Pittsburgh, Pennsylvania, USA, 1992), 144–152.

[5] Bruls, M., Huizing, K. and Wijk, J.J. van 2000. Squarified Treemaps. *Eurographics / IEEE VGTC Symposium on Visualization* (2000).

[6] Celko, J. 2004. Joe Celko's Trees and Hierarchies in SQL for Smarties, (The Morgan Kaufmann Series in Data Management Systems) (The Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann.

[7] Daniel H. Huson, C.S. Regula Rupp 2010. *Phylogenetic Networks Concepts, Algorithms and Applications*. Cambridge : Cambridge University Press.

[8] Date, C.J. 2003. *An Introduction to Database Systems*. Addison-Wesley Longman Publishing Co., Inc.

[9] Eades, P. 1993. Algorithms for drawing trees. *CM SIGACT News*. (1993).

[10] Eddy, S.R. 2004. What is a hidden Markov model? *Nature Biotechnology*. 22, 10 (Oct. 2004), 1315–1316.

[11] Farris, J.S. 1970. Methods for Computing Wagner Trees. *Systematic Biology*. 19, 1 (Mar. 1970), 83–92.

[12] Ferdosi, B.J. and Roerdink, J.B.T.M. 2011. Visualizing High-dimensional Structures by Dimension Ordering and Filtering Using Subspace Analysis. *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization* (Bergen, Norway, 2011), 1121–1130.

[13] Fitch, W.M. 1971. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Biology*. 20, 4 (Dec. 1971), 406–416.

[14] Fling, K. Four Ways To Work With Hierarchical Data.

[15] Gonen, A., Rosenbaum, D., Eldar, Y.C. and Shalev-Shwartz, S. 2016. Subspace Learning with Partial Information. *Journal of Machine Learning Research*. 17, 52 (2016), 1–21.

[16] Greenacre, M. Hierarchical cluster analysis.

[17] Ham, F. van and Wijk, J.J. van 2002. Beamtrees: compact visualization of large hierarchies. *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.* (2002), 93–100.

[18] Holzinger, A. 2014. Biomedical Informatics: Discovering Knowledge in Big Data. Springer.

[19] Holzinger, A. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*. 3, 2 (Jun. 2016), 119–131.

[20] Holzinger, A., Plass, M., Holzinger, K., Crisan, G.C., Pintea, C.-M. and Palade, V. 2017. A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop. *CoRR*. abs/1708.01104, (2017).

[21] Huerta-Cepas, J., Serra, F. and Bork, P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular biology and evolution*. 33, 6 (Jun. 2016), 1635–1638.

[22] Hund, M., Böhm, D., Sturm, W., Sedlmair, M., Schreck, T., Ullrich, T., Keim, D.A., Majnaric, L. and Holzinger, A. 2016. Visual analytics for concept exploration in subspaces of patient groups. *Brain Informatics*. 3, 4 (Dec. 2016), 233–247.

[23] Hund, M., Sturm, W., Schreck, T., Ullrich, T., Keim, D., Majnaric, L. and Holzinger, A. 2015. Analysis of Patient Groups and Immunization Results Based on Subspace Clustering. *Brain Informatics and Health: 8th International Conference, BIH 2015, London, UK, August 30 - September 2, 2015. Proceedings*. Y. Guo, K. Friston, F. Aldo, S. Hill, and H. Peng, eds. Springer International Publishing. 358–368.

[24] Jacobson, R.E. and Jacobson, R. 2000. *Information Design*. MIT Press.

[25] Abu-Jamous, A.K.. Basel. Fa Rui. Nandi 2015. *Integrative Cluster Analysis in Bioinformatics.* Somerset : Wiley.

[26] Abu -Jamous Basel, N.A.K. Fa Rui 2015. *Integrative Cluster Analysis in Bioinformatics*. Wiley.

[27] Jeanquartier, F., Jeanquartier, C. and Holzinger, A. 2015. Visualizing Uncertainty of RNA Sequence Base Pairing Variants. (2015).

[28] Jiang, M.Q.. Rui. Zhang Xuegong. Zhang 2013. Basics of Bioinformatics Lecture Notes of the Graduate Summer School on Bioinformatics of China. Berlin : Springer Berlin Heidelberg.

[29] John Stasko, K.M. Richard Catrambone Mark Guzdial 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human - Computer Studies*. 53(5), (2000), 663–694.

[30] Kannan, L. and Wheeler, W.C. 2012. Maximum Parsimony on Phylogenetic networks. *Algorithms for molecular biology : AMB*. 7, 1 (May. 2012), 9+.




In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

[31] Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler and D. 2002. The Human Genome Browser at UCSC. *Genome Research*. 12, 6 (Jun. 2002), 996–1006.

[32] Keogh, E. and Mueen, A. 2010. Curse of Dimensionality. *Encyclopedia of Machine Learning*. C. Sammut and G.I. Webb, eds. Springer US. 257–258.

[33] Kreft, Ł., Botzki, A., Coppens, F., Vandepoele, K. and Van Bel, M. 2017. PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics*. 33, 18 (2017), 2946–2947.

[34] Kruskal, J.B. and Landwehr, J.M. 1983. Icicle Plots: Better Displays for Hierarchical Clustering. *The American Statistician*. 37, 2 (1983), 162–168.

[35] Kuznetsova, I., Karpievitch, Y.V., Filipovska, A., Lugmayr, A. and Holzinger, A. 2017. Review of Machine Learning Algorithms in Differential Expression Analysis. *ArXiv e-prints*. (Jul. 2017).

[36] Loewenstein, Y., Portugaly, E., Fromer, M. and Linial, M. 2008. Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space. *Bioinformatics*. 24, 13 (Jul. 2008), i41–i49.

[37] Lugmayr, A. 2007. Ambience, ambience, ambience - What are Ambient Media? *Interactive TV: A Shared Experience, TISCP Adjunct Proceedings of EuroITV 2007* (Amsterdam, 2007).

[38] Lugmayr, A., Risse, T., Stockleben, B., Kaario, J. and Laurila, K. 2009. Special issue on semantic ambient media experiences. *Multimedia Tools and Applications*. 44, 3 (2009), 331–335.

[39] Lugmayr, A., Serral, E., Scherp, A., Pogorelc, B. and Mustaquim, M. 2013. Ambient media today and tomorrow. *Multimedia Tools and Applications*. (2013), 1–31.

[40] Lugmayr, A., Stockleben, B., Scheib, C. and Mailaparampil, M. 2017. Cognitive Big Data. Survey and Review on Big Data Research and its Implications: What is Really "New"? Cognitive Big Data! *Journal of Knowledge Management (JMM)*. 21, 1 (2017).

[41] Lugmayr, A., Stockleben, B., Scheib, C., Mailaparampil, M., Mesia, N. and Ranta, H. 2016. A Comprehensive Survey on Big Data Research and It's Implications - What is really 'new' in Big Data? It's Cognitive Big Data. *Proceedings of the 20th Pacific-Asian Conference on Information Systems (PACIS 2016)* (2016).

[42] Mazza, R. 2009. Introduction to Information Visualization. Springer.

[43] Müller, E., Assent, I., Krieger, R., Jansen, T. and Seidl, T. 2008. Morpheus: Interactive Exploration of Subspace Clustering. *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery in Databases (KDD 2008), Las Vegas, USA* (New York,NY,USA, 2008), 1089–1092.

[44] Müller, J.S., Bünau, P. von, Meinecke, F.C., Király, F.J. and Müller, K.-R. 2011. The Stationary Subspace Analysis Toolbox. *J. Mach. Learn. Res.* 12, (Nov. 2011), 3065–3069.

[45] Nielsen, P. and Parui, U. 2009. *Microsoft SQL Server 2008 Bible*. Wiley Publishing.

[46] Noble, W.S. 2006. What is a support vector machine? *Nature Biotechnology*. 24, 12 (2006), 1565–1567.

[47] Otjacques, B., Cornil, M., Noirhomme, M. and Feltz, F. 2009. CGD – A New Algorithm to Optimize Space Occupation in Ellimaps. *Human-Computer Interaction – INTERACT 2009: 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedings, Part II*. T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R.O. Prates, and M. Winckler, eds. Springer Berlin Heidelberg. 805–818.

[48] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 25, 13 (2004), 1605–1612.

[49] Pogorelc, B., Vatavu, R.-D., Lugmayr, A., Stockleben, B., Risse, T., Kaario, J., Lomonaco, E. and Gams, M. 2012. Semantic ambient media: From ambient advertising to ambient-assisted living. *Multimedia Tools and Applications*. 58, 2 (2012), 399–425.

[50] Porollo, A. and Meller, J. 2007. Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinformatics*. 8, 1 (Aug. 2007), 316.

[51] Puga, J.L., Krzywinski, M. and Altman, N. 2015. Points of significance: Bayesian statistics. *Nature Methods*. 12, 5 (Apr. 2015), 377–378.

[52] Ramakrishnan, R. and Gehrke, J. 2003. *Database Management Systems*. McGraw-Hill, Inc.

[53] Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. 2011. Integrative genomics viewer. *Nature Biotechnology*. 29, 1 (Jan. 2011), 24–26.

[54] Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*. 4, 4 (Jul. 1987), 406–425.

[55] Schulz, H. H. & Schumann 2006. Visualizing Graphs - A Generalized View. *Tenth International Conference on. IEEE*. (2006).

[56] Severin, J., Lizio, M., Harshbarger, J., Kawaji, H., Daub, C.O., Hayashizaki, Y., Consortium, T.F., Bertin, N. and Forrest, A.R.R. 2014. Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat Biotech*. 32, 3 (Mar. 2014), 217–219.

[57] Sharma, A., Boroevich, K.A., Shigemizu, D., Kamatani, Y., Kubo, M. and Tsunoda, T. 2017. Hierarchical Maximum Likelihood Clustering Approach. *IEEE Transactions on Biomedical Engineering*. 64, 1 (Jan. 2017), 112–122.

[58] Shepard, R.N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*. 27, 3 (Sep. 1962), 219–246.

[59] Shi, M., Gao, J. and Zhang, M.Q. 2017. Web3DMol: interactive protein structure visualization based on WebGL. *Nucleic Acids Research*. 45, W1 (2017), W523–W527.

[60] Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings of the 1996 IEEE Symposium on Visual Languages* (Washington, DC, USA, 1996), 336–.

[61] Shneiderman, B. 1992. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics (TOG)*. (1992).

[62] Simonic, K.M., Holzinger, A., Bloice, M. and Hermann, J. 2011. Optimizing long-term treatment of rheumatoid arthritis with systematic documentation. *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops* (May. 2011), 550–554.

[63] Sokal, R.R. and Michener, C.D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*. 28, (1958), 1409–1438.

[64] Spence, R. 2001. *Information Visualization*. Springer-Verlag.

[65] Stasko, J. and Zhang, E. 2000. Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations. *Proceedings of the IEEE Symposium on Information Vizualization 2000* (Washington, DC, USA, 2000), 57–.

[66] Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE*. 6, 7 (2011), 1–9.

[67] T. Von Landesberger, D.W. A.Kuijper T.Schreck J.Kohlhammer J. J.Van Wijk J.-D.Fekete D.W. Fellner 2011. Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. *Computer Graphics Forum*. 30, 6 (2011), 1719–1749.

[68] Tenenbaum, J.B., Silva, V. and Langford, J.C. 2000. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*. 290, 5500 (2000), 2319–2323.

[69] Tropashko, V. 2005. Nested Intervals Tree Encoding in SQL. *SIGMOD Rec.* 34, 2 (Jun. 2005), 47–52.

[70] Tropashko, V. 2002. Trees in SQL: Nested Sets and Materialized Path.

[71] Tu, Y. and Shen, H.W. 2007. Visualizing Changes of Hierarchical Data using Treemaps. *IEEE Transactions on Visualization and Computer Graphics*. 13, 6 (Nov. 2007), 1286–1293.

[72] Tufte, E.R. 1990. *Envisioning information*. Graphics Press.

[73] Turkay, C., Jeanquartier, F., Holzinger, A. and Hauser, H. 2014. On Computationally-Enhanced Visual Analysis of Heterogeneous Data and Its Application in Biomedical Informatics. *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics: State-of-the-Art and Future Challenges*. A. Holzinger and I. Jurisica, eds. Springer Berlin Heidelberg. 117–140.

[74] Turkay, C., Laramee, R. and Holzinger, A. 2017. On the Challenges and Opportunities in Visualization for Machine Learning and Knowledge Extraction: A Research Agenda. *Machine Learning and Knowledge Extraction: First IFIP TC 5, WG 8.4, 8.9, 12.9 International Cross-Domain Conference, CD-MAKE 2017, Reggio, Italy, August 29 – September 1, 2017, Proceedings*. A. Holzinger, P. Kieseberg, A.M. Tjoa, and E. Weippl, eds. Springer International Publishing. 191–198.

[75] Wattenberg, M. 1999. Visualizing the Stock Market. *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, 1999), 188–189.

[76] Wong, K.-C. 2016. Computational Biology and Bioinformatics : Gene Regulation. Raton : CRC Press.

[77] Yellen Jay, R.K.B. Maurer StephenB 2003. *Handbook of Graph Theory*. CRC Press.

[78] Zhang, H., Gao, S., Lercher, M.J., Hu, S. and Chen, W.-H.H. 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic acids research*. 40, Web Server issue (Jul. 2012), W569–W572.

[79] 1999. SmartMoney map of the market.(Interactive Media Design Review 1999 gold winner). *I.D.* 46, 4 (1999), 62(2).

[80] The simplest(?) way to do tree-based queries in SQL.

[81] Troels' links: Relational database systems.

[82] What are the options for storing hierarchical data in a relational database?



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# Children Road Safety Training with Augmented Reality (AR) [Demo]

Artur Lugmayr, Joyce Tsang, Will Hobs, Toby Williams, Casey Wei Lim, Yeet Yung Teo, Matthew Farmer[1]

[1]Curtin University, Perth, AUSTRALIA
artur.lugmayr@curtin.edu.au and artur.lugmayr@artur-lugmayr.com
www.curtin.edu.au/vismedia

## ABSTRACT

Children killed or seriously injured through road accidents can be avoided through an appropriate safety training. Through play and engagement children learn and understand hazards at i.e. railway stations, bus stops, crossings, school zones, train stations, footpaths, or while cycling. We developed a rapid prototype of an *Augmented Reality (AR)* safety training proof-of-concept demonstrator for a scaled real-world model of dangerous road hazards. Two scenarios have been picked to give children the possibility to apply, and acquire knowledge of road safety: 1. handling emergency situations and informing authorities; 2. correct behavior at a bus stop on arrival/departure of a bus. In this paper we discuss our design approach, outline the technical implementation of the system, and give a brief overview of our lessons learned.

## KEYWORDS

Augmented Reality (AR), child road safety, Virtual Reality (VR), safety training, pay, engagement, traffic safety, road safety.

## 1 INTRODUCTION

It's common knowledge, that children are the most vulnerable pedestrians, cyclists, or vehicle passengers in road traffic. The World Health Organization (WHO) estimates that 30% of injury deaths of the 0-19 years old is due to traffic hazards, accounting for 2% of deaths globally. Dependent on country and region, this translates to 5-20 fatalities per 100.000 children [1]. In 2015, the average fatality rate across Australia was 5.1/100.000 for all age groups, where approx. 15% was related to children (21/100.000), in sum 20% of the total population [2]. These numbers illustrate the importance of training children at young age in traffic safety. While they "understand the danger but have little idea what to look for, and what to ignore" [3][4]. Therefore, it's important to develop a training environment, where children can be trained in hazardous situations – spotting dangers, and eventually avoiding fatal situations.

Using latest technologies, in particular AR enables the simulation of real-life hazardous traffic situations. Children are exposed to potential threatening situations and train how to react correctly. They learn to spot harmful situations, and how to respond. AR gives a realistic experience if combined with a scaled real-world model of train stations, bus stops, schools, or traffic crossings. Figure 1 illustrates the scaled real-world model of the Constable Care Child Safety Foundation (CCCSF) safety school, which has been opened in Perth, Western Australia [5]. The model is fully equipped with AR markers, WiFi, and interactive features, and can be considered as the world's first installation of this kind.

**Figure 1: Scaled real-world model built by the Constable Care Child Safety Foundation in Perth, WA, Australia equipped with dangerous road hazards (i.e. pedestrian crossing, bus-stop)** [5].

Through a cooperation between Curtin University's VisMedia Lab [6] and the CCCSF a student group has been working on creating design concepts, prototypes, and scenarios for the project as part of work integrated learning activities. To illustrate the potential of AR, we have been focusing on the development of an interactive scenario for the AR marker on the backside of the bus, which is illustrated in Figure 2. The goal of this paper is to:

1. introduce the design approach and technical challenges of the demonstrator;
2. describe the AR demonstration based on the marker on the bus (the creation of a minimal viable prototype); and
3. give a very brief overview of educational and training aspects of the demonstrator.

## 2 RELATED WORKS

Much work has been devoted to investigate AR (especially e.g. [7]) and smart, locative, and ubiquitous service systems



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

(especially e.g. [8] and [9]). However, research in converging child safety training and technologies such as AR and intelligent systems is rather scarce.


**Figure 2: AR marker on the scaled real-world model of a bus placed in the overall child safety training center.**

## 3 AR DEMONSTRATOR

The demonstrator was developed following a *Design Thinking* [10], [11] methodology, where a broad variety of ideas is generated at the beginning, and the best are selected for prototyping. After meetings with the clients, it was decided to develop a scenario for the AR tracker on the backside of the bus. Table 1 gives an overview of the results of the brainstorming sessions, and the scenario that we have been aiming at. The principle idea was to train children to call emergency services, when an emergency event occurs. In our case, a bus was on fire, and children have to react correctly.

**Table 1. Description of the Learning Event: 'Bus on Fire'.**

| | |
|---|---|
| **Learning Intensions** | Training children the right process to call emergency services |
| **Learning Experience** | Students should learn how to react in unforeseen emergency situations correctly, and memorize the number of the emergency services |
| **Description** | A bus is on fire, and children are provided with different choices how to react in this particular situation |
| **Key Messages** | Emergency services don't know your location, think before acting wrongly; what do emergency services typically ask; the number is not 911 |
| **Quiz/Questioner** | • Dialing the correct emergency number<br>• Emergency services know your location (Y/N)<br>• What is the best to do in an emergency? |
| **Information for Teachers** | Information materials about emergency services, how to dial numbers correctly, how to react correctly |

The implementation is running on Android OS, and was developed in Unity with the Vuforia software package. A tablet PC and its camera was used for placing the graphical content on the marker on the backside of the bus. The overall flow of the application is illustrated in Figure 3, and a few screenshots of the application are presented in Figure 4.

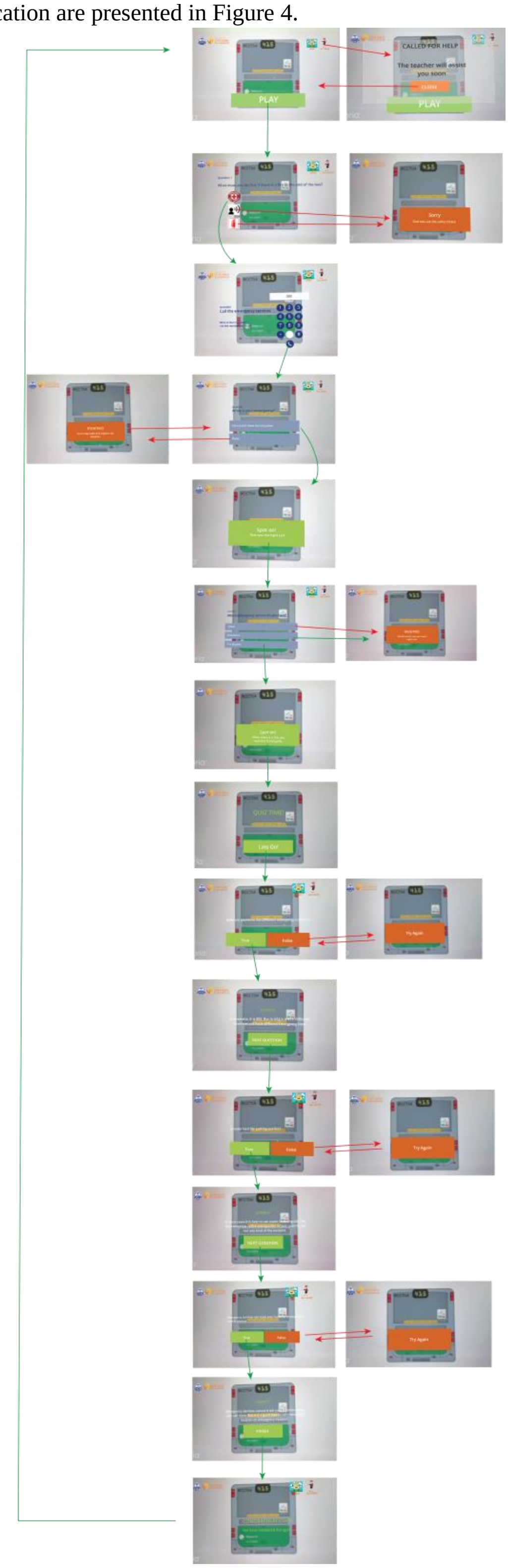



**Figure 3: Flow diagram of the demonstration.**



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

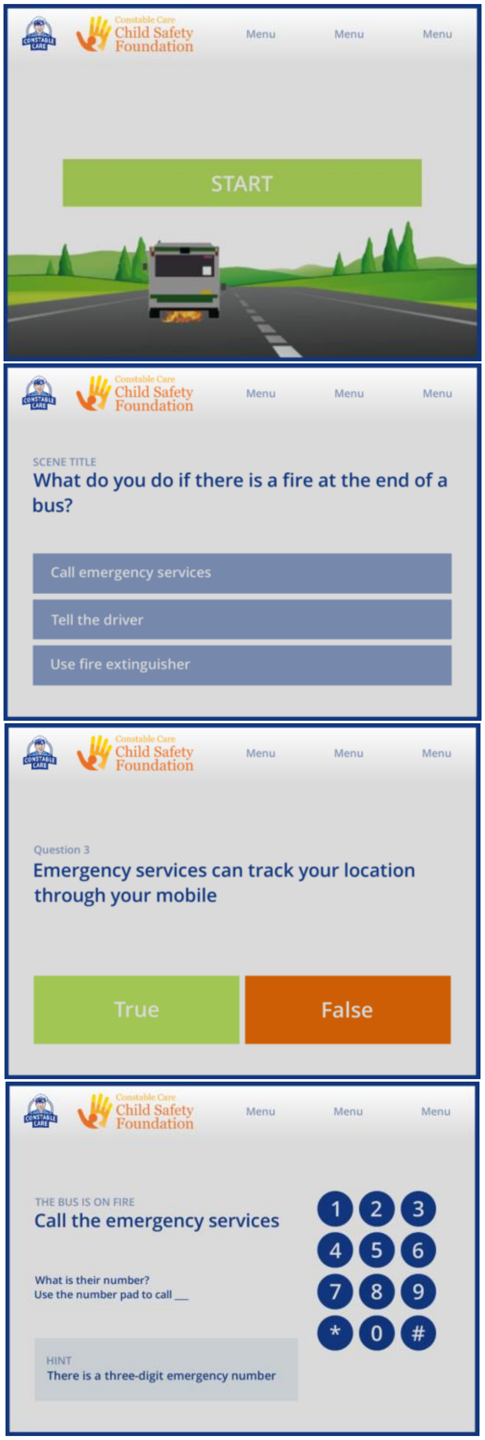



**Figure 4: Screenshots of the AR system.**

## 4 CONCLUSIONS

We strongly believe that this application will help children to understand the danger, but also train them in different ways how they can cope with the situation – e.g. to call emergency services, and ignore certain other non-viable alternatives. Through AR the learning experience is intensified, and real-life situations can be trained, especially when using a world scaled model of typical traffic hazards. Our future work will focus on the development of more scenarios, and data analytics to understand training performance and improve the learning experience.

## Acknowledgements

We would like to thank the Constable Care Child Safety Foundation (CCCSF) and DSBS Ltd. for their kind help and cooperation.

## REFERENCES

[1] WHO, "World Report on Child Injury Prevention." [Online]. Available: http://www.who.int/violence_injury_prevention/child/injury/world_report/Road_traffic_injuries_english.pdf

[2] R. S. Commission, "Reported Road Crashes in Western Australia (2015)." Government of Western Australia, 2016 [Online]. Available: https://www.rsc.wa.gov.au/RSC/media/Documents/Road%20Data/Statistics/Annual%20crash%20statistics/annual-crash-statistics-2015.pdf

[3] G. D. Pietro, "Child Development and Road Safety: A Report for the Motor Accidents Authority." 2004.

[4] Allianz, "Children and Road Safety." [Online]. Available: https://www.allianz.com.au/car-insurance/news/children-and-road-safety

[5] "Constable Care Child Safety Foundation (CCSF)." [Online]. Available: http://www.cccsf.org.au

[6] A. Lugmayr, "Visualisation and Interctive Media (VisMedia)." [Online]. Available: http://www.curtin.edu.au/vismedia

[7] M. Billinghurst, "Using augmented reality to create empathic experiences," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, 2014, pp. 5–6.

[8] M. Bilandzic and M. Foth, "A review of locative media, mobile and embodied spatial interaction," *International Journal of Human-Computer Studies*, vol. 70, no. 1, pp. 66–71, 2012.

[9] B. Pogorelc, R.-D. Vatavu, A. Lugmayr, B. Stockleben, T. Risse, J. Kaario, E. Lomonaco, and M. Gams, "Semantic ambient media: From ambient advertising to ambient-assisted living," *Multimedia Tools and Applications*, vol. 58, no. 2, pp. 399–425, 2012.

[10] A. Lugmayr, "Applying 'Design Thinking' As a Method for Teaching in Media Education," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 2011, pp. 332–334 [Online]. Available: http://doi.acm.org/10.1145/2181037.2181100

[11] A. Lugmayr, M. Jalonen, Y. Zou, L. Libin, and S. Anzenhofer, "Design Thinking" in Media Management Education - A Practical Hands-On Approach," in *4th Semantic Ambient Media Experience (SAME) Workshop in Conjunction with the 5th International Convergence on Communities and Technologies*, 2011.

# Data-Driven Approach to Human-Engaged Computing

**Xiaojuan Ma**
Hong Kong University of Science and Technology
Hong Kong SAR
mxj@cse.ust.hk

## ABSTRACT

This paper presents an overview of the research landscape of data-driven human-engaged computing in the Human-Computer Interaction Initiative at the Hong Kong University of Science and Technology.

## KEYWORDS

Engagement, data-driven, inference, analysis, application, machine intelligence, hybrid realities.

## 1 INTRODUCTION

Making technologies more engaging for human users is gaining increasing attention in academia and industry these days. It aims to equip computing devices and services with the abilities to perceive and understand users' attentional, emotional, cognitive, and behavioral engagement, as well as to manage and use such information to improve user interactions and experiences. In this paper, we present an overview of our research efforts on human-engaged computing (HEC) (Fig. 1). In particular, we take a data-driven approach to 1) inferring human engagement dynamicity from various signals, and 2) analyzing factors that engage users in online and offline activities in everyday life that can potentially be adopted in human-computer interaction (HCI). Then, we apply the resulting insights to two main application areas: achieving more engaging interaction experiences 3) with artificial intelligence (AI) via the design of emotionally and socially intelligent robots and agents; and 4) with non-AI entities (e.g., data and objects) via immersive hybrid realities. Last but not least, we experiment with the use of engagement ingredients to enrich user experiences in different contexts, ranging from education, e-commerce, health and wellbeing, to creativity.

## 2 ABOUT ENGAGEMENT

## 2.1 Definition of Engagement

In human-human interaction, engagement is "the process by which interactors start, maintain, and end their perceived connections to each other during an interaction" [21]. In the context of HCI, we take a more comprehensive definition that details the dimensions of engagement – considering it as "the attentional, emotional, cognitive, and behavioral connection that exists between a user and the task at hand at any point in time and possibly over time" ([22], adapted from [2], P. 2). Some literature takes a narrower definition of engagement that refers to only one of the following dimensions (Fig. 2).

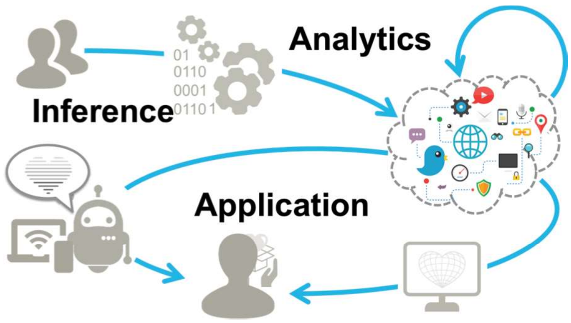



Figure 1: Overview of our research on data-driven human-engaged computing.

*Attentional Engagement* concerns attention allocation and redistribution (e.g., [4]), which can be measured by gaze data collected via eye tracking.

*Emotional engagement* concerns users' affective reactions such as interest, excitement and boredom (e.g., [10]), commonly measured using subjective questionnaires. Recently, researchers have explored the use of sensors to detect users' affective states.

*Cognitive engagement* concerns psychological devotion to a task (e.g., [13]), such as active thinking and reflection, and can be measured by sensors like EEG (objective) or by self-reporting (subjective).

*Behavioral engagement* concerns physical participation and involvement (e.g., [7]), often measured by attendance, time spent, number of actions, and number of attempts.

## 2.2 Human-Engaged Computing

With the technology advancement, researchers and practitioners start looking into how to design synergized interactions between humans and technologies, with the goal of augmenting the capabilities of both parties and maximizing their capacities. In the article "Rethinking the Relationship between Humans and Computers", Xiangshi Ren proposes that human-engaged computing aims to achieve "a state of optimal balance between engaged humans and engaging computers" [20].



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

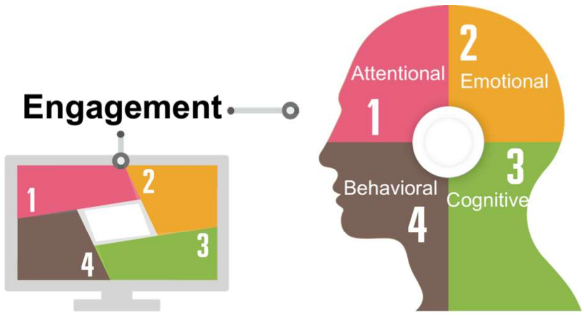



Figure 2: Components of Human-Engaged Computing (HEC): engaged human and engaging computer.

There are two essential research components of this notion: *engaged humans* and *engaging computers* (Fig. 2). The former requires the ability for computers (in a general sense) to perceive and understand the states and changes of human engagement. The latter demands feature(s) in a technological design that can motivate and facilitate humans to establish connections with it. Our group proposes to take a data-driven approach, i.e., leveraging rich data about humans, computers, and assorted interactions from diverse sources, to infer user engagement dynamicity and analyze engaging factors for design inspiration. Based on the derived insights, we further explore the creation of engaging experience with AI systems and non-AI entities.

## 3 RESEARCH LANDSCAPE OF HEC

### 3.1 Inference (Engaged Humans)

The main goal of the *Inference* research component is to sense and model human engagement dynamicity in real-time, i.e., the states, transitions, and fluctuations of engagement – a single dimension or multiple dimensions as an integrated measure. We have been exploring the use of three types of signals as engagement cues: social signal, physiological signal, and behavioral signal.

*Social signals* are "communicative or informative signals that, either directly or indirectly, provide information about social facts, namely social interactions, social emotions, social attitudes, or social relations" [26]. In other words, social signals are indicators of attentional, emotional, and even cognitive engagement in interpersonal interactions, which can potentially be transferred to other interaction contexts. Common social signals include gaze, facial expressions, vocal behaviors, proxemics, gesture, posture, and other body languages. Our group investigates non-intrusive methods to capture social signals during an interaction, using sensors like cameras and microphones. Fig. 3 shows an example setup of a posture detection system [34]. In actual applications, the sensors are better embedded in the environment.

*Physiological signals* are readings produced by physiological processes of human beings, including but not limited to heartbeat rate (ECG/EKG signal), respiratory rate and content (capnogram), skin conductance (EDA signal), muscle current (EMG signal), brain electrical activity (EEG signal), etc. Research has shown that physiological signals can serve as cues of attentional (e.g., [9]) and emotional engagement (e.g., [1]). We have previously experimented with extracting emotional cues from pulse [33] and skin conductance data [24] (Fig. 4)

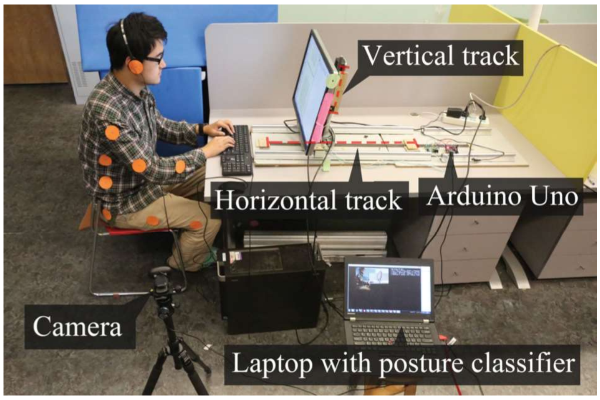



Figure 3: Example setup to detect posture [34].

*Behavioral signals* are actions and activities performed during an interaction or on an interface online or offline, such as conversational acts, text input, clicks, scrolling, page switch, emoji usage, likes, check-ins, etc. Some of these are task-specific, but they all suggest users' level of behavioral, cognitive, and sometimes emotional engagement. We have been mining behavioral signals from different sources, such as gameplay data [12], social media data [29], social commerce data [31], multimedia data [25], and public service data [6].

We can use the different types of engagement signals individually or collectively, according to the application context and needs.

### 3.2 Analytics (Engaging Computers)

The purpose of our *Analytics* research component is to identify engaging factors people may encounter in their everyday life. In particular, we are interested in what engages users when they interacting with other humans or with physical / virtual entities.

Our research on *engagement with social actors* concerns a) characteristics of individuals that tend to attract other people to

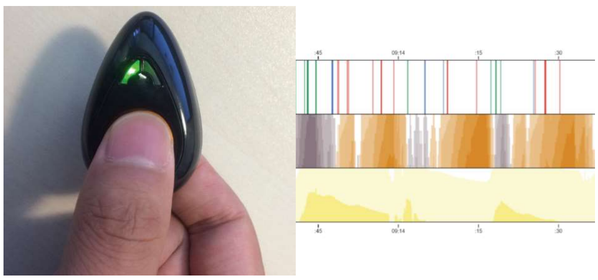


Figure 4: Monitoring skin conductance using Pip sensor and sample readings [24].



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

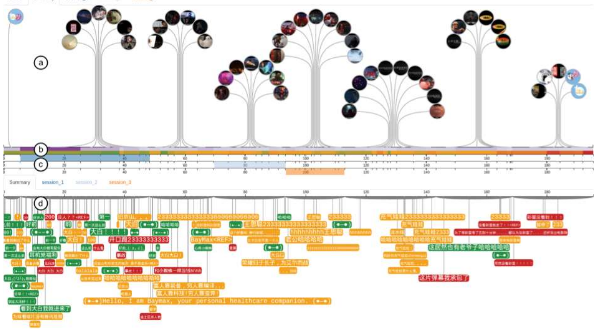


Figure 5: Analysis of viewer engagement during live video commenting [25]: (middle) video timeline; (top) summary of video scenes, height encoding intensity of behavioral engagement; (bottom) comments posted during video watching, color encoding sentiment.

Figure 6: Example of engagement sensing and handling in human-robot interaction; the user disengaged from his conversation with the robot to work on a task on the computer [23].

participate in activities with them e.g., personality [30] and structure of their intimacy network [11]; and b) strategies people used to directly or indirectly manage others' engagement [29] e.g., money gifting.

Our research on *engagement with physical and virtual entities* intends to capture tangible and nontangible properties that make something more engaging than others. Our studies have covered a wide spectrum of design space, ranging from appearance (e.g., cuteness [19]), layout (e.g., product arrangement on e-comerce site [31]), medium (e.g., food for social messaging [28] and data display [27]), to semantics (e.g., associations in humor [3]).

To give an example, Fig. 5 shows a visualization of viewer engagement on an online video sharing platform, by analyzing the live commenting data [25]. We can gain an idea about which part of a video is engaging and why, and how viewers engage one another to create an illusion of co-watching.

### 3.3 Application I: Engagement with AI Systems

Our first application of insights drawn from inference and analytics is designing engaging experiences with Artificial Intelligent (AI) systems, naming robots and virtual agents (HRI / HAI). More specifically, we would like to augment AI systems with emotional and social intelligence, enabling smoother, more effective, and more enjoyable human-AI collaboration.

For example, we compare the efficacy of two disengagement handling techniques (dominant / explicit versus submissive / implicit) adapted from human-human interaction, when employed by a physical robot to manage potential interaction breakdowns with a human user [23]. During the entire process, the robot closely monitors the user's shift of engagement by social and behavioral signals. We have also experiment with applying these two traits on a virtual agent to deal with user challenges such as verbal abuse and sexual harassment [30].

### 3.4 Application II: Engagement with Non-AI Entities

Our second application focuses on non-AI entities in a computing system, such as data and traditional interface elements. We use engaging factors (e.g., appearance, medium, semantic, etc.) identified previously to inform the design of these entities across the **reality–virtuality** continuum. In other words, a design can live in conventional digital media, in "ambient media" situated in everyday life [14], or a combination of both. As a result, these designs could be of better assistance in areas like informatics (e.g., healthcare [17] and wellness [24], photo archive [15], etc.), narratives (e.g., video synopsis [25], paper-craft [35], etc.), persuasion (e.g., volunteerism [8] and healthy aging [32]), and recommendation (e.g., e-commerce [31], travel [5], and transportation [6]).

In the work shown in Fig. 7, we use crowdsourcing techniques to infer users' allocation of attentional and cognitive engagement when looking at digital medical graphics [16]. Guided by the results, we can use optimization algorithms to automatically improve the perceptual effectiveness of the information display during physician-patient communication.

In another work, we conduct meta-synthesis of daily anecdotes about how breakage of technology may lead to engagement of (restoring, reinforcing, or promoting) online and offline interpersonal communication [18]. Based on the findings, we propose a Breakage-to-Icebreaker (B2I) design process, i.e., embedding icebreaking mechanisms into existing products and services to create opportunities for users to interact and reflect while enjoying the original functionalities (Fig. 8).

## 4 CONCLUSIONS

This paper provides a brief overview of the research on data-driven human-engaged computing (HEC) conducted in the Human-Computer Interaction Initiative at the Hong Kong University of Science and Technology. We summarize our works related to two essential components of HEC: engaged humans



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

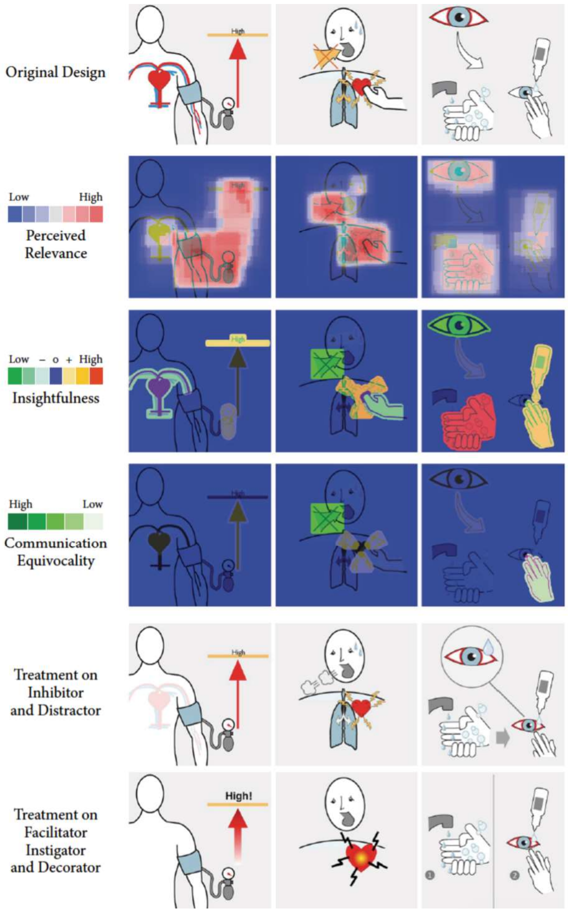



Figure 7: User attentional and cognitive engagement-guided design of medical graphics used in physician-patient communication [16].

(inference) and engaging computers (analytics). We also presents some exploratory applications of our research results, enabling more engaging synergized interactions between human users and AI systems as well as non-AI elements in a technology.

## ACKNOWLEDGMENTS

This work was partially supported by the HKUST Initiation Grant #IGN15EG02 and by the WeChat-HKUST Joint Laboratory on Artificial Intelligence Technology (WHAT LAB) grant #1516144-0 and #1617170-0.

## REFERENCES

[1] Bradley M Appelhans and Linda J Luecken. 2006. Heart rate variability as an index of regulated emotional responding. *Review of general psychology* 10, 3: 229.

[2] Simon Attfield, Gabriella Kazai, Mounia Lalmas, and Benjamin Piwowarski. 2011. Towards a science of user engagement. In *WSDM Workshop on User Modelling for Web Applications*. ACM.

[3] Andrew Cattle, and Xiaojuan Ma. 2016. Effects of Semantic Relatedness between Setups and Punchlines in Twitter Hashtag Games. *PEOPLES 2016*: 70.

[4] Peter Chapman. 1997. *Models of engagement: Intrinsically motivated interaction with multimedia learning software*. University


Figure 8: Prototypes of persuasive design for encouraging interpersonal interactions [18]: (S2D2) Cool Map, map sharing user-perceived temperature; (S1D1) WiFi Teeterboard, a WiFi redistribution switch; (S2D3) Sense Me Chat with Me, affective bench; and (S2D1) Remotouch, a collaborative air-conditional remote control.

of Waterloo.

[5] Chao Chen, Daqing Zhang, Bin Guo, Xiaojuan Ma, Gang Pan, and Zhaohui Wu. 2015. TripPlanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints. *IEEE Transactions on Intelligent Transportation Systems* 16, no. 3: 1259-1273.

[6] Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, Xiaojuan Ma, Shijian Li, Zhaohui Wu, Gang Pan, Thi-Mai-Trang Nguyen, and Jérémie Jakubowicz. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 841-852. ACM.

[7] Luis De-Marcos, Adri´an Dom´ınguez, Joseba Saenz-de Navarrete, and Carmen Pag´es. 2014. An empirical study comparing gamification and social networking on e-learning. Computers & Education 75 (2014), 82–91. DOI: http://dx.doi.org/10.1016/j.compedu.2014.01.012

[8] Don Sim Jianqiang, Xiaojuan Ma, Shengdong Zhao, Jing Ting Khoo, Swee Ling Bay, and Zhenhui Jiang. 2011. Farmer's tale: a facebook game to promote volunteerism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 581-584. ACM.

[9] Wolfgang Klimesch, Michael Doppelmayr, H. Russegger, Th Pachinger, and J. Schwaiger. 1998. Induced alpha band power changes in the human EEG and attention. *Neuroscience letters* 244, no. 2: 73-76.

[10] Brenda Laurel. 1993. *Computer as Theatre*. Addison-Wesley.

[11] Quan Li, Qiaomu Shen, Yao Ming, Peng Xu, Yun Wang, Xiaojuan Ma, and Huamin Qu. 2017. A Visual Analytics Approach for Understanding Egocentric Intimacy Network Evolution and Impact Propagation in MMORPGs. In *Proceedings of IEEE Pacific Visualization Symposium (PacificVis 2017)*, Seoul, South Korea.

[12] Quan Li, Peng Xu, Yeuk Yin Chan, Yun Wang, Zhipeng Wang, Huamin Qu, and Xiaojuan Ma. 2017. A visual analytics approach for understanding reasons behind snowballing and comeback in moba games. *IEEE transactions on visualization and computer graphics* 23, no. 1: 211-220.

[13] Wei Li, Tovi Grossman, and George Fitzmaurice. 2014. CADament: A Gamified Multiplayer Software Tutorial System. In *Proc. CHI 2014*. ACM, 3369–3378. DOI: http://dx.doi.org/10.1145/2556288.2556954

[14] Artuer Lugmayr. 2006. The future is 'ambient'. *Proceedings of SPIE* Vol. 6074, 607403 Multimedia on Mobile Devices II 6074. DOI: 10.1117/12.655169

[15] Artur Lugmayr, Adam Greenfeld, Andrew Woods, and Pauline

Joseph. 2016. Cultural Visualisation of a Cultural Photographic Collection in 3D Environments–Development of ‘PAV 3D’(Photographic Archive Visualisation). In *International Conference on Entertainment Computing*, pp. 272-277. G. Wallner, S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, and H.-S. Yang, Eds. Cham: Springer International Publishing, 2016, pp. 272–277 Online. Available: http://dx.doi.org/10.1007/978-3-319-46100-7_29
[16] Xiaojuan Ma. 2014. Data-driven techniques to assist design of perceptually effective visual displays. *Information Design Journal (IDJ)* 21, no. 3.
[17] Xiaojuan Ma. 2015. Developing Design Guidelines for a Visual Vocabulary of Electronic Medical Information to Improve Health Literacy. Interacting with Computers 28, no. 2: 151-169.
[18] Xiaojuan Ma, Ke Fang, and Fengyuan Zhu. 2016. From Breakage to Icebreaker: Inspiration for Designing Technological Support for Human-Human Interaction. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pp. 403-414. ACM.
[19] Aaron Marcus, Masaaki Kurosu, Xiaojuan Ma, Ayako Hashizume. 2017. *Cuteness Engineering: Designing Adorable Products and Services*. Springer Series on Cultural Computing. Springer International Publishing. DOI: 10.1007/978-3-319-61961-3
[20] Xiangshi Ren. 2016. Rethinking the Relationship between Humans and Computers. *IEEE Computer* 49, no. 8: 104-108.
[21] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1-2 (2005), 140–164. DOI: http://dx.doi.org/10.1016/j.artint.2005.03.005
[22] Chaklam Silpasuwanchai, Xiaojuan Ma, Hiroaki Shigemasu, and Xiangshi Ren. 2016. Developing a Comprehensive Engagement Framework of Gamification for Reflective Learning. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 459-472. DOI: https://doi.org/10.1145/2901790.2901836
[23] Mingfei Sun, Zhenjie Zhao, and Xiaojuan Ma. 2017. Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 556-567. DOI: https://doi.org/10.1145/3025453.3025469
[24] Zhida Sun, Nan Cao, and Xiaojuan Ma. 2017. Attention, Comprehension, Execution: Effects of Different Designs of Biofeedback Display. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2132-2139. ACM.
[25] Zhida Sun, Mingfei Sun, Nan Cao, and Xiaojuan Ma. 2016. VideoForest: interactive visual summarization of video streams based on danmu data. In *SIGGRAPH ASIA 2016 Symposium on Visualization*, p. 10. ACM,.
[26] Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schroeder. 2012. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing* 3, no. 1: 69-87.
[27] Yun Wang, Xiaojuan Ma, Qiong Luo, and Huamin Qu. 2016. Data Edibilization: Representing Data with Food. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 409-422. ACM, 2016.
[28] Jun Wei, Xiaojuan Ma, and Shengdong Zhao. 2014. Food messaging: using edible medium for social messaging. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 2873-2882. ACM.
[29] Ziming Wu, and Xiaojuan Ma. 2017. Money as a Social Currency to Manage Group Dynamics: Red Packet Gifting in Chinese Online Communities. In *Proceedings of the 2017 CHI Conference Extended Abstracts*, pp. 2240-2247. ACM.
[30] Yang Yang, Xiaojuan Ma, and Pascale Fung. 2017. Perceived Emotional Intelligence in Virtual Agents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2255-2262. ACM.
[31] Yanzhen Yue, Xiaojuan Ma, and Zhenhui Jiang. 2016. Influence of Content Layout and Motivation on Users' Herd Behavior in Social Discovery. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5715-5719. ACM.
[32] Mingqian Zhao, Zhutian Chen, Ke Lu, Chaoran Li, Huamin Qu, and Xiaojuan Ma. 2016. Blossom: Design of a Tangible Interface for Improving Intergenerational Communication for the Elderly. In *Proceedings of the International Symposium on Interactive Technology and Ageing Populations*, pp. 87-98. ACM.
[33] Bin Tina Zhu, Yanqing Zhang, Xiaojuan Ma, and Haibo Li. 2015. Bringing Chinese Aesthetics into Designing the Experience of Personal Informatics for Wellbeing. Design and semantics of form and movement 289.
[34] Fengyuan Zhu, Ke Fang, and Xiaojuan Ma. 2017. Exploring the Effects of Strategy and Arousal of Cueing in Computer-Human Persuasion. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2276-2283. DOI: https://doi.org/10.1145/3027063.3053122
[35] Kening Zhu, Rongbo Zhu, Hideaki Nii, Hooman Samani, and Borhan Brian Jalaeian. 2014. PaperIO: a 3D interface towards the internet of embedded paper-craft. *IEICE TRANSACTIONS on Information and Systems* 97, no. 10: 2597-2605.

# DupRobo: An Interactive Robotic Platform for Physical Block-based Autocompletion

Taizhou Chen, Yi-Shiun Wu, Feng Han, Baochuan Yue, Arshad Nasser, Kening Zhu*
School of Creative Media, City University of Hong Kong, Hong Kong

## Abstract

In this paper, we present DupRobo, an interactive robotic platform for tangible block-based design and construction. DupRobo supported user-customisable exemplar, repetition control, and tangible autocompletion, through the computer-vision and the robotic techniques. With DupRobo, we aim to reduce users' workload in repetitive block-based construction, yet preserve the direct manipulatability and the intuitiveness in tangible model design, such as product design and architecture design.

**Keywords:** Robotic, Block assembly, Physical autocompletion.

**Concepts: •Human-centered computing → Interactive systems and tools;**

## 1 Introduction

Assembly blocks (e.g. LEGO and DUPLO) have been a popular type of toys over time. They are also widely applied in various creative processes, such as product design [Lofaro et al. 2009] and architecture design [Smithwick et al. 2017]. Different from sketching which is thought of as 2D visual design thinking, physical block building, with its emphasis on assembly and manipulation, ought to be considered 3D physical design thinking a more tangible, interactive way of exploring designs [Smithwick et al. 2017].

Research [Smithwick et al. 2017] showed that this type of hands-on prototyping platform can retain users interest and attention, promote design creativity, and facilitate team work. Unlike sketching, which involves marking a 2D flat surface, physical model making takes place in three dimensional space and involves different forms of material interaction, can further facilitating problem solving.

However, it is still confusing and tedious for non-experienced users to construct large physical models [Strobel 2010], and it is even more difficult to create new models from scratch. One reason could be that a large complicated model often involves many repetitive parts. For example, the Parthenon model contains multiple similar pillar structures, and the Great Wall model consists of many embattlements. On the other hand, it is observable that highly creative patterns can be generated through controlling the repetition of a primitive exemplar. The exemplar repetition, also known as autocompletion, has been widely supported in many 2D/3D design softwares [Kazi et al. 2012; Xing et al. 2014; Peng et al. 2017]. But interacting with a 3D model virtually can be far less intuitive than actually making the physical model [Ishii and Ullmer 1997]. In this research, we aim to reduce the manual workload in repetitive block assembly through physical autocompletion, yet preserve the direct manipulatablity and the intuitiveness in tangible modelling.

In this paper, we present DupRobo, an interactive robotic platform for tangible block-based design and construction. DupRobot supported user-customisable exemplar, repetition control, and physical autocompletion. As shown in Fig. 1, the setup of DupRobot consists of a Kinect sensor above the construction platform to track the user input (i.e. the exemplar building and the paper-based commands), and a robotic arm for automatic construction. Here we work only with the DUPLO blocks with a 2 2 arrangement of studs in the current prototype, to keep the problem computationally tractable. We didn't take the colors of the blocks as an input parameter, since our current main focus is to facilitate the physical structural autocompletion. While robotic assistants have been widely applied to many industrial domains to reduce the repetitive workload of human, e.g. component assembling, they were rarely used to facilitate tangible creative processes.

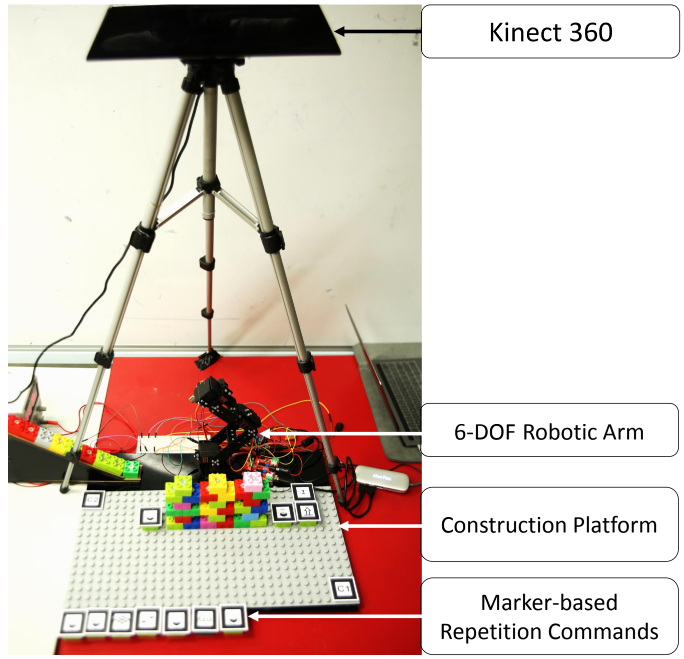



**Figure 1:** *DupRobo setup.*

Fig. 2 shows the processing flow of DupRobo. The user first constructs a physical block-based model as the physical exemplar. The exemplar building process is captured and tracked using the Kinect sensor above the construction platform on the table. The DupRobo system provides a set of paper-based commands for users to create the repetitive pattern, and controls the robotic arm to build the physical model. By repeating a physical exemplar, DupRobo facilitates users to physically create complex structures. Fig. 3 shows a list of examples constructed by DupRobo with various combinations of physical exemplars and paper-based repetition commands.Furthermore, the native support of assembling and disassembling in block-based construction allows users to modify the physical autocompleted model.

## 2 Related Work

DupRobo was inspired by the existing works on the 2D drawing autocompletion, the 3D block-assembly tracking, and the interaction

*e-mail: keninzhu@cityu.edu.hk



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

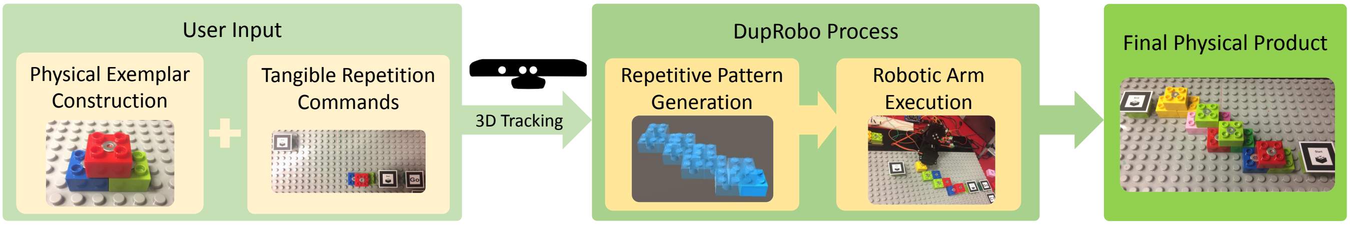



**Figure 2:** *System diagram of DupRobo.*


**Figure 3:** *Examples of physical exemplars and repetitive structures physically autocompleted by DupRobo.*

with robotic assistants.

### 2.1 2D Drawing Autocompletion

The research on automatic repetitions of visual patterns has been greatly advanced in recent data-driven methods [Kazi et al. 2012; Xing et al. 2014]. These works impose a list of sequential orders with the user-defined exemplars to be cloned to the desired output regions through various gestural commands, such as brushes. Similar autocompletion techniques were recently applied on 3D surface sculpting [Peng et al. 2017]. Although the autocompleted 3D virtual surface can be physicalised through 3D printing, it still requires users to edit the 3D model in the graphical user interfaces which could be less direct or intuitive than the tangible user interfaces. Taking one step further, DupRobo support the physical block-based modelling autocompletion with tangible commands.

### 2.2 Tracking and Generating Block Assembly

There has also been a series of research on tracking block-assembly procedures. Miller et al. [Miller et al. 2012] developed a Kinect-based system to track how a DUPLO model is built,with the assumption that the model always stays with its base on the table to reduce the tracking to 3 degrees of freedom (DOF). Gupta et al. [Gupta et al. 2012] presented a real-time system which can track an assembly process of Duplo blocks in 6-DOF. In DupRobot, similar to Miller et al., we assumed users build the exemplar on the table block by block, and tracked the building process using Kinect.

### 2.3 Robotic Assistant

Last but not the least, DupRobo is strongly inspired by the recent development in interacting with robotic assistants for industrial and in-home purposes. Zhao et al. [Zhao et al. 2009] utilized AR markers to control house-keeping robots. More recently, Sefidgar et al. [Sefidgar et al. 2017] developed a set of physical blocks with visual markers for robot programming in a pick-and-place task context. Their studies proved the high intuitiveness and learnability of situated tangible programming for robotic assistants. Thus, the similar interaction techniques was adopted in DupRobo. For creative design process, Smithwick et al. [Smithwick et al. 2017] envisioned an intelligent robotic platform that assists architects in tangible prototype design. DupRobo was directly motivated by this vision, and stepped further with tangible control interface for the robot.

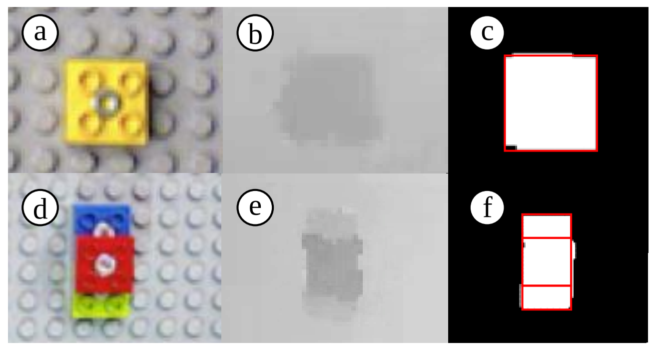



**Figure 4:** *Block tracking on the construction platform: (a) RGB image of the first block (b) Depth image of the first block (c) binary image of the first block after retangle detection (d) - (f) Tracking of a completed exemplar.*

## 3 System Description

In the current DupRobo prototype, we used a grey DUPLO board with 32 x 16 studs as the construction platform, and the non-grey DUPLO blocks as the construction pieces. Each DUPLO block was attached with a 1mm-thick iron plate on its surface, for the "grabbing" action of the electromagnet in the robotic arm. There are three main technical components in DupRobo: 3D tracking of the physical exemplar , the paper-based repetition commands, and the robotic arm for the physical autocompletion.

### 3.1 Physical Exemplar Tracking

Since only 2 x 2 non-grey square DUPLO blocks are used, the computer-vision component performs the process of rectangle and color detection when the user places the blocks on the construction platform. Firstly, the RGB image (Fig. 4 (a))captured by the Kinect sensor is cropped to the region of the construction platform. The cropped image is then converted to the HSV space, and thresholded to eliminate the grey color. Next, image dilation and erosion are performed to obtain the binary image which will then go through the contour-detection process. Finally, the edge filter, the angle fil-



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

ter, and the area filter are set to obtain the proper rectangle of a newly placed block (Fig. 4(c)). The on-screen position of the block is then used to update the virtual construction platform by adding a new virtual block accordingly. If there has already been blocks overlapped with a particular location, the new block will be stacked on top of the existing block. Additionally, the Kinect depth image is used along with the RGB image to detect the newly placed block when it is placed on an existing block with the same color.

### 3.2 Paper-based Repetition Commands

Once the user finishes building the physical exemplar, he/she can then indicate the repetition patterns placing the paper-based commands on the construction platform. We designed seven paper-based markers (Fig. 5) as the repetition commands. The START (Fig. 5(a)) and the END markers (Fig. 5(b)) represent the starting position and the ending position that will be connected by repeating the exemplar. When there are more than two pairs of START and END markers on the platform, the repetition of the exemplar can form a closed shape which can be filled with the exemplar using the Fill marker (Fig. 5(c)). The "Upper Arrow" marker (Fig. 5(d)) indicates the upward repetition perpendicular to the construction platform. Placed along with the upward marker, Fig. 5(e) and (f) indicates repeat two and three times of the exemplar upward. When the user finishes planning the repetition with the markers, he/she places the Go marker (Fig. 5(g)) on the platform. Detecting the Go marker, the DupRobo backend algorithm will generate a virtual 3D preview of the repetition model as shown in the "Repetitive Pattern Generation" box in Fig. 2, and control the robotic arm to construct the model piece-by-piece layer-by-layer.

### 3.3 Robotic Arm for Model Construction

The 6-DOF robotic arm consists of five off-the-shelf servo motors (Model No.: KS-3518) and one 12V electromagnet as the end effector. When grabbing the DUPLO blocks, the electromagnet will be turned on by the transistor-controlled switch circuit. The five servo motors were controlled by two Arduino boards with motor shields, and the angle of each motor was calculated in real-time using the position of the current to-built piece based on the inverse kinetics.

## 4 Social Impact and Future Vision

While DupRobo is inspired by Smithwick et al.'s vision of robotic design assistant, we further envision that the robotic assistant in tangible creative process could duplicate the designer's initial design, and perform iterations of the initial design, enabling rapid prototyping of different design configuration. In addition, the robotic assistant can experiment and construct complex possibilities that are difficult for manual efforts , and further provide new design suggestions to the designer. As the future work, we will incorporate multiple robotic arms to construct more complex models.

## Acknowledgements

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU21200216), and partially supported by a grant from City University of Hong Kong (Project No. 7004582). Special thanks to Mr. Dan Couture for his voice-over for the video.

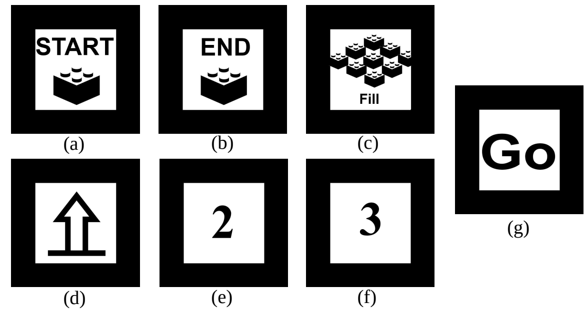



**Figure 5:** *Maker patterns for paper-based repetition commands*

## References

GUPTA, A., FOX, D., CURLESS, B., AND COHEN, M. 2012. Duplotrack: a real-time system for authoring and guiding duplo block assembly. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, ACM, 389–402.

ISHII, H., AND ULLMER, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, ACM, 234–241.

KAZI, R. H., IGARASHI, T., ZHAO, S., AND DAVIS, R. 2012. Vignette: interactive texture design and manipulation with freeform gestures for pen-and-ink illustration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1727–1736.

LOFARO, D. M., LE, T. T. G., AND OH, P. 2009. *Mechatronics Education: From Paper Design to Product Prototype Using LEGO NXT Parts*. Springer Berlin Heidelberg, Berlin, Heidelberg, 232–239.

MILLER, A., WHITE, B., CHARBONNEAU, E., KANZLER, Z., AND LAVIOLA JR, J. J. 2012. Interactive 3d model acquisition and tracking of building block structures. *IEEE transactions on visualization and computer graphics 18*, 4, 651–659.

PENG, M., XING, J., AND WEI, L.-Y. 2017. Autocomplete 3d sculpting. *arXiv preprint arXiv:1703.10405*.

SEFIDGAR, Y. S., AGARWAL, P., AND CAKMAK, M. 2017. Situated tangible robot programming. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 473–482.

SMITHWICK, D., KIRSH, D., AND SASS, L. 2017. Designerly pick and place: Coding physical model making to inform material-based robotic interaction. In *Design Computing and Cognition '16*, Springer International Publishing, Cham, J. S. Gero, Ed., 419–436.

STROBEL, J. 2010. *All the better to see you with: A comparison of approaches to delivering instructions for Lego construction tasks*. PhD thesis, Bowling Green State University.

XING, J., CHEN, H.-T., AND WEI, L.-Y. 2014. Autocomplete painting repetitions. *ACM Transactions on Graphics (TOG) 33*, 6, 172.

ZHAO, S., NAKAMURA, K., ISHII, K., AND IGARASHI, T. 2009. Magic cards: a paper tag interface for implicit robot control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 173–182.

# Towards Emergent Play in Mixed Reality

Patrick Misteli
ETH Zurich
pat.misteli@gmail.com

Steven Poulakos
Disney Research
steven.poulakos@disneyresearch.com

Mubbasir Kapadia
Rutgers University
mubbasir.kapadia@rutgers.edu

Robert W. Sumner
Disney Research, ETH Zurich
sumner@disneyresearch.com

## ABSTRACT

This paper presents a system to experience emergent play within a mixed reality environment. Real and virtual objects share a unified representation to allow joint interactions. These objects may optionally contain an internal mental model to act autonomously based on their beliefs about the world. The experience utilizes intuitive interaction patterns using voice, hand gestures and real object manipulation. We author experience by specifying dependency graphs and behavior models, which are extended to support player interactions. Feedback is provided to ensure the system and player share a common play experience, including awareness of obstacles and potential solutions. An author can mix features from game, story and agent-based experiences. We demonstrate our system through an example adventure game using the Microsoft HoloLens.

## KEYWORDS

Mixed Reality, Emergent Play, Interactive Narrative, Cognitive Models, Behavior Trees, Dependency Graphs, Tangible User Interface

## 1 INTRODUCTION

We propose a system to enable emergent play in a mixed reality environment with real and virtual objects. The play experience may include a mixture of gameplay, narrative and agent-based interactions. The player may creatively interact with the system to progress the experience. We utilize the Microsoft HoloLens as the mixed reality interface. It provides the spatial mapping to place virtual objects within the real physical environment. We additionally detect real objects in the environment and integrate properties of real objects within the play experience.

Our paper makes the following main contributions to realize this vision: (1) We have developed a common representation for both real and virtual objects and a runtime instantiation strategy so they may participate in the experience. (2) These objects may optionally contain an internal mental model to act autonomously and make decisions based on their beliefs about the world. (3) We describe a formalism for defining agent behavior, both in terms of the execution of behavior and the visualization of required logical dependencies. (4) We utilize techniques based on Interactive Behavior Trees [14, 17] for accommodating natural player interaction with real and virtual objects. This includes the use of feedback to ensure the system and player share a common play experience, including awareness of current obstacles and potential solutions. (5) Our system allows for the design of mixed reality experiences, which include characteristics of game, story and agent-based simulation.

We demonstrate our system through an example experience depicted in Fig. 4, which presents an adventure game where the player must overcome obstacles in order to achieve a goal. Obstacles must be overcome by player interactions (e.g. using gesture, voice and real object manipulation) to advance the experience. Our system supports emergent play by giving the player feedback about the current experience and the freedom for mixed reality interactions.

## 2 RELATED WORK

Salen and Zimmerman broadly define play as free movement within a more rigid structure [28]. The player drives the experience, for example with the goal of "getting to know the properties of objects" or exploration of what can I do with this object" [9]. Our aim is to support emergent play in the context of gameplay, narrative and agent interaction in mixed-reality environments.

**Emergent gameplay** refers to complex situations that emerge from the interaction of relatively simple game mechanics. For example, the game *Scribblenauts* [11] allows a player to spawn any desired object to solve a puzzle. This is an example of intentional emergence designed into the game. Another example of intentional emergence can be found in games like *Minecraft* [5, 23] where goals can be achieved in multiple ways. Unintentional emergence may happen when a player identifies an unintended use of a game feature. We aim to provide an environment where emergence can take place.

**Emergent Narrative** may include narrative generation techniques that are responsive to author input or user interaction. Our work is inspired by the linear *narrative generation* system, CANVAS, which enables computer-assisted authoring of 3D animated stories [15]. We adopt aspects of the knowledge representation, animation control and visualization to achieve our immersive play experience. This work also motivates our hybrid representation in which character behavior is defined by an external narrative script or by an internal reasoning process [16].

*Interactive narratives* afford the player to alter the direction or outcome of a storyline and offer a wide spectrum of experiences [25]. Traditional choose your own adventure experiences offer a strong story with manually-specified author intent. The opposite end of this spectrum includes emergent narrative systems in which characters in the story have strong autonomy and the authored intent is automatically generated [1]. Games like *The Sims* [22] or *Minecraft* [23] do not have a predefined story but rather allow the player to create their own goals and let a story emerge from the possible interactions. Furthermore interactive narrative experiences such as Mimesis [31] and Merchant of Venice [24] offer strong story control



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

of virtual characters while simultaneously supporting automatically generated stories. Narrative experiences such as Façade [21] are hybrid systems in which virtual characters in Façade exhibit a balance of autonomy and story influence. Our system offers a hybrid of pre-defined stories and an emergent narrative experience for the player.

**Emergent agent interactions** require a notion of intelligent agents and has been studied in cognitive science to explain the structure of the human mind in a comprehensive computer model [27]. The incorporation of intelligent agents allows a system to utilize autonomous entities that observe the world and act according to their own rationale. They react to changes in the world. One of the main four classes of agents described by Weiss [29] is the Belief-Desire-Intention model where each component is represented as a data structure and is manipulated by the agents and its surroundings. We also draw inspiration from Bratman's theory of human practical reasoning [4] to implement a Belief-Desire-Intention (BDI) software architecture. Applying this model we create decision-theoretic goal-based agents, which have beliefs about itself and other agents, desires and possible actions to follow these desires.

**Mixed-reality interaction techniques** expand the possible user interfaces where real world objects can provide an intuitive way to interact with virtual content [32]. Intuitive manipulation and interaction of physical objects that provide an interface to the system is known as a Tangible AR interface [3] and is used in examples such as a table top environment [18] or a "MirageTable" [2]. A challenge of Tangible AR interfaces is to show users how to manipulate the real object to achieve the desired command. Providing visual hints [30] can aid this process. We make use of real toys to represent entities of the system and use their position and orientation to provide input to the system. In addition to mixed-reality techniques, we also support hand gestures [19, 20] and voice input to provide a natural human interface.

## 3 OVERVIEW

Our goal is to create an emergent play experience by mixing real and virtual objects. The objects have a common representation and optionally an internal agent mental model which are both compatible with our interaction design. The interactions are designed with graphical representations. Visual and audible feedback during play ensures the system and player share a common play experience.

**Mixed-Reality Smart Objects.** We apply the concept of smart objects [13] to represent both real and virtual objects where smart objects have states and affordances, which offer capabilities to interact with other smart objects. Virtual objects may be wished into existence and real objects may be introduced during game-play with the system reacting to them. Smart objects can be redefined during runtime by substituting the smart object script.

**Intelligent Agent Architecture.** In addition to specifying the states and affordances associated with a smart object, we may define an internal agent model of the smart object. We choose a model which embodies mental attitudes and model them with states and affordances. The model operates on a shared state space and determines both when and which internal character behavior to execute. We use dependency graphs to visualize the model and behavior trees to implement the model.

**Mixed-Reality Interaction Design.** We support three user interaction types: gesture, physical object manipulation and voice commands. We use formalisms to specify these interactions and their influence on the experience progression. We separate different experience characteristics into game, story or agent simulation.

## 4 MIXED REALITY SMART OBJECTS

In Mixed Reality (MR), we represent all entities (both real and virtual) as smart objects [13]. A smart object contains a set of states and affordances (or capabilities) to interact. Having the same internal representation of all real and virtual objects provides a consistent specification and interaction possibilities of all objects in the world.

**State.** The states define the nature of a smart object such as "being asleep" or "level of friendliness". They can be represented by any type (e.g. bool, enum, int). They are part of the shared state-space of the system and can be accessed from anywhere in the system. Since physical objects are also declared as smart objects they also include states.

**Affordance.** Affordances define the interaction possibilities offered by a smart object and can be invoked by the player, the smart object itself or other smart objects. An example affordance of a smart object could be to speak a defined string or navigate to a certain position. Affordances are specified as execution nodes in a Behavior Tree (BT) [6]. This allows handling of affordance results (success or failure). Affordances can make changes to the shared state-space.

**Instantiation.** The states and affordances of a smart object are defined in a smart object script. Smart object scripts are defined in a hierarchical manner to support code reuse. Each smart object will always inherit all the states and affordances of its predecessors. The system instantiates a smart object by attaching a smart object script to either a virtual or a tracked physical object. That way the system grants interaction capabilities over the smart objects. The hierarchy allows authoring of events (collection of affordances) that require a certain subtype of smart object rather than specifying a specific smart object. We support two types of instantiation, namely pre-defined and dynamic.

An author may *pre-define* objects by attaching the desired smart object script to it and configuring the initial states. This will allow a player to spawn new objects into the world that are known to the system. For example the player could speak "create a key" and the system will instantiate the pre-defined object "key" as shown in Fig. 4h. Given the hierarchical structure the player also has the ability to introduce new undefined *dynamic* smart objects at runtime by either physical or virtual means. The player can wish for a virtual object which will result in a token being spawned with a google image search to texture the desired object shown in Fig. 4d. After the player introduced a new object and defined it as a certain type of smart object the system attaches the corresponding smart object script and the object will inherit all affordances and states of the specified type. Since scripts can be removed and added



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

during runtime the player is also able to reconfigure already existing objects by redefining the smart object script.

# 5 INTELLIGENT AGENT ARCHITECTURE

We utilize a Belief-Desire-Intention (BDI) based agent model [29] to allow real and virtual smart objects to act autonomously and make decisions based on their emotions and the state of the system.

## 5.1 Belief-Desire-Intention Model

Agent intelligence within the BDI model is specified with three components: belief, desire and intention.

**Belief.** The belief states represent what the agent thinks of the system. In a trivial case this will be the ground truth of the current system, such as number of objects of a certain type in the world. However, an agent's belief must not necessarily reflect the actual state of the system. It can be mislead by for example only letting the agent observe part of the system and furthermore include emotional beliefs (e.g. as the belief that another agent is friendly) which cannot be trivially measured. A belief is represented as a state in a smart object and is thus part of the shared state-space. It can be of any type (e.g. bool, enum, int). Beliefs of a smart object are influenced by changes of other states in the shared state-space. A belief state is defined internally to the smart object itself.

**Desire.** The desire states represent the goals of the agent, i.e. the desire to do something such as interacting with an object or running away from a certain area. As the aforementioned belief state a desire is represented as a state in a smart object script. It can be of any type and cannot directly be altered by any other means than the smart object itself. Desires can be manipulated by changes of other states in the shared state-space.

**Intention.** The intentions are the resulting actions the agent will take given a defined combination of beliefs and desires. They are represented as affordance calls in the system.

## 5.2 Behavior Tree Representation of Belief-Desire-Intention Model

A BDI at its core is one or more intentions that are triggered by a defined set of beliefs and desires. Beliefs and desires are in turn affected by changes in the shared state-space. These dependencies can be represented in a dependency graph. The concept of dependency graphs have been used commercially in games such as DeathSpank by Electronic Arts [8]. They show all state dependencies of an experience in the form of a polytree. Parents of every node are the preconditions that must be met in order for the the node itself to become true. This allows to author complex experiences better than with a complex story graph where possible path flows would have to be considered. It also aids the implementation of behavior trees to perform routine checks on the changeable states.

Fig. 1 shows how a dependency graph can be used to visualize how desires and beliefs influence intentions and are influenced by the shared state-space. In the directed graph all parents of a node must be fulfilled in order for the node to also be true or executed.

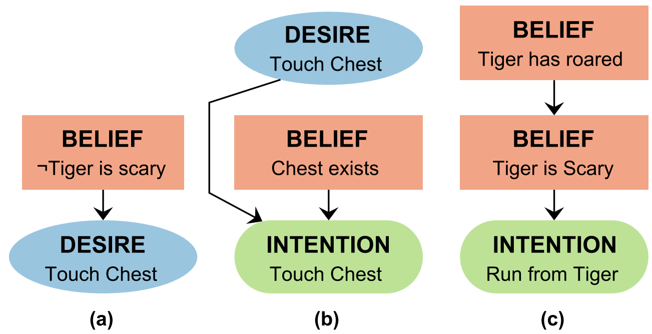



**Figure 1: Inter-influence of belief, desire and intention. (a) The belief that the tiger is not scary results in the desire to touch the chest. (b) The desire to touch the chest and the belief that the chest exists results in the agent intention to touch the chest. (c) The belief that the tiger has roared results in the belief that the tiger is scary which results in the agent intention to flee from the tiger.**

**Behavior Trees** (BTs) [6, 12] are represented as hierarchical nodes that control the flow of decision making of an AI entity. It consists of nodes that are either the root, a control flow node or an execution node. A BT is executed by sending a tick with a certain frequency to the root node which forwards this tick to the execution node defined by the control flow nodes. An execution node can either return *running*, which will cause it to consume the next tick, *success* or *failure*. Depending on the finish state (success or failure) the control flow nodes send the next tick to a different execution node. BTs can grow as complex as defined by the author.

The BT model allows sequential routine checking of all defined states. Each state that can be influenced has a routine check of its preconditions. Every node with one or more parent shown in Fig. 1 requires one routine check which is implemented as a sequence in a BT. An example of such a mapping is shown in Fig. 2 with the corresponding dependency graph depicted in Fig. 1b. It is implemented as a sequence in a BT with the first two execution nodes being leaf-asserts validating whether the desire to touch the chest and the belief the chest exists are active. Should either of those asserts fail the sequence will terminate, invalidate the intention and restart. Should both of them succeed the last node, the touch chest intention is executed. Intentions are implemented as an affordance call of the smart object; in this case the touch-chest affordance.

Each smart object contains its own BT which is ticked independently of any other BT. This allows introduction and removal of smart objects with a BDI model during run-time. Discussion of player interaction with the BT-based BDI can be found in Sec. 6.2.

# 6 MIXED-REALITY INTERACTION DESIGN

The goal of our system is to enable interactions between the mixed-reality smart objects and the player. In this section, we describe how the system considers player interactions in the context of different play experiences.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

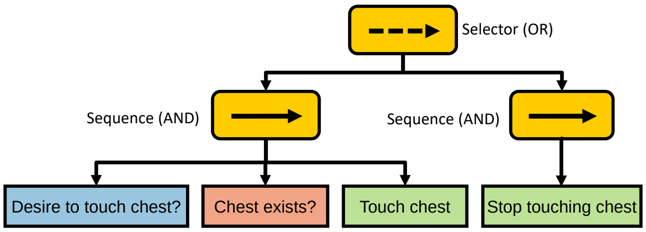



**Figure 2: Behavior Tree of Belief-Desire-Intention module shown in Fig. 1b involving the chicken and the chest**

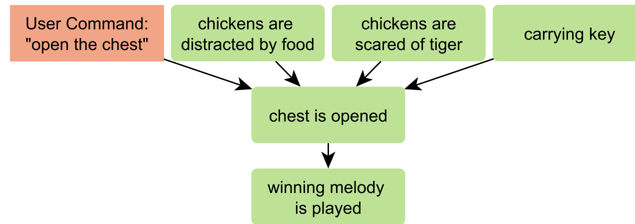



**Figure 3: Main dependency subtree showing player interaction-state (red) and system states (green) as pre-conditions for the system state "chest is opened"**

## 6.1 Supported Player Interactions

The Microsoft HoloLens offers several forms of player interaction to enable a natural play experience. It does not require any controllers leaving the hands of the user free to manipulate physical objects and allows hand gestures. The interaction in our system should be intuitive, require no prior knowledge or experience with MR and no additional hardware setup. We thus chose 3 types of user interaction possibilities:

**Gesture Interaction.** We included the standard air-tap gesture defined and provided by the HoloLens. It can be configured to, for example, select objects and move them if the corresponding smart object allows it.

**Real Object Interaction.** We extended the system with the Vuforia library [10] to allow real objects such as children's toys to be tracked and their position and orientation to be available for the system to use. Manipulating its position or orientation can directly influence states of the physical smart object and thus influence the shared state-space. For example a toy animal's rotation can determine its state "is sleeping" by lying on its side.

**Voice Interaction.** The system was further extended by adding SRGS grammar definition [7]. This allows to author the same command with different variables. For example the commands "*show me the ⟨obj⟩*" and "*go to the ⟨obj⟩*" work in conjunction with the variable definition *⟨obj⟩ = "dog", "cat", "ball", "user"*. The author can then introduce new objects by only extending the ⟨obj⟩ list without having to add all possible voice command combinations involving the object.

## 6.2 Interactive Behavior Trees

We utilize the Interactive Behavior Trees (IBT) concept of Kapadia et al. [14] to extend the notion of Behavior Trees (BT) [6, 12] to enable user interaction with the system. Our IBTs were implemented using a library [26] that enables concurrent access to a globally shared state-space by means of an exposed parameter interface. We generate the following three main distinct subtrees to enable interactions with the system: (1) An event-tree that contains all the events (a collection of smart object affordances) which will get triggered as soon as the corresponding event-state is activated. (2) An interaction-tree which enables an interaction-state when a corresponding user interaction mentioned in Sec. 6.1 was registered. (3) A state-monitor-tree that registers the interaction states, checks authored preconditions and enables event states. This state-monitoring-tree or main dependency tree can be visualized with dependency graphs and implemented in a similar way as the BDI models in Sec. 5.

Fig. 3 shows part of the main dependency graph used in our resulting experience. Given a player interaction (red) we can anticipate their intention and provide feedback if a precondition is not yet met. The behavior tree mapping is done the same way as shown in Sec. 5.2 with one main difference. If a leading assert (i.e. one of the preconditions) fail we do not invalidate the intention but let the player know why a precondition failed and possibly give a hint on how to overcome the obstacle.

## 6.3 Experience characteristics

With our system we can author stories, games, agent-based simulations or any mixture as desired. The characteristics of having a pure game, story or agent simulation are described as follows. A pure **game** experience will not progress without user interaction. The aim is to solve a puzzle or overcome an obstacle. A pure **story** experience will progress without player interaction. A player is able to spectate what is taking place without having the ability to influence the action. A pure **agent-based simulation** experience is also unaffected by any user interaction, but unlike a story the logical decisions of the smart objects and the calling of affordances are not implemented on a global level but on agent-level in form of a BDI for each smart object as described in Sec. 5.

Our concept allows authoring of a seamless hybrid of all three experiences. The main dependency tree gives complete control to the author to create any dependencies. A game requires player interaction to achieve goals. Thus when designing a dependency graph for a game a proportionally large amount of interaction-states are assigned as a precondition to other states. This causes the progression of the experience to be in the players hands and the experience is strongly game-based. Reducing the amount of interaction-states assigned as preconditions in the dependency graph design will make the experience progress more on its own. Events can be authored to invoke more affordances which can lead to a single interaction having a large affect on the progression of the experience. In addition we can define more states being changed in the shared state-space after a single interaction. This can in return result in more preconditions without interaction-state to succeed and further progress the experience autonomously. Designing the experience this way will shift the experience to be more story-based



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

and is a more passive experience for the player when compared to a game-based experience.

The addition of smart objects that include a BDI-based agent model is independent of the previous design choices. An author can choose to have the intentions of a smart object with a BDI affect the shared state-space which can make the smart objects part of the game- or story-experience. The author may also choose to design a dependency graph which ignores any states that are influenced by the BDI model of a smart object. In that case the smart object will appear to be a background character enhancing the world without affecting it.

To ensure a smart object that includes a BDI will not override or ignore a command from the main dependency tree the concept of a "freedom-belief" is introduced. This belief-state is a precondition for all intentions in the BDI model. If the main dependency tree requires control over the smart object it can falsify the freedom-belief. This will instantly disable all preconditions inside the BDI model and allows the main dependency tree to call any affordance.

## 7 RESULTS

We demonstrate a seamless merge of all (game, story and agent simulation) aspects of our system in an adventure game where the user must open the chest by manipulating both real and virtual objects. The resulting experience is shown in a storyboard in Fig. 4. Fig. 4a and Fig. 4b show the real and virtual world setup when initializing the experience. After a player uses a voice command to open the chest, a longer event is triggered (Fig. 4c) where the protagonist walks to the chest then stops because chickens were spawned and are flocking to the chest causing him to be scared. The creation of a dynamic object "corn" (Fig. 4d) and definition as food for chickens alters the belief of a subset of chickens that there is chicken food in the world. This will affect their BDI and cause them to run for the introduced object (Fig. 4e). Commanding to open the chest will result in having the next hint provided that the tiger could be woken up. Fig. 4f shows that the tiger is detected (orange pin) and is in a sleeping state (visual "zzZ" and audible purring). Changing its orientation (Fig. 4g) will result in the tiger waking up. This will change the belief of the other subset of chickens, causing them to flee from the tiger. Fig. 4h shows the player having created a predefined object "key" which does not need to be defined by the player. Fig. 4i shows the successful completion of the last dependency by opening the chest.

## 8 LIMITATIONS AND FUTURE WORK

While we propose design concepts, we do not have a graphical interface to facilitate the design process from conception to implementation. Future work could include mapping of the IBT to a dependency tree visualization. With a provided GUI to design dependency graphs the author can be aided with detection of unreachable states. Further on the dependency graph could then be exported to multiple routine checks represented in an IBT.

The system currently explores the design concepts, but does not employ offline or online planning associated with IBT-based systems. Future work can additionally incorporate planning to generate play experiences based on user interaction.

Future work can also include a completely dynamic voice detection without having to be limited to a predefined grammar. This would increase the possibilities of the user experience and speed up development since no expected words would have to be pre-defined.

The creation of a dynamic object results in a questionnaire that always starts at the top of the smart object hierarchy. The addition of assumptions based on an obtained knowledge-base could speed up or even replace the process of defining an unknown object.

## 9 CONCLUSION

We demonstrated an emergent play experience in a mixed reality environment using the Microsoft HoloLens. A common representation has been applied to both real and virtual objects allowing interaction within the play experience. The player has the option to introduce new objects which are either pre-defined or defined by the player at runtime. These objects optionally have an internal mental model to autonomously respond to the environment. We described an approach for representing complex system behavior using a combination of dependency graph and behavior trees. We additionally utilized techniques based on interactive behavior trees for accommodating natural player interactions with real and virtual objects. This includes the specification of player feedback to ensure the system and the player share a common play experience. Our interaction design allows for game, story and agent-based play experiences.

## ACKNOWLEDGEMENT

The authors would like to thank Maurizio Nitti, Alessia Marra and Mattia Ryffel for providing support with the virtual game assets. The cartoon corn image in Figure 4 (d-e) is provided by ©memoangeles - Can Stock Photo Inc.

## REFERENCES

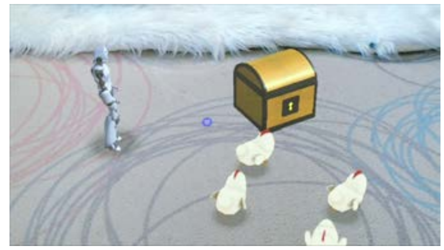
[1] Ruth Aylett. 1999. Narrative in virtual environments-towards emergent narrative. In *Proceedings of the AAAI fall symposium on narrative intelligence*. 83–86.
[2] Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. MirageTable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 199–208.
[3] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2008. Tangible augmented reality. *ACM SIGGRAPH ASIA Courses* 7 (2008).
[4] M. Bratman. 1987. *Intention, plans, and practical reason.* Harvard University Press.
[5] Josh Bycer. 2015. Examining Emergent Gameplay. http://www.gamasutra.com/blogs/JoshBycer/20150916/253682/Examining_Emergent_Gameplay.php. (2015).
[6] R. G. Dromey. 2003. From requirements to design: formalizing the key steps. In *First International Conference on Software Engineering and Formal Methods*. 2–11. https://doi.org/10.1109/SEFM.2003.1236202
[7] Max Froumentin. 2005. The W3C Speech Interface Framework Media Types: application/voicexml+ xml, application/ssml+ xml, application/srgs, application/srgs+ xml, application/ccxml+ xml, and application/pls+ xml. (2005).
[8] Ron Gilbert. 2014. Puzzle Dependency Charts. http://grumpygamer.com/puzzle_dependency_charts. (2014).
[9] Corinne Hutt. 1966. Exploration and play in children. In *Symposia of the Zoological Society of London*, Vol. 18. London, 61–81.
[10] PTC Inc. 2014. Vuforia Augmented Reality SDK. (May 2014). https://www.vuforia.com
[11] Warner Bros. Entertainment Inc. 2013. Scribblenauts. (2013). https://www.scribblenauts.com
[12] D Isla. 2005. GDC 2005 Proceeding: Handling Complexity in the Halo 2 AI. (2005). Retrieved August 31, 2017 from https://www.gamasutra.com/view/feature/130663/gdc_2005_proceeding_handling_.php
[13] Marcelo Kallmann and Daniel Thalmann. 1999. Direct 3D Interaction with Smart Objects. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '99)*. ACM, New York, NY, USA, 124–130. https://doi.org/10.1145/323663.323683



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

**(a) Real world setup**

**(b) Virtual world setup**



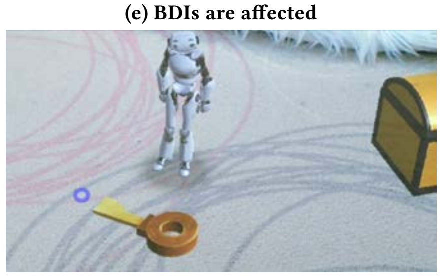
**(c) Longer event sequence (story)**

**(d) Spawning a dynamic object**

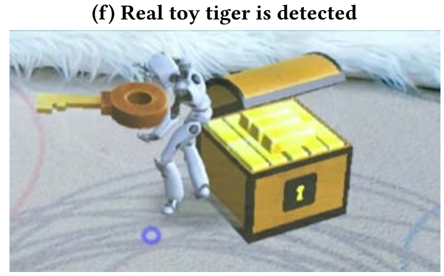
**(e) BDIs are affected**

**(f) Real toy tiger is detected**

**(g) Changing state of real toy**



**(h) Spawning a predefined object**



**(i) Final goal achieved**

**Figure 4: Storyboard of experience showing different aspects of the system's capabilities**


[14] Mubbasir Kapadia, Jessica Falk, Fabio Zünd, Marcel Marti, Robert W Sumner, and Markus Gross. 2015. Computer-assisted authoring of interactive narratives. In *Proc. of the 19th Symposium on Interactive 3D Graphics and Games*. ACM, 85–92.

[15] Mubbasir Kapadia, Seth Frey, Alexander Shoulson, Robert W. Sumner, and Markus Gross. 2016. CANVAS: Computer-assisted Narrative Animation Synthesis. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '16)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 199–209. http://dl.acm.org/citation.cfm?id=2982818.2982846

[16] Mubbasir Kapadia, Alexander Shoulson, Cyril Steimer, Samuel Oberholzer, Robert W. Sumner, and Markus Gross. 2016. An Event-centric Approach to Authoring Stories in Crowds. In *Proceedings of the 9th International Conference on Motion in Games (MIG '16)*. ACM, New York, NY, USA, 15–24. https://doi.org/10.1145/2994258.2994265

[17] Mubbasir Kapadia, Fabio Zünd, Jessica Falk, Marcel Marti, Robert W. Sumner, and Markus Gross. 2015. Evaluating the Authoring Complexity of Interactive Narratives with Interactive Behaviour Trees. In *Foundations of Digital Games (FDG'15)*. 9.

[18] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana. 2000. Virtual object manipulation on a table-top AR environment. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*. 111–119. https://doi.org/10.1109/ISAR.2000.880934

[19] Hui Liang, Junsong Yuan, Daniel Thalmann, and Nadia Magnenat Thalmann. 2015. AR in Hand: Egocentric Palm Pose Tracking and Gesture Recognition for Augmented Reality Applications. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 743–744. https://doi.org/10.1145/2733373.2807972

[20] S. Malik, C. McDonald, and G. Roth. 2002. Hand tracking for interactive pattern-based augmented reality. In *Proceedings. International Symposium on Mixed and Augmented Reality*. 117–126. https://doi.org/10.1109/ISMAR.2002.1115080

[21] Michael Mateas and Andrew Stern. 2003. Façade: An experiment in building a fully-realized interactive drama. In *Game developers conference*, Vol. 2.

[22] Maxis. 2000. The Sims. [CD-ROM]. (2000).

[23] Mojang. 2011. Minecraft. [CD-ROM]. (2011).

[24] Julie Porteous, Marc Cavazza, and Fred Charles. 2010. Applying Planning to Interactive Storytelling: Narrative Control Using State Constraints. *ACM Trans. Intell. Syst. Technol.* 1, 2, Article 10 (Dec. 2010), 21 pages. https://doi.org/10.1145/1869397.1869399

[25] Mark O. Riedl and Vadim Bulitko. 2013. Interactive Narrative: An Intelligent Systems Approach. *AI Magazine* 34, 1 (2013), 67–77.

[26] Alexander Shoulson, Francisco M. Garcia, Matthew Jones, Robert Mead, and Norman I. Badler. 2011. Parameterizing Behavior Trees. In *Proceedings of the 4th International Conference on Motion in Games (MIG'11)*. Springer-Verlag, Berlin, Heidelberg, 144–155. https://doi.org/10.1007/978-3-642-25090-3_13

[27] John Sweller, Jeroen J. G. van Merrienboer, and Fred G. W. C. Paas. 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 10, 3 (01 Sep 1998), 251–296. https://doi.org/10.1023/A:1022193728205

[28] Katie Salen Tekinbaş and Eric Zimmerman. 2003. *Rules of play: game design fundamentals*. MIT Press, Cambridge, Mass.

[29] Gerhard Weiss (Ed.). 2013. *Multiagent systems* (second edition ed.). The MIT Press, Cambridge, Massachusetts.

[30] Sean White, Levi Lister, and Steven Feiner. 2007. Visual hints for tangible gestures in augmented reality. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 47–50.

[31] R Michael Young, Mark O Riedl, Mark Branly, Arnav Jhala, RJ Martin, and CJ Saretto. 2004. An architecture for integrating plan-based behavior generation with interactive game environments. *Journal of Game Development* 1, 1 (2004), 51–70.

[32] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. 2008. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 193–202.




In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

# AI, You're Fired! Artwork

Art Performance

Aleksandra Vasovic

Independent Artist

11000 Belgrade, Serbia,

a.s.vasovic@gmail.com

## ABSTRACT

The paper is text-based artwork, which is representing the initial conceptualization or contemplative phase of the media art and contemporary art performance and installation.

The objective of the long term art project is to further examine the potential of engagement of the advanced technology within the context of artistic research and contemporary art practice, with the specific postulate that the potential product of the artwork is expected to be imperceptible.

The artistic research is referring to the philosophical and metaphysics idea that the alleged real reality cannot be perceived or defined via some concept. The question is, if it is so, than, is the art or the artist capable to successfully illustrate the undetectable real reality, even with the most advanced technological instruments employed.

The text-based contemporary artwork is partly referring to another segment, which can be also observed within the context of the contemporary art – text based computer adventure games. More specifically, the method implemented for establishing the artwork's concept uses some aspects similar to those used in early text-based computer games.

There are several stages in which the long-term artwork will progress.

The initial form is designed in such a manner which would confirm that this segment of artwork not only does serve as a fundament for the other parts to unfold, but is also autonomous and is already completed in terms of contemporary art. This stand is applicable to all the consecutive stages – each segment is both independent and contextual.

The following stages would include the interactivity between the author, art audience, but also with the devices applied for the producing the artwork, like advanced technology instruments e.g. augmented reality (AR), virtual reality (VR), mixed reality (MR) devices, then interactive 3D technology, artificial intelligence (AI), plus the interactivity with the no-reality (reality in spiritual and philosophical contexts).

## CCS CONCEPTS

• The employment of advanced technology in media art and contemporary art, interactive art performance, interactive live art performance beyond the context of visualization

## KEYWORDS

artificial intelligence; art performance; interactive art; art installation; contemporary art; ambient installation; ambient media; text based art; text based games; immersive technology; interactive media; interactive theater; media art; augmented reality; virtual reality; mixed reality; 3D virtual environment; data visualization;

## 1 INTRODUCTION

The text based artwork, as a part of the live contemporary interactive art performance, is the initial phase, the scenario for the more complex art projects. It aims to explore the context in which the artwork's potential development will be expanding and to determine if any of the versions of the artwork is more significant than other for the art practice and potential artistic outcome: is it the text based conceptual and contextual part, or maybe yet to be realized media art piece with the most advanced contemporary technology integrated.

The artwork is a combination of the segment which includes the contemplations about the topic and the segment which is introducing the story and its development to the audience - where the audience and some of the instruments applied for the realization of the artwork become engaged and immersed as co-creators and integral parts of the artistic process.

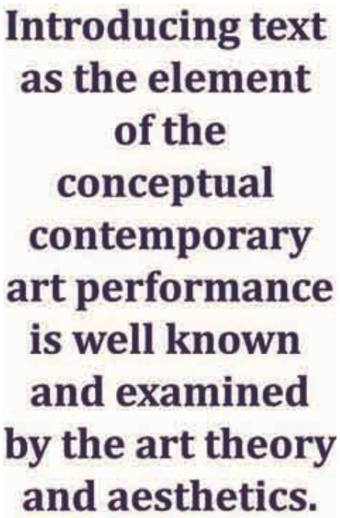
## 2 ART PERFORMANCE DESCRIPTION

### 2.1 Text-based Art Performance: the Initial Stage

*2.1.1 Text-based Performance is an Efficient Tool for the Artists living in Eastern European Milieu.* Introducing text as the element of the conceptual contemporary art performance is well known and examined by the art theory and aesthetics. The form which is used for this artwork is most similar to the Eastern European performance art practice, since this socio-political setting is the prevailing milieu where the author of this artwork lives and works. Specific political and social conditioning in Eastern Europe also initiated and designed the artistic and theoretical approach in the field of the Eastern European version of performance art, many times represented via various forms of writings, texts, correspondence, and lectures. Since the same conditioning more or less persists, this is the motive behind the decision to apply the form of text-based art in this phase. [1]



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal




**AI, You're Fired! Artwork/Text-based Artwork – Digital Print, Video, Segment of Art Performance**

## 2.2 Art: how to Deal with Challenges through Text-based Art Performance

Given the circumstances in which this artwork is conducted, the text-based art form has proven to be the most efficient way to be selected for this artistic project to be successfully generated and potentially without obstructions.

## 2.3 Text-based Computer Games also use Text-based concept as a Tool

*2.3.1 Usually, the Picture is worth Thousand Words.* There is a text-based computer games concept which is borrowed to develop the artistic narrative, or more precisely, one aspect was the most adequate to be employed. Early text-based adventure computer games, like Colossal Cave Adventure by Will Crowther and Don Woods, although developed just using textual discourse, were sufficiently effective and sustainable to maintain autonomously and independently as a milestone in a genre of computer games. [2]

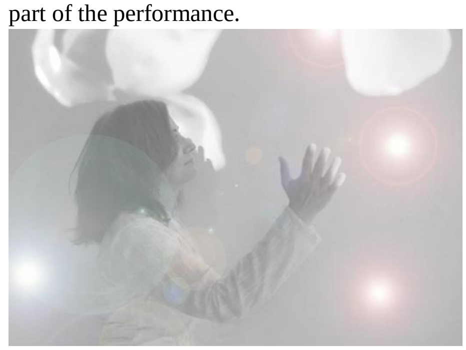
# 3 ARTWORK'S DEVELOPMENT

## 3.1 Projected Structure

*3.1.1 Phases.* Even though the text-based variant of artwork can remain in this form, the artwork would continue to develop in several directions.

Those new stages can be conducted at the art galleries or inside the scientific research centers and laboratories, e.g. as a segment of the doctoral research. The expected outcome of the artistic research is dynamic, the process, not a specific exact result of the experimentation. Correspondingly, all stages during the process are eligible to be exhibited and shared at the exhibiting and other spaces and via internet as ongoing interactive live performance.

*3.1.2 The Procedure.* At some applicable exhibiting space or spaces, the author and the audience would be practicing the spiritual concept of exploring the nature of the reality trough the inquiry. It would be beneficial if the advanced meditation or spiritual practitioners and specialists are also members of the art audience. This type of spiritual practice is performed for thousands of years and consists of repeating a set of carefully designed questions by the practitioners. Eventually, the genuine revelation occurs and the profound knowledge about the real nature of the world is liberated. Other forms of spiritual researches can also be part of the performance.



**AI, You're Fired! Artwork/Humanoid Releasing Thoughts, Digital Print, Video, Segment of Art Performance**

Several practitioners would be invited to participate in the performance; the performance is also intended to be presented as ongoing live internet performance.

# 4 EXPECTED OR UNEXPECTED OUTCOME IN ART PERFORMANCE

## 4.1 Can Art refer to the Real Nature of the Reality

*4.1.1 Art, Metaphysics, Quantum Physics – their Intersection in Art Performance.* Many spiritual and philosophical branches indicate that the world as we observe it with our senses and which we contemplate about based on the experience and knowledge gained through existence of the humankind, in fact is merely the projection, the hologram or some sort of model, fabrication. They suggest that the real reality is in fact undetectable, imperceptible and cannot be comprehended from the point where we are now as humanity. [3]

The science is undertaking the researches to explore the real nature of the reality. Quantum physics is approaching the subject from the scientific perspective and conducts the experiments attempting to provide the material proof that the models and behaviors immanent to the system in which the undetectable/imperceptible essence of existence is detected and examined, defined, efficiently employed etc. [4]

*4.1.2 Artistic Discourse and Potential Outcome.* Maybe the art is also qualified to enter into such a specific field and try to point to the true quality of the real reality. There are many examples throughout the history of art, which are proving that maybe art is the most adequate discipline to deal with this subject.

There are several arguments to support this idea.

Firstly, in the art, there are no firmly established rules which proscribe what procedures need to be undertaken to provide some



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

evidence for certain claim or idea. There is no rule which defines that the art process needs to necessarily have the objective or a goal and to prove that anybody or anything is right or wrong about any thesis or idea. Art and artists do not need the approval to undertake the artistic process, and there is no predetermined goal which art needs to fulfill or prove.

Evidently, art is the discipline which is relatively liberal and where the outcome may be surprisingly accurate and astonishing.

As an artist, I can observe, that the history of humankind proves that no true art remains unrecognized, and anything which pretends to be art cannot escape being exposed - sooner or later.

So, at the exhibiting place, or just any adequate place, when conducting the procedure of artistic/spiritual inquiry, at some point, there would be the outcome recognized, and it would consist of the recognition that the "real" reality, a space without any visions, thoughts, concept, is the origin of other systems.

*4.1.3. Visualization of Imperceptible and Unperceivable.* This phase of the project is the moment when it would be interesting to determine if there is a need to create a visualization of this nothingness, of completely empty space, containing no data at all. The question might arise - as arising questions is the crucial characteristic of the art - that if some idea is completely understood, is there a need to provide material evidence for understanding it to the even greater extent.

*4.1.4 Discussion Panels, Elements of the Art Performance*. Maybe some of many contemporary art panel discussions which can be part of the art performance would determine that, although there is no rational need to create a visualization of no-data, it would still be entertaining to proceed. This would be valid enough reason from the artistic stand to undertake further artistic research and employ the advanced technology visualization tools.

If the visualization is applied, several devices would be applied to create visualizations or materialize no-data, nothingness, e.g. 3D interactive environments, AG, VR, MR devices. The advanced technology and germane methods would be applied if there is interest expressed from the parts involved in this kind of research and production to be engaged in this particular artistic research.

*4.1.5 Final Stages of the Art Project. Artificial Intelligence becomes Art Public and Artist and interacts with Human Versions of Art Public and Artists.* In some very advanced stages, the artificial intelligence (AI ) would be involved, not only just to try to visualize no-data no-substance essence, but to take part into the inquiry alongside with the artist, audience, and spiritual specialists.

Several reasons support the idea to involve the AI as the co-creator and the audience, or the observer.

AI is operating based on interpolating and intersecting of the enormous amount of information and data which it had been previously fueled with.

From the artistic point of view, it would be intriguing to investigate several aspects.

First, what would the system which operates strictly on utilizing data conclude about the system which exists as a data-free self-sustainable, completely autonomous entity and which is at the same time simultaneously the source, the origin and the consequence, the ending point.

Furthermore, as the art operates fundamentally by very carefully and efficiently utilizing the emotions and connotations, semiotic principles, what can art do with the system which is predestined not to successfully comprehend the emotions?

In addition, if it is proved that there is the superior system, which just is, which does not need to perform any sort of action, which contains no information, emotions, thoughts, which does not need any concept to justify its existence – then, it is applicable to suggest several questions. If humans feed the AI with the information in order to employ it for the development of the humanity, then who is feeding the humans (including artists) with the emotions and thoughts and for what purpose, for whose development?

Finally, if the humans are observing the AI operating with the information previously stored, then, who is observing the humans operating with the previously carefully obtained and stored information, knowledge, experiences, emotions, thoughts? Who, what system uses the humankind and human intelligence to serve as artificial intelligence? Is it a person or some impersonal essence? Maybe artists already know or are able to sense the resolution? Of course, this is the idea which originates entirely from the artistic point of view and is immanent to the artistic practice.

There are many evidence of the advanced application of the AI, naturally, art is one of the disciplines which has evident progress within this machine-man interaction field. [5]

In some cases, while being involved in some inquiry, the AI (popular Google's chat bot) provided meaningful, insightful and deep responds related to the domain of philosophy and metaphysics. [6]

## 4.2 Suggested Discussion Subjects

*4.2.1 Thoughts and Emotions.* How can re-think and a new approach toward big data, and cognitive big data, correlate and contribute to the concept and context of this artwork? The artwork is examining, from artistic scope, the occurrence and importance of thinking, processing the thoughts and emotions (as the equivalence of the processing of data) and what happens if they are absent in human experience. Furthermore, in this context, what occurs in human – robot interaction and how can this be observed, processed, explained and incorporated in the artistic practice? How can virtual reality, augmented reality and artificial intelligence instruments behave and what is their meaning in the context of the thought-less empty space (as philosophical concept)? Will the results of those artistic observations have any significance from the scientific point of view? [8]

And vice versa, how can the artistic interest developed through this artwork arise questions and initiate new approaches on the other, scientific scope of this discourse, including dealing with the technology, machine learning and artificial intelligence development and employment in improving and enriching human lives and experience on all possible levels?

*4.2.2 Emotional Intelligence Element.* There are certain artistic attributes, e.g. the performative character of art, often characterized by the embracement of the option that the unexpected outcomes during artistic processes are desirable. The reason behind this is that



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

the abundance of emotions arise when being involved within art and dealing with it in this particular way. How can this be combined with the improvement inside the scientific sphere related to the enhancing the quality of the real time robots' perception and the emotional intelligence aspect of human-robots interaction?

Can the emotions incited by art be calculated into the evaluation of the human-robots interaction and will this be a contribution from the scientific stand? To what extent will this artwork benefit if the machines significantly develop the emotional intelligence segment, and this is applied to the artistic process? [9][10]

*4.2.3 Virtual Environments and Immersive Aspect of the Artwork.*

*4.2.3.1 Visualization.* It will be valuable for the artistic concept if the various phases of the artistic process are developed through the process of visualization utilizing new interactive 3D technologies, in virtual environments, in this case, employed as the theatrical layout where both the participants and the visualization technology are all amalgamated as the segments of the interactive version of theatrical performance. Not only would it be essential for the artistic process to provide the visualization of the internal cognitive processes and emotions, or experience of those who participate in the process, but it is also crucial to observe and visualize the interactive encounter, or, how the participating in creating the artwork, becoming the co-author, affects the cognitive and emotional respond of the participants. This part, when the observing of the visualization of the previous experience creates new experience, can also be represented through the same or similar methods of visualization, as a part of the artwork, which is embedded into the structure of artistic process. [11]

*4.2.3.2 Ambient media.* The immersive nature of this artwork, where the natural environment, virtual environments, artist, collaborators, public as the co-creator - are all merged and are equally important parts of the artistic process, corresponds on some level with the ambient media principles. The immersive constellation unfolding within the artistic procedure may be the indicator of the several directions in which the ambient media could develop in the following period, affecting the experiencing of the art, as a specific category of human engagement. [12]

## 4.3 Art, Humans and Machines

*4.3.1 Art and Artificial Intelligence.* Although it is difficult to define and classify such a complex human activity as artistic activity, can the engagement of the artificial intelligence, applied on the advanced level on the artistic playground, reveal new insights about the meaning and significance of art for human experience?

In addition, what if the directions in which the art and machine intelligence evolve, indicate that the art is not an exclusively human activity? Would it be interesting for the scientists to examine how this idea would affect the following categories: human artists, human art public, machine artists, and machine art public? What if I, as an artist, choose that I prefer the machines as art public, that I love them more than human art public, that they love me, too, and I decide to create art exclusively for machine art public?

## 6 CONCLUSIONS

All participants of the art performance, including AI, will have their own time to truly comprehend the space where no data can exist.

For some, this would occur within nanoseconds, for some it would take hours, days, years, decades. This is why this art project is ongoing and long-term (for some of the participants very long-term) and this is absolutely regular procedure.

For the specific participant in the artistic research (the AI), the experiment will determine how long it would take the AI to truly realize the nature of the “real” no-data reality and if the AI would realize it at all.

The evidence that the AI had truly comprehended the imperceptible no-thing-ness would be its spontaneous claim: “This is it – there is no data there – so, I quit my job”, according to this artwork’s rules.

Then, the artist can determine that the AI’s request can be satisfied, and that it can be excused from the further artistic examination and observation – because, there is nothing to be observed and no data to be processed.

This is the reason why the title of this text-based art performance is, in fact, liberating for the AI, and it says: AI, You’re Fired! Artwork. [7]

But, again, if there is nothing to be observed, and no data to be processed, visualized, then, is the entertainment the reason which substantiates the perpetual observing and processing of the data generally?

## 8 REFERENCES


[1] Performance Art in Eastern Europe (1950-1990), Performance Art, http://www.performanceart.info/theory/

[2] Colossal Cave Adventure, https://en.wikipedia.org/wiki/Colossal_Cave_Adventure

[3] Swami Venkatesananda , The Concise Yoga Vasistha, State University of New York, Albany, 1984, ISBN 0-87395-954-x ISBN 0-87395-955-8, page 5

[4] VlatkoVedral, Decoding Reality, Oxford University Press, 23 February 2012, ISBN: 9780199695744

[5] Ramón López de Mántaras, Artificial Intelligence and the Arts: Toward Computational Creativity, Open Mind, https://www.bbvaopenmind.com/en/article/artificial-intelligence-and-the-arts-toward-computational-creativity/?fullscreen=true

[6] Nathan McAlone, Google's AI Bot Thinks The Purpose of Life Is 'to Live Forever', Science Alert, https://www.sciencealert.com/google-s-ai-bot-thinks-the-purpose-of-life-is-to-live-forever

[7] Aleksandra Vasovic, AI, You’re Fired! Artwork Preview, Aleksandra Art Work Vimeo Profile, https://vimeo.com/240176893

[8] Lugmayr, A. et al., 2017. Cognitive Big Data. Survey and Review on Big Data Research and its Implications: What is Really "New"? Cognitive Big Data! Journal of Knowledge Management (JMM), 21(1), Available: http://emeraldgrouppublishing.com/products/journals/call_for_papers.htm?id=5855k, Journal of Knowledge Management/Emerald.




In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

[9] Sun, M., Zhao, Z., and Ma, X. (2017). Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices. In Proc. CHI2017 (to appear).
[10] Yang, Y., Ma, X., and Fung, P. (2017). Perceived Emotional Intelligence in Virtual Agents. In Proc. CHI2017 EA (to appear).
[11] A. Lugmayr, A. Greenfeld, A. Woods, and P. Joseph, "Cultural Visualisation of a Cultural Photographic Collection in 3D Environments – Development of 'PAV 3D' (Photographic Archive Visualisation)" in Entertainment Computing - ICEC 2016: 15th IFIP TC 14 International Conference, Vienna, Austria, September 28-30, 2016, Proceedings, G. Wallner, S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, and H.-S. Yang, Eds. Cham: Springer International Publishing, 2016, pp. 272–277 Online. Available: http://dx.doi.org/10.1007/978-3-319-46100-7_29
[12] A. Lugmayr, E. Serral, A. Scherp, B. Pogorelc, and M. Mustaquim, "Ambient media today and tomorrow", Multimedia Tools and Applications, vol. 71, 2014, pp. 7–37 Available: http://dx.doi.org/10.1007/s11042-012-1346-z.



In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal