Wolfgang Aigner
Kerstin Blumenstein
Michael Iber
Thomas Moser
Matthias Zeppelzauer
Grischa Schmiedl
(Eds.)

FMT 2017

Proceedings of the
10th Forum Media Technology and
3rd All Around Audio Symposium

November 29-30, 2017
St. Pölten, Austria

Editors' addresses:
St. Pölten University of Applied Sciences
Institute of Creative\Media/Technologies
Matthias Corvinus-Straße 15
3100 St. Pölten
Austria

{wolfgang.aigner | thomas.moser | kerstin.blumenstein | matthias.zeppelzauer |
michael.iber | grischa.schmiedl}@fhstp.ac.at

Preface

Audiovisual, interactive and mobile media interweave with our everyday life, a major aspect of the so-called megatrend digitalization. In almost all our areas of life, be it private or professional, they change our behavior and affect our perception, thinking and feeling. The ‘Forum Media Technology (FMT)’ is a scientific event that aims at providing an intensive dialogue of experts working at agencies, studios, and companies of different industries with students, lecturers, researchers, and developers in the field of digital media.

This year in its 10^{th} jubilee edition, the Forum Media Technology was held on November 29-30, 2017 at St. Pölten University of Applied Sciences, Austria. In conjunction with the main conference, the 3^{nd} edition of the trans-disciplinary symposium ‘All Around Audio (AAA)’ as well as the 2^{nd} edition of the ‘Graduate Consortium’ were part of the two day program.

All Around Audio Symposium

Although, audio has not abandoned its status as a standalone discipline, its trans-disciplinary participation at the conception and design of products and environments has become more and more essential. In this sense, All Around Audio not only addresses specialists in the audio domain but particularly encourages researchers and designers from other fields to participate in the symposium. In its third edition, 16 international speakers were invited for talks on a wide spectrum of topics reaching from media economy, audio technology, auditory display to music and media arts.

Graduate Consortium

The FMT 2017 Graduate Consortium session was intended to provide an opportunity for graduate and PhD students to explore, discuss and develop their research topic in an interdisciplinary workshop, under the guidance of a panel of distinguished researchers.

Forum Media Technology Conference Track

Submissions for the FMT conference track were accepted in two categories of full and short papers. All submitted papers underwent a double-blind review process, during which each paper was reviewed by at least three members of the international program committee (IPC) of the conference. Based on the written reviews, final decisions were made by the paper chairs and finally 13 papers (8 full + 5 short) were accepted for presentation at the conference (acceptance rate: 50%). In addition, a total of 7 poster contributions were presented in an interactive sessions, of which 2 posters are part of the special track of the OpenGLAM hackathon, which took place in September at St. Pölten University of Applied Sciences. The accepted papers can be clustered into four areas: computer vision, research design & digital healthcare, data visualization, as well as human-computer interaction.

For the second time this year, also one best paper and three honorable mention awards were given to the top four contributions of the FMT. The award is based on the scientific excellence of the paper as well as the presentation performance of the authors at the conference. The selection process for the best paper award followed a three-step process. First, IPC members indicated whether a paper should be considered for the award. Based on that, the paper chairs decided on a short list of four nominated papers for the award and the selected papers’ authors were informed about their nomination. The final decision was made by an anonymous best paper committee consisting of three members who took into account both the paper as well as the presentation to make their final decision.

November 2017

Wolfgang Aigner
Thomas Moser
Kerstin Blumenstein
Matthias Zeppelzauer
Michael Iber
Grischa Schmiedl

Best Paper Award Sponsor:

WISSENSCHAFT · FORSCHUNG NIEDERÖSTERREICH

Organizing Committee

Conference Chairs

Wolfgang Aigner, St. Pölten UAS, AT
Thomas Moser, St. Pölten UAS, AT

Paper Chairs

Kerstin Blumenstein, St. Pölten UAS, AT
Matthias Zeppelzauer, St. Pölten UAS, AT

Graduate Consortium Chair

Grischa Schmiedl, St. Pölten UAS, AT

AAA Symposium Chair

Michael Iber, St. Pölten UAS, AT

Local Organizers

Julia Klaus, St. Pölten UAS, AT
Ulrike Wieländer, St. Pölten UAS, AT

Best Paper Committee

Paolo Buono, University of Bari Aldo Moro, IT
Hans-Christian Jetter, UAS Upper Austria, AT
Michael Zeiller, UAS Burgenland, AT

Program Committee

External

Albert Amor-Amoros, TU Wien, AT
Christoph Anthes, UAS Upper Austria, AT
Thomas Artner, TU Wien, AT
Arnold Baca, Vienna University, AT
Till Becker, University Bremen, DE
Alessio Bertone, Dresden University of Technology, DE
Christian Bors, TU Wien, AT
Nadia Boukhelifa, INRIA, FR
Paolo Buono, University of Bari Aldo Moro, IT
Mario Döller, UAS Kufstein, AT
Thomas Görne, Hamburg UAS, DE
Theresia Gschwandtner, TU Wien, AT
Svenja Hagenhoff, Friedrich-Alexander University Erlangen-Nürnberg, DE
Georg Hajdu, Hochschule für Musik und Theater Hamburg, DE
Christoph Heinzl, UAS Upper Austria, AT
Hans-Christian Jetter, UAS Upper Austria, AT
Isabel John, UAS Würzburg-Schweinfurt, DE
Johannes Kehrer, Technical University Munich, DE
Andreas Kerren, Linnaeus University, SE
Simone Kriglstein, TU Wien, AT
Tim Lammarsch, Researcher, AT
Robert Laramee, Swansea University, GB
Wilfried Lepuschitz, Practical Robotics Institute, AT
Thomas Lidy, TU Wien, AT
Christian Märtin, UAS Augsburg, DE
Munir Merdan, Practical Robotics Institute, AT
Paul Modler, Karlsruhe University of Arts and Design, DE
Annette Mossel, TU Wien, AT
Paul Murschetz, Alpen-Adria-University Klagenfurt, AT
Peter Plessas, Institute of Electronic Music and Acoustics Graz, AT
Cornelius Pöpel, Ansbach UAS, DE
Michael Riegler, Simula Research Laboratory, NO
Michael Roither, UAS Burgenland, AT
Paul Rosenthal, University of Rostock, DE
Markus Schledl, Johannes Kepler University Linz, AT
Sebastian Schlund, TU Wien, AT
Raphael Schneeberger, PocketScience GmbH, AT
Klaus Schöffmann, Klagenfurt University, AT
Tobias Schreck, TU Graz, AT
Michael Sedlmair, Vienna University, AT
Anselm Spoerri, Rutgers University, US
Marc Streit, Johannes Kepler University Linz, AT
Markus Tauber, UAS Burgenland, AT
Catagay Turkay, City University London, GB
Manuela Waldner, TU Wien, AT
Edgar Weippl, SBA Research, AT
Florian Windhager, Danube University Krems, AT
Kerstin Witte, Otto von Guericke University Magdeburg, DE
Michael Zeiller, UAS Burgenland, AT

Internal

Jakob Doppler, St. Pölten UAS, AT
Astrid Ebner-Zarl, St. Pölten UAS, AT
Franz Fidler, St. Pölten UAS, AT
Andreas Gebesmair, St. Pölten UAS, AT
Christian Gradl, St. Pölten UAS, AT
Florian Grassinger, St. Pölten UAS, AT
Stefanie Größbacher, St. Pölten UAS, AT
Mario Heller, St. Pölten UAS, AT
Brian Horsak, St. Pölten UAS, AT
Matthias Husinsky, St. Pölten UAS, AT
Michael Iber, St. Pölten UAS, AT
Andreas Jakl, St. Pölten UAS, AT
Thiemo Kastel, St. Pölten UAS, AT
Christoph Musik, St. Pölten UAS, AT
Christina Niederer, St. Pölten UAS, AT
Lars Oertel, St. Pölten UAS, AT
Tassilo Pellegrini, St. Pölten UAS, AT
Johannes Pfegerl, St. Pölten UAS, AT
Julia Püringer, St. Pölten UAS, AT
Alexander Rind, St. Pölten UAS, AT
Gernot Rottermanner, St. Pölten UAS, AT
Grischa Schmiedl, St. Pölten UAS, AT
Djordje Slijepcevic, St. Pölten UAS, AT
Rosa von Suess, St. Pölten UAS, AT
Markus Wagner, St. Pölten UAS, AT

Contents

Keynotes & Capstone

Keynote: Stops in Motion – Animation as Meta-cinematographic Concept

Franziska Bruckner, St. Pölten UAS, AT

Abstract

Stop-motion is an animation technique, in which objects are shifted in small increments between individually photographed frames, creating the illusion of movement when the series of photos are projected as a continuous sequence. In filmic contexts stop-motion reaches a specific potential of expression, but already includes historically grown aspects of intermediality by combining artistic spheres like photography, music, fine arts, theater or puppetry. More comprehensive ideas of animation are featured from several theorists, such as Lev Manovich (1995), Alan Cholodenko (1991) or Suzanne Buchan (2013), which also emphasize the variety of manual, mechanical and conceptual possibilities of stop-motion beyond the medium film.

The talk "Stops in Motion" focuses on an expanded notion of stop-motion, its potential before, in, and beyond its filmic boundaries and aims to explore a brief history of this aesthetically diverse animation technique. Beginning with precinematic devices, the lecture outlines important steps of stop-motion in film history and explores innovative prospects since its digitalization. This not only includes possibilities of established stop-motion software, but also focuses on opportunities in virtual and augmented reality applications. As stop-motion vitalizes objects in a visible fragmented way, it is an ideal concept for investigating new understandings of cinematic perception. Viewed from this perspective, stop-motion functions not anymore as a technique but as a "meta-cinematographic" concept and becomes a tool to fragment and recompose the world.

Biographie

Dr. Franziska Bruckner (Salzburg, 1981) is head of the research group Media Creation at St. Pölten University of Applied Sciences, co-coordinator of the Focus Group Animation within the German-speaking Society of Media Studies and board member of ASIFA-Austria.

She graduated in Theater-, Film- and Media Studies at the University of Vienna as well as Painting and Animation at the University of Applied Arts Vienna. From 2009 to 2013 she was a university assistant at the department of Theater-, Film- and Media Studies in Vienna, from 2013 to 2017 she worked as lecturer for animation theory and practice at the University of Vienna, University of Tübingen and University of Applied Sciences Upper Austria.

Keynote: Collaborative Data Experiences: Novel designs for visualizing and exploring data together

Hans-Christian Jetter, UAS Upper Austria, Campus Hagenberg, AT

Abstract

We are witnessing an unprecedented exponential growth in the data that we create and that we are exposed to in our daily lives. This trend towards "Big Data" promises novel applications that could revolutionize business, administration, policy making, and science. To let users experience and make sense of this data, there is already a lot of research on the algorithmic side, e.g., new methods for data mining, machine learning, etc. There is, however, much less work on how to visually communicate and present results in an "intuitive" and interactive manner, especially to groups of casual or non-expert users.

I will show different examples from my research work that demonstrate how the careful design of interaction and visualization techniques can substantially improve our human-data interaction with visualizations, for example by enabling groups of users to collaborate using visual-tangible user interfaces on interactive tabletops or by working seamlessly across many mobile devices in "bring your own device scenarios". I will illustrate how a combination of applied informatics, design, and user research can help us to better understand how humans interact with data and achieve a much improved collaborative human-data experience.

Biographie

Hans-Christian Jetter is a computer scientist and Professor of User Experience and Interaction Design at the University of Applied Sciences Upper Austria, Campus Hagenberg. Before joining Hagenberg, Christian worked as a post doc with Yvonne Rogers at the University College London in the Intel Collaborative Research Institute for Sustainable Connected Cities. Christian also worked as research intern and research visitor at Microsoft Research Cambridge where he explored the use of novel collaborative tools for scientists of the NanoPhotonics Centre of the University of Cambridge. He received his PhD (summa cum laude) and M.Sc. & B.Sc. in Information Engineering from Harald Reiterer at the Human-Computer Interaction Group of the University of Konstanz.

Capstone: Pervasive Technologies to Enrich People Experience in Visiting Cultural Heritage sites

Paolo Buono, University of Bari Aldo Moro, IT

Abstract

Various empirical programs have been carried out worldwide with the aim of kindling people's interest in visiting Cultural Heritage (CH) sites. This talk narrates the research conducted at Interaction, Visualization and Usability (IVU) Lab of the University of Bari (Italy) that investigates the use of different technologies to ensure more engaging visit experiences at Cultural Heritage (CH) sites and to increase the appropriation of CH content by visitors. Initially, pervasive games using mobile devices were developed to break away from the usual static paradigm of room play and go towards a more dynamic and social experience. Such games addressed young students and aimed at stimulating them to acquire knowledge during visits at CH sites. Technological advances led us to introduce large multitouch displays in school activities to reflect and deepen gained knowledge.

Furthermore, the growing availability of smart objects has stimulated us to use the Internet of Things technologies in the CH domain. There are very few approaches trying to facilitate the adoption of such technologies by end users, who are required to define the behavior of smart objects but they might not have any skill in programming. A visual composition paradigm that allows non-programmers to synchronize the behavior of smart objects was defined, in order to comply with the need of curators and guides of CH sites to define smart visit experiences. A serious game has been used to show the potential of the visual composition paradigm approach. The talk ends by discussing technological solutions and future challenges.

Biographie

Paolo Buono is an Assistant professor and a member of the IVU (Interaction, Visualization, Usability and UX) lab at the Department of Informatics of the University of Bari Aldo Moro. His current research focuses on HCI, specifically in information visualization, mobile applications, time series. His research has been also involved in other domains such as visual analytics, video analysis and telementoring. He has been involved, at different levels of responsibility, in several European, national and regional projects in various domains, including: environment, logistics, cultural heritage, healthcare. He holds a PhD in Computer Science from the University of Bari, Italy. He has been visiting scientist at various research centers including: AVIZ research group (F), Human-Computer Interaction Lab of the University of Maryland (USA), Fraunhofer IPSI (D). He has organized several HCI conferences, such as AVI (2016, 2004), IS-EUD 2011, INTERACT 2005. He is one of the inventors of the Patent No. 1401512 concerning a multimedia framework and a method to support the visit of a site of interest, such as an archaeological park.

Session 1: Computer Vision

Towards Automated Real Estate Assessment from Satellite Images with CNNs

Valentin Muhr
Kufstein University of Applied Sciences
Email: v.muhr@a1.net

Miroslav Despotovic
Kufstein University of Applied Sciences
Email: miroslav.despotovic@fh-kufstein.ac.at

David Koch
Kufstein University of Applied Sciences
Email: david.koch@fh-kufstein.ac.at

Mario Döller
Kufstein University of Applied Sciences
Email: mario.doeller@fh-kufstein.ac.at

Matthias Zeppelzauer
St.Pölten University of Applied Sciences
Email: matthias.zeppelzauer@fhstp.ac.at

***Abstract*—A driving factor for real estate prices is the location quality. Models for location quality are usually built from available price information and distinct GIS information. In this paper, we present a first approach towards the automated assessment of location quality from satellite images using computer vision. For this purpose, we first introduce a novel dataset generated from publicly available data sources with suitable ground-truth annotations for location assessment. Next, we adapt a state-of-the-art convolutional neural network (CNN) and adapt it to predict different land covers and objects from satellite images. Finally, we feed information derived from the recognized land covers into a regression-based price model which acts as a proxy for the assessment of location quality. Our results show that (i) land cover classification can be performed with high accuracy and demonstrates that automatic classification could further be used in the future for the detection of mis-aligned and erroneous GIS data; (ii) our adapted network reaches state-of-the-art performance in much less training time compared to our reference network; (iii) the automatically extracted visual information improves the prediction of real estate prices and thereby shows clear potential for the description of location quality.**

I. Introduction

One of the most important criteria for assessing real estate is its location and its neighborhood. The by the authors of this paper specified research objective is to assess the *quality and livability* of urban geographical locations to support the automatic assessment of real-estate objects and properties. To realize this goal, we aim at leveraging the capabilities of automated land cover analysis in satellite images.

Land Cover Analysis focuses on the extraction of information from remote sensing satellite image data and GIS data with the goal to study and monitor geological resources and their dynamic changes [1]. In this aspect, remote sensing has become more and more important due to the increasing amount of available geographical data [2], provided through satellite recordings world-wide and comprehensive descriptive information collections from GIS. These data allow better understanding and development of the earths global and physical processes [3] and in combination with GIS have fostered several applications and meaningful results in the fields of agriculture, environment, and eco-environment assessment [4], [5], [6], [7]. Due to these comprehensive data former limitations to recognize fine-grained structural patterns of objects (e.g. buildings) and background are dissolved [8], and thereby enable a number of novel approaches, as the one presented in this work.

A rich literature on satellite image analysis and land cover classification exists [9], [10] and numerous datasets have been introduced [11], [12] for developing automatic classification and detection algorithms. Furthermore, recently, deep learning has made a significant impact on the field, especially Convolutional Neural Networks (CNNs) [13] have improved classification performance of satellite imagery significantly [9], [14], [15], [16], [17]. Although there is work on many different tasks in the context of satellite image analysis, such as building detection, road network analysis, terrain classification, flood detection, and even poverty detection, the detection of location quality has, to the knowledge of the authors, not been investigated so far [18], [19], [20], [21], [22].

For a proper development and implementation of our research objectives, we had to build our own satellite image dataset, which suits the requirements for location quality assessment. For this purpose, we used publicly available satellite image data from the Tyrolean Tiris Database[1] and the freely available Open Government Data[2] (OGD). For location quality assessment we have defined a set of eight relevant land covers, which reflect geological characteristics and also considerable infrastructure in Austria. Based on this novel dataset, we designed a CNN for location assessment and compared it to an existing network (VGG network model) [23]. The networks were designed (and adapted respectively) to classify the land covers and objects in our dataset.

This work represents a first step towards our main research objective to assess location quality by providing the necessary basis information for this task, namely pre-classified land

[1]Orthofoto Tirol - data.gv.at. [Online] Available: https://www.data.gv.at/katalog/dataset/35691b6c-9ed7-4517-b4b3-688b0569729a. Accessed on: Aug. 30, 2017

[2]Datenauftritt Land Tirol — data.gv.at. [Online] Available: https://www.data.gv.at/auftritte/?organisation=land-tirol&katFilter=geographie-und-planung#showresults. Accessed on: Aug 30, 2017

covers, terrains and objects in RGB satellite images. The training of regression models for location quality based on this analysis data is currently under investigation. The hedonic pricing method, provides the theoretical background in real estate and is widely used in housing research and appraisals [24]. Originally developed for automobiles by [25], hedonic price models describe how the quantity and quality of these characteristics determine its price in a particular market. In a formal way, the general hedonic price function takes the form

$$P_i = f(S_i, L_i, N_i) \tag{1}$$

where P_i is the *log* of the price or rent of house i, S_i is a vector of structural housing characteristics, L_i is a vector of location variables and N_i is the neighborhood characteristics. In the field of hedonic pricing, only the aspects that are available as quantifiable data, e.g. for location variables, such as distances to schools, public transport, etc., are considered. For instance, if one wants to evaluate sociodemographic data within the scope of hedonic pricing on a certain geographical level, an appropriate statement can be developed only on that particular aggregated level. Therefore, the exact location or the immediate neighborhood can be valuated only conditionally in a fully automatic way. Furthermore, the demarcation of micro-residential areas based on e.g. predefined spatial units, and price similarity cannot always lead to an adequate qualitative distinction. Pattern recognition from satellite images circumvents this problem, as the factual delineation of the macro-location is ignored and only the features within selected micro-location are being investigated.

II. Related Work

The major objective of image analysis is to extract discriminative visual features from images [26]. An early approach, which is sometimes still applied today, is the application of manually designed visual features based on color- and edge-descriptors [27], [28], e.g. histograms and correlograms. Later local features (e.g. SIFT, SURF, ORB) were introduced and became popular within the Bag-of-Visual-Word model for image classification [29]. During the last years, increasingly more automated methods for feature extraction and entirely end-to-end learned approaches based on neural networks (deep learning, DL) have been introduced and achieved remarkable improvements in image analysis and classification [30], [31], [32].

In remote sensing different image types are utilized, ranging from grayscale and RGB images, to multi-spectral images, and Synthetic Aperture Radar (SAR) images, as well as measuring procedures such as Light Detection and Ranging (LiDAR) [33] enabling 3-dimensional geographical surface scanning. RGB satellite images are the most widely distributed resources in remote sensing, since region-wide captures are broadly available.

In RGB satellite image analysis we can differentiate between low-level tasks and (more sophisticated) higher-lever tasks, which often build upon low-level tasks. Low-level tasks focus on detection of certain objects, such as buildings [34], [35], [36], [37], road networks [38], [39], [40], and the segmentation or classification of different land covers, such as grassland, forest, and water [41], [42], [43], [44]. Satellite images are usually cut into smaller patches before analysis, in order to provide local spatial information (at a particular location) together with its surrounding context (neighborhood) to the subsequent analysis. Afterwards, traditionally feature extraction is employed followed by machine learning [28]. Deep Learning, in contrast, combines feature extraction and classification into one process, which is demonstrated by an early work by Mnih and Hinton [45]. This work is based on a neural network with multiple local and fully-connected layers, designed to detect roads and road networks. In recent years many different methods for feature learning based on neural networks have been introduced [46], [47], [41], [31], [48]. In these methods, the extracted features (neural activations at intermediate layers of the networks) are further processed by other classifiers, such as Support Vector Machines (SVM) to obtain a final classification. In other works the networks combine feature extraction and learning into one supervised learning process by adding a classification layer on top of the network [45], [22], [44], [49], [50], [51], [32], [43].

Higher-level tasks often build upon a basic land cover classification or segmentation and include, for example building type classification, i.e., the classification of residential buildings, apartment buildings, and industrial and factory building [52], [53] to support urban development and proper planning. Higher-level tasks exist for different scaling levels. An example for a very fine-grained analysis is the segmentation of rooftops of buildings into areas of different angles [35] and thereby the evaluation of the suitability of these rooftops for photo-voltaic systems [54], [55]. Other works focus on higher-level tasks at a much larger scale, such as the analysis of abstract demographic and socio-economic parameters from satellite images. An example is the work of [56] where the authors try to derive regions of poverty through satellite image data. They exploit the light intensity in areas of settlement in night satellite images as an indicator. Based on this information they apply transfer learning to identify characteristical visual features which indicate poverty. In a follow-up work, the authors were able to explain 75% of the variation of economic outcomes from satellite images with a CNN [19].

The assessment of location quality is a new high-level task for satellite image analysis introduced in this paper, which builds upon low-level land cover classification. To facilitate the characterization of location quality from land covers, quality-relevant land covers need to be selected and recognized in the images. We have performed a survey on satellite image datasets to identify suitable annotated data collections for location assessment.

One of the most recent dataset is the AID dataset. It has been introduced in early 2017 by [57] and covers 10.000 object images divided into 30 individual categories. The set entails high intra-class variations, since visible lighting conditions and time or season differ from sample to sample. Also due to multi-resolution collections, the images range widely in their spatial resolution of 50 cm and up to eight meters. For our initial experiments on location assessment, we aim at a more controlled dataset with less degrees of freedom (i.e. different scales and different seasons) to reduce the number of influencing factors. We consider the AID dataset as a challenging dataset for future experiments, once a first location assessment method has been established.

The RSSCN7 dataset, introduced by [58], contains 2800 satellite images with eight different annotated land cover classes. With 400 images in the categories of grassland, forest, farmland, parking lot, residential region, industrial region, and river and lake, these classes are closely resembled to the introduced dataset of current reading. Challenging is the fact that the samples are captured under different seasons, weather conditions and resolutions. The overall size, however, is much smaller than our dataset (10k images per class vs. 400 images per class).

The UC Merced Land Use Dataset by [59], is one of the most utilized datasets [60], [57], [61], [62], [63] and has been introduced in 2010. This set covers 21 object classes with 100 images each, in a spatial resolution of 30 cm. Although, this dataset covers mostly all of the, for this research determined, land covers and objects, the research project of this reading requires far more image samples per category.

Due to the reasons mentioned above, we have created a new satellite dataset from publicly available satellite and GIS data to better address our research objectives.

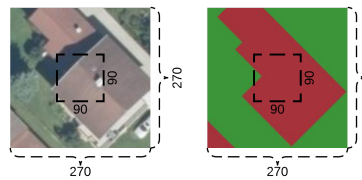
III. Dataset

The satellite image dataset for our investigation has been retrieved from the orthophoto map of the Tyrolean Tiris Database in combination with Open Government Data (OGD). Images together with the OGD have been loaded into the GIS software QGIS (version 2.18.2), which enables to align the satellite images and the vector layers of the OGD and to process them together. The OGD covers multiple different land cover and object categories. From the rich available data, we have selected those categories that are most relevant for our research purpose of location quality assessment. A total of eight object and land cover categories has been selected:

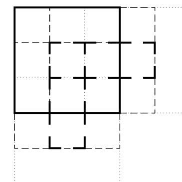
Residential Buildings	Local Roads
Non-Residential Buildings	Grassland
Highway	Water
Federal Highway	Forest

These eight classes are assumed to be most descriptive and provisioning for the visual representation of local land covers and infrastructures in Austria. Together they represent the ground truth classes of our dataset. By using the scripting capabilities of QGIS we have exported large-scale image tiles from the satellite images at the highest available resolution. The area from which tiles were extracted covers the entire region of Tyrol in Austria. Since large areas in Tyrol are hardly or non-settled (covered settlement area 11.8% of whole Tyrol [64]) we have selected those tiles which cover settlement areas, resulting in a total of 21,076 tiles. Each obtained tile (4,050x4,050 px) covers 500x500 m^2. The tiles have a ground resolution of 8.1 cm^2 per pixel. The vector layers were rasterized with the same resolution as the satellite image tiles for further processing. Figure 1(a) shows a part of an image tile with the corresponding ground-truth. The dataset has then been partitioned randomly into a training set (80%), including a hold-out validation set (10%), and a test set (20%). This initial split guarantees a clear separation of training and test data in our experiments.

(a) extracted satellite image with ground-truth



(b) patch to class assignment



(c) sliding window approach

Fig. 1. (a) (left) extracted satellite image; (right) extracted ground truth. (b) Patch extraction algorithm, localizing the centered object and identifying its corresponding class. (c) Sliding window approach, visualizing the right and downward movement by 90px.; Credit: Land Tirol - data.tirol.gv.at

Next, every satellite image and ground truth layer has been processed in parallel by a sliding window operation to cut out patches from the images. The patch size is determined by $4050px/15 = 270px$. A single patch thus covers 478 m^2 of landscape and thereby provides a significant amount of visual context for the detection of land covers and objects.

For each image patch a ground-truth label has been computed. The ground-truth label is derived from the center area of a patch (90x90 px) and is specified as the label of the most frequently occurring land cover in this area. The remaining area of the patch is considered as contextual information and is not used to define the class label. Figures 1(b) illustrates the labeling process of an image patch, showing a residential building.

The patch extraction is performed with a step size of 90 px for the sliding window operation. In this way a dense coverage of the input tiles is obtained, see Figure 1(c). Thereby, each tile is cut into 1,849 patches.

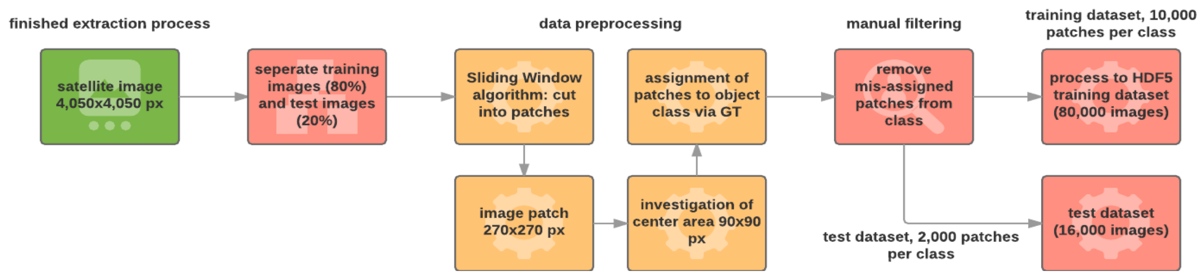


Fig. 2. The process workflow of database construction.

Fig. 3. An OGD polygon which is indicating the shape of a federal highway.; Credit: Land Tirol - data.tirol.gv.at

TABLE I
NUMBERS AND PERCENTAGES OF MIS-ASSIGNED PATCHES INSIDE EACH CATEGORY DUE TO INACCURATE OGD.

Category	**% mis-assigned**
Residential Buildings	1% (132/15,000)
Non-Residential Buildings	1% (147/15,000)
Highway	57% (8,578/15,000)
Federal Highway	20% (2,991/15,000)
Local Road	6% (913/15,000)
Grassland	22.5% (3390/15,000)
Water	35.5% (5,332/15,000)
Forest	10% (1,489/15,000)

During construction of the dataset it showed that the OGD shape files (vector layers) frequently exhibit poor accuracy. Distorted and inaccurate polygons occur frequently, which do not match with the satellite image. Figure 3, demonstrates an example of an inaccurate OGD polygon. The visualized polygon should indicate the shape of a federal highway but actually overlaps to a large degree with the nearby forest. Similar observations have been made in all other categories. To clean the dataset, we have manually investigated the input tiles and the patches and removed those with inaccurate OGD polygons. For each category 15,000 patches have been manually investigated. The amount of patches with inaccurate ground-truth are listed in Table I. The percentage reaches from 6% for local roads to 57% for highway. The only exception are the two building classes with a tolerable deviation of only 1%. Since category of highway and water consisted of so many wrong images, additional images to the previous 15,000 have been investigated. The final training set consists of 10,000 patches per category and the test set contains another 2,000 patches per category. Thus, there are 80,000 patches in the training set and 16,000 patches in the. The complete process of dataset generation is summarized in Figure 2. Finally, the training dataset is converted to HDF5 format for easier processing.

IV. METHOD

For the classification of the different land covers and objects in our dataset we propose - in accordance with the current state of the art - end-to-end learning by deep convolutional neural networks (CNNs). As described in Section III, the satellite images have been pre-processed, i.e. cut into patches by applying a sliding window operation to the large-scale image tiles. The patches are stored in a HDF5 image database and fed into the employed networks

A. Center Patch Approach

The center patch approach is inspired by the work of [22] and has been considered already during database construction. The idea of this approach is that the networks should learn to classify only the center area of a given patch and to use the surrounding context information as additional input. For this reason, we label each patch according to the majority class in the center of the patch (see Section III). The visual context is used for making more accurate decisions and is not being directly predicted. By following this approach the networks are tuned to classify the image center by using additional information from the neighborhood. This is reasonable since some objects are more frequently surrounded by certain land covers than others, e.g. gardens around residential buildings, woods and bushes next to highways and water. Additionally, context information helps to disambiguate larger objects, that cover the entire center area. For the network training this means that the entire patch is fed in as input and the label for the center region is learned, independent of the land covers present in the neighborhood.

B. Network Models

For the training process, two network models have been employed for land cover classification. First, the VGG Network [23], which acts as a baseline and second our own model, called SatNet-8 in the following. The VGG model is used with its original architecture. Only minor adaptations have been made, i.e. increase image input size to 270x270, set number of output neurons in the last fully-connected layer to the number of classes in our dataset and application of the 'adam' optimizer instead of 'RMSProp'. SatNet-8 is a variation of VGG with the same input and output dimensions. For training, the same parameters and number of epochs are used to enable a fair performance comparison.

1) VGG Network: The VGG network is a CNN with 16 layers, i.e. 13 convolutional (conv) layers and three fully-connected (fc) layers, which have been formed to logical groups. The first group represents two conv layers and one max pooling operation. The conv layers operate with 64 filters and a filter size of 3 and ReLU as activation function. Max pooling is set to a stride of 2. In the second group the number of filters per conv layer increases to 128. The third group consists of three conv layers with 256 filters each and one max pooling operation. The fourth and the fifth group exhibit 512 filters for every conv layer. After the conv layers, two fc layers follow. These fc layers consist of 4096 neurons each and employ again ReLU for activation. Followed by a dropout function with a 50% dropout rate after the first and second fc layer. Dropout helps to avoid overfitting by randomly suppressing weight updates during learning [65]. The third and final fc layer defines the output and has been changed to a number of $K = 8$ neurons to be compatible to our dataset. This layer uses a softmax activation function. See Table II for a overview of the architecture. Finally a logistic regression function is applied.

2) SatNet-8: During the development of SatNet-8 the VGG Network has been used as a reference model. Multiple structural modifications have been made and each training session has been investigated in relation to its previous performance and structural design. Our basic assumption for the optimization is that we assume satellite images to have a lower overall complexity than arbitrary (object-related or scene-related) images, which were originally used to train VGG and similar networks. Thus, we expect that the complexity of the network can be reduced without loosing discriminative power. Therefore, most modifications have been made with the goal to reduce the models complexity and to keep training performance at a high level.

After evaluating several different modifications and strategies for the reduction of complexity, we decided for the following architecture. The SatNet-8 consists of only 10 conv layers (instead of 13 as in VGG) and three smaller fc layers, which have been arranged into five groups. Every group represents two conv layers and a max pooling operation, followed by a batch normalization (bn) layer. This layer acts as a regularizer for the distribution of the inputs to a given layer. It normalizes the layers input for each batch during the training process. As a result, it allows higher learning rates and improves accuracy. The first two groups of conv layers operate with 128 filters with a filter size of 3 and ReLU activation. Since the first conv layers of a network model recognize simple edge and shape information, the number of filters has been increased because many object types in our dataset are characterized by simple shapes, such as lines (e.g. boarders of buildings and roads). The max pooling is set to a stride of 2 and thereby always halves the input dimension. The following two groups exhibit 256 filters, which is again doubled to a total of 512 in the fifth group. The first and the second fc layers consist of only 1048 neurons with ReLU activation. Since, the number of classes in our dataset is much smaller than that used for the original VGG network, this reduction in complexity is reasonable. To avoid overfitting dropout functions with a 50% dropout rate are employed for the fc layers. The final output layer is similar to VGG Networks output layer, including the logistic regression function. The adaptations performed for the SatNet-8 network lead to a reduction of parameters by approximately a factor of four. Both networks are initialized with random weights and trained from scratch in our experiments.

TABLE II
COMPARISON OF ARCHITECTURE BETWEEN VGG AND SATNET-8

Layer Grp.	VGG	SatNet-8
Group 1	conv1 - 64 filters	conv1 - 128 filters
	conv2 - 64 filters	conv2 - 128 filters
	max pool	max pool
		batch normalization
Group 2	conv3 - 128 filters	conv3 - 128 filters
	conv4 - 128 filters	conv4 - 128 filters
	max pool	max pool
		batch normalization
Group 3	conv5 - 256 filters	conv5 - 256 filters
	conv6 - 256 filters	conv6 - 256 filters
	conv7 - 256 filters	
	max pool	max pool
		batch normalization
Group 4	conv8 - 512 filters	conv7 - 256 filters
	conv9 - 512 filters	conv8 - 256 filters
	conv10 - 512 filters	
	max pool	max pool
		batch normalization
Group 5	conv11 - 512 filters	conv9 - 512 filters
	conv12 - 512 filters	conv10 - 512 filters
	conv13 - 512 filters	
	max pool	max pool
		batch normalization
	fc1 - 4096 nodes	fc1 - 1024 nodes
	dropout - 50% rate	dropout 50% rate
	fc2 - 4096 nodes	fc2 - 1024 nodes
	dropout - 50% rate	dropout 50% rate
	fc3 - 8 nodes	fc3 - 8 nodes
	regression function	regression function
	# parameters: 180.8 M	**# parameters: 43.8 M**

C. Implementation

We employ Tensorflow with GPU support by CUDA Toolkit (version 7.5) and cuDNN (version 5.1) for training. On top of Tensorflow, the higher-lever API TFLearn[3] has been employed, which allows for rapid prototyping of experimental setups while still retaining full access to Tensorflows capacities. The training hardware has been an Intel Core i7-7700K with an Nvidia GeForce GTX 1080 Ti.

[3]A. Damien, TFLearn — TensorFlow Deep Learning Library. [Online] Available: http://tflearn.org. Accessed on: Aug 30, 2017

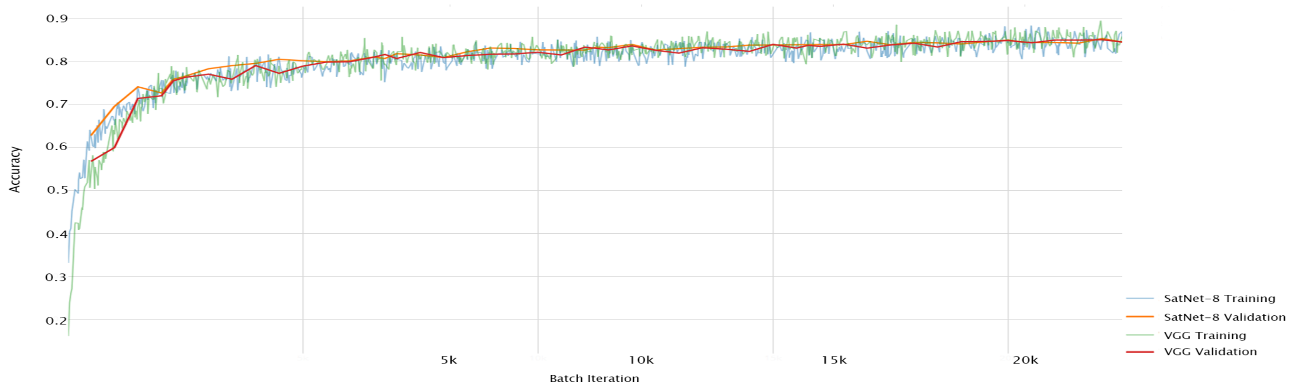


Fig. 4. Training and validation accuracy over the number of batch iterations.

V. Experiments & Results

A. Training Setup

Prior to training, the networks' input layers have been adjusted to an input dimension of 270x270 px (the size of the image patches in our dataset) and three color channels.We performed different pre-processing steps on the input images. First, we zero-center the images to normalize them regarding illumination and contrast variations. We zero center each image by the measured mean across all three RGB color channels only. This normalization facilitates the networks to converge faster. Second, we apply augmentation to the training images. Augmentation adds additional variation to the training set by performing image transformations on the training patches. We employ two transforms to the patches: rotations by 0, 90, 180, and 270 degree and vertical flipping. While the training data is loaded into the system, the methods randomly rotate every image and subsequently perform vertical flipping.

The learning rate is a key parameter in training and influences the convergence of the training significantly. A too high rate can lead to an unstable training while a too low rate unnecessarily increases the required training time. We employ stochastic gradient descent (adam optimizer) for learning with an initial learning rate of 0.0001. This optimizer computes adaptive learning rates for each parameter during the training process[66] and yields good and stable results in practice. To measure classification loss, we utilize the categorical cross-entropy loss function.

To monitor the training process, we have employed a validation set. This set has contained 10% of all training samples and has been chosen randomly thereof. These samples serve the network to validate its current performance after a training iteration and are never used for training itself. The training batch size has been adjusted to 32 images. A value of 32 represents a good performance trade-off for the training hardware employed in our experiments. We train both the networks from scratch for a number of 10 epochs.

Figure 4 visualizes the learning performance of both networks over the entire training. In direct comparison, both network models reach a similar performance level after 10 epochs. Interestingly, however, SatNet-8 faster learns, especially in early epochs. The VGG takes longer and more iterations to reach the same accuracy as the SatNet-8. Also the SatNet-8 seems to be more robust on the validation set. The main reason for the faster learning of SatNet-8 is the significantly lower number of parameters. This is also reflected by the training times. The training of the VGG network has taken 122.6 hours, whereas the SatNet-8 finished training after approx. 79.6 hours. As shown in Figure 4, both networks have potentially not reached their best performance capabilities after 10 epochs as the validation accuracy continuously increases which further shows that the network does not overfit on the data. Anyway, due to the total amount of training time, we stoped training after epoch 10. The final result is an overall training accuracy of VGG of 86.55% and SatNet-8 of 84.98%. The validation accuracy of VGG is 84.51% and that of SatNet-8 is 84.48%. There is no significant difference between the performance measures, except for the significantly lower training time of SatNet-8.

B. Classification Results

After both networks have been trained on our dataset, we have evaluated them on the so far unseen test set. Therefore, every test image patch has been passed through the network and the most probable class from the eight possible classes has been taken as the final prediction for the patch. To assess the performance, we have computed the overall accuracy as well as a confusion matrix. The confusion matrix provides insight into the quality of classification as it displays mis-classifications between all categories and thereby reveals the networks weaknesses and strengths. Figure 5 and Figure 6 present the confusion matrices with the percentages of correctly and incorrectly classified test image patches. The top row indicates the predicted categories (pc) and the left row provides the true label (tl). The diagonal contains the achieved classification accuracy in percent (in bold letters). The off-diagonal values represent mis-classifications. The column 'false positive' contains the percentage of false predictions between the two building-related categories (residental building

tl\pc	residential building	non-residential building	highway	federal highway	local road	grassland	water	forest	false positive	% mis-classification
residential building	**74**	**23.7**	0.05	0.05	1.45	0.25	0.45	0.05	23.7	26
non-residential building	**18.6**	**75.4**	0.35	0.3	2.15	0.15	2.9	0.15	18.6	24.6
highway	0	1.2	**91.05**	**6.3**	0.7	0.05	0.6	0.1	7	8.95
federal highway	0.4	1.6	6.8	**73.35**	**14.7**	0.5	1.95	0.7	21.5	26.65
local road	2	6.25	1	**9.45**	**75.9**	1.45	3.25	0.7	10.45	24.1
grassland	0.55	0.2	0	0.35	**1.75**	**94.75**	1.4	1		5.25
water	0.05	0.45	0	0.1	0.5	**1.05**	**95.95**	1.9		4.05
forest	0	0	0.05	0.15	0.35	0.3	**1.55**	**97.6**		2.4

Fig. 5. Confusion Matrix of the VGG Network. Values in percentage terms. 'false positive' describes percentage of false predictions between related object classes. '% mis-classification is the sum of all wrongly predicted images.

tl\pc	residential building	non-residential building	highway	federal highway	local road	grassland	water	forest	false positive	% mis-classification
residential building	**81.7**	**15.65**	0.05	0.1	1.65	0.15	0.65	0.05	15.65	18.3
non-residential building	**26**	**69.05**	0.4	0.15	2.35	0.3	1.45	0.3	26	30.95
highway	0	0.6	**92.65**	**5.1**	1.1	0	0.45	0.1	6.2	7.35
federal highway	0.5	0.95	4.95	**61.85**	**29.45**	0.85	0.7	0.75	34.35	38.15
local road	2.55	3.9	0.25	**5.25**	**83.15**	1.6	2.15	1.15	5.5	16.85
grassland	0.45	0.2	0	0.15	1.15	**95.75**	1	**1.3**		4.25
water	0.15	0.45	0.05	0.5	0.8	0.85	**94.75**	**2.45**		5.25
forest	0	0	0.05	0.05	0.25	0.1	**0.6**	**98.95**		1.05

Fig. 6. Confusion Matrix of the SatNet-8. Values in percentage terms. 'false positive' describes percentage of false predictions between related object classes. '% mis-classification is the sum of all wrongly predicted images.

and non-residental building, first two rows) and the percentage of false predictions among the three street categories (highway, federal highway, and local road, rows 3-5). Column '% of mis-classifications' sums up the percentages of wrongly predicted images per class. The VGG Network achieved its best results on the classes grassland, water, and forest. With an accuracy of 94.75% to 97.6%. VGG is capable of predicting 1,895 / 2,000 grassland samples, 1,918 / 2,000 water samples, and 1,952 / 2,000 forest samples correctly. When it comes to the prediction of street types, the category of highway achieved 91.05% accuracy, where only some samples tend to be mis-classified as federal highway. Since in some areas these two street types look quiet similar, these errors are comprehensible. A similar pattern are observed for federal highway and local road. The building-related classes can be detected with a lower accuracy of 74% and 75,4% for residential building and non-residential building, respectively. A total amount of 1,480 / 2,000 and 1,508 / 2,000 samples have been correctly allocated to their corresponding class. Mis-classifications of buildings, however, mostly occur between the two building categories. If we join both categories, an overall accuracy for building detection of approximately 96% would be reached.

The SatNet-8 has achieved similar results to the VGG Network with slightly different prediction accuracies. The biggest difference is the class of federal highway, which is only predicted correctly in 61.85% of all cases, whereas the VGG

TABLE III
PERFORMANCE COMPARISON BETWEEN VGG AND SATNET-8

	VGG	SatNet-8
average accuracy	84.75%	84.73%
building detection	95,85%	96,2%
building classification	74,7%	75,375%
road detection	93,08%	94,58%
road classification	80,1%	79,21%

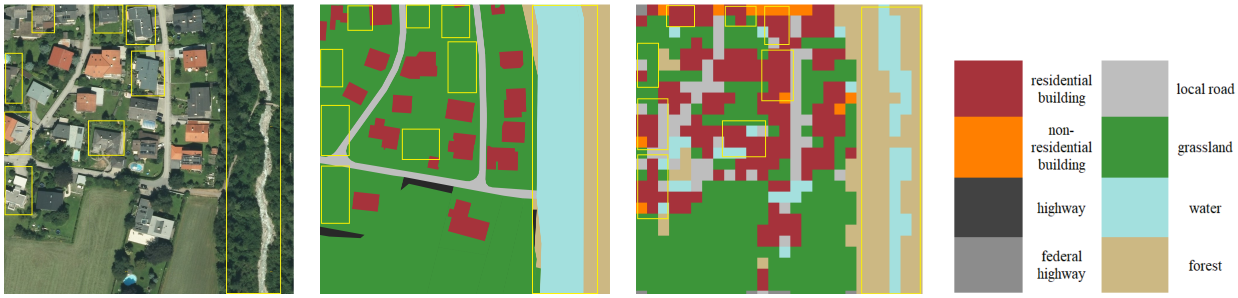
Network achieved 73.35%. However, the SatNet-8 performs notable better in the categories of residential building 81.7% and local road 83.15%. The overall prediction accuracy and predicted values of related object groups reveal that both networks perform similar and at a very high performance level, see Table III.

For location quality assessment the building-related categories have an especially high relevance. Thus, we want to compare the performance level of our building type classification (see Table III) to related works. An approach for building classification is presented in [53]. The authors classify buildings from Light Detection and Ranging (LiDAR) data [67] and consider three building types: single-family houses, multiple-family houses, and non-residential buildings. A overall accuracy of 70% is achieved for all three classes which is slightly lower than our accuracy of 74.7% to 75.4%. This may on the one hand be due to the larger number of classes and thus the more complex tasks but on the other hand LIDAR data provides 3D information which is not available for our method. Thus, the performance level achieved by our method can be considered similar. Another approach for building classification is presented in [52]. The authors investigate the classification of buildings into three main classes: Residential/Small Buildings, Apartment Buildings, and Industrial and Factory Building by a Random Forest (RF) classifier and report an average F1 score of 69.56% over all three classes. Although F1 score cannot be directly compared to classification accuracy, it shows that the performance level is similar.

C. Detection of Erroneous GIS Data

As mentioned already in Section III, we have observed partly poor accuracy of the OGD polygons with the visual content in the satellite images, which has led to some serious complications during dataset generation. The detection of wrong OGD polygons could be automated by applying a reliable land cover classification or segmentation algorithm to satellite image and by finding inconsistencies between the prediction and the polygons. Thereby, mis-aligned polygons of buildings, roads, rivers, lakes, forest, and other object categories could be identified and in a further step corrected (semi-)automatically.

Figure 7 provides an example image with the ground-truth information derived from the available OGD polygons. It can be observed that a number of buildings are missing in the ground-truth which is maybe due to the fact that the polygons



(a) original satellite image (b) original OGD GT (c) SatNet-8 predicted labels (d) color legend

Fig. 7. (a) Original satellite image from the Tiris Database; (b) Ground-truth derived from the original GIS Information (OGD polygons); (c) Predicted patch-wise labels via SatNet-8. (d) Color legend for land cover classes. (The yellow boundaries indicate buildings and land cover objects, which have been detected correctly by the SatNet-8, but have not been displayed accurately in the OGD GT.); Credit: Land Tirol - data.tirol.gv.at

TABLE IV
HEDONIC PRICE MODEL: FOR CONVENIENCE WE SHOW ONLY PARAMETERS FROM THE SATNET-8 PREDICTIONS. COLUMN TWO SHOWS THE EFFECT OF THE COEFFICIENTS ON THE LOGARITHMUS OF THE PRICE. R^2 IST THE ADJUSTED COEFFICIENT OF DETERMINATION AND RMSE MEANS THE ROOT-MEAN-SQUARE ERROR

	Hedonic Price Model
Residential Buildings	0.004*
Non-Residential Buildings	0.005
Highway	−0.012**
Federal Highway	−0.007**
Local Road	−0.001
Water	0.001
Grassland	0.002***
Forest	0.002**
Adj. R^2	0.66
Num. observations	2739
RMSE	0.34

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

are rather old. In Figure 7(c) the patch-wise prediction of the SatNet-8 is depicted. We can see that several buildings that were not present in the ground-truth were correctly detected (highlighted in yellow boundaries). We thus, see a great potential for the automatic detection of inaccurate GIS data by automated land cover classification.

D. Hedonic Pricing

Location quality and livability are usually approximated by the price or price-level of a location. This means that by predicting the price of a location, conclusions about the location quality can be drawn. A popular approach for the modeling of prices of real estate is hedonic pricing (which we briefly described in the introduction). We evaluated the utility of our land cover classifications for hedonic price modeling by feeding parameters, derived from the SatNet-8 predictions, to the hedonic (regression) model as additional inputs. Therefore, we used 2739 valuations of residential buildings, as we focus only on the location and neighborhood characteristics. In our model, for the location variables, we used only the municipalities in Tyrol, Austria. Additional to this information, we used the information from our eight land cover clusters. Therefore, we calculated the share (in terms of covered area) of each category within a 100 meter radius from each residential building. The result of the regression is displayed in Table IV. For convenience we cut out the location coefficients. All municipality dummies are significant and reflect the expected magnitude, but are not in the focus of this paper. The categories from the SatNet-8 reflect the expected magnitude and most are statistically significant. For example the category grassland shows that a higher portion of grassland results in a higher price. On the other side, the presence of highway or federal highway in the near neighborhood indicates a negative impact which is related to a decline in price. This is in line with real estate theory. Our investigation provides first insights that confirm a positive contribution of the automatically extracted visual parameters and demonstrates that this novel type of modeling location quality has promising potential.

VI. CONCLUSION

In this paper, we presented a first approach towards the automated assessment of location quality from satellite images. We adapted the SatNet-8 to predict different land covers and objects from our novel dataset. There out, we have fed information derived from the recognized land covers into a regression-based price model which acts as a proxy for the assessment of location quality. Our results show that (i) land cover classification can be performed with high accuracy and demonstrate that automatic classification could further be used in the future for the detection of mis-aligned and erroneous GIS data; (ii) our adapted SatNet-8 reaches state-of-the-art performance in much less training time compared to the reference VGG Network; (iii) the automatically extracted visual information improves the prediction of real estate prices and thereby shows clear potential for the description of location quality. Future work will include the extension of the proposed approach to larger areas including urban areas like cities and additional land covers.

ACKNOWLEDGMENT

This work was supported by the Austrian Research Promotion Agency (FFG), Project No. 855784 and Project No. 856333.

References

[1] C. Berlanga-Robles and A. Ruiz-Luna, "Land use mapping and change detection in the coastal zone of northwest mexico using remote sensing techniques," vol. 18, pp. 514–522, 06 2002.

[2] Z. Wang and L. Liu, "Assessment of coarse-resolution land cover products using casi hyperspectral data in an arid zone in northwestern china," *Remote Sensing*, vol. 6, no. 4, pp. 2864–2883, 2014. [Online]. Available: http://www.mdpi.com/2072-4292/6/4/2864

[3] A. T. Hudak and C. A. Wessman, "Textural analysis of historical aerial photography to characterize woody plant encroachment in south african savanna," *Remote sensing of environment*, vol. 66, no. 3, pp. 317–330, 1998.

[4] X. Li and A. G. O. Yeh, "Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the pearl river delta," vol. 19, pp. 1501–1518, 05 1998.

[5] H. Chang and W. S. Yoon, "Improving the classification of landsat data using standardized principal components analysis," *KSCE Journal of Civil Engineering*, vol. 7, no. 4, pp. 469–474, Jul 2003. [Online]. Available: https://doi.org/10.1007/BF02895842

[6] H. Long, X. Wu, W. Wang, and G. Dong, "Analysis of urban-rural land-use change during 1995-2006 and its policy dimensional driving forces in chongqing, china," *Sensors*, vol. 8, no. 2, pp. 681–699, 2008. [Online]. Available: http://www.mdpi.com/1424-8220/8/2/681

[7] P. K. Mallupattu and J. R. S. Reddy, "Analysis of land use/land cover changes using remote sensing data and gis at an urban area, tirupati, india," vol. 2013, p. 268623, 05 2013.

[8] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325–3337, June 2015.

[9] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images." *ISPRS Journal of Photogrammetry & Remote Sensing*, 2016.

[10] D. Lu, Q. Weng, E. Moran, G. Li, and S. Hetrick, *Remote sensing image classification*. CRC Press/Taylor and Francis: Boca Raton, FL, USA, 2011.

[11] C. Homer, J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown, "Completion of the 2011 national land cover database for the conterminous united states–representing a decade of land cover change information," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 5, pp. 345–354, 2015.

[12] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2017.

[13] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[14] T. Ishii, R. Nakamura, H. Nakada, Y. Mochizuki, and H. Ishikawa, "Surface object recognition with CNN and SVM in Landsat 8 images," in *Proceedings of the 14th IAPR International Conference on Machine Vision Applications, MVA 2015*, 2015.

[15] P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," *Pattern Recognition and Image Analysis*, vol. 26, no. 1, pp. 9–15, 2016. [Online]. Available: http://link.springer.com/10.1134/S1054661816010065

[16] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," 2016.

[17] G. Cheng, P. Zhou, and J. Han, "Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, 2016.

[18] S. Ahmed, M. Liwicki, M. Weber, and A. Dengel, "Improved Automatic Analysis of Architectural Floor Plans," in *2011 International Conference on Document Analysis and Recognition*. IEEE, sep 2011, pp. 864–869. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6065434

[19] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, 2016. [Online]. Available: https://pdfs.semanticscholar.org/1b3a/c4b4187a3dbc9373869e7774b1dc63f748d2.pdf

[20] L. J. Zhao, P. Tang, and L. Z. Huo, "Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, 2014.

[21] Y. Cao, X. Niu, and Y. Dou, "Region-based convolutional neural networks for object detection in very high resolution remote sensing images," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016*, 2016.

[22] S. Saito and Y. Aoki, "Building and road detection from large aerial imagery," vol. 9405, 02 2015.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[24] W.-C. Liao and X. Wang, "Hedonic house prices and spatial quantile regression," *Journal of Housing Economics*, vol. 21, no. 1, pp. 16–27, 2012.

[25] A. T. Court, "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*, C. F. Roos, Ed. New York: General Motors, 1939, pp. 99–117.

[26] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *2014 Fourth International Conference on Advanced Computing Communication Technologies*, Feb 2014, pp. 5–12.

[27] J. A. dos Santos, O. A. B. Penatti, and R. da Silva Torres, "Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification." in *VISAPP (2)*, 2010, pp. 203–208.

[28] Y.-W. Seo, C. Urmson, and D. Wettergreen, "Exploiting Publicly Available Cartographic Resources for Aerial Image Analysis."

[29] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *null*. IEEE, 2003, p. 1470.

[30] K. Nogueira, O. A. B. Penatti, and J. A. Dos Santos, "Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification," 2016.

[31] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015, pp. 44–51.

[32] K. Nogueira, W. O. Miranda, and J. A. D. Santos, "Improving spatial feature representation from aerial scenes by using convolutional networks," in *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug 2015, pp. 289–296.

[33] J. Zhao and S. You, "Road network extraction from airborne lidar data using scene context," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 9–16.

[34] J. P. Cohen, W. Ding, C. Kuhlman, A. Chen, and L. Di, "Rapid building detection using machine learning," *CoRR*, vol. abs/1603.04392, 2016. [Online]. Available: http://arxiv.org/abs/1603.04392

[35] Y. E. Merabet, C. Meurie, Y. Ruichek, A. Sbihi, and R. Touahni, "Building roof segmentation from aerial images using a line-and region-based watershed segmentation technique," in *Sensors*, 2015.

[36] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular vhr optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, March 2013.

[37] F. Dornaika, A. Moujahid, Y. E. Merabet, and Y. Ruichek, "Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors," *Expert Systems with Applications*, vol. 58, pp. 130 – 142, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417416301154

[38] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, "A higher-order crf model for road network extraction," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1698–1705.

[39] J. A. Montoya-Zegarra, J. D. Wegner, L. Ladickỳ, and K. Schindler, "Mind the gap: modeling local and global context in (road) networks," in *German Conference on Pattern Recognition*. Springer, 2014, pp. 212–223.

[40] C. Poullis, "Tensor-cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 95, pp. 93 – 108, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271614001543

[41] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14 680–14 707, 2015. [Online]. Available: http://www.mdpi.com/2072-4292/7/11/14680

[42] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. R. Nemani, "Deepsat - A learning framework for satellite imagery," *CoRR*, vol. abs/1509.03602, 2015. [Online]. Available: http://arxiv.org/abs/1509.03602

[43] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *CoRR*, vol. abs/1508.00092, 2015. [Online]. Available: http://arxiv.org/abs/1508.00092

[44] M. Längkvist, A. Kiselev, M. Alirezaie, and A. Loutfi, "Classification and segmentation of satellite orthoimagery using convolutional neural networks," vol. 8, p. 329, 04 2016.

[45] V. Mnih and G. E. Hinton, *Learning to Detect Roads in High-Resolution Aerial Images*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 210–223.

[46] J. Wang, J. Song, M. Chen, and Z. Yang, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," *International Journal of Remote Sensing*, vol. 36, no. 12, pp. 3144–3169, 2015.

[47] A. Lagrange and B. Le Saux, "Convolutional neural networks for semantic labeling," Tech. Rep., Onera–The French Aerospace Lab, Tech. Rep., 2015.

[48] A. Lagrange, B. L. Saux, A. Beaupère, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo, and M. Ferecatu, "Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 4173–4176.

[49] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. V.-D. Hengel, "Effective semantic pixel labelling with convolutional networks and conditional random fields," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015, pp. 36–43.

[50] O. Firat, G. Can, and F. T. Y. Vural, "Representation learning for contextual object and region detection in remote sensing," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 3708–3713.

[51] Q. Lv, Y. Dou, X. Niu, J. Xu, and B. Li, "Classification of land cover based on deep belief networks using polarimetric radarsat-2 data," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 4679–4682.

[52] M. Belgiu, I. Tomljenovic, T. J. Lampoltshammer, T. Blaschke, and B. Höfle, "Ontology-based classification of building types detected from airborne laser scanning data," *Remote Sensing*, vol. 6, no. 2, pp. 1347–1366, 2014. [Online]. Available: http://www.mdpi.com/2072-4292/6/2/1347

[53] Z. Lu, J. Im, J. Rhee, and M. Hodgson, "Building type classification using spatial and landscape attributes derived from lidar remote sensing data," *Landscape and Urban Planning*, vol. 130, pp. 134 – 148, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169204614001601

[54] N. Srećković, N. Lukač, B. Žalik, and G. Štumberger, "Determining roof surfaces suitable for the installation of PV (photovoltaic) systems, based on LiDAR (Light Detection And Ranging) data, pyranometer measurements, and distribution network configuration," *Energy*, vol. 96, pp. 404–414, feb 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544215017235

[55] S. Szabó, P. Enyedi, M. Horváth, Z. Kovács, P. Burai, T. Csoknyai, and G. Szabó, "Automated registration of potential locations for solar energy production with Light Detection And Ranging (LiDAR) and small format photogrammetry," *Journal of Cleaner Production*, vol. 112, pp. 3820–3829, jan 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959652615010483

[56] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping," p. 16, 2015. [Online]. Available: http://arxiv.org/abs/1510.00098

[57] G. S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2017. [Online]. Available: https://arxiv.org/pdf/1608.05167.pdf

[58] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," vol. 12, pp. 1–5, 11 2015.

[59] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 270–279. [Online]. Available: http://doi.acm.org/10.1145/1869790.1869829

[60] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–19, 2017.

[61] O. A. B. Penatti, K. Nogueira, and J. A. Dos Santos, "Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?"

[62] K. Nogueira, W. O. Miranda, and J. A. D. Santos, "Improving Spatial Feature Representation from Aerial Scenes by Using Convolutional Networks," in *Brazilian Symposium of Computer Graphic and Image Processing*, 2015.

[63] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land Use Classification in Remote Sensing Images by Convolutional Neural Networks," *Arxiv preprint arXiv:1508.00092*, aug 2015. [Online]. Available: http://arxiv.org/abs/1508.00092

[64] Amt der Tiroler Landesregierung. Flaechennutzung. Accessed: 2017-08-29. [Online]. Available: https://www.tirol.gv.at/statistik-budget/statistik/flaechennutzung/

[65] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[66] S. Ruder, "An overview of gradient descent optimization algorithms," *CoRR*, vol. abs/1609.04747, 2016. [Online]. Available: http://arxiv.org/abs/1609.04747

[67] S. Reutebuch, H. Andersen, and R. Mcgaughey, "Light detection and ranging (lidar): An emerging tool for multiple resource inventory," vol. 103, pp. 286–292, 09 2005.

Fashion and Apparel Classification using Convolutional Neural Networks

Alexander Schindler
Austrian Institute of Technology
Digital Safety and Security
Vienna, Austria
alexander.schindler@ait.ac.at

Thomas Lidy
Vienna University of Technology
Institute of Software Technology
Vienna, Austria
lidy@ifs.tuwien.ac.at

Stephan Karner, Matthias Hecker
MonStyle
Vienna, Austria
matthias.hecker@monstyle.io

***Abstract*—We present an empirical study of applying deep Convolutional Neural Networks (CNN) to the task of fashion and apparel image classification to improve meta-data enrichment of e-commerce applications. Five different CNN architectures were analyzed using clean and pre-trained models. The models were evaluated in three different tasks *person* detection, *product* and *gender* classification, on two small and large scale datasets.**

I. Introduction

The recent progress in the image retrieval domain provides new possibilities for a vertical integration of research results into industrial or commercial applications. Based on the remarkable success of Deep Neural Networks (DNN) applied to image processing tasks, this study focuses on the task of fashion image classification. Photographs of clothes and apparels have to be classified into a set of pre-annotated categories such as skirt, jeans or sport-shoes. Online e-commerce companies such as Asos-EU [1], Farfetch [2] or Zalando [3] provide access to the data of their products in stock including item-meta-data and images. Especially the provided meta-data varies in quality, granularity and taxonomy. Although, most of the companies provide categorical descriptions of their products, the applied terminology varies as well as the depth of the categorical hierarchy. Fashion image classification is thus used to consolidate the meta-data by enriching it with new generalized categorical labels.

This is a traditional image processing task with domain specific challenges of large variating styles, textures, shapes and colors. A major advantage is the image quality which are professionally produced high quality and high resolution images. There are generally two categories of photographs. The first arranges products in front of a white background. The second portraits a person or parts of a person who is wearing the products. While the first category reduces semantic noise of the images, the second one introduces it, because a person wearing multiple items such as jeans, t-shirt, shoes and belt is only assigned to a single label. Clothing and apparel retrieval has been addressed to find clothes similar to a photograph [1] or a given style [2]. The main challenge these studies faced was the definition and extraction of relevant features to describe the semantic content of the images with respect to the high variability and deformability of clothing items. Recent approaches harness the potential of Deep Neural Networks (DNN) to learn the image representation. In [3] a siamese network of pre-trained Convolutional Neural Networks (CNN) is used to train a distance function which can be used to asses similarities between fashion images.

In this study we present an empirical evaluation of various DNN architectures concerning their classification accuracy in different classification tasks. These tasks are evaluated on two different datasets on further two different scales. First, a wide evaluation is performed on a smaller scale dataset and the best performing models are then applied to large scale datasets. The remainder of this paper is organized as follows. In Section II we review related work. In Section III the datasets used for the evaluation are presented. Section IV provides an overview of the evaluated neural network architectures. Section V describes the evaluation setup and summarizes as well as discusses the results. Finally, conclusions and outlooks to future work are given in Section VI.

II. Related Work

Recently, CBIR has experienced remarkable progress in the fields of image recognition by adopting methods from the area of deep learning using convolutional neural networks (CNNs). A full review of deep learning and convolutional neural networks is provided by [4]. Neural networks and CNNs are not new technologies, but with early successes such as LeNet [5], it is only recently that they have shown competitive results for tasks such as in the ILSVRC2012 image classification Challenge [6]. With this remarkable reduction in a previously stalling error-rate there has been an explosion of interest in CNNs. Many new architectures and approaches were presented such as *GoogLeNet* [7], *Deep Residual Networks (ResNets)* [8] or the *Inception Architecture* [7]. Neural networks have also been applied to metrics learning [9] with applications in image similarity estimation and visual search. Recently two datasets have been published. The MVC Dataset [10] for view-invariant clothing retrieval (161.638) images and the DeepFashion Dataset [11] with 800.000 annotated real life images.

[1] http://www.asos.de/
[2] https://www.farfetch.com
[3] https://www.zalando.de/

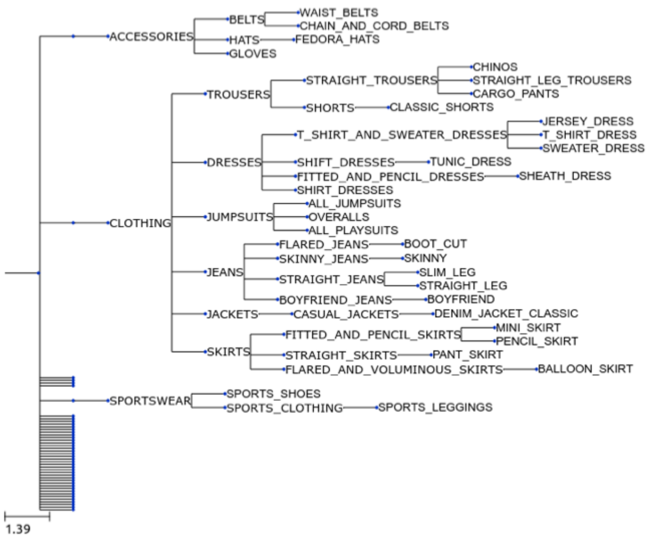


Fig. 1. Fashion categories hierarchy.

III. Data

The data provided was retrieved from online e-commerce companies such as Asos-EU, Farfetch or Zalando.

Person: The persons dataset consists of 7833 images and the corresponding ground truth assignments. 5.669 images are labeled as *Person* and 2.164 images are labeled as *Products*.

Products: The product dataset consists of 234.884 images and their corresponding ground-truth assignments. These images belong to 39.474 products where each product is described by 5,95 images on average. Ground-truth labels are provided for categories category, gender and age. All labels, including age, are provided on a categorical scale. The provided ground-truth assignments consists of 43 classes for the category attribute. These categories are based on a hirarchical taxonnomy. The hirarchy for the provided dataset is depicted in Figure 1. Its largest class *SPORTS SHOES* contains 66.439 images (10.807 products) and its smallest class *JUMPSUITS* 6 images (1 product). To facilitate more rapid experimentation, the provided dataset was sub-sampled to approximately 10% of its initial size. Further, due to the class imbalance of the provided category labels, an artificial threshold has been applied to the class sizes of the assignments. All classes with less than 100 images have been skipped. The remaining classes have been sub-sampled to a 10% subset. The sub-sampling adhered to further restriction. First, stratification was used to ensure that the frequency distribution of class labels in the subsample corresponds to that of the original one. Second, sub-sampling was performed on product-level. This ensured the consistency of product-images and that there are no products with only one image. Finally, sub-sampling of a class was stopped when a minimum of 100 images was reached. This resulted in a subset of 23.305 instances, ranging from 5.659 images for *SPORTS_SHOES* (922 products) and 103 images for *STRAIGHT_LEG_TROUSERS* (19 products).

IV. Deep Neural Network Models

In this study we compared five different DNN architectures which varied in depth and number of trainable parameters, including three winning contributions to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [12] and two compact custom CNNs with fewer trainable parameters. The following architectures were evaluated:

Vgg16 and Vgg19: very deep convolutional neural networks (VGGnet) [13] with 16/19 layers and 47/60 million trainable parameters, reaching an ILSVRC top-5 error rate of 6.8%.

InceptionV3: high performance network at a relatively modest computational cost [7] with 25 million trainable parameters reaching an ILSVRC top-5 error rate of 5.6%.

Custom CNN and Vgg-like: compact convolutional neural network with only 10 million trainable parameters.

The models were implemented in Python 2.7 using the keras[4] Deep Learning library on top of the Theano[5] backend.

V. Evaluation

The Convolutional Neural Networks were evaluated towards their classification accuracies in the tasks of differentiating *persons from products* as well as classifying product images according their product *category* and *gender*. We performed three-fold cross-evaluation and calculated the accuracies on a per-image and a per-product scale. To calculate the per-product

[4]https://github.com/fchollet/keras
[5]https://github.com/Theano/Theano

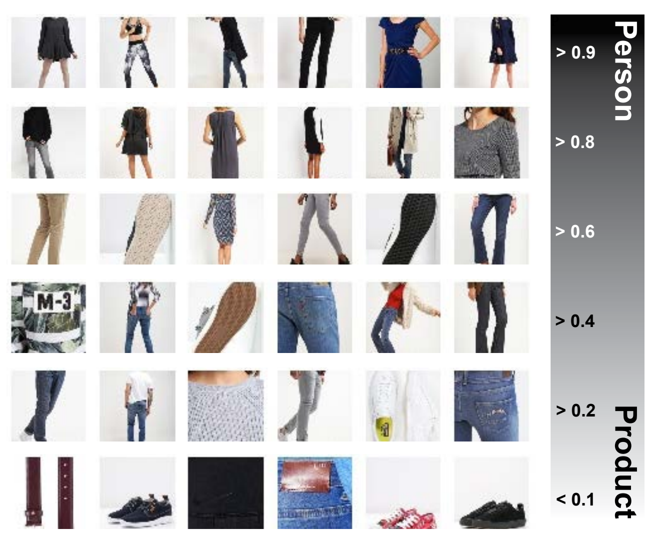


Fig. 2. Examples predictions of the *person detector*. Prediction was realized as binary classification. Values above a values of 0.5 are classified as *persons* and values below as *products*. Images in the first line thus represent images predicted as *persons* with high confidence.

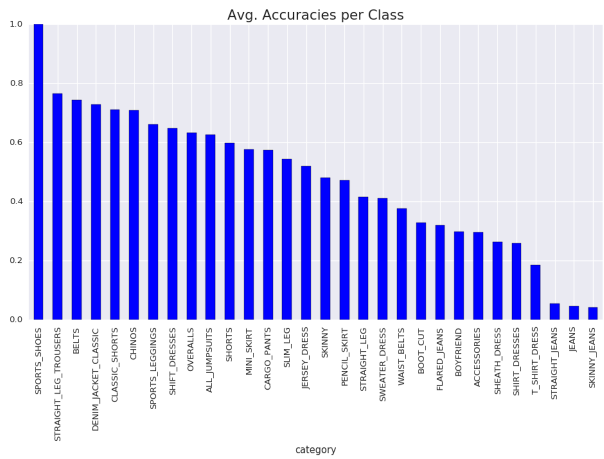


Fig. 3. Prediction accuracies on a per-image level for the best performing model - a fine-tuned *InceptionV3* on the 234K dataset.

accuracy the cumulative maximum of all predicted product images was taken into account.

A. Detecting Persons

Person detection was introduced based on the observation that products are presented in two general types. First, there are images of products placed in front of a white background or table. The other type of images are worn products. Because persons on these images are wearing more than one product such as trousers, shirts, shoes and belts, it is hard for a classifier to learn the right boundaries. Thus, the intention was to train a person detector and to either filter person images, or to use this additional information as input for further models.

We applied a custom VGG-like CNN with three layers of batch-normalized stacked convolution layers with 32, 64 and 64 3x3 filters and a 256 units fully connected layer with 0.5 dropout. We realized this task as a binary classification problem by using a sigmoid activation function for the output layer. Predictions greater or equal 0.5 were classified as persons. This approach already provided an accuracy of 91.07% on the *person* dataset. Figure 2 shows example images of the person detection model. Images on the bottom row were predicted with values below 0.1 and are thus categorized as *products*, whereas images in the top-row are considered to be *persons*.

B. Product Classification

The product classification experiments were conducted using the different CNN architectures presented in Section IV on two different scales. First, a broad evaluation was performed on the small-scale subset of 23.305 images. Then, the best performing models were evaluated on the large scale 234.408 images dataset. All models, except those where explicitly mentioned, were trained using image data augmentation, including horizontal flipping of the image, shifting it by 25% in height and width as well as a 25% zoom range.

1) Train from scratch or Fine-tune: This part of the evaluation deals with the question of whether to train a model from scratch or to fine-tune a pre-trained model. The availability of a large collection of high quality images and a relative small number of classes suggests that models can be effectively fitted according the specific domain.

The results presented in Table I show that pre-trained models outperform clean models that have been specifically trained from scratch using only the images of the fashion image collection. Additionally, we evaluated the two different types of applying pre-trained models: a) resetting and training only the top fully connected layers while keeping all other parameters fixed, and b) continued fitting of all parameters on the new data - which is also referred to as fine-tuning. In either way the 1000 unit output layer of the pre-trained models had to be replaced with a 30 units layer representing the 30 product categories.The results of the evaluation show that fine-tuning outperforms the fitting of clean fully connected layers by 5.9% (VGG16) to 7.9% (InceptionV3). The smaller custom models did provide an advantage concerning processing time of fitting and applying the model, but their accuracy results differ by 16.1% to the top performing fine-tuned model.

Figure 3 shows the prediction accuracy per class for the best model (fine-tuned IncepionV3) on the 234.408 images dataset. The most reliably predicted classes are *SPORT_SHOES*, *STRAIGHT LEG TROUSERS* and *BELTS*, the least reliable classes are *STRAIGHT JEANS*, *JEANS* and *SKINNY_JEANS*. These results indicate the problem of different granularity within the provided ground-truth assignments. Root- and leaf-

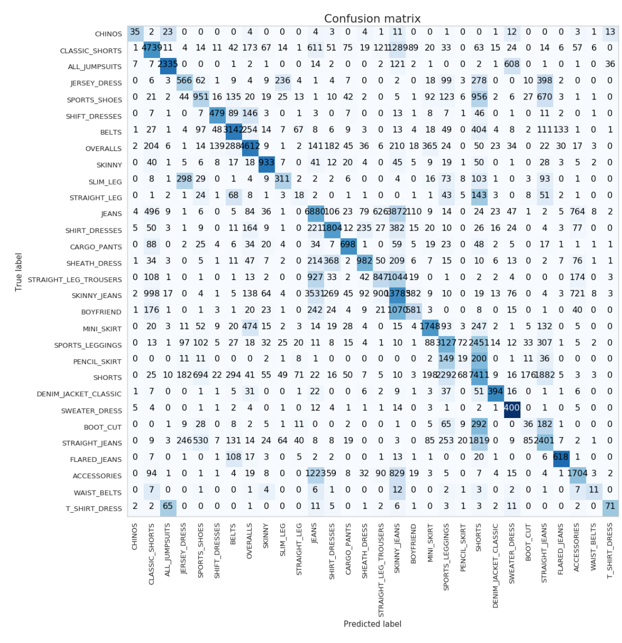


Fig. 4. Confusion matrix on a per-image level for the best performing model - a fine-tuned *InceptionV3* on the 234K dataset. The vertical axis represents the annotated category, the horizontal the prediction.

Description	best fold	best fold cum max	Mean cum max
InceptionV3, pretrained, fine-tuned	0.706	0.794	0.791
InceptionV3, pretrained, fine-tuned	0.658	0.729	0.716
VGG16, pretrained, fine-tuned	0.646	0.711	0.691
InceptionV3, pretrained, fine-tuned, person filter model as layer	0.569	0.685	0.658
VGG19, pretrained, fine-tuned	0.579	0.673	0.634
InceptionV3, pretrained, fine-tuned, no augmentation	0.564	0.673	0.647
VGG19, pretrained, train only top-layers	0.578	0.669	0.343
VGG16, pretrained, train only top-layers	0.603	0.652	0.368
InceptionV3, pretrained, train only top-layers	0.585	0.650	0.643
InceptionV3, pretrained, fine-tuned - person filtered metadata	0.640	0.636	0.614
InceptionV3, clean	0.492	0.594	0.580
Custom CNN, augmentation	0.506	0.568	0.538
Custom CNN	0.463	0.556	0.523
Custom VGG-like	0.438	0.549	0.519
VGG16, clean	0.439	0.455	0.443
VGG19, clean	0.437	0.447	0.430
VGG19, pretrained, train only top-layers	0.819	0.887	0.880
InceptionV3, pretrained, fine-tuned	0.798	0.863	0.836
VGG19, pretrained, fine-tuned	0.762	0.846	0.830

TABLE I

CLASSIFICATION RESULTS FOR THE *product category* CLASSIFICATION TASK. RESULTS SUMMARIZE PER IMAGE ACCURACY OF THE BEST FOLD, PER PRODUCT ACCURACY OF THE BEST FOLD, MEAN PER PRODUCT ACCURACY OF ALL FOLDS.

nodes are used interchangeably which results from the aggregation of different e-commerce catalogs using different taxonomies. Although confusion a child- with a parent-class is semantically not wrong, but the trained models do not take this hierarchy into account and predict each label individually. Thiseffect can be seen in the confusion matrix in Figure 4 where spezialized classes such as *JEANS* and *SKINNY_JEANS* or *SKINNY* and *SKINNY_JEANS* or are confused frequently.

C. Gender Prediction

The aim of the gender prediction task was to predict the intended gender of the product into the classes *MALE*, *FEMALE* and *UNISEX*. The results are comparable to the product classification task in the sense that pre-trained and fine-tuned models provide the highest accuracies with a best performing value of 88%.

VI. CONCLUSIONS AND FUTURE WORK

In this study we presented an empirical evaluation of different Convolutional Neural Network (CNN) architectures concerning their performance in different tasks in the domain of fashion image classification. The experiments indicated that dispite the large amount and high quality of provided fashion images, pre-trained and fine-tuned models outperform those which were trained on the given collections alone. Future work will concentrate on analyzing models on a scale of two million images.

REFERENCES

[1] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan, "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3330–3337.

[2] W. Di, C. Wah, A. Bhardwaj, R. Piramuthu, and N. Sundaresan, "Style finder: Fine-grained clothing style detection and retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 8–13.

[3] A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie, "Learning visual clothing style with heterogeneous dyadic co-occurrences," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4642–4650.

[4] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[5] Y. LeCun, J. S. Denker, S. A. Solla, R. E. Howard, and L. D. Jackel, "Optimal brain damage." in *NIPs*, vol. 2, 1989, pp. 598–605.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[9] P. Jain, B. Kulis, J. V. Davis, and I. S. Dhillon, "Metric and kernel learning using a linear transformation," *Journal of Machine Learning Research*, vol. 13, no. Mar, pp. 519–547, 2012.

[10] K.-H. Liu, T.-Y. Chen, and C.-S. Chen, "Mvc: A dataset for view-invariant clothing retrieval and attribute prediction," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '16. New York, NY, USA: ACM, 2016, pp. 313–316. [Online]. Available: http://doi.acm.org/10.1145/2911996.2912058

[11] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, 2015.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

Session 2: Research Design and Digital Healthcare

Gendergerechtes Forschungsdesign für Digitale Medien

Dorothea Erharter
ZIMD
Zentrum für Interaktion, Medien & soziale Diversität
Wien
d.e@zimd.at

***Abstract*—The use of digital media is one of the core themes of gender-relevant technological research. The development of innovative media is always research at the interface with humans. However, although the skills of the target groups may not be more different in media usage, there are often blind spots in media development, and the requirements of the target groups are not consistently well met. Based on best practice projects, the authors explain how the inclusion of gender in research design can close such gaps and make better use of potentials.**

***Zusammenfassung*—Die Nutzung digitaler Medien ist eines der Kernthemen genderrelevanter technologischer Forschung. Entwicklung innovativer Medien ist immer auch Forschung an der Schnittstelle zum Menschen. Doch obwohl die Kompetenzen der Zielgruppen in der Mediennutzung unterschiedlicher nicht sein könnten, gibt es in der Medienentwicklung häufig blinde Flecken, und den Anforderungen der Zielgruppen wird nicht überall gleichbleibend gut entsprochen. Anhand von Best Practise Projekten erläutert die Autorin, wie durch die Berücksichtigung von Gender im Forschungsdesign derartige Gaps geschlossen und Potenziale besser genutzt werden können.**

I. Einleitung

Die Implementierung von Gender in Organisationen gilt als Gradmesser für Innovationsfähigkeit [1]. Vielfach ist aus diesem Grund die Berücksichtigung von Gender in technologischen Forschungsprojekten bereits eine Anforderung und ein wichtiges Kriterium der Forschungsförderung, beispielsweise in allen Basisprogrammen der Österreichischen Forschungsförderungsgesellschaft.

Als gender-relevant gilt jegliche Forschung an der Schnittstelle zum Menschen [2]. Mediennutzung ist damit das Kerngebiet genderrelevanter Projekte im IKT-Sektor und in allen technologischen Bereichen. Während im biomedizinischen Sektor die Physis des Menschen im Zentrum steht und damit biologische Unterschiede eine wichtige Rolle spielen können, sind diese in der Mediennutzung vernachlässigbar. Ein auf *geschlechtsspezifische* Unterschiede fokussierendes Forschungsdesign birgt daher die Gefahr in sich, Geschlechterstereotype zu verstärken. Um das zu vermeiden ist der Fokus auf Gender, also das soziokulturelle Geschlecht zu legen. Mit diesem Artikel werden auf Basis des Forschungsstands und der Erfahrungen der Autorin in informationstechnologischen Forschungsprojekten die wichtigsten Elemente gendergerechten Forschungsdesigns für digitale Medien zusammengefasst und mit Beispielen belegt.

II. Thematische Eingrenzung

In welchen Projekten sollen Gender-Aspekte berücksichtigt werden, und was kann es dort bringen? Diese Frage ist zunächst für alle Projekte zu stellen.

A. Gender-Relevanz

Der Begriff der Gender-Relevanz wurde 2006 von Bührer & Schraudner eingeführt. Als genderrelevant gelten Forschungen und Produktentwicklungen prinzipiell an der Schnittstelle zum Menschen, wobei hier sowohl körperliche als auch soziokulturelle und Nutzungszusammenhänge berücksichtigt werden sollen [2]. Für digitale Medien bedeutet das, dass *alle Projekte gender-relevant sind, die Nutzungsschnittstellen* haben. *Nicht* gender-relevant wären zum Beispiel Untersuchungen, mit welchen Protokollen Daten einfacher oder schneller übertragen werden können, oder welche Materialien sich als Leitersysteme besser eignen.

B. Innovationspotential

Dass die Berücksichtigung von Genderaspekten ein technologisches Innovationspotenzial mit sich bringt, ist mittlerweile gut belegt. Gemäß Bührer und Schraudner [2] können dadurch neue Zielgruppen und Nutzungszusammenhänge für technologische Produkte erschlossen werden. Laut Schiebinger können Gender-Aspekte Impulse für die Entwicklung neuer Produkte und Dienstleistungen liefern, die den Anforderungen komplexer NutzerInnengruppen gerecht werden und damit das menschliche Wohlbefinden einschließlich der Gleichstellung der Geschlechter fördern. Damit werde insgesamt die globale Wettbewerbsfähigkeit und Nachhaltigkeit gefördert [3].

Die Berücksichtigung von Gender-Aspekten in Forschungsprojekten bringt nach Meinung der Autorin vor allen Dingen eines: Die Öffnung der Projekte für unterschiedliche Sichtweisen. Denn ein gelungen angewendeter Gender-Fokus öffnet den Blick auch für weitere Diversity-Dimensionen und insgesamt für die Lebensrealitäten unterschiedlicher Menschen, ohne dabei die Kategorie „Gender“ aus den Augen zu verlieren oder zu verwässern.

Damit schleicht sich durch den Gender-Fokus ein sehr menschzentrierter Ansatz in die Forschung ein, der zugleich auch Fremd- und Selbstzuschreibungen im Blick hat, und die Konsequenzen, die sich daraus für das alltägliche Leben ergeben. Vom reinen Human Centered Design unterscheidet sich dieser Ansatz dadurch, dass die sehr wirkmächtige Kategorie „Gender“ nie aus den Augen verloren wird.

Dass allein dadurch Innovationen gefördert werden, zeigen mittlerweile zahlreiche Forschungsprojekte. Zwei Beispiele seien hier genannt: Im Projekt Con Bioenergy wurde nach Befragungen, Beobachtungen in Haushalten und Fehlersimulationen erkannt, dass es häufig Frauen sind, die im Fehlerfall an Heizthermen agieren. Es wurde ein Support-Button an den Geräten eingeführt und Smartphones, Internet und Tablet-PC werden seither zur Störungsbehebung verwendet. [4]. Es muss nicht extra betont werden, dass diese Funktionen allen NutzerInnen zu Gute kommen.

Im Projekt FEMroute sollte ein Fußgänger-Navigationssystem für Frauen geschaffen werden. Aus dem zunächst rein geschlechtsspezifischen Ansatz wurde ein Projekt, das das Innovationspotenzial durch Genderfokus sehr gut verdeutlicht. Ca. 20 gut ausgewählte Testpersonen (siehe I) wurden entlang einer Testroute bei jeder Kreuzung gefragt, wo sie weitergehen würden und warum. Aus den vielfältigen Antworten entwickelte das Team Kategorien für Routen. Neben den üblichen Wahlmöglichkeiten „schnell“ und „attraktiv“ konnten weitere wesentliche Routenarten identifiziert werden: „sicher“ und „komfortabel“ [5].

III. Forschungsstand

Im Gegensatz zu einem solch umfassenden Zugang wird – auch in Forschungsprojekten – häufig implizit ein engerer Zugang gewählt, in dem die eigenen Bedürfnisse und Anforderungen auf andere projiziert werden.

Für diesen Sachverhalt wurde der Begriff I-Methodology von Corinna Bath [6] im deutschen Sprachraum eingeführt: Technologische Entwicklungen in Europa werden von relativ homogenen Teams aus Männern mittleren Alters dominiert, was dazu führt, dass vor allem die Bedürfnisse und Anforderungen dieser Gruppe berücksichtigt werden und andere KundInnengruppen vernachlässigt werden [7]. Dies hat ernste Konsequenzen:

„It decreases the innovative power and inventiveness because of missing opponent, ambiguous or even conflicting viewpoints. It increases the pitfalls of „I-methodologies“ which means that the producers‘ assumptions become more or less consciously the leading benchmarks for technological developments instead of real users‘ needs and demands.“ [8]

Mittlerweile gibt es zahlreiche Beispiele für Forschungs- und Medienprojekte, in denen es gelungen ist, durch einen Gender-Fokus den Blick zu erweitern und damit auch innovativere Ergebnisse zu erzielen.

Das Projekt Discover Gender, dessen Federführung beim Fraunhofer Institut für Innovationsforschung lag, war der erste Beitrag, Gender in Forschungsprojekten systematisiert aufzubereiten. Es wurde ein Leitfaden zur Berücksichtigung von Gender-Aspekten in Forschungs- und Entwicklungsvorhaben entwickelt [9]. Der Leitfaden wurde anhand einer Reihe von Fallbeispielen aus sehr unterschiedlichen technologischen Richtungen überprüft und für NaturwissenschaftlerInnen und TechnikerInnen verständlich dargestellt [2]. Abweichend vom aktuellen Stand der Gender-Forschung wurde in den Beispielen des Leitfadens das Gender-Konzept allerdings auf eine „*strikt binäre Logik verengt*“ [6]. Bei einer geringeren Überlebenschance von Frauen im Fall eines Herzinfarkts ist diese biologisch-dualistische Sichtweise sinnvoll. Es soll gemäß Leitfaden aber auch nach unterschiedlichen Nutzungszusammenhängen von Frauen und Männern oder deren unterschiedlichen Anforderungen an die Gestaltung gefragt werden, ohne dabei auf Lebensrealitäten zu fokussieren. Damit läuft der Leitfaden Gefahr „*Geschlechterstereotype zu verstärken und die Vielfalt der sozialen Welt nur ungenügend zu adressieren.*“ [10]. Die Gender-Studies setzen im Gegensatz dazu eine intersektionale Überlagerung verschiedener Faktoren voraus und fordern beispielsweise eine Berücksichtigung der Lebensrealitäten entlang physischer und soziokultureller Unterscheidungsmerkmale ein.

Mittlerweile ist das von Londa Schiebinger geleitete, an der University of Stanford initiierte, internationale Projekt Gendered Innovations ist *das* Leuchtturmprojekt zum Thema Gender in Forschungsvorhaben. Mehr als 60 ExpertInnen aus ganz Europa, den Vereinigten Staaten und Kanada erarbeiteten von 2009 bis 2012 Materialien zur Integration von Gender-Aspekten in verschiedene naturwissenschaftliche und technische Disziplinen.

Auf der Plattform http://genderedinnovations.stanford.edu werden die in diesem Projekt erarbeiteten Methoden und inhaltlichen Ergebnisse präsentiert und laufend ergänzt. Sie bauen in grundlegenden Konzepten teilweise auf den Ergebnissen von Discover Gender auf und stehen im Web zur freien Nutzung zur Verfügung [3]. Inzwischen wurde fast die gesamte Plattform von der TU Wien unter der Leitung von B. Ratzer ins Deutsche übertragen und steht unter www.geschlecht-und-innovation.at zur Verfügung.

Die große Qualität der beiden Projekte Gendered Innovations und Discover Gender liegt in einer Sammlung von Leitfragen zur Reflexion der Forschungskultur, von Standards und Prämissen der jeweiligen Disziplin, zum Forschungsdesign und zur sprachlichen und visuellen Repräsentation.

Ein Vorgehensmodell, mit dem Gender- und Diversity-Forschung für die Informatik nutzbar gemacht werden kann, stellte die Hochschule Bremen mit GERD (Gender Extended Research and Development) vor. Sie unterscheiden zwischen Kernprozessen, die an ein Wasserfallmodell angelehnt sind, und Reflexionsaspekten, die als Kontext in diese Kernprozesse einfließen. Diese Reflexionsaspekte sind relativ allgemein formuliert und erfordern nach Angaben der ProjektmitarbeiterInnen für die Anwendung zunächst Know-How-Transfer [10].

IV. Zuschreibungen

Zuschreibungen begleiten – und erleichtern – das menschliche Leben, sie sind „normal“. Es geht also nicht darum, Zuschreibungen zu verhindern oder auszulöschen, sondern darum, sie zu reflektieren und bewusst zu machen. Denn unbewusste Zuschreibungen verstellen den Blick auf die tatsächlichen Bedürfnisse und Anforderungen. Neben Zuschreibungen, die auf I-Methodology [7] beruhen, sind Stereotypen und Vorurteile sowie Selbstzuschreibungen für Forschungsprozesse relevant.

A. Stereotype und Vorurteile

Zuschreibungen über die Bedürfnisse und Anforderungen anderer an Produkte entstehen zum einen auf Basis dieser Projektionen, zum anderen durch stereotype Vorstellungen von Gruppen, also so genannte Stereotypen undVorurteile.

„*Stereotypen dienen dazu, einen Gegenstand, eine Person oder eine Gruppe zu charakterisieren. Ein Vorurteil ist ein Urteil, das ohne vorherige Erfahrung über etwas gefällt wurde. Beide erfüllen für die Menschen die Funktion, Unsicherheit und Bedrohung psychisch abzuwehren. Sie dienen dazu, die Welt überschaubar zu machen, Komplexität zu reduzieren. Sie schaffen Sicherheit für das eigene Handeln*." [11]

Das Wort „Vorurteil" ist üblicherweise negativ besetzt. Um Vorurteile besprechbar zu machen, muss meist erst diese negative Konnotation aufgelöst werden. Vorurteile haben ursprünglich eine wichtige lebenspraktische Funktion und funktionieren oft ohne unser aktives Zutun bzw. sind auch kognitiv nur begrenzt zugänglich: Sie helfen, möglicherweise gefährliche Situationen rasch einzuschätzen und Entscheidungen zu treffen. Daher wäre es ein Defizit, wenn jemand keine Vorurteile hätte. Allerdings: Die meisten Situationen, in denen Menschen sich heute befinden, enthalten keine potenzielle Gefahr, – und dann sind Vorurteile ein Hindernis, weil sie eine differenzierte Betrachtung unmöglich machen. Es ist also im Zuge der Entwicklung von Medien-Produkten von zentraler Bedeutung, dass das Team sich über die eigenen Vorannahmen bewusst wird.

Stereotypen sind Charaktarisierungen von Menschen aufgrund eines *Merkmals*– z.B. Männer/Frauen, ältere Menschen, Jugendliche, Arbeitende, Arbeitslose, Menschen mit fremdem/nationalem Hintergrund, StädterInnen, Menschen vom Land. Mit diesem Merkmal werden weitere Eigenschaften verknüpft, die diesen Personen dann zugeschrieben, bzw. auf die sie reduziert werden. Haben Menschen in der Realität andere Eigenschaften als dem Stereotyp entspricht, werden diese häufig abgewertet, beispielsweise als „unweiblich" oder „unmännlich". Damit wird ein gesellschaftlicher Druck aufgebaut, diesen Stereotypen zu entsprechen.

Stereotypen und Vorurteile stimmen meistens nur sehr bedingt. Sie schränken die Kommunikation auf das ein, was erwartet wird, und werden damit zur selbsterfüllenden Prophezeiung: Man nimmt nur das wahr, was man schon gewusst hat.

In der Entwicklung von digitalen Medien schränkt man durch Stereotypen und Vorurteile häufig die Zielgruppen ein. Jemand könnte zum Beispiel davon ausgehen, dass ältere Menschen keine Computer nutzen und daher als Zielgruppe für den Webshop ausgeschlossen werden können. Dann wäre es auch nicht nötig, den Webshop so zu gestalten, dass er von älteren Menschen verstanden und flüssig genutzt werden kann. Wird der Webshop dann gestaltet, ohne ältere Menschen einzubeziehen, wird er möglicherweise für diese Gruppe wirklich schwerer zu verwenden sein, weil die Schrift vielleicht zu klein ist, oder die Buttons schwer verständlich. Damit werden ältere Menschen dann de facto als Zielgruppe ausgeschlossen und die Zuschreibung bestätigt sich. Bedenkt man, dass die Altersgruppe 50 Plus in Österreich 44 % der Kaufkraft besitzt, kann man den enormen Schaden ermessen, den solche Zuschreibungen auslösen. Und umgekehrt lässt sich auch das Potenzial erkennen, das darin steckt, die eigenen Stereotypen - die wie gesagt jede/r hat - zu reflektieren und damit in der Lage zu sein, Zielgruppen adäquat anzusprechen.

Im Projekt G-U-T haben wir Leitfragen entwickelt, anhand derer eine produktspezifische Selbstreflexion durchgeführt werden kann. Dabei geht es neben der Reflexion von Stereotypen auch um die den Projekten zugrundeliegende Zielsetzung, die Zielgruppen, deren Bedürfnisse und Interessen, physiologische Unterschiede und Nutzungsszenarien etc. [12].

B. Selbstzuschreibungen und Technikhaltungen

Es gibt aber auch Selbstzuschreibungen. Diese sind ebenfalls durch Rollenerwartungen geprägt und beeinflussen das Selbstbild und damit insbesondere auch die Haltungen gegenüber Technik und digitalen Medien. Eine solche Selbstzuschreibung könnte sein: „Das kann ich nicht, dafür bin ich schon zu alt" oder „Alles Technische macht bei uns der Karli."

Selbstzuschreibungen und Technikhaltungen beeinflussen sich gegenseitig, so dass „männlich" konnotierte technische Artefakte (Bohrmaschine) eher als „Technik" eingestuft werden als weiblich konnotierte technische Artefakte (Mixer).

V. Gendergerechtes Forschungsdesign

Die Autorin hat in den letzten Jahren im ZIMD einige Forschungsprojekte im Bereich digitaler Medien mit Gender-Fokus im durchgeführt, die als Basis für diesen Artikel dienen. In all diesen Projekten befindet sich gendergerechtes Forschungsdesign immer im Spannungsfeld von geschlechtsspezifischen Unterschieden und der Kritik daran. Dies hat damit zu tun, dass es beim Versuch Gender zu berücksichtigen, ja vordergründig um Unterschiede zwischen Männern und Frauen geht. Und da Männlichkeit und Weiblichkeit auf den ersten Blick biologisch bedingt scheinen, wird zunächst von Forschungsteams fast immer auf biologische Unterschiede fokussiert.

Im Projekt G-U-T (http://g-u-t.zimd.at) haben wir durch vergleichende Analyse überprüft, welche Gender-Aspekte und Diversity-Dimensionen in der Praxis für das Design und Development von Apps und Websites relevant sind, und eine Guideline zu deren Berücksichtigung entwickelt [13]. Im Projekt MOBISENIORA (www.mobiseniora.at) wurde die Nutzung von Smartphones und Tablets durch Seniorinnen und Senioren unter Berücksichtigung von Gender-Aspekten untersucht und Leitfäden für App-Entwicklung, Bildungsangebote sowie Verkaufsberatung und Support durch Telekom-Anbieter entwickelt [14]. Im Projekt GEMPLAY (www.gemplay.at) wurden gendergerechte Spielkonzepte für Videogames zur Bewegungsförderung entwickelt.

In GENSISYS (http://mc.fhstp.ac.at/projects/gensisys) wurde ein Methodenset zur Evaluation von Gender und Diversity Dimensionen für Ergonomie und Usability an Arbeitsplätzen entwickelt. Im Folgenden werden die methodischen Erkenntnisse zu gendergerechtem Forschungsdesign aus diesen Forschungsprojekten zusammengefasst und mit Beispielen belegt.

A. Besonderheiten von Medien-Projekten

Nur sehr wenige biologische Unterschiede wirken sich auf die Nutzung digitaler Medien aus. Diese sind biologisch meist nur mit-beeinflusst und häufig stärker durch Diversity-Faktoren wie das Alter geprägt. Diversity-Faktoren können in Medienprojekten eine große innovationstreibende Kraft entwickeln [15].

Im Gegensatz zu biomedizinischer Forschung spielen biologische Unterschiede in den Medienprojekten und in den Bereichen, in denen es um Nutzungsschnittstellen geht, fast ausschließlich hinsichtlich Beeinträchtigungen und Behinderungen eine Rolle, wie sie durch Accessibility (Inclusive Design, Design for All) adressiert werden.

Im Projekt G-U-T konnte das Team der Autorin festestellen, dass biologische Unterschiede zwischen Frauen und Männern für die Gestaltung von Websites und Apps keine nennenswerte Rolle spielen. Im Projekt MOBISENIORA wurde darüber hinaus bestätigt, dass selbst die Unterscheidung zwischen Jung und Alt in biologischer Hinsicht auf einige wenige mögliche Beeinträchtigungen hinausläuft, die – zumindest in diesem Projekt – eine weitaus geringere Rolle spielten als vom Team zuvor angenommen [16].

In den genannten Projekten hat sich gezeigt, dass mit wenigen Ausnahmen, in denen ergonomische Aspekte eine Rolle spielen, in der Medien-Forschung und -Entwicklung über den Bereich der Accessibility hinausgehende biologische Unterschiede vernachlässigt werden können. Der Umfang der Fragestellungen in Materialien zur Berücksichtigung von Gender- und Diversity-Aspekten, wie den unter II.C genannten, konnte für diesen Bereich daher deutlich verringert werden.

B. Doing Gender

In der Geschlechterforschung gibt es den Begriff des Doing Gender [17]. Damit ist gemeint, dass das soziale Geschlecht im Wesentlichen erst durch das Tun zustande kommt. Verhalte ich mich als Frau, werde ich als Frau wahrgenommen, verhalte ich mich als Mann, werde ich als Mann wahrgenommen. Dies inkludiert Kleidung, Aufmachung, aber auch Körperhaltung, sowie Handlungen, Arbeitsteilung etc. Die Lebensrealitäten der Menschen sind mehr oder weniger stark durch Doing Gender geprägt, also durch den Vollzug dessen, was das jeweilige Umfeld von Männern bzw. Frauen erwartet. Die Lebensrealitäten können sich in vielfältiger Weise unterscheiden, hier spielen auch andere Diversity-Faktoren eine Rolle. Wichtig ist, sich bewusst zu machen, dass Unterschiede zwischen Männern und Frauen viel mehr als durch biologische Unterschiede durch Doing Gender zustande kommen [17].

Beispielsweise wirkt es sich massiv auf Nutzungsszenarien aus, ob jemand Betreuungspflichten hat, da diese Personen im Laufe eines Tages deutlich mehr Wege zurück legen, stärker abgelenkt sind, etc. Dass sie dadurch andere Anforderungen, zum Beispiel an Navigationssysteme, haben können, ist *nicht* an das *biologische* Geschlecht gekoppelt, es ist eine Gender-Thematik [18]. In der Medien-Forschung ist daher nicht nur von Männern und Frauen auszugehen, sondern von Menschen mit vielen verschiedenen Merkmalen und in vielen verschiedenen Nutzungskontexten, da sonst die Gefahr der Stereotypisierung besteht. Davon profitieren auch Männer, wie Raewyn Connell gezeigt hat, die von verschiedenen "Männlichkeiten" spricht [19].

C. Berücksichtigung der Lebensrealitäten

Wie gezeigt wurde, betreffen für Medienprojekte relevante Genderaspekte vor allem die unterschiedlichen Lebensrealitäten und Einstellungen von Männern und Frauen. Hier sind insbesondere die folgenden Bereiche relevant:

- Raum-zeitliche Rahmenbedingungen und Wege, die wesentlich von den Lebensumständen und eventuellen Betreuungspflichten geprägt sind [18];
- Werthaltungen und Einstellungen gegenüber Technik, die wesentlich durch Sozialisationsprozesse geprägt sind [16];
- Technikerfahrung und Technikwissen, die wesentlich durch die berufliche Biographie geprägt sind [16].

Um diese Einflussgrößen strukturiert zu erfassen, muss offen gefragt und die Nähe zu den NutzerInnen gesucht werden. **Partizipative Methoden** sind dafür unerlässlich.

Egal, welche Methoden aus dem vielfältigen Vokabular des Participatory Design gewählt werden: *Die Fragestellungen sollten die oben genannten Bereiche abdecken bzw. dafür offen gehalten und in der Auswertung auf diesbezügliche Unterschiede geachtet werden.*

Im Projekt MOBISENIORA wurden beispielsweise mittels semistrukturierter Interviews die Technikhaltungen und das Nutzungsverhalten von SeniorInnen und ihre Erwartungen an Smartphones/Tablets im Kontext ihrer Sozialisationsprozesse betrachtet. Dabei konnten sehr vielfältige Aspekte einfließen. In der Auswertung konnten einige *gender*spezifische, also durch soziokulturelle Faktoren geprägte Unterschiede sichtbar gemacht werden. Im Projekt GENSISYS kamen die aufschlussreichsten und spannendsten Ergebnisse durch qualitative Methoden wie *Kontextanalyse*, *Arbeitsplatzbeschreibung* und *Tagebuch* zustande, auch wenn sich darin kaum genderspezifische Unterschiede spiegelten. Dennoch waren diese Ergebnisse durch die Berücksichtigung von Gender-Faktoren getriggert.

D. Geschlechtsspezifische Unterschiede

Zunächst: Wir unterscheiden geschlechtsspezifische und genderspezifische Unterschiede. Unter geschlechtsspezifisch verstehen wir Unterschiede auf biologischer Ebene (Sex).

Unter genderspezifisch verstehen wir Unterschiede, die durch soziokulturelle Faktoren zustande kommen.

Geschlechtsspezifische Unterschiede wurden jahrtausendelang missbraucht um Frauen von gesellschaftspolitischer Beteiligung und Bildung abzuschneiden. Auch heute gibt es noch Literatur, die auf Basis angeblicher (evolutions)biologischer Unterschiede soziokulturelle Verhaltensweisen zu legitimieren versucht [20]. Eine kritische Auseinandersetzung damit ist daher wesentlich [21].

Tatsächlich sind geschlechtsspezifische Unterschiede meist viel geringer als gemeinhin angenommen. Männer und Frauen liegen in der Ausprägung der meisten Merkmale sehr nahe beisammen. Die Unterschiede *innerhalb* eines Geschlechts sind (beispielsweise bei der räumlichen Wahrnehmung) viel größer als die Unterschiede *zwischen* den Geschlechtern. Dennoch wird die Verteilung der Geschlechterunterschiede häufig so wahrgenommen, als gäbe es riesige Unterschiede zwischen den Geschlechtern [22].

E. Quellenkritik

Ein wichtiger Ausgangspunkt jeglicher Forschungsprojekte ist die vorhandene Literatur. Eine Fülle von Studien scheinen geschlechtsspezifische Unterschiede zwischen Frauen und Männern zu belegen. Auf den ersten Blick erscheinen jedoch häufig Unterschiede als biologisch, die in Wirklichkeit vor allem auf Lernerfahrungen, also auf soziokulturellen Unterschieden beruhen. Solche scheinbar biologischen Unterschiede kommen häufig durch ein unsauberes Forschungsdesign zustande, und bilden dann eher die Zuschreibungen und Vorannahmen der Forschenden ab, als real existierende Unterschiede [20]. Es kann davon ausgegangen werden, dass es sich bei den meisten dokumentierten geschlechtsspezifischen Unterschieden vor allem im Bereich der Evolutionsbiologie entweder um erlernte Unterschiede handelt. Darüber hinaus hat sich gezeigt, dass Studien, die Unterschiede zwischen Frauen und Männern belegen, sehr viel häufiger zitiert werden, als Studien, die belegen, dass es keine Unterschiede gibt [23].

Eine günstige Vorgangsweise um geschlechtsspezifische Unterschiede in der Literatur einer kritischen Betrachtung zu unterziehen, richtet sich zum einen auf die Studien selbst: Wie wurden solche Studien gemacht? Sind sie prinzipiell in sich sauber gemacht? Mit welcher Brille wurde das Forschungsdesign gemacht? Welche Sprache wird verwendet? Welche Vorannahmen spiegeln sich in den Ergebnissen? Zum anderen in der Recherche kritischer Perzeption: Wie werden die Quellen in der Literatur diskutiert? Was wird kritisiert und von wem? Was wird zitiert, und von wem?

Ein wichtiger Aspekt betrifft auch die Wiedergabe von Forschungsergebnissen, die „Forschungs-Stille-Post“: Durch verallgemeinernde und reduzierte Wiedergabe werden Forschungsergebnisse verfälscht dargestellt, wie „Frauen haben Schulterbeschwerden“ anstelle von „51 % der Frauen und 26 % der Männer haben Schulterbeschwerden“ [24].

Im Projekt GENSISYS wurden zahlreiche für Nutzungsschnittstellen relevante Studien gesichtet, die geschlechtsspezifische Unterschiede reflektieren. Die meisten Unterschiede waren entweder recht dünn belegt oder durch soziokulturelle Faktoren zu erklären. Bewährt hat sich in diesem Projekt, aus den in der Literatur gefundenen Unterschieden Bereiche zu identifizieren, in denen es prinzipiell größere Unterschiede geben kann, als bisher vielleicht berücksichtigt, unabhängig davon, welche Dimensionen (Alter, Geschlecht, Kultureller Hintergrund etc.) zu diesen Unterschieden führen.

F. Hypothesenbildung

Fokussiert das Forschungsdesign auf geschlechtsspezifische Unterschiede, wird auf diese ein Vergrößerungsglas gerichtet: Sie erscheinen größer als sie tatsächlich sind. Damit reproduzieren sie Ungleichheiten, die zusammengenommen zu gesellschaftlichen Ungleichwertigkeiten führen können. Für Medien-Forschung relevante Unterschiede können darüber hinaus sehr vielfältige Ursachen haben.

Für die Formulierung von Hypothesen in der anwendungsorientierten technologischen Forschung bedeutet das, den Fokus auf die Inhalte der festgestellten Unterschiede zu richten und die Frage, wodurch diese Unterschiede verursacht sind, erst in zweiter oder dritter Linie zu stellen.

Im Projekt GENSISYS beispielsweise wurden Hypothesen, die in der Einreichphase geschlechtsspezifisch formuliert waren (A), von uns folgendermaßen zu Forschungsfragen umformuliert (B):

A. Räumliche Wahrnehmung: Durch größere Bildschirme kann die räumliche Wahrnehmung zugunsten der Frau im Vergleich zum Mann verbessert werden.

B. Es gibt Studien, die zeigen, dass die räumliche Wahrnehmung von Frauen durch größere Bildschirme mehr verbessert werden kann als von Männern. ***Wie wirken sich größere Bildschirme in der konkreten Situation aus?***

Größere Bildschirme können für verschiedene Personen Vor- und/oder Nachteile haben, aufgrund verschiedener Faktoren, von denen einer das Geschlecht sein kann. Wichtig ist, die Inhalte vom biologischen Geschlecht zu lösen. Die Fragestellung wird dadurch auch offener für weitere Faktoren. Eine entsprechende Hypothese könnte also sein:

Durch größere Bildschirme kann die räumliche Wahrnehmung bei verschiedenen Gruppen (Geschlecht, Alter, kulturelles Umfeld, …) verbessert werden.

In der Auswertung kann dann innerhalb der Personen mit besserer, gleicher und ev. schlechterer räumlicher Wahrnehmung bei größeren Bildschirmen wieder die Kategorie Geschlecht betrachtet werden, wie auch die anderen Kategorien, die betrachtet wurden. Siehe dazu auch H.

G. Formulierung qualitativer Fragestellungen

An diesem Beispiel zeichnet sich darüber hinaus ein weiteres Merkmal gendergerechter Forschung ab: die Formulierung qualitativer Fragestellungen. Mit quantitativen

Fragestellungen kann überwiegend auf vordefinierte Kategorien und Merkmale eingegangen werden, die per se dazu angetan sind, Vorannahmen zu bestätigen. Qualitative Fragestellungen sind zwar aufwändiger auszuwerten, aber ergebnisoffener und daher prinzipiell besser geeignet für gendergerechtes Forschungsdesign.

H. Gendergerecht Clustern

Ein wesentliches Merkmal *quantitativer* Studien sind Cluster, die gebildet werden, um Aussagen über bestimmte Gruppen zu treffen. Angesichts der Gefahr der Verstärkung von Stereotypen empfiehlt es sich, nicht entlang vorgegebener Merkmale wie Geschlecht, Alter etc. zu clustern, sondern ergebnisoffen entlang von Faktoren, die sich aus der Studie selbst ergeben. Damit kann die Fortschreibung stereotyper Zuschreibungen vermieden werden.

Innerhalb dieser Cluster soll dann die Geschlechterverteilung betrachtet werden. Dies ist zentral, da ohne diesen Schritt die Kategorie „Gender“ aus dem Blick geraten würde. Vorhandene Ungleichheiten würden dadurch unsichtbar gemacht. Wolffram & Winker haben beispielsweise in einer Studie über technische StudienanfängerInnen zunächst nach Technikhaltungen geclustert (z.B. „einseitig technikzentrierte Haltung“, „distanzierte Technik- und Computerhaltung“), und erst *danach* die Geschlechterverteilungen innerhalb der fünf Gruppen betrachtet. Mit einer solchen Vorgangsweise fällt es viel leichter, der Falle vorschneller Zuschreibungen zu entgehen [25].

Ist dies nicht möglich, und müssen Cluster auf Basis vorher festgelegter Merkmale gebildet werden, so empfiehlt es sich Merkmale zu wählen, die weniger durch Stereotypen geprägt sind. Die inneren Persönlichkeitsmerkmale: Geschlecht, Alter, soziale Herkunft, Ethnie, geistige/körperliche Fähigkeiten/Einschränkungen, sexuelle Orientierung [26] und zusätzlich auch Religion/Weltanschauung sind häufiger mit Vorannahmen verknüpft und damit sind stereotype Zuschreibungen (die es immer gibt) schwerer erkennbar.

Im Projekt GEMPLAY wurde beispielsweise nach „Spielertypen“ geclustert, *innerhalb* derer dann die Geschlechter- und Altersverteilung betrachtet wurde.

Eine gute Möglichkeit besteht darin, bereits in das Design einer quantitativen Befragung mögliche Auswirkungen von Lebensrealitäten einfließen zu lassen. So wurde im Projekt GEMPLAY beispielsweise nach Zeitverwendung und Freizeitverhalten gefragt. Dies kann allerdings qualitative Befragungen nicht ersetzen.

I. Partizipative Methoden und Auswahl von Testpersonen

Wie dargestellt, sind partizipative Methoden, die NutzerInnen in den Forschungs- und Designprozess einbeziehen, für gendergerechte Forschung unerlässlich.

Zentral ist hier die Auswahl der Testpersonen. Mindestens ebenso wichtig wie eine ausgewogene Zusammensetzung nach biologischem Geschlecht ist die Berücksichtigung der drei oben genannten Merkmale raumzeitliche Rahmenbedingungen, Technikhaltungen sowie Technikaffinität und –erfahrung als Auswahlkriterien für Testpersonen bzw. als Auswertungskriterien (z. B. viel/mittel/wenig Technikerfahrung). Darüber hinaus kommt hier ebenfalls die Diversity zum Zug. Es empfiehlt sich die (insbesondere inneren) Persönlichkeitsmerkmale (vgl. IV.H) auf Relevanz für das Projekt hin zu betrachten.

Im Projekt MOBISENIORA wurden die Testpersonen z. B. entlang einer Matrix ausgewählt, in der neben Geschlecht auch Alter, Region und Nutzungserfahrung berücksichtigt war. Im begleitenden Fragebogen zu den Usability-Tests wurden die Testpersonen dann nach der wöchentlichen Gerätenutzung in Stunden gefragt. In der Auswertung wurde der Median der jeweiligen Gerätenutzung gebildet und daraus dann die Gruppen gebildet (viel/wenig/keine Gerätenutzung), für die nach Korrelationen zu anderen Ergebnissen gesucht wurde.

J. Personas

Personas werden in der Softwareentwicklung genutzt, um ein user-oriented Design zu gewährleisten. Sie sollen den EntwicklerInnen helfen, ihr Produkt durch „die Maske der UserInnen“ zu sehen [27]. Aus gendersensibler Sicht ist dieser „Perspektivenwechsel“ besonders wichtig, da die meisten Teams in der Softwareentwicklung hauptsächlich aus Männern mittleren Alters bestehen. Es besteht die Gefahr, dass z.B. die weiblichen und älteren UserInnen nicht beachtet werden [28].

Personas wären deshalb theoretisch sehr gut geeignet, um ein gender- und diversitysensibles Design zu gewährleisten, da sich die EntwicklerInnen dadurch in die UserInnen hineinversetzen können. Das Problem dabei ist, dass Personas immer eine gewisse Vereinfachung beinhalten müssen und deshalb auch zu Stereotypisierung neigen. Die große Herausforderung ist deshalb, die nötige Vereinfachung der Komplexität und die Beschreibung der individuellen Diversitätsfaktoren im richtigen Maß gegeneinander abzuwägen [29/30].

Personas stellen Entwicklungsteams also vor eine schwierige Aufgabe: Zum einen sollen sie typische NutzerInnen abbilden, müssen also „geclustert“ sein und bei den einzelnen Merkmalen mittlere Werte darstellen. Zum anderen widerspricht eine Clusterung der Zielgruppen ganz prinzipiell der Forderung nach einer Berücksichtigung der Interessen und Lebensrealitäten der einzelnen Zielpersonen, da dabei immer Details verloren gehen müssen. Sie sollen also typisch sein, aber trotzdem nicht stereotyp.

Eine wichtige Voraussetzung für die Entwicklung von Personas ist damit die oben beschriebene Selbstreflexion der eigenen Zuschreibungen und stereotyper Vorstellungen über Personengruppen. Auch Personas sollen die Lebensrealitäten der Zielpersonen möglichst realistisch darstellen. Eine weitere sinnvolle Vorgangsweise besteht darin, die Clusterungen nicht entlang vorher festgelegter Merkmale vorzunehmen, sondern anhand von Merkmalen, die sich aus der User Reseach ergeben, in der Weise wie oben am Beispiel der Studie von Wolffram beschrieben.

VI. Fazit

Die Vorschläge der AutorInnen für ein gendergerechtes Forschungsdesign im für digitale Medien betreffen unterschiedliche Aspekte der quantitativen und qualitativen Forschung:

- die qualitative Formulierung von Forschungsfragen und Hypothesen in einer offenen, nicht geschlechtsspezifischen Weise;
- die Blickrichtung auf raum-zeitliche Rahmenbedingungen und Lebensrealitäten (z.B. Betreuungspflichten); auf Selbstzuschreibungen, und Technikhaltungen;
- die kritische Quellenbetrachtung, insbesondere von geschlechtsspezifischen Ergebnissen außerhalb von biologischer Forschung;
- eine genaue Sprache bei der Wiedergabe geschlechtsspezifischer Forschungsergebnisse;
- für quantitative Studien eine stereotypenresistente Vorgehensweise zu Clusterung und Auswertung;
- für qualitative Studien die Ausschöpfung des Potenzials partizipativer Forschungs- und Designmethoden;
- eine diversitätsbewusste Auswahl von Testpersonen;
- das explizite Beschreiben von nicht-vorhandenen Unterschieden, wenn keine gefunden wurden.

Wichtig ist, diese Vorschläge zu gendergerechtem Forschungsdesign in Medien-Projekten nicht als Einzelmaßnahmen, sondern als Paket zu betrachten und anzuwenden. Natürlich kann schon die Anwendung einer einzelnen Maßnahme wichtig und sinnvoll sein, doch erst im Paket können die Vorschläge ihre volle Innovationskraft entfalten.

References

[1] Rat für Forschung und Technologieentwicklung, FAS.research (Hg.), Excellente Netzwerke (2005), http://www.fas.at/download-document?gid=253, zuletzt abgerufen 2014-02-28. S.37.

[2] Bührer, S., & Schraudner, M. (Eds.) (2006). Wie können Gender-Aspekte in Forschungsvorhaben erkannt und bewertet werden? Karlsruhe: Fraunhofer Verlag.

[3] Schiebinger, L. & Klinge, I. (2013). *Gendered Innovations. How Gender Analysis Contributes to Research.* Report oft he Expert Group „Innovation through Gender“. European Commission.

[4] http://www.aee-now.at/cms/index.php?id=8&L=1&tx_ttnews[pointer]=15&tx_ttnews[pS]=1474236999&tx_ttnews[tt_news]=116&tx_ttnews[backPid]=7&cHash=077feda4d8, zuletzt abgerufen am 28.10.2017

[5] Häusler, E., Steinmann, R., Gartner, G., Schmidt, M. (2010): The FEMroute Project – A Gender-Sensitive Approach to Route Planning Systems for Pedestrians Keynote Lecture, 7th International Symposium on Location Based Services & TeleCartography, Guangzhou, China In: Proceedings of the 7th International Symposium on Location Based Services & TeleCartography", G. Gartner, Y. Li (ed.), 124 - 128

[6] Bath, C. (2007). Discover Gender in Forschung und Technologieentwicklung. *Soziale Technik 4/2007*. www.sts.aau.at/ias/content/download/1656/7774. Zuletzt abgerufen 5.10.2014. S.2.

[7] Joost, Gesche, Bessing, Nina und Buchmüller, Sandra, G – Gender Inspired Technology, (2010). In: Ernst, Waltraud (Hrsg.). (2010). Geschlecht und Innovation. Gender-Mainstreaming im Techno-Wissenschaftsbetrieb. Internationale Frauen- und Genderforschung in Niedersachsen. Teilband 4. Lit-Verlag. Berlin.

[8] Buchmüller, Sandra, Joost, Gesche, Bessing, Nina und Stein, Stephanie, Bridging the gender and generation gap by ICT applying a participatory design process. In: Personal and Ubiquitous Computing 15 (7), (2011), S. 743–758. DOI: 10.1007/s00779-011-0388-y.

[9] SCHRAUDNER, M. "Discover Gender: Das Potenzial von Gender für die Forschung, Gender Kompetenz Zentrum, 2006b." (2006).

[10] Maaß, S., Draude, C. & Wajda, K. (2014). Das GERD-Modell. In. Marsden, N. & Kempf, U. (Hrsg.): *Gender-UseIT - HCI, Web-Usability und User Experience unter Gendergesichtspunkten.* Oldenbourg. S. 127-141. S.68.

[11] Friesenhahn, Günter. (2016). Stereotypen und Vorurteile. Modul „Interkulturelles Lernen“ auf dem Portal der Fachstelle für Internationale Jugendarbeit der Bundesrepublik Deutschland eV IJAB. https://www. dija.de/fileadmin/medien/downloads/Dokumente/Guenter2IKL.pdf, zuletzt abgerufen am 28.10.2017. S. 1.

[12] Erharter, Dorothea, G-U-T Guideline. http://www.zimd.at/gut-guideline-checklist, (2013), zuletzt abgerufen 2017-09-17.

[13] Erharter, D. (2014). Gender- und Diversity-Faktoren in interaktiven Medien. In: Markus Seidl & Grischa Schmiedl (Hrsg.): Forum Medientechnik - Next Generation, New Ideas. Glückstadt.

[14] Amann-Hechenberger, B., Buchegger, B., Erharter, D., Felmer, V., Fitz, B., Jungwirth, B., Kettinger, M., Schwarz, S., Knoll, B., Schwaninger, T. & Xharo, E. (2015). Tablet & Smartphone: Seniorinnen und Senioren in der mobilen digitalen Welt. Forschungsbericht zum Projekt „mobi.senior.A“. Wien.

[15] Erharter, D. & Xharo, E. (2014). Gendability. Gender & Diversity bewirken innovative Produkte. In: Marsden, N. & Kempf, U. (Hrsg.): Gender-UseIT 2014 (#GUI2014). HCI, Web-Usability und UX unter Gendergesichtspunkten. S. 127–141.

[16] Erharter, D., Jungwirth, B., Knoll, B., Schwarz, S., Posch, P. & Xharo, E. (2014). Smartphones, Tablets, App für Seniorinnen und Senioren. In: Kemper, Guido (Hrsg.): Assistenztechnik für betreutes Wohnen. AAL Testregion Westösterreich. Tagungsband zum uDay XII. S. 221-235

[17] Bidwell-Steiner, M. & Krammer, S. (2010). (Un)Doing Gender als gelebtes Unterrichtsprinzip: Sprache - Politik – Performanz. Wien.

[18] Scambor, E. & Zimmer, F. (2012). *Die intersektionelle Stadt.* Bielefeld: transcript.

[19] Connell, R. (1999). Der gemachte Mann: Konstruktion und Krise von Männlichkeiten. Berlin: Springer.

[20] Precht, R. D. (2010). Liebe. Ein unordentliches Gefühl. München.

[21] Hammerl, E. (2009). Was ist Gender? Geschlechtsspezifische Rollenzuweisungen und ihre alltäglichen Auswirkungen. In: Hochleithner, M. (Hg): Gender Medicine. Ringvorlesung an der Medizinischen Universität Innsbruck. Band 2. Wien. S. 9-17.

[22] Neyer, F. J. & Asendorpf, J. B. (2012). Geschlechtsunterschiede. Grafik abgerufen von http://link.springer.com/chapter/10.1007/978-3-642-30264-0_7#page-1 am 26.02.2014.

[23] Fine, Cordelia (2012). Die Geschlechterlüge. Stuttgart: Klett-Cotta Verlag.

[24] Erharter, Dorothea (2015): Gender- und Diversity-Dimensionen in der Entwicklung von IKT-Projekten. In: Barke, Helena, Siegeris, Juliane, Freiheit, Jörn & Krefting, Dagmar (Hrsg). Gender und IT-Projekte: Neue Wege zu digitaler Teilhabe. Leverkusen. Budrich UniPress Ltd. ISBN 978-3863887094.

[25] Wolffram, A. & Winker, G. (2005). Technikhaltungen von Studienanfängerinnen und -anfängern in technischen Studiengängen. Auswertungsbericht der Erstsemesterbefragung an der TUHH im WS 03/04.

[26] Gardenswartz, L. & Rowe, A. (2002). Diverse Teams at Work. Society for Human Resource Management. 2002.

[27] Nielsen, Lene. (2012). Personas - User Focused Design. ISBN-13: 978-1447140832.

[28] Oudshoorn, Nelly, Rommes, Els & Stienstra Marcelle (2004): Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. In: Science Technology & Human Values 29 (1), S. 30-63.

[29] Turner, Phil, & Turner, Susan. (2011). Is stereotyping inevitable when designing with personas? Design Studies, 32(1), 30-44. doi: http://dx.doi.org/10.1016/j.destud.2010.06.002

[30] Marsden, Nicola, Link, Jasmin, & Büllesfeld, Elisabeth. (2014). Personas und stereotype Ge-schlechterrollen. In Nicola Marsden & Ute Kempf (Eds.), Gender-UseIT - HCI, Usability und UX unter Gendergesichtspunkten (pp. 91-104). München: De Gruyter Oldenbourg.

RegionAAL

The Styrian AAL-test-region in Graz, Deutschlandsberg and Leibnitz

Kurt Majcen
Institute DIGITAL
JOANNEUM RESEARCH Forschungsgesellschaft mbH
Graz, Austria
kurt.majcen@joanneum.at

Kerstin Löffler
Albert Schweitzer Institute for Geriatrics and Gerontology
Geriatric Health Care Centres Graz
Graz, Austria
kerstin.loeffler@stadt.graz.at

***Abstract* — As a result of demographic change and increasing life expectancy, challenges in providing care for elderly people is expected. Furthermore, older people seek to remain in their own living environment as long as possible in order to retain a certain quality of life. Due to chronic conditions, multi-morbidity and social isolation our care system has to face challenges in order to accomplish the above mentioned. Active and assisted living (AAL) technologies try to meet these challenges but only a few technologies have been effective so far. This study measures the usability and acceptability of existing AAL-technologies which are bundled in a package for older people in home care or assisted living. Possible impacts of using AAL-technologies on the quality of life and the independence in older age are monitored between 07/2017 and 09/2018 in a two-armed randomized controlled trial (N=240).**

Keywords — active and assisted living, elderly people, independence, usability, acceptability

I. Introduction

Many older people wish to remain in their own environment enjoying the highest possible quality of life for as long as possible. This is also an aim which is echoed in many official bodies and organizations, such as the WHO, OECD and EU. However, the ageing process is typically associated with an increase in, and worsening of, chronic conditions, both physical and mental, including diabetes, high blood pressure and depression. This multi-morbidity among older people is often accompanied by social isolation and associated with real or perceived need for help and assistance. Thus, in order to be able to remain living at home, the need for external support – in terms of personal care or social needs – has been rising.

Active and assisted living (AAL) as well as innovative ICT technologies supporting older people have been developed and tested on a small-scale within research projects. However, as evaluated by the European AAL program (about 13 examples mentioned [11]), so far only a few technologies have been successful. Areas covered by these solutions range from wearable protections systems and expanded rollators to communication facilities and mechanisms to access care information. So why have relatively few of these AAL technologies been successful so far? Could one reason be that they were not researched and developed with the specific needs and requirements of this particular population group in mind?

RegionAAL is an AAL-test region project focused on the urban centres of Graz, Deutschlandsberg and Leibnitz. Its aim is to support older people in being able to reside for longer in their own surroundings. In order to achieve this goal information and communication technologies (ICT) which are be accepted by elderly people shall be used.

Besides research on the acceptance and use of these technologies, information regarding possibilities of these new systems and services shall be widespread in media and the region's population in order to foster the uptake of AAL technologies in the future.

The benefit program aims to make AAL technologies available, to get some kind of evidence on the efficacy and usability of the technologies and to bring information about these developments into our society on different levels. Similar to other AAL-test regions in Austria – there are already six – some elements comprise this project to underpin these major aims of the benefit program. These elements are:

- identifying the technologies and services to be used
- implementing and integrating a prototype of the package to be used in the test households
- preparing a study design which allows evaluation of the chosen technical approach
- implementing the prototype package in each test household
- running the system and services in a larger number of test households over the defined test period
- evaluation of the acceptance of the solutions and the efficacy of the introduced interventions
- disseminating results and information about the test region

II. Related work

The benefit programme (Austria's pendant to the European AAL programme) has included AAL test regions within its calls since 2012 whereas three regions (i.e. Modulaar, West-AAL and ZentrAAL) within this framework, even started before the RegionAAL project.

The aim of Modulaar (2012-2014, [12]) was to equip about 50 households in assisted living facilities in Burgenland with AAL technologies (in the areas safety, comfort, health and social interactions) adapted to the special needs of the households. The evaluation focused on acceptance of the new technologies, influence on quality of life and socio-economic aspects.

West-AAL (2014-2016, [13]) aimed at equipping approx. 70 households in urban and rural areas in Tyrol and Vorarlberg with AAL technologies to prolong independent life of participants in their own homes. About 80 different components / products were offered to involved care organizations and participants could choose individual settings with 4-5 components. The testing was foreseen for a period of about 18 months. Technologies concerned the management of risk factors (falls, emergency cases), chronical diseases (like dementia), mobility assistance, smart home components and attached services (e.g. transport or shopping). Used components were developed in prior projects or have been already available on the market.

ZentrAAL (2015-2017, [14]) aimed at offering an expandable software system to assist independent living. The focus group were younger seniors (aged 60 to 75 years) who were still fit but with some need for assistance, for example because of diabetes type 2 or due to mild cognitive impairment. Around 60 test households in the central area of Salzburg could be equipped. Simultaneously, a same sized control group was recruited for a period of approx. 15 months. The evaluation also included market barriers and market attractiveness of the software systems with the goal of successfully entering the market with the offered solutions.

In several workshops with those three regions, lessons learned were identified and taken into account in the various work packages of RegionAAL. Findings concerned very different aspects:

- Technologies (availability of internet; even widespread commercial products such as tablets, offer lots of technical and not reproducible problems, e.g. WLAN in half of devices was deactivated after reboot)
- Recruiting (difficulties in recruiting members of the control group and how to reach them best)
- Test phase (little use of hotlines in the test phase due to inhibitions of participants; still few social interactions but not due to technical difficulties, but rather due to non-existing contacts in many cases; Other stakeholders like general practitioners did not make use of possibilities as these kinds of services are not billable; performing surveys very often requires assistance through caregivers)

Besides the common aspects among the various AAL test regions in Austria, there are differences between all those projects according to their definition, their evaluation approaches etc.

III. METHOD

A. *Study design*

Instead of having a fully defined set of technologies as part of the project proposal, the first phase of the RegionAAL project was dedicated to an intense analysis on the evidence of AAL technologies (see [1] for requirements on analysis methods) and to the definition of a potential user population for the study phase of the project. Thus, the project intended to use only those interventions which were already seen as useful by elderly people and caregivers.

Based on the findings of the analysis, a potential test household was planned. Furthermore, a scientific enquiry was initiated in order to find already available technologies (e.g. in [2][3][4][5] but also in online electronic shops) which can be integrated and are able to fulfil the identified needs. Important aspects of the technology research were: technologies have to be on a product state rather than prototypes from projects; technologies shall be implementable in new buildings as well as in existing homes; the usage and maintenance shall be easy; and an ongoing functionality is guaranteed after the project end. Parallel to the technology research, workshops with caregivers and potential users were held. The aims of the workshops were to figure out which technologies may be interesting for potential users, what the usefulness would be and what are potential drawbacks when introducing the technologies. All these points were discussed with caregivers and potential users to receive aspects from different points of view.

After finalizing the actual set of interventions and technologies, two activities were started: defining / preparing the evaluation processes and implementing / adapting products and software for the use in test households. The implementation and adaptions happened in close cooperation with the care organizations represented in the RegionAAL project group. Developments and adaptions were discussed with caregivers in regular meetings to get continuous feedback about details for the technical work.

The test of technologies in households took place in two phases. The first phase was performed with eight friendly users. They tested the mobile technologies (tablet and smart watch) for two weeks. On the one hand, this created information on the instruction process and on the other hand, feedback on the use of the technologies could be collected. After the friendly user tests had taken place, the large field trial has started as the second phase. The households were recruited by the care organizations and randomized to achieve an intervention and a control group with similar size.

In addition, a baseline, a 6-months intermediate and a 12-months follow up questioning are also part of the study. The participants are interviewed on their health status, quality of life, autonomy and their experience according to electronic devices. The participants of the intervention group are asked to rate the usability of the AAL-technologies on a scale from 1 to 5 in the 6-months intermediate and 12-months follow up and are also questioned on their satisfaction and usability of the devices used in the study. The caregivers are interviewed on the care situation as well as on their satisfaction and usability with the technologies used. [7] The control group as well as their informal caregivers also take part in the baseline and the follow up questioning, however, they are not tested on any AAL-technologies during the one-year intervention. The results from the control group are then used to point out the differences of the use of AAL-technologies on their health status, quality of life and autonomy and therefore state the relevance of such technologies for older people.

B. Subjects

The evaluation was defined as a randomized controlled trial (RCT) with the aim to include 240 participants for a period of one year. An appropriate protocol with accompanying documents (questionnaires, participants' information, informed consent etc.) were prepared and proposed to the ethics committee.

The following inclusion and exclusion criteria have to be met: The participants are over 60 years old, have a care level between 0 and 4 (max. of 160 hours of care needed per month) and do not have any kind of cognitive impairment or procuration. [6] [7] They are living at home, in an assisted living facility or visiting a day care centre in the area of Graz, Leibnitz or Deutschlandsberg and do not use any kind of senior-adjusted tablets or smart watches. Written informed consent is obtained from all participants. Furthermore, the informal caregivers or mobile care services are willing to participate in the study as well.

C. Recruiting

The participants were recruited by the three care organizations of the RegionAAL project group from March 2017 until September 2017. Already existing clients of their assisted living facilities, day care centres and mobile services were asked to participate. Furthermore, announcements in local newspapers were made. Through senior clubs and presentations on exhibitions further participants were recruited. The older people stated their interest via telephone, were screened according to the inclusion and exclusion criteria and got first information about the one-year-intervention. Afterwards, the more detailed information materials and the informed consent were sent to them. If required personal information was given by project staff. [7]

In the meantime, training programs on the applications and devices used took place for projects' staff.

D. Implementation

After signing the informed consent, appointments with participants from the intervention group were made in order to install the technical devices and to get the intervention started. The installation was realized by the technical partner in the project. At the same time, instructions on the applications and on the different devices were conducted by the project staff.

IV. Preliminary Results

The aim of the study should be achieved through the implementation of those ICT elements (fig. 1) that are likely to be accepted and used by end users. Evidence of ICT that is potentially useful for and acceptable to this group was identified. Existing technologies were adapted, extended and integrated - with parallels to Smart City. The one year scientific evaluation includes 219 test households (intervention group = 110; control group = 109) and assesses the effectiveness of ICT in meeting stated aims. [7]

Preliminary results from the literature review indicate that safety, support in Instrumental Activities of Daily Living (IADL) and Activities of Daily Living (ADL) and social interaction are important factors regarding the quality of life. The results of the workshops with caregivers show similarities: Technologies should assist in safety, fall prevention, ADLs, prevention and training. [8]

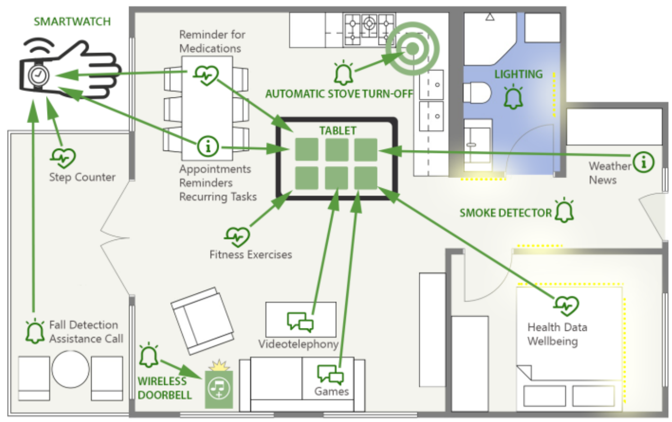


Fig. 1. Interventions and devices in use

A. Interventions

Thus, a set of core functionalities was fixed preparing a prototype of a technical setting which is used in the test households of the intervention group. Due to these intermediate findings, the AAL-technologies which are going to be tested in the one-year-study, are divided into the following four fields of application: health and wellbeing-related technologies, safety, information and communication, and entertainment. The main functions available contain: [9] [10]

- Health and wellbeing
    - vital data at a glance
    - various reminder (e.g. medication)
    - training videos and fitness trackers to stay fit
- safety
    - electric stove shutdown
    - wireless doorbell
    - fall detection
    - emergency call
    - individual light assistance
- Information
    - calendar and daily agenda

  - events and meetings
- communication and entertainment
  - video phone calls
  - photos from relatives
  - games and entertainment

The main functions (fig. 1) are implemented on a number of technologies which are already available on the market, such as a Samsung Tab A 2016 tablet, a Finow X3 Smartwatch, asina platform for mobile devices, diverse apps from the PlayStore (games, video player/conferencing, for configuring the tablet etc.), an electric stove shutdown from Hager or Bedlight as part of the light assistance.

The intervention group, which is testing the technologies mentioned above, is accompanied by monthly meetings in small groups discussing and answering questions according the handling of the applications and technologies. Furthermore, a hotline was installed that is available from 8 a.m. to 3 p.m. on weekdays. Besides that, a 24/7 hotline exists for urgent matters according the stove shutdown and lightning.

V. Discussion

Lessons learned that were made so far concern the recruiting process, technical installation and instruction but also practical matters like purchasing and rollout of a larger number of equipment. Especially the recruitment claimed all forces within the project team and extended the project time by several months. Local connection and various media announcements in approved newspapers were needed. However, the sample size of 240 participants was nearly met. Furthermore, the participants have very diverse knowledge on technological devices. Therefore, the planned instruction process and instruction material did not fit everyone. Individual adaptations had to be made.

The main success factors until now are

- that the used technologies can be installed in every existing household,
- that there are regular meetings to answer questions and
- a telephone hotline,
- as well as individual support for participants.

The final results will be available in January 2019 after the one year trial took place and all data are analysed.

Acknowledgment

The project RegionAAL is partially funded under the program benefit by the Austrian Research Promotion Agency (FFG) and the Austrian Ministry for Transport, Innovation and Technology (bmvit). The authors of this paper want to thank all other partners in the RegionAAL consortium for their cooperation in the described study.

References

[1] PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

[2] The Central Managment Unit (CMU), Ambient Assisted Living Joint Programme – Catalogue of Projects 2014

[3] I-stay@home, ICT-Lösungen für eine alternde Gesellschaft – Forschung, Tests und Erfahrungen, August 2015.

[4] Demenz Support Stuttgart – Zentrum für Informationstransfer, dess@work_4.1: Technische Unterstützung bei Demenz – Fokus eigene Häuslichkeit: Produktkatalog, März 2015

[5] Webseite „Wegweiser Alter und Technik“ www.wegweiseralterundtechnik.de

[6] Land Steiermark, Pflegegeld, 2016, http://www.gesundheit.steiermark.at/cms/beitrag/11645011/72574865

[7] RegionAAL consortium, Study protocol, 2017, unpublished

[8] JOANNEUM RESEARCH Forschungsgesellschaft, PotenziAAL-Pflege - Abschätzung des Marktpotenzials von Technologien aus dem Bereich „AmbientAssisted Living“, 2015, Studienabschlussbericht

[9] RegionAAL consortium, Bedarfs- und Evidenzanalyse, 2016, unpublished

[10] RegionAAL consortium, Spezifikation des Prototyps, 2016, unpublished

[11] AAL Europe programme, Success stories – AAL innovations creating real impact, 2nd edition, 2016, http://www.aal-europe.eu/wp-content/uploads/2016/10/AALSuccess2016_HIRES.pdf

[12] Project Modulaar, www.modulaar.at

[13] Project West-AAL, www.west-aal.at

[14] Project ZentrAAL, www.zentraal.at

LifeStream: Design and Prototypical Implementation of a Monitoring System for Dispatch Life Support

Florian Grassinger, Jakob Doppler, Markus Wagner and Wolfgang Aigner
Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences, Austria
first.lastname@fhstp.ac.at

***Abstract*—Most laypersons who reanimate for the first time do it inappropriately. Until now the only way to review the ongoing reanimation was verbal feedback by the dispatcher on the phone, who has only limited resources in order to review the reanimation process. To overcome this issue, we designed and implmemented LifeStream, a system using current smartphone technologies in order to measure reanimation parameters: chest compression rate (CCR) and chest compression depth (CCD). The system is based on a server, web client and mobile application, which gathers, processes and transfers the data. The development of algorithms for CCR and CCD detection as well as the evaluation of the system functionality is part of this paper. We conducted a 2-day user test, where we compared the guided standard reanimation process to the application supported process. The results of the tests showed that it is possible to develop an application, which runs for at least ten minutes (crucial time till ambulance arrives) and enhances the whole reanimation cycle for laypersons and dispatchers [1].**

I. Introduction

Our fast aging population results in an increase of out-of-hospital cardiac arrest situations. Often dispatch life support and Cardiopulmonary Resuscitation (CPR) interventions are performed by untrained laypersons and bystanders rather than medical professionals [2]. The fear of making bad decisions often restrains people from helping and saving life's or bridge the critical minutes until the ambulance arrives [3]. Time critical medical emergency situations are situations where a proper execution of all steps in the chain of survival is crucial and therefore every second counts [4]. A CPR often requires immediate reaction and even if the chest compressions are not totally appropriate, the attempt is crucial to save a person's life. Over the past years, cardiopulmonary resuscitation has continuously improved and was further investigated by Roessler et al. [5].

Today's smartphones are equipped with multimodal sensors to measure important context and even vital parameters that can be used to assess the situation during a reanimation. For the development of a functional prototype, which assists laypersons or unexperienced people by performing chest compressions, we used the accelerometer sensor of the smartphone. Additionally, we utilized the network connectivity, as well as maintain an ongoing phone call and perform background tasks such as transmitting real-time data to develop an effective algorithm for chest compression rate and depth detection.

The main goal of this research is to present a prototypical implementation of a system which uses algorithms for chest compression rate (CCR) and chest compression depth (CCD) detection and compares them to existing standards and therefore enhance the overall reanimation process for the dispatcher and the layperson.

The major contribution of this paper is a functional prototype that was tested during user tests and evaluated as well as a straightforward experimental implementation.

II. Related work

There are a number of tools and research work, that deal with the quality of CPR and its enhancement. However, none of them transmits the data in real time to an emergency medical dispatcher (EMD). *PocketCPR* [1] is a mechanical device that enhances the quality of CPR by simple audiovisual feedback in real time, which was already evaluated [6]. The mobile version is called *ZOLL PocketCPR*, [2] which gives real time feedback of an ongoing CPR through the smartphone. It uses smartphone sensors to give the user audio-visual feedback and introduces the user to the whole process of CPR. *CPREzy* [3], is designed for CPR assistance and offers a simple interaction. It has an audible chirp and visual light pacing system with a metronome to guide the CPR. In a study the device was compared with a normal reanimation and the results have shown, that there was no significant difference in compression rate or duty cycles between the techniques [7]. Song et al. [8] describe the usage of an inbuilt accelerometer sensor in smartphones to enhance the quality of a CPR by directly measuring the CCR and CCD. The main difference is that the feedback is restricted to the user and not an EMD. Up to now, like the ones mentioned, have begun to examine how to enhance the quality of CPR. But none of these studies concentrate on direct user feedback and on the dispatch of crucial CPR parameters to the EMD.

III. Background

Using an accelerometer the following physical and technical backgrounds should be considered:

A. Physical considerations of spatio-temporal parameters

The algorithms for CCR and CCD detection are based around the physical concept of acceleration and its first and second order integral velocity and distance. Any change in the velocity of an object results in acceleration. So acceleration

[1]http://www.zoll.com/de/produkte/pocketcpr/, Accessed: April 11, 2016
[2]https://goo.gl/B1pdMR, Accessed: April 11, 2016
[3]http://www.heartworkscpr.com/cprezy-facts.html, Accessed: April 16, 2016

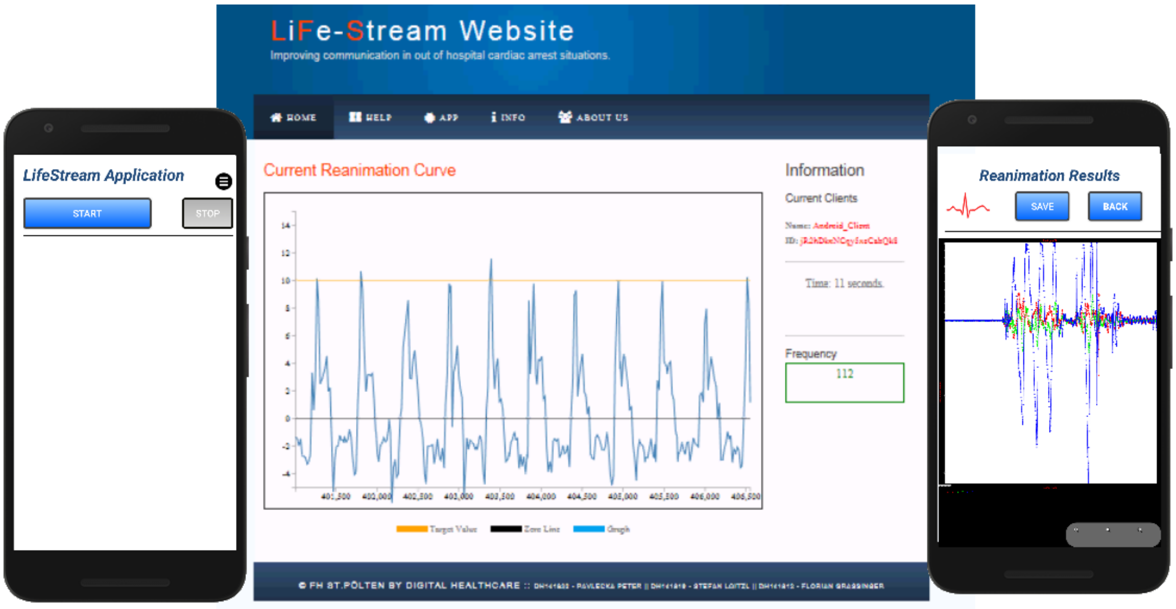


Fig. 1. The website client and the application. On the left side is the starting point of the application and on the right side the results view after the application was finished (mainly for testing purpose and recording).

is related to velocity, or depends on the change of it [9]. The relation of acceleration, displacement and velocity is important, as all of these three quantities are vector quantities (give information about direction) [10]. Using the example of displacement (the directed distance between two points *A* and *B*), it is theoretically possible to determine the final position of the mobile phone, if it is used for CCD detection. The distance would be wrong, as it's only a scalar and counts up the traveled way and not the direct way between two points. Frequency detection is also possible, because after a certain push threshold is exceeded, the push is correct and this counts to the total frequency.

B. Technical considerations

The accelerometer is a powerful mechanical low cost sensor, which is implemented into nearly every smartphone. It offers the possibility to measure the acceleration in a specified direction. The values measured by the smartphone are in m/s^2 and always include the acceleration and deacceleration [11]. In essence the accelerometer measures force that is applied not acceleration. Acceleration just causes an inertial force that is captured by the force detection mechanism of the accelerometer or acceleration is the amount of force needed to move each unit of mass. All calculations take place directly on the smartphone and are processed further to the server, redirecting them to the lifestream website (see 1 for visualization. For the prototype the visualization is restricted to one mobile client. Later, each EMD has his own implementation in the already established call taking system where it shows the visualization of an ongoing CPR.

Fig. 2. Reanimation with phone and LifeStream-App.

IV. Design & Implementation of LifeStream

Based on the input of project members and partners the requirements for the main prototype were formulated: A mobile client with a medical dispatch visualization server to handle clients and a visualization website for visualizing data.

A. Usage scenario

When receiving an emergency call, the EMD advises the caller to open the application (if not open already). Then the application registers at the server endpoint of the medical dispatch center and starts the streaming session. The EMD then instructs over the phone and guides the reanimating layperson through the process. The phone has to be placed between the hands and the victim. Although, many people hesitate (results of user studies) to push directly on the phone, it won't crack in most cases as the hands are laying flat on the phone. Figure 2 shows the placement of the hands.

During the usage scenario definition and the continuous collaboration with the project partners and members, the following four main design considerations were defined:

1) **Simple usability**: The application must be simple and has to gather data, perform calculations, transmit it and stop the whole data acquisition and transmission process.
2) **Restricted functionality**: Frantic laypersons require an application that is protected against unwanted termination. This means the buttons, which are normally used to terminate the application or go back, were disabled. The application is also running in full screen mode and it stays in wakeup state the whole time (10 minutes minimum till ambulance arrives [1]).
3) **Easy configuration**: A simple and non-intrusive menu is used, which allows the change of parameters for the calculation and termination of streaming.
4) **Fast transmission**: As a stable network connection cannot be granted the transmission has to be optimized. Therefore a small and simple data format (JSON [12]) is used, which already contains calculations.

B. Server, Website & Mobile Client

Server: The server is a basic NodeJS server which serves the website and redirects the mobile clients. It allows bidirectional communication, so real-time communication is possible. The server distinguishes between normal clients (web) and mobile clients (Android), who transmit data to it.

Website: The website combines various web technologies. D3.js [4] is used for visualizing the data in a running line graph in real time. The website can be reached over the domain *lifestream.fhstp.ac.at*. At the current prototype state every web client receives the website and while an Android client is connected and streaming data, he can view the reanimation data (seen in Figure 1). On the website the visualization is separated into two major parts. First, the dynamically updating line chart that constantly plots the acquired reanimation data from the smartphone. The color codes used for the frequency, or pushes per minute, are abstractions based on the frequency range:

- Red indicates a very bad frequency (all under 90 or above 130).
- Yellow indicates an average frequency (from 90 to 100 & 120 to 130 pushes).
- Green indicates an optimal frequency (from 100 to 120 pushes).

Second, the information section includes information about the registered clients and the reanimation parameters, e.g., the *frequency*.

Mobile Client: The mobile client is available for Android devices and opens a stream to the server and transmits data of an ongoing reanimation. The data acquisition, calculation and transmission can be started with a simple button press, while the stop functionality is hidden in a small menu above along other configuration options.

[4] https://d3js.org/, Accessed: April 28, 2016

C. Calculation restriction:

Physically and theoretically it should be possible to calculate the traveled distance of the phone by using the accelerometer. If the acceleration is integrated once, the result is the velocity of the object (in this case the smartphone). After a second integration the result is the traveled distance [9], [13]. Despite these equations seem fairly straightforward to implement, they are practically not possible. The natural spread error propagates problematically after each integration as well as the included gravitational force that applies to the phone. A solution to this problem is the usage of a *linear accelerometer*, a sensor fusion of various other sensors that factors out the gravitational force. The main issue with using the above mentioned method is that accelerometers are bad at dead-reckoning (continuous position determination). Accelerometers have some noise which varies from smartphone to smartphone as each has its own manufacturer and device type. The noise can be filtered using various filter types, but normal accelerometers produce raw data, which is not filtered or smoothed. This noise will usually result in a non-zero mean, that is continuously added and accumulates in the resulting velocity signal and later of course in the distance integration. This behavior is called sensor drift, as the integration starts fairly well, but quickly accumulates the errors and the resulting values drift away.

Using the linear accelerometer of the Android system leads to better results, as the gravity is already removed and the resulting values are much smoother. After the gravity is removed and the values are read and filtered with a respective filter, it is advised to calculate the magnitude of the acceleration values before continuing with further calculations [11].

D. Calculation Solution:

By taking all the previous problems and considerations into account, a final functional prototype was developed. The algorithm is a very basic but powerful *peak detection* and *frequency* estimator. After 15 seconds (an adequate update time, based on expert feedback) the frequency on the website is updated based on the average reanimation frequency during this time. The frequency is calculated for these 15 seconds or any other interval, approximated to one minute and then transmitted to the server along with other values (e.g. approximate pressure depth).

The optimal frequency of 100 pushes per minute should theoretically be achieved by pushing always at least five centimeters into the chest of the victim [2]. As CCD detection with the given sensor and technology is not really possible, the approach with frequency seemed more promising as well as an approximation of the distance based on the z-axis acceleration. As the performed reanimation of the user normally changes over time, especially when the power ceases, the frequency detection is very difficult. The requirement to the algorithm must be to detect hard pushes as well as faint pushes. Therefore, peak detection is implemented. According to previous studies and extensive acceleration data logging and plotting, the following concept was devised. Once the

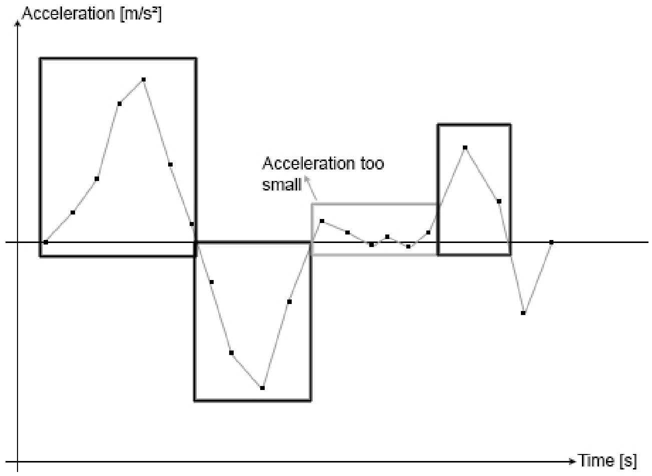


Fig. 3. Peak detection and minimum threshold (along z-axis).

acceleration values or signal traverse the zero line, a change in acceleration happens and a peak can be detected (fig. 3).

Thus, no matter how weak or hard the user pushes, the peak can be detected by its zero-line crossing and change of acceleration with a minimal threshold of applied acceleration. The algorithm is still based on the basic equation and calculation of the magnitude,

V. Evaluation & Validation of prototype

The algorithm was evaluated during a two-day user study, which involved 25 laypersons between 18-50 years (14 lay rescuers and eleven paramedic experienced). They have been selected randomly during an open experiment at St. Pölten University of Applied Sciences, Austria. Each volunteer was instructed before by two professional paramedics. The setup included a reanimation phantom as well as a professional EMD, which was in his actual workplace. The participants were filmed during the whole process and the reanimation phantom also recorded the reanimation process for later comparison to the algorithm. Each participant was not further instructed in the CPR process and they had to reanimate (guided) for full ten minutes. Further randomization happened as some of the laypersons were just reanimating on the phantom without the mobile application. For further insight in the evaluation and test scenario refer to [14] and [15].

A. Results

The outcome of the tests clarified that a guided CPR by using the system is far more efficient for both sides, the EMD and the layperson, rather than a standard phone guided CPR. Some other interesting outcomes as well are:

- Most people hesitate to push on a phone, as it could crack.
- The application detects the frequency very well and is comparable to a professional reanimation phantom that is used for training purposes. However, the accuracy is not as high as a professional sensor, compared to a smartphone accelerometer.
- Displacement can be detected over a short amount of time (movement along the z-axis).
- Displacement detection is not possible with the low cost accelerometer.

VI. Conclusion & Limitations

The performed tests have shown that available smartphone accelerometers along with their embedding systems vary widely and often heavily rely on the hardware and the algorithm used. The accelerometer sensor is often erroneous and creates a non-zero mean that adds up to further calculations. The only solution is filtering and using a linear accelerometer. Often enough the sensor samples slightly slower than the actual sampling frequency as other tasks are more important for the operating system in the background. That means for any calculation it is problematic to rely on fixed time intervals as they are often slightly shorter or longer. The errors are adding up over time and contribute heavily to the whole calculation. A possible solution was to wait an offset time before calculating and estimating depth. At least 100 ms proved to be useful (discarding all before). A remaining problem is that some time stamps are 100 ms or longer and also the fact that numerous important values are lost during the defined pause. During peak detection this can be fatal, as a global maximum could be skipped.

During development it turned out that frequency detection is much easier than continuous position determination, especially in smaller unit ranges (like cm). The theoretic (and physically correct) equations are not feasible for usage in real world applications. Accelerometers measure the acceleration in a body-fixed reference frame, where normally displacement in earth-fixed reference frames is necessary. Therefore, it is not possible to only integrate the accelerometer twice and find the displacement, except it is rotated into the earth fixed frame before the integration takes place. The project showed, that with the given premises of only using the low cost accelerometer sensor in smartphones, it is not possible to make a sturdy point about the displacement. Nevertheless, it is possible to make a point about the current reanimation frequency very well by using the developed peak detection algorithm. Even a position determination could be possible by using the peak detection and the currently viewed values during the peak detection (a so-called window of values) for the integration. As the values are always restricted to a certain amount and interval, a double integration of those values would contain less errors that could add up over time.

Acknowledgments

The authors would like to thank Stefan Loitzl and Peter Pavlecka who have been former project members, contributed during development and lead the testing part. We also would like to give a special thanks to our project partners from Nortuf Niederösterreich especially Heinz Novosad and Raphael Van Tuldar for their support. This work was supported by the Austrian Science Fund (FWF) via the “VisOnFire” project (P27975-NBL).

References

[1] M. Ljunggren, M. Castrn, and K. Bohm, "Abstract 231: The Effect of Dispatcher Call Processing Interval and Ambulance Response Interval for Survival From Cardiac Arrest," *Circulation*, vol. 128, no. Suppl 22, pp. A231–A231, Jan. 2016.

[2] R. O. Cummins, J. P. Ornato, W. H. Thies, and P. E. Pepe, "Improving survival from sudden cardiac arrest: the "chain of survival" concept. a statement for health professionals from the advanced cardiac life support subcommittee and the emergency cardiac care committee, american heart association." *Circulation*, vol. 83, no. 5, pp. 1832–1847, 1991.

[3] L. Wik, J. Kramer-Johansen, H. Myklebust, H. Sreb, L. Svensson, B. Fellows, and P. A. Steen, "Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest," *JAMA*, vol. 293, no. 3, pp. 299–304, Jan. 2005.

[4] C. Ro, "The "chain of survival" concept: how it can save lives." *Heart disease and stroke : a journal for primary care physicians*, vol. 1, no. 1, pp. 43–45, 1991.

[5] B. Roessler, R. Fleischhackl, H. Losert, C. Wandaller, J. Arrich, M. Mittlboeck, H. Domanovits, and K. Hoerauf, "Cardiopulmonary resuscitation and the 2005 universal algorithm: Has the quality of CPR improved?" *Wiener klinische Wochenschrift*, vol. 121, no. 1-2, pp. 41–46, Jan. 2009.

[6] S. Kirkbright, J. Finn, H. Tohira, A. Bremner, I. Jacobs, and A. Celenza, "Audiovisual feedback device use by health care professionals during CPR: A systematic review and meta-analysis of randomised and non-randomised trials," *Resuscitation*, vol. 85, no. 4, pp. 460–471, Apr. 2014.

[7] G. D. Perkins, C. Augr, H. Rogers, M. Allan, and D. R. Thickett, "CPREzy: an evaluation during simulated cardiac arrest on a hospital bed," *Resuscitation*, vol. 64, no. 1, pp. 103–108.

[8] Y. Song, Y. Chee, J. Oh, and S. Lee, "Development of android based chest compression feedback application using the accelerometer in smartphone," in *Proceedings of the International Conference on Biomedical Engineering and Systems*, pp. 130–1.

[9] M. Mounts, "Physics hypertextbook," vol. 44, pp. 1946–1947, 01 1970.

[10] J. Rybach, *Physik fr Bachelors: mit 92 durchgerechneten Beispielen, 176 Testfragen mit Antworten sowie 93 bungsaufgaben mit kommentierten Musterlsungen*, 2nd ed. Mnchen: Fachbuchverl. Leipzig im Carl-Hanser-Verl, 2010.

[11] G. Milette and A. Stroud, *Professional Android sensor programming*, ser. Wrox Programmer to programmer. Indianapolis, Ind: Wiley, 2012.

[12] N. Nurseitov, M. Paulson, R. Reynolds, and C. Izurieta, "Comparison of JSON and XML data interchange formats: a case study." *Caine*, vol. 2009, pp. 157–162, 2009.

[13] P. Becker, "Erfassung und Verarbeitung von Sensordaten," Hochschule Bonn-Rhein-Sieg, 2013.

[14] P. Pavelcka, "LiFeStream - Validation of a mobile application for visual assessment of cardiopulmonary resuscitation (CPR) acceleration data compared with a life-support manikin sensory," Masterthesis, FH St. Pölten, St. Pölten, 2016.

[15] S. Loitzl, "Validating a sensor-based mobile health service with real-time data acquisition for improving MPDS communication for lay-rescuer in out-of-hospital cardiac arrest based on a randomized simulation trial," Masterthesis, FH St. Pölten, St. Pölten, 2016.

Session 3: Interactive Session - Poster

Enabling Decision-Making for Situation-Aware Adaptations of Interactive Systems

Christian Märtin
Faculty of Computer Science
Augsburg University of Applied Sciences
86161 Augsburg, Germany
Christian.Maertin@hs-augsburg.de

Christian Herdin
Faculty of Computer Science
Augsburg University of Applied Sciences
86161 Augsburg, Germany
Christian.Herdin@hs-augsburg.de

***Abstract*—SitAdapt is a new pattern- and model-based architecture and development platform for enabling situation-aware real-time adaptation of media-rich interactive web and mobile applications in e-business and technical contexts. This paper gives an overview of the SitAdapt system, discusses situation patterns, and examines, how instances of this new pattern-type can be mined during lab-based usability tests for task accomplishment and improved user experience. It also demonstrates, how such patterns can be exploited in order to facilitate the generation of different adaptation types at runtime.**

Keywords—situation-awareness; situation analytics; adaptive systems; situation patterns, HCI-patterns; MBUID environments

I. Introduction

SitAdapt is an integrated software system for enabling situation-aware real-time adaptations for web and mobile applications that were developed with the PaMGIS framework [4], [5]. Major application areas are the individualization of digital marketing activities and the contextual support of operators in complex technical environments.

An observer component synchronizes and records the signals from the interfaces to a Tobii eye-tracker, the Noldus FaceReader visual emotion recognition software, advanced wearables like the Empatica E4 wristband, and application meta data (Fig. 1). These data are interpreted by the situation analytics component. A decision component then concludes whether a dynamic adaptation is necessary or not and controls the generation of an appropriate modification of the target software at runtime.

After having demonstrated with a prototype, how the observation, decision-making, and adaptation components of such a system are collaborating [11], we have defined the SitAdapt architecture and adaptation process [12] in detail. The present paper focuses on the operation and the needed knowledge categories of the decision-making component.

Using a MBUID (model-based user interface development) environment [13] for constructing interactive applications offers many advantages before, during, and after target system construction. For instance, models at different abstraction levels can even be accessed after the target system was implemented or generated. This can be extremely helpful for enabling runtime-adaptations of the interactive target system.

For this purpose PaMGIS, a pattern- and model-based MBUID environment that was developed in accordance with the CAMELEON reference framework (CRF), had to be structurally and functionally extended by integrating and interfacing the SitAdapt architecture and its components. The SitAdapt module is linked with various PaMGIS models (Fig. 2). They are exploited for enabling a dynamic and model-driven adaptation process.

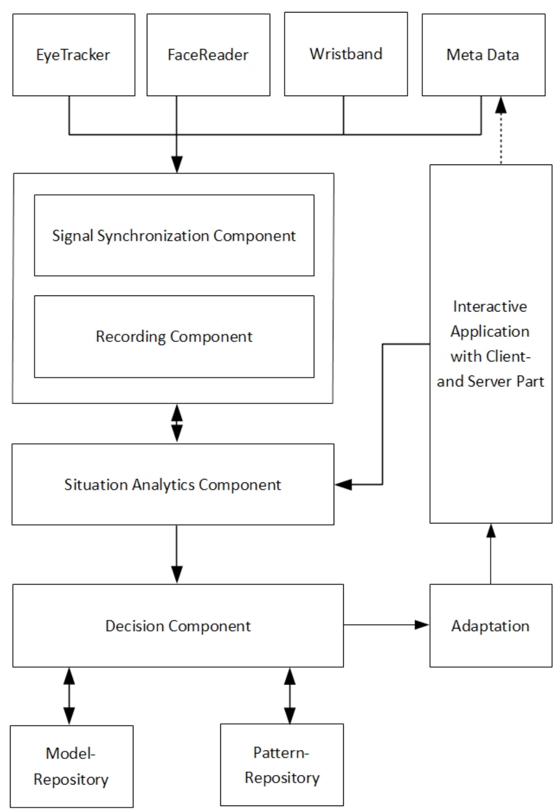


Fig. 1. SitAdapt Architecture

PaMGIS offers pattern and model repositories that can be re-used for the development of other applications. These repositories are continuously extended by patterns and model fragments that are created during target system construction and target system evaluation. The resources available in the repositories, e.g. user interface patterns, low-level user interface templates, or glue code, linking the user interface to business objects, play an important role as artifacts during the dynamic adaptation process.

Apart from serving as a platform for creating adaptive and situation-aware applications, the PaMGIS framework has contributed to the evolution of the MBUID field by adding several types of software patterns and HCI-patterns to the model-based development process. The framework supports automated generation of model refinements and final user interface code. To specify, organize and apply patterns and models, several software tools have been integrated into the framework. An interface for improving the resource base with usability evaluation results is also provided by the framework. Another main contribution of our approach is the PPSL (PaMGIS Pattern Specification Language [6]). PPSL is an extended superset of the major modeling methods for HCI-pattern languages.

In chapter II we discuss the challenges and requirements for enabling situation-aware adaptation. Due to space limitations references to related work are given directly in this chapter. Chapter III focuses on the decision component. As a new contribution, structure and functionality of situation-patterns that are exploited for finding and generating adaptations, are introduced. Chapter IV concludes the paper.

II. Situation-Aware Adaptation

The CRF, a de-facto standard architecture for the model-driven construction of interactive systems [2], includes some model categories and use-cases that allow for adapting the target software in pre-defined ways. However, in order to design interactive systems that are able to adapt dynamically to situational and contextual changes in a way completely tailored to the specific needs of the individual user, a new approach had to be engineered.

The two main goals, supported by this new situation-aware adaptation approach are the following:

- Improved task accomplishment and quality of work. By observing the user on her way to reaching a goal or possibly failing to reach this goal, the system can compare the actual way taken by the user with the workflow proposed by the task model. If necessary, the system can offer help or dynamically restructure the user interface or the task workflow to support the successful completion of the task.
- Better user experience. By observing the current emotional and physical state of the user, the system can propose or generate user interface or other software modifications in order to improve the individual user's sentiments and the overall user experience.

A. Situation-Awareness

Central to this new approach is the concept of *situation-awareness.* Since the introduction of intelligent human-machine interfaces and smart mobile devices HCI research has started to take into account the various new usability, interaction and device-to-device communication requirements of application software running on smaller or embedded hardware devices with touch-screen or speech interaction, e.g in cars, on smartphones or wearables.

Mobile applications that migrate smoothly from one device type to another need special support for responsiveness and user interface quality. Several of the necessary requirements for these apps targeted at different platforms and devices can be specified and implemented using the models and patterns already existing in advanced MBUID systems.

Even runtime support for responsiveness with the interactive parts distributed or migrating from one (virtual) machine to the other and the domain objects residing in a cloud can be modeled and managed by CRF-conforming development environments [14].

When discussing adaptive user interface modifications more generally, three different types of adaptation have to be distinguished [1], [20]:

- Adaptable user interfaces. The user interface is a-priori customized to the person*al preferences of the user.*
- Semi-automated adaptive user interfaces*. The user interface provides recommendations for adaptations, which can be accepted by the user or not.*
- Automated adaptive user interfaces. *The user interface automatically reacts to changes in the context of the interactive application*

In order to arrive at interactive systems that can be modified depending on changing situations at runtime, semi-automated and automated adaptivity of the user interface have to be supported. Note, however, that for situation-aware adaptation this is not sufficient, because a reaction to changing situations may involve more than a user interface modification. For instance, modifications of the interface between the business domain classes and the user interface or even of the business domain classes and the representing task and concept models (see fig. 2) might be necessary to reach an adequate level of system intelligence.

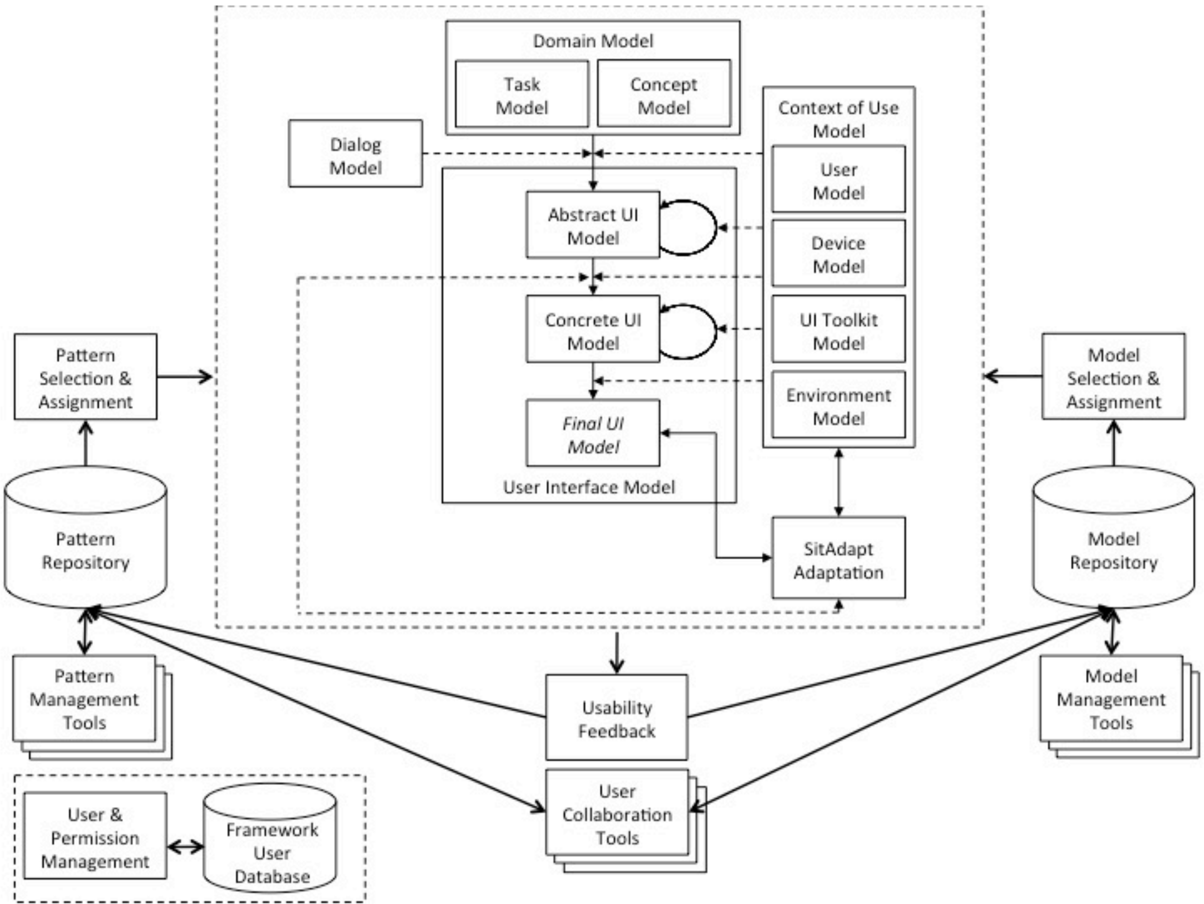


Fig. 2. PaMGIS MBUID framework with SitAdapt extensions and model-interrelations for situation-aware dynamic adaptation

The concept of context-aware computing was first proposed for distributed mobile computing in [17]. In addition to software and communication challenges to be solved when dynamically migrating an application to various devices and locations within a distributed environment, the definition of context also included environmental and social aspects (e.g. lighting and sound environment, are there other people around?, who are these people?, etc.). At the same time period early definitions of the term situation-awareness appeared in psychology and the cognitive sciences, with the aim to support human operators in complex situations, e.g. pilots during the landing phase, by defining situation-dependent requirements for allowing a smooth and correct task accomplishment [7], [8].

Since then, interactive software has made huge steps towards understanding and reacting to varying situations. To capture the individual requirements of a situation, Chang [3] proposes that a situation specification must cover the user's operational environment *E*, the user's social behavior *B* by interpreting his or her actions, and a hidden context *M* that includes the users' mental states and emotions. A situation *Sit* at a given time *t* can thus be defined as $Sit = \langle M, B, E \rangle_t$. A user's intention for using a specific software service for reaching a goal can then be formulated as temporal sequence $\langle Sit_1, Sit_2, \ldots, Sit_n \rangle$, where Sit_1 is the situation that triggers the usage of a service and Sit_n is the goal-satisfying situation. Chang also proposes a framework *(Situ)* that can be used for modeling and implementing applications that are situation aware and adapt themselves to the users' changing needs over runtime.

Our own work on SitAdapt was inspired by [3]. However, our main goal was to design a high-quality and practicable software engineering approach for building situation-aware target systems. Therefore, we maintained the model-based approach of the PaMGIS framework by linking the domain and user interface models with the user-centric situation-aware adaptation component.

The architectural details of this integrated solution and the steps of the adaptation process implemented by SitAdapt are discussed in depth in [12]. SitAdapt records situations in user-specific situation profiles. The PaMGIS context of use model, mainly the user sub-model, accesses the situation profiles in order to gain insight into the user state at any time during the observation period.

B. Observing the User

For implementing the emotion recognition functionality that can be exploited for inferring the desires and sentiments of individual users while working with the interactive application, the current version of SitAdapt captures both visual and biometric data signals. In its current version user monitoring within several e-business scenarios (e.g. travel booking, finding and ordering beauty products) is implemented in an advanced usability lab environment. The user is observed already before starting to interact with the application, during interaction, and until after the session is closed.

In [9] we discuss the interplay of the various recognition approaches used in the SitAdapt system. Our work was influenced by several current research approaches for visual and bio-physical emotion recognition techniques, e.g., [15], [18], [16]. We have also studied the practical aspects of integrating runtime analytics, e.g. [10], and the consequences for sharing such information and privacy issues, e.g. [19].

We are currently beginning to evaluate the large data sets obtained by our user tests using big data analytics methods in order to extract typical emotion- or stress-correlated patterns in the usage behavior. The lab results are also interpreted in order to mine typical situation patterns that will be used by the decision component to trigger individual adaptations. Chapter III gives examples of situation patterns for different adaptation types.

III. Situation Patterns for Implementing Dynamic Adaptation

To give readers an idea of how the SitAdapt adaptations are performed by the decision component, we present three synthetic situation patterns for different adaptation types. Future situation patterns will be detected by analyzing the lab data from scenario-based user tests as described above.

Situation patterns *(SitPat)* must not be confused with design patterns. Each SitPat consists of a pattern-recognition and an action part. The action part may contain modifications of attribute settings in the PaMGIS models, UI pattern and model fragment activations for all modeling levels, or other user interface actions.

Note, that the complete adaptation process involves the dynamic interaction of the SitAdapt decision and adaptation components with PaMGIS components and models on all abstraction layers and is discussed elsewhere [12]. Also note, that the PaMGIS/SitAdapt architecture offers a high degree of flexibility and is also open to solutions for adaptation implementation that are not based on situation patterns.

A. Pre-Runtime Adaptation

This pattern uses the FaceReader attribute *age,* to set attribute values concerning the legal capacity and some presentation attributes for the current user in the concrete UI model.

```
<SitPat> YoungUserConditioning
    FOR <Situation_i>
    <FaceReader> <Age> (<18)
    <Action> <UserModel:UserLegalCapacity>
                 := NO
    <Action> <CUI:FontSize> := SMALL
    <Action> <CUI:Coloring> := YOUNG
```

B. Runtime Adaptation of the User Interface

In this pattern, several successive situations give hints to an inattentive car driver. An attention assist pattern and a sound-signal is activated in the user interface.

```
<SitPat> TiredUser
    FOR N <Situation_i>
    <Eye_Tracking> Not Focused
    <Gaze_Tracking> Rotating
    <FUI> WindshieldView
    <Pulse> Low
    <Stress Level> Green
    <Emotion> Neutral
    <Action> SHOW AttentionAssistFUIPattern
    <Action> ACTIVATE FUIAttentionSound
```

C. Domain-Dependent Runtime Adaptation with Task Model Interaction

This pattern recognizes the user's interest in a certain product in a web-shop. After three minutes a text is displayed, e.g. notifying the user that in case of the purchase of product (Id) within the next 10 minutes, a voucher of $10 is granted for the user's next purchase. A link to the voucher processing task in the task model is activated.

```
<SitPat> OfferingVoucher
    FOR N <Situation_i> IN 180s
    <Eye_Tracking> Field Product Product(Id)
    <Gaze_Tracking> Contains Field
           Product(Id) (>5)
    <Pulse> (85-100)
    <PulseRate> rising
    <Emotion> excited
    <StressLevel> orange
    <Action> SHOW AT 180s VoucherText1FUI
    <Action> WAIT VoucherText1FUIInput
    <Action> LINK VoucherText1FUIInput
           TaskModel VoucherProcessingTask
```

IV. Conclusion

In this paper we have presented the current state of our SitAdapt project. Since the start of the project we have built a demonstrator prototype, designed the system architecture and its integration into the PaMGIS MBUID environment, and specified the detailed process for situation-aware dynamic adaptation of the user interface and necessary interactions with other PaMGIS models.

In the present paper we have discussed the concepts of situation-awareness, and, for the first time, how situation patterns can easily be exploited to automate the adaptation process and at the same time preserve the model-driven nature of the PaMGIS development paradigm.

We are now beginning to evaluate the SitAdapt approach in the lab with a real-world e-business portal for beauty and health products. Here we are both looking for easy to handle situation patterns, but also trying to mine usage patterns that give us the directions to design a SitAdapt version for the end-user without having to go the full observation procedure, saving user privacy, but still getting some relevant emotional data.

V. References

[1] Akiki, P.A., et al.: Integrating adaptive user interface capabilities in enterprise applications. In: Proceedings of the 36th International Conference on Software Engineering (ICSE 2014), pp. 712-723. ACM (2014)

[2] Calvary, G., Coutaz, J., Bouillon, L. et al., 2002. "The CAMELEON Reference Framework". Retrieved August 25, 2016 from http://giove.isti.cnr.it/projects/cameleon/pdf/CAMELEON%20D1.1RefFramework.pdf

[3] Chang, C.K.: Situation Analytics: A Foundation for a New Software Engineering Paradigm, IEEE Computer, Jan. 2016, pp. 24-33

[4] Engel, J., Märtin, C., Forbrig, P.: A Concerted Model-driven and Pattern-based Framework for Developing User Interfaces of Interactive Ubiquitous Applications, Proc. First Int. Workshop on Large-scale and Model-based Interactive Systems, Duisburg, pp. 35-41, (2015)

[5] Engel, J., Märtin, C., Forbrig. P.: Practical Aspects of Pattern-supported Model-driven User Interface Generation, To appear in Proc. HCII 2017, Springer (2017)

[6] Engel, J., Märtin, C., Forbrig, P.: A Unified Pattern Specification Formalism to Support User Interface Generation. To appear in Proc. of HCII 2016, Toronto, 17-22 July, Springer LNCS, 2016

[7] Flach, J.M.: Situation awareness: The emperor's new clothes, in Mouloua, M., Parasuaman, R. (eds.): Human performance in automated systems: Current research and trends, pp. 241-248, Erlbaum, (1994)

[8] Flach, J.M., Mulder, M, Van Paassen, M.M.: The Concept of the Situation in Psychology, in: Banbury, S. and Tremblay, S. (eds): A Cognitive Approach to Situation Awareness: Theory and Applications, Ashgate Publisching, Oxon (UK), pp. 42-60, (2004)

[9] Herdin, C., Märtin, C., Forbrig, P.: SitAdapt: An Architecture for Situation-aware Runtime Adaptation of Interactive Systems. To appear in Proc. HCII 2017, Springer (2017)

[10] Lee, Y., Balan, R.K.: The Case for Human-Centric Personal Analytics, Proc. WPA '14, pp. 25-29, ACM (2014)

[11] Märtin, C., Rashid, S., Herdin, C.: Designing Responsive Interactive Applications by Emotion-Tracking and Pattern-Based Dynamic User Interface Adaptations, Proc. HCII 2016, Vol. III, pp. 28-36, Springer (2016)

[12] Märtin, C., Herdin, C., Engel, J.: Model-based User-Interface Adaptation by Exploiting Situations, Emotions and Software Patterns, Proc. CHIRA 2017, Funchal, Madeira, Oct. 31 – Nov., 2, SCITEPRESS (2017)

[13] Meixner, G., Calvary, G., Coutaz, J.: Introduction to model-based user interfaces. W3C Working Group Note 07 January 2014. http://www.w3.org/TR/mbui-intero/. Accessed 27 May 2015

[14] Melchior, J., Vanderdonckt, J., Van Roy, P.: A Model-Based Approach for Distributed User Interfaces, Proc. EICS '2011, pp. 11-20, ACM (2011)

[15] Picard, R.:"Recognizing Stress, Engagement, and Positive Emotion", Proc. IUI 2015, March 29-April 1, 2015, Atlanta, GA, USA, pp. 3-4

[16] Qu, F., Wang, S.-J. et al. CAS(ME)2: A Database of Spontaneous Macro-expressions and Micro-expressions, M. Kuroso (Ed.): HCI 2016, Part III, NCS 9733, pp. 48-59, Springer, (2016)

[17] Schilit, B.N., Theimer, M.M.: Disseminating Active Map Information to Mobile Hosts, IEEE Network, vol. 8, no. 5, pp. 22–32, (1994)

[18] Schmidt, A. Biosignals in Human-Computer Interaction, Interactions Jan-Feb 2016, pp. 76-79, (2016)

[19] Warshaw, J. et al.: Can an Algorithm Know the "Real You"? Understanding People's Reactions to Hyper-personal Analytics Systems, Proc. CHI 2015, pp. 797-806, ACM (2015)

[20] Yigitbas, E., Sauer, S., Engels, G.: A Model-Based Framework for Multi-Adaptive Migratory User Interfaces. In: Proceedings of the HCI 2015, Part II, LNCS 9170, pp. 563-572, Springer (2015)

HoloMuse

A Concept for Augmented Learning in Museums

Dr. Kasra Seirafi, Florian Wiencek
Fluxguide
Vienna, Austria
kasra@fluxguide.com, florian@fluxguide.com

Providing accessible ways to our cultural heritage is a key responsibility for modern societies. This paper outlines ideas for using new Augmented and Virtual Reality technologies at museums. First, we will give a short introduction on the latest developments of AR/VR technology and their inherent potentials for museums. Secondly, the ongoing R&D project "HoloMuse" will be introduced which aims at developing new forms of Augmented Learning for cultural heritage venues. HoloMuse is a collaborative effort of technology partners, academic institutions, and worldwide renowned museums to make the next steps towards real innovative and meaningful museum visitor experiences based on AR/VR technologies and cutting-edge learning theories. We will outline several deployment scenarios of AR/VR in museums giving a prospective on the future of cultural learning.

Keywords—Augmented Reality, Virtual Reality, cultural learning, museum, exhibition

I. Introduction And Related Work

Cultural heritage institutions are not only responsible for preserving but also for mediating the abundant treasure of worldwide art, architecture, history, and technology. Visiting a museum should be an inspiring experience. It should enable us to go beyond our own horizon, to foster creativity and new ideas. Thus, one main challenge is to provide ways for non-expert visitors to gain a deeper understanding of museum exhibitions. This is the field of „museum interpretation", which for a long time was narrowed down to wall labels and museum tours, later on extended by „audio guides". In recent past, mobile technologies like smartphones, tablets, and mobile apps came up offering richer content, explaining museum exhibits with multimedia such as text, images, audios, etc. However, the traditional concept of museum interpretation remained, i.e. to conceptualize the visitor as a learner which has to be „fed" with information. This passive understanding of learning has massively been refuted by learning theory and educational sciences (see e.g. [1] [2] [3]). Sustainable museum experiences cannot be accomplished via an information transfer from a curator to a passive visitor. True knowledge about topics of an exhibition cannot be mediated to visitors by mere information transfer, be it via an audio guide or smartphone app. Therefore, reception processes are essential in order for an exhibition item to unfold its full creative potential. Only *active thinking* and *doing* enables museum visitors to create meaning and to dwell into the richness of what museums all over the world have to offer. This paper introduces ways in exploiting new technology developments to provide active experience spaces improving the way we experience our heritage.

Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR) offer unprecedented new possibilities for inspiring experiences as well as learning from and with art and culture as well as for curating and designing exhibitions [4,5]. AR is a technology that generates an interactively augmented view of reality in which the physical space can be enriched and overlaid with digital elements. Or, as Lev Manovich states it: AR is a physical space transformed into a dataspace by either "extracting data from it […] or augmenting it with data" [6]. VR is a computer-generated environment that either simulates the physical reality and possible experiences within it, or has the potential to go beyond it creating a fictive (other) reality with its own rule-system. The visitor is immersively embedded in this environment. MR is a mixture of both modalities, in which real and virtual objects co-exist in a space and interact with each other, where it is not clear if the physical or virtual environment is the predominant one [7][1]. Therefore, these technologies show high potential for mediation of art and culture that works via an emotional, visual and haptic layer. They enable explicit information mediation, which gives visitors not only access to information but also a better understanding of situations and processes [8].

With more affordable VR devices such as Oculus Rift, Samsung VR or Google Cardboard VR immersive experiences are on the rise as well as massively adapted in current media art[2]. Museums are starting to take up this technology in order to give their audience more immersive experiences and learning opportunities[3]. Where for VR dedicated devices are used for providing the experience, currently most AR applications are geared towards smartphones and tablets as end devices. But the

[1] It expands for example AR with an extended understanding of the physical space around the user and therewith enables the user to physically move through space and explore for example a virtual 3D scene.

[2] From early works in CAVE and other immersive environments such as "World Skin" by Jean-Baptiste Barrière & Maurice Benayoun to current computer animation works) and for editorial content (e.g. "Zero Days VR" by Scatter or "Out of Exile" by Nonny de la Peña / Emblematic Group)

[3] see [9]; e.g. "We are like vapours" by Jeffrey Shaw, Sarah Kenderdine & Cédric Maridet or experiences designed by Woofbert VR

really interesting jump in technology is the introduction of AR-glasswear. Big manufacturers such as Vuzix or Epson already have been working on the development of market-ready AR devices that are currently mainly used in industry. The most advanced AR-glasswear at the moment is Microsoft Hololens. The breakthrough of more market-ready devices for mass market is expected in the upcoming years and will lead to broad usage of AR head mounted applications.

II. THE HOLOMUSE PROJECT: FROM MOBILE LEARNING TO AUGMENTED LEARNING

The following questions drive the R&D project "HoloMuse": How can AR and VR be employed for individual guiding and mediation, for cultural learning, as well as for personal mediation? How can these modes enable co-creative knowledge generation of and with visitors? How can it support group activities, for example of researchers and students involving physical items, cultural data as well as research data? And how to keep the focus in the gallery on the exhibited items and their context and not on a technological device? What kind of novel experiences with and of cultural data can be created that foster sustainable cultural learning? These questions led to the proposal of several use cases and innovative mediation modes for co-creative knowledge generation and cultural learning.

The R&D project HoloMuse will be carried out by *Fluxguide* (http://www.fluxguide.com/) and several partners[4]. *Fluxguide* already put many efforts in developing inspiring learning environments to activate visitors, to interact and reconstruct art, science and knowledge worldwide [10]. By expanding the *Fluxguide©* Museum Solution [11] with the "HoloMuse" module (see Fig. 1), museums and cultural institutions will be able not only to offer mobile mediation via smartphones and tablets but also to implement mediation as AR/VR applications. The aim is to develop a software solution for realizing HoloMuse mediation experiences ("HoloMuse.APP" runnable on mobiles as well as AR and VR devices). It will be accompanied by a didactically and scientifically sound mediation and learning concept ("HoloMuse.TOOLBOX") for cultural learning which enables institutions worldwide to effectively and meaningfully conceptualize and employ AR and VR in their mediation processes.

HoloMuse expands the experience inside and outside of museums with alternative and other realities (in time and place), and with visions for the future or fictive realities. The solution will also be able to integrate research data from the Digital Humanities directly into the mediation process.

III. HOLOMUSE LEARNING MODES – IDEASKETCHES

The following paragraphs will outline concepts of the main HoloMuse "Modes" and give an overview of possibilities provided by AR/VR based museum mediation.

Fig. 1: HoloMuse –overview showing the components of the overall HoloMuse modules including envisaged learning modes.

A. *HoloMuse.immersive.spaces*

The mode HoloMuse.immersive.spaces provides users with engaging, immersive experiences inside and outside of the museum. By using digital and immersive storytelling – so-called explicit information mediation [8] – not only a transfer of information takes place. The physical space is augmented with virtual elements and visitors are sent on a journey into "another world" – for example through a reconstruction of historic places, a living vision for the future, or via bringing alive the fictive world of paintings or literature.

Thus, this mode allows to *transform every possible existing location into a learning space*, augmenting the physical location with new and old knowledge, stories, possible histories, presences and futures. HoloMuse allows to immerse into the depicted fictive world of an artwork. For example, a character of a painting could step out into the gallery space and act as co-present gallery guide introducing the (hi)story of the artwork and context of the depicted scene. The fictive character becomes an „authentic" authority to give interactive and contextual information and to represent the artistic idea (see Fig. 2).

Fig. 2: HoloMuse.immersive.spaces – A fictional character from a painting steps out of the artwork via Augmented Reality technology, acting as a guide giving insights about the work.

4 Renowned museums like Albertina or Deutsches Technikmuseum, technology partners like Microsoft, and scientific partners like the University of Applied Arts Vienna.

Moreover, one can take a virtual fieldtrip in the physical world and explore, for example, a virtual reconstruction of a monument that has become merely a ruin in present times (see Fig. 3). This allows immersive experiences of a possible past. Also, future scenarios can be experienced: one could for example visualize the transformation of an existing landscape due to climate change 100 years from now, and therewith illustrate the meaning and specific interpretations of research data. Therefore, this mode can also be useful for science communication and vividly mediating research data. The experience of the visitor's physical environment is altered and formed.

Fig. 3: Holomuse.immersive.spaces enable the exploration of a historic place or a future scenario in the present physical environment.

B. HoloMuse.exploration

This mode allows museum visitors to explore exhibitions with an end device like smartphone or glass wear. Users receive interactive, 3D-multimedia information about key objects and navigate within the exhibition space by gestures. An example could be enabling visitors to read hieroglyphs and provide translations to them in relation to old Egypt objects. Virtual content-pins on objects reveal detail information and therewith foster personal interpretation. This may prompt the visitor to fill mental gaps, to ask questions by engaging with the exhibited object or trigger associations and memories related to visual hooks or entry points (see Fig. 4). A learning mode enables to trigger interactive challenges and explorative task while directly engaging the visitor with the original physical or a digitized virtual object (see Fig. 6).

Fig. 4: HoloMuse.exploration augments virtual and physical objects with visual hooks displaying information and leading the visitors' gaze.

C. HoloMuse.tour

Humans learn from humans. Therefore, HoloMuse will be integrated with human mediation and guided tours. During a guided tour through the physical museum space a guide may pass around virtual objects to individuals or a group. This enables visitor interaction e.g. with closer explanations of fragile museum artefacts which otherwise could not be touched in a museum environment (see Fig. 5; a similar feature was also presented by [12]). Additional information or media content can be augmented and discussed directly in relation to the object.

Fig. 5: A guide discusses a virtual object with a visitor during a guided Holomuse.tour.

The inclusion of interactive exercises in interaction with the virtual or physical object fosters learning processes for groups and individuals (see Fig. 6). It also offers promising opportunities for the important area of education mediation, e.g. for museum school class visits. Visitors could join the tour in the museum but also in virtual co-presence.

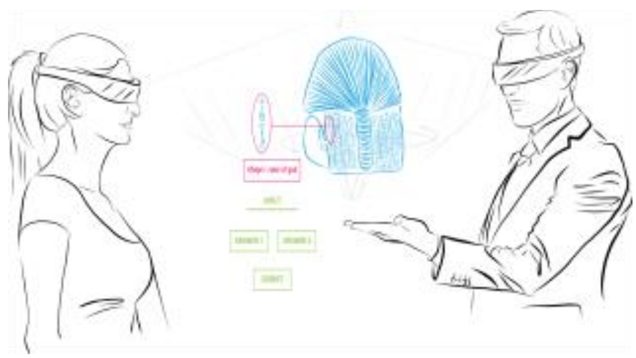


Fig. 6: Learning mode in HoloMuse.exploration and HoloMuse.tour. A learning task is solved in direct interaction with a virtual or physical museum object by individuals or groups.

D. HoloMuse.virtual.laboratory

Especially for science communication, science centers and technology museums, a virtual laboratory enables to conduct interactive experiments in VR as well as AR. Simulations of complex processes could be observed and tested in physical or virtual co-presence in an exhibition space, a museum laboratory, a university or at home. Instructed or independently. This opens up a whole range of possibilities for interlinking digital humanities research directly with museums, their collections, exhibitions as well as informal cultural learning. Latest research becomes interactively available to life-long learners, involving the visitors in co-creative knowledge generation processes around the data on multiple devices (see also [13]). This offers benefits for both museums and researchers, binding back the research to the museum as trusted information source.

Fig. 7: The HoloMuse.virtual.laboratory enables a co-present exploration of research data or simulations in physically dispersed spaces as well as co-creative knowledge generation.

E. HoloMuse.curating.room

In HoloMuse.curating.room users design their own exhibitions on basis of digitized cultural data from the museum collection or favorites collected during the museum visit on a mobile (see Fig. 7).

One scenario could be preceded by a real museum visit where users mark their favorite objects on a mobile museum guide. At the end of the exhibition the visitor receives AR glasses, e.g. a Microsoft HoloLens, and places objects in a "curating room" – also in collaboration with others. For visitors, the design of such a personal exhibition increases the active engagement with cultural artifacts. It could also be taken home as memory of a museum visit which can be displayed on smart devices as well as AR and VR devices and shared with friends and family.

Fig. 8: In the HoloMuse.curating.room visitors and museum professionals are able to place collection items in an augmented physical museum space. Either as learning tool or for planning new exhibitions.

For museum professionals, this "curating.room" tool addresses the issue that it is very hard to design an exhibition at a specific venue, taking into account the specific spatial situation of a gallery space as well as relations between objects and conservatory prerequisites. Even if exhibition designers are on-site for the design process, it takes huge efforts to do iterative positioning of artworks and other exhibition items. Smart use of new technologies may help here as well.

At the same time a virtual exhibition can also enable a re-experience of historic exhibition settings [14] as well as the meta-experience [15] of endangered (also natively digital) media artworks, that otherwise cannot be shown physically due to technological obsolescence. This creates a unique experience for the museum visitor. For the museum the digital reconstruction, documentation and meta-experience additionally acts as preservation of these artworks [16, 17].

IV. Conclusion

The scenarios above only outline a first glimpse of possibilities, opportunities, but also challenges, of utilizing AR/VR technology for cultural interpretation. The concepts are based on the assumption that AR devices will be widely adopted by the mass market in near future outside of purely industrial applications. This is necessary to go beyond mobile devices as end-devices for AR applications and therewith enable a more immersive and seamless experience. First results show that it will be crucial to use those technologies to *enhance* cultural exhibitions, not to *substitute* them. That is why especially AR technology seems very promising. This is intertwined with the challenge that interactive AR/VR content really should give added value and new learning possibilities, and therefore should be integrated and backed in sound didactical and learning concepts. Thus, the project HoloMuse will go on focusing on both technological and conceptual challenges, together with museum professionals and technology partners.

References

[1] S. Downes, "Connectivism and Connective Knowledge. Essays on meaning and learning networks," 2012. [Online]. Available: http://www.downes.ca/files/Connective_Knowledge-19May2012.pdf.

[2] A. Cutler, "What is to be done – Sandra? Learning with Young People in Cultural Institutions of the 21st Century," in *Kulturelle Bildung im Museum. Aneignungsprozesse – Vermittlungsformen – Praxisbeispiele*, H. Kuntz-Ott, S. Kudorfer, and T. Weber, Eds. Bielefeld: transcript Verlag, 2009, pp. 57–74.

[3] C. Mörsch, "Allianzen zum Verlernen von Privilegien: Plädoyer für eine Zusammenarbeit zwischen kritischer Kunstvermittlung und Kunstinstitutionen der Kritik," in *medien kunst vermitteln*, N. Lüth and S. Himmelsbach, Eds. Berlin: Revolver Publishing, 2011, pp. 19–31.

[4] B. Freeman, A., Adams Becker, S., Cummins, M., McKelroy, E., Giesinger, C., and Yuhnke, NMC Horizon Report. 2016 Museum Edition, Austin, TX: The New Media Consortium, 2016.

[5] Tscheu, Frances, and Dimitrios Buhalis. "Augmented reality at cultural heritage sites." Information and Communication Technologies in Tourism 2016. Springer, Cham, 2016. 607-619.

[6] L. Manovich, "The Poetics of Augmented Space," 2002. [Online]. Available: http://manovich.net/content/04-projects/034-the-poetics-of-augmented-space/31_article_2002.pdf

[7] A. De Souza, "Space and Culture From Cyber to Hybrid Mobile Technologies as Interfaces," 2006.

[8] T. Moesgaard, M. Witt, J. Fiss, C. Warming, J. Klubien, and H. Schoenau-Fog, “Implicit and Explicit Information Mediation in a Virtual Reality Museum Installation and its Effects on Retention and Learning Outcomes,” *Eur. Conf. Games Based Learn.*, pp. 387–395, 2015.

[9] B. Freeman, A., Adams Becker, S., Cummins, M., McKelroy, E., Giesinger, C., and Yuhnke, NMC Horizon Report. 2016 Museum Edition, Austin, TX: The New Media Consortium, 2016, p. 42.

[10] Fluxguide, “Museum Apps. Best Practices,” 2017. [Online]. Available: http://bit.ly/2hhjSND.

[11] Fluxguide, “Fluxguide Modules & References. Guiding for the next Museum,” 2016, [Online]. Available: http://bit.ly/2eHrXuP.

[12] Christina Pollalis, Whitney Fahnbulleh, Jordan Tynes, and Orit Shaer. 2017. HoloMuse: Enhancing Engagement with Archaeological Artifacts through Gesture-Based Interaction with Holograms. In Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (TEI '17). ACM, New York, NY, USA, 565-570. DOI: https://doi.org/10.1145/3024969.3025094

[13] Blumenstein, Kerstin, Markus Wagner, and Wolfgang Aigner. "Cross-Platform InfoVis Frameworks for Multiple Users, Screens and Devices: Requirements and Challenges." Proceedings of the DEXiS 2015 Workshop on Data Exploration for Interactive Surfaces. Workshop in conjunction with ACM ITS. Vol. 15. 2015.

[14] Stedelijk Museum Amsterdam, „History and Archive, “ 2017, [Online]. Available: http://www.stedelijk.nl/en/artours/history-and-archive.

[15] Wiencek, F., „The Meta-Experience of (Media) Art. Activating the Archive,“ 2013, [Online]. Available: http://www.slideshare.net/fwiencek/the-metaexperience-of-media-art-activating-the-archive.

[16] Muñoz Morcillo, Jesús & Faion, Florian & Zea, Antonio & Hanebeck, Uwe & Trotha, Caroline. “e-Installation: Synesthetic Documentation of Media Art via Telepresence Technologies,” 2016. 173-191. 10.1007/978-3-319-24942-1_11.

[17] Wiencek, F. & Lauke, S. S., „The Remediation of Experience. A Case Study“. Proceedings of the ISEA 2011 conference, 2011. [Online]. Available: http://isea2011.sabanciuniv.edu/paper/remediation-experience-case-study.

Innovative and Intuitive Hands-on Interaction with RFID to Enhance Digital Media Experience of Exhibits

Sandra Schadenbauer, Alexander Nischelwitzer, Robert Strohmaier, Gerhard Sprung
Information Management
FH JOANNEUM
Graz, Austria
sandra.schadenbauer@fh-joanneum.at, alexander.nischelwitzer@fh-joanneum.at, robert.strohmaier@fh-joanneum.at, gerhard.sprung@fh-joanneum.at

***Abstract*—RFID (Radio Frequency Identification System) technology is very popular today and is used in everyday life. However, RFID hasn't won much recognition in museums yet, especially not in the field of hands-on experience. This paper shows that RFID can be used for innovative and intuitive hands-on interaction to enhance digital media experience of exhibits in museums and presents several projects in museums, which uses RFID for the exhibits' interaction.**

***Keywords*—RFID, digital media; tangible objects; invisible tagging; exhibit; exhibition; museums; showrooms; hands-on**

I. Introduction

The research field Digital Media Technologies (DMT) of the Institute of Information Management at the University of Applied Sciences FH JOANNEUM Graz develops multimedia implementations for different fields of application. DMT is active in numerous projects with interactive multimedia installations and stations for museums and exhibitions.

The DMT team strives to choose the right technology for the right purpose. Over the years, we applied the RFID (Radio Frequency Identification System) technology at various multimedia projects.

At the moment, museums use RFID mainly for security (protecting from theft) [1][12], artefact tracking (inventory) [1][2][10][12], ticketing[1] and tracking visitors' flow [4][5] – which is also a topic of our own research [13]. RFID is also implemented for enhancing the visitors' experience by providing additional information in different languages about exhibits with a "personalized" museum guide [1][5][6][7][11][13], which in some cases also collects data during the visit, so that visitors can take a part of the exhibitions home for the post-visit experience [1][7][11][13]. This is used for example in the travelling exhibitions "Heart over Heals" or "Show Me the Money", which we developed in cooperation with the Graz Children's Museum FRida & freD[1].

[1] http://fridaundfred.at/

Most of the approaches and projects, which aim to increase the visitors' experience by using RFID technology, provide additional content about an exhibit or the whole exhibition, but do not offer "hands-on" experience. Therefore, DMT uses RFID for interactive multimedia exhibits, where objects, equipped with RFID tags, are essential to use and understand the exhibit.

This paper is organized as follows: Section II introduces the RFID technology and compares it to other methods of identification in the context of hands-on exhibits. Section III describes the software interface between RFID readers and exhibition software. Finally, section IV gives an idea of innovative and intuitive hands-on interaction with RFID by explaining some of our projects.

II. RFID

A. General

RFID stands for Radio Frequency Identification System and basically consists of a reader and several transponders (also called tags) [9]. Every RFID tag has a worldwide UID (Unique ID).

RFID is contactless and uses radio frequency signals for communication [8]. Readers emit radio waves, which are received by the tags and in turn, they response with data, which is read by the reader [9].

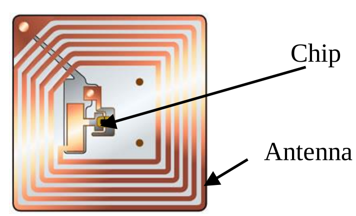


Fig. 1. RFID Tag[2]

[2] Image Source: https://www.phidgets.com/SpecSensorSeries3/html/858x-/25/rfid-label-phidgets-usb-sensing-and-control.jpg

The RFID technology is very popular today and is used in everyday life, for example:

- Cashless/wireless payment [1][8]
- Ticketing [1]
- Access control [1][8]
- Smart home controls [1]
- Medical purposes [8]
- Warehouse inventory [8]
- Sport events [8]

RFID readers and tags are available in many forms and can be categorized by frequency (Low Frequency 9-135 KHz, High Frequency 13.56 MHz, Ultra High Frequency 0.3-1.2 GHz and Microwaves 2.45-5.8 GHz), size, tag technology (Mifare, ICODE, etc.), interface (USB, Ethernet, Serial, Bluetooth, etc.) and active (transponders with own power supply) or passive (transponders are powered by a reader's inductive field) systems. The reading range depends on these characteristics, and can cover a few millimeters to several meters [9].

Fig. 2. Variety of RFID Tags

The selection of the system depends on the requirements. In our projects, we have gained positive experience with passive ICODE tags from NXP[3], which operate on 13.56MHz (HF) supporting ISO 15693 and ISO 18000-3 [14]. The reasons are that ICODE tags come in several dimensions and forms and can also be processed by many readers with different sizes of antennas. We develop multimedia applications that mostly run on personal computers. The exhibition software communicates via an XML socket application with the reader, which is connected to the USB interface of the PC. Wireless communication would also be possible, but USB is more reliable.

Our RFID implementations can be classified by following characteristics:

- Mobility: stationary or moving readers
- Quantity: one or more readers
- Range: from 2 mm to 50 cm
- Usage: single user and multiuser applications

[3] https://www.nxp.com/

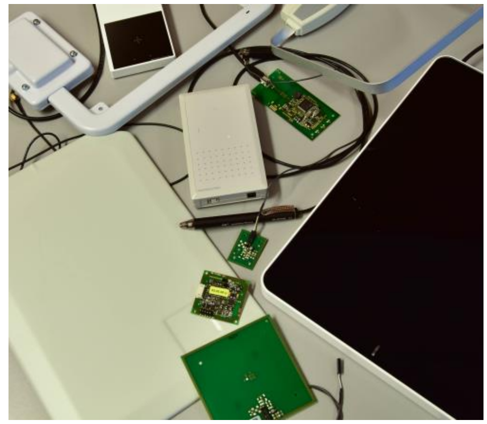


Fig. 3. Several RFID reader and antennas operating on 13.56 MHz

B. Differences to other technolgies like barcodes, QR codes and fiducials

Barcodes, QR (Quick Response) codes and fiducials are visual codes printed on paper or other material. They have one characteristic in common, they all need a line of sight to be identified. In many cases a conventional camera (webcam, smartphone cam, etc.) is used to identify these optical 1D or 2D codes.

A line of sight is not necessary with RFID tags. They can be directly built into different objects, for example in 3D printed bones. Therefore, various tangible objects with an invisible identification feature can be created.

Fig. 4. 3D printed bone with RFID tag inside

In contrast to RFID tags, barcodes, QR codes and fiducials are not unique; it is possible to print endless copies of them.

RFID tags can store more information than just the UID. Readers are not only able to read this data, but can also save new data on a RFID tag. However, readers only know that one or more RFID tags are near the antenna, but there is no information about the distance and so forth. Another challenge dealing with RFID is that metal and power supply lines can interfere the signal.

In contrary, fiducials additionally allow tracking the position and the rotation of an object. The downside of this technology is that the camera needs a stable light source to recognize the fiducials.

Our team uses all these technologies in different projects, but the overall experience shows, that the RFID technology offers one of the most reliable identification methods.

III. XML SOCKETS

We developed several XML socket applications to ensure the communication between the exhibition software and the API of different RFID readers. The XML sockets deliver information about the reader and the identified tags with a certain XML declaration. This allows us to switch to a different brand of RFID reader with no need to adapt the exhibition software. Furthermore, it is possible to connect multiple readers to a personal computer at the same time. This allows us to identify the position of the tags (for example near reader A or reader B).

IV. Projects

For more than 10 years, the DMT research group implemented the RFID technology in several projects. This section shows a selection of exhibits, which use RFID as a user interface.

A. Projects with stationary readers

In the following exhibits, we integrated one or more fixed readers and antennas, which cannot be displaced by the visitors. Users interact with the exhibit by moving the transponders. In these examples, we mostly use one or more ID ISC.MR102 readers and ID ISC.ANT310/310 antennas from Feig[4] which operate on 13.56 MHz (ISO15693).

1) Paint Walls: At this exhibit, walls can be virtually painted with a selected color. Multiple RFID readers are hidden in the paint buckets and walls, the brushes are equipped with RFID tags. For the paint buckets, we used smaller readers and antennas. The exhibit was developed in cooperation with the Graz Children's Museum FRida & freD for the travelling exhibition "Archinature".

Fig. 5. "Paint Walls" with RFID antennas behind the walls

2) Building A House: This exhibit is another development in cooperation with the Graz Children's Museum FRida & freD for the travelling exhibition "Archinature". Children get to know the process of building a house by placing the required tools and machines (which are equipped with RFID tags) on a property (with a RFID reader mounted beneath).

Fig. 6. "Building A House" with tangible objects

[4] http://www.feig.de/

3) The History of Writing: For the traveling exhibition "The Inventive Geniuses" we develop in cooperation with the Graz Children's Museum FRida & freD a book which tells about the history of writing. The book's pages are equipped with RFID tags and two RFID readers are placed beneath the bookshelf. With this information, the animated content of the book is projected on the blank pages.

Fig. 7. "The History of Writing" with RFID tags inside the animated book

4) Inventory: Children can recreate themselves with this exhibit, which was developed in cooperation with the Graz Children's Museum FRida & freD for the exhibition "My Family". RFID cards represent different body parts, clothes and colors.

Fig. 8. "Inventory" with RFID cards

5) Medical Doll: The goal of this project is to inform visitors about different medical child diseases and their treatments. A doll in the size of a child lies on an examining table with a large display. The doll is equipped with RFID antennas and responds to medical items (with build in RFID tags), like stethoscope, otoscope or magnifier. Feedback is given with a large LCD screen and also directly with a projection on the doll itself. This exhibit was developed in cooperation with the Graz Children's Museum FRida & freD.

Fig. 9. "Medical Doll" with build in RFID antennas

6) Effective Nature Conservation: This exhibit explains various tools of nature conservation. Wodden objects, which represent the tools, are equipped with RFID tags. RFID readers are mounted beneath the information- and the tool-pad. Visitors not only receive information about the tools, but they can also apply them in different scenarios. We developed this multimedia application in cooperation with the Styrian Nature Parks[5] for the exhibtion "Nature in Human Hands".

Fig. 10. "Effective Nature Conservation" with two RFID readers

7) Boundaries of Mobility: Visitors play a board game, which is equipped with multiple RFID tags (inside the pieces) and readers (beneath the board). During the game, videoclips are shown depending on the pieces' position on the board.

Fig. 11. "Boundaries of Mobility" with multiple RFID readers

B. Projects with moving readers

In contrast to the projects above, in the following cases the RFID reader is moving, while the tags are static.

1) Graz Fairy Tale Train: While the Graz Fairy Tale[6] train is on its tracks, audio is automatically played and interactive stations are switched on by default. The two locomotives are equipped with RFID readers and RFID tags are mounted along the trail. The whole timing can easily be changed by relocating the RFID tags or adapting the configuration files. The installed mid range readers have a antenna with the size of 30 x 30 cm and the reading range covers about a 50 cm radius around the antennas. The readers are mounted in a heigh of 20-30 cm above the tags. In order to increase the accuracy of the identification, we use 3-4 tags per station. However, if the trains are going too fast, the tags cannot be identified by the readers, because it is not possible to communicate with them in such a short time. The Graz Fairy Tale Train was realized in cooperation with the Graz Children's Museum FRida & freD.

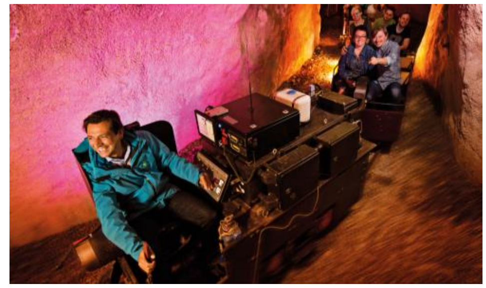


Fig. 12. "Graz Fairy Tale Train" (© GMB – Hannes Loske)

2) Chrismas All Around The World: Christmas is celebrated different across the globe. RFID tags are hidden in a globe behind every capitol city. By pointing on the city with a pen-like RFID reader the visitors receive information about the Christmas traditions in the chosen country. This exhibit is current in development in cooperation with the cultural club "Blaues Fenster". In this application, we can identify spots on the interactive globe within 2x2 mm^2.

Fig. 13. Christmas All Around The World (demonstrator)

3) Coin Magnifier: Visitors use magnifiers to receive additional information about coins in the Coin Cabinet of the Universalmuseum Joanneum in the Eggenberg Palace in Graz[7]. Magnifying glasses, which are common in coin cabinets, are used to take a closer look at the coins, but they also act as an interface with the invisible computer. RFID antennas are built into the magnifiers and the RFID tags are placed beneath the coins. Therefore, customized antennas have been designed and built inside the magnifiers. This allows to show visitors further information on the screens without an additional interaction.

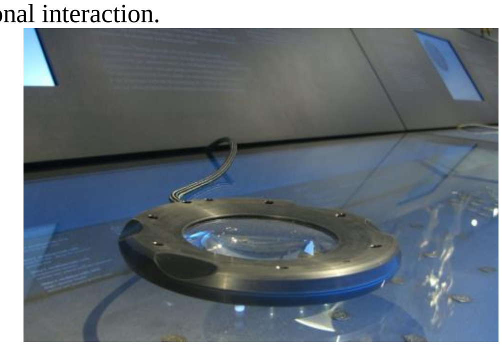


Fig. 14. Coin Magnifier with integrated antenna

[5] http://www.naturparke-steiermark.at/
[6] http://www.grazermaerchenbahn.at/
[7] https://www.museum-joanneum.at/muenzkabinett

V. Conclusion

The presented projects show that RFID technology can be applied in various use cases. The users interact with objects, without the need to know anything about the technology behind it. This allows an innovative and intuitive hands-on interaction to enhance digital media experience of exhibits in museums.

References

[1] M. Patchett, A. Flack, "RFID in Museums and Galleries: A Review", 2014, http://old.react-hub.org.uk/objects-sandbox/projects/2014/curpanion/journal/rfid-in-museums-and-galleries-a-review/ (visited on 2017-09-14)

[2] Kinser, Owen, "Passive Radio Frequency Identification (RFID) and Museum Collections", 2016. Master's Projects, http://repository.usfca.edu/capstone/463 (visited on 2017-09-14)

[3] Wang, Hong-You & Liu, Gi-Zen & Hwang, Gwo-Jen, „Integrating socio-cultural contexts and location-based systems for ubiquitous language learning in museums: A state of the art review of 2009-2014" , 2017. British Journal of Educational Technology. 48. 653-671.

[4] T. Kälin, L. Müller, "Auswertung von Besucherbewegungen in Museen mittels RFID", 2009, Wireless Communication and Information, 217-230.

[5] S.H. Hashemi, J. Kamps, W. Hupperetz, "Busy versus Empty Museums: Effects of Visitors' Crowd on Users' Behaviors in Smart Museums", 2017, Proceeding UMAP '17 Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, 333-334

[6] A. Fevgas, N. Fraggogiannis, P. Tsompanopoulou, P. Bozanis, "The iMuse Virtual Museum: towards a cultural education platform", 2014, The 5th International Conference on Information, Intelligence, Systems and Applications, Chania, 2014, 171-175

[7] A. Fevgas, P. Tsompanopoulou, A. Tsiovoulos, G. Drasidis, P. Bozanis, "Utilizing UHF RFIDs to Enhance Museum Visiting Experience", 2010 14th Panhellenic Conference on Informatics, Tripoli, 16-20

[8] M.A.Khan, M. Sharma, B. Prabhu, "A Survey of RFID Tags", 2009, International Journal of Recent Trends in Engineering, Vol 1, No. 4, May 2009

[9] T. Hassan, S. Chatterjee, "A Taxonomy for RFID," 2006, Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), 184b-184b.

[10] C. Xiao, N. Chen, D. Li, Y. Lv, J. Gong, "SCRMS: An RFID and Sensor Web-Enabled Smart Cultural Relics Management System", 2017, Sensors, Volume 17, Issue 1, doi:10.3390/s17010060

[11] S. Hsi, H. Fait, "RFID enhances visitors' museum experience at the Exploratorium", 2005, Commun. ACM. 48. 60-65. 10.1145/1081992.1082021.

[12] M Buzzi, C. Senette, N. Masini, F. Soldovieri, „RFID Sensors and Artifact Tracking", 2017, Sensing the Past: From artifact to historical site, Springer International Publishing, 435-451

[13] R. Strohmaier, G. Sprung, A. Nischelwitzer, S. Schadenbauer. "Using visitor-flow visualization to improve visitor experience in museums and exhibitions." MW2015: Museums and the Web 2015. Published January 15, 2015. Consulted September 7, 2017.

[14] NXP, "ICODE", https://www.nxp.com/products/identification-and-security/smart-label-and-tag-ics/icode:MC_42024 (visited on 2017-09-14)

ISIDOR – Ein auditiver HCI-Prototyp

Josefine Riedel, Thomas Böck, Julian Fischer, Felix Rauchwarter

St. Pölten University of Applied Sciences
E-Mail: dm161537@fhstp.ac.at

***Abstract*—The work shown engages with the conception and realization of the human-computer interaction (HCI-) sound installation 'Isidor' which gives the user the possibility to conduct a virtual string quartet. The control system appears via a leap motion camera (motion tracking) and a clearly arranged user interface (GUI) implemented on a touch screen. Emanating from the GUI parameters as volume, soli and mutes, room impressions, tonal pitches (Hz), moods (emotional), the changing of instrumental player positions and the punctual launch from a certain position are possible. The leap motion camera acts like a translator from the conductor movements of the user's digital workable control data. Therefor working in real time is essential for tempo changes. The Max/MSP-object Bachscore is accountable for the score implementation. MIDI data is transferred to VST-instruments which eventually are responsible for the acoustic sound. Max/MSP patches connect the different elements: Leap motion tracking, DMX light control, graphics and Bachscore to an interactive sound installation.**

***Zusammenfassung*—Die vorliegende Arbeit beschäftigt sich mit der Konzeption und Umsetzung der Human-Computer Interaction (HCI)-Klanginstallation "Isidor". Sie ermöglicht es Nutzern, ein virtuelles Streichquartett zu steuern. Der Benutzer interagiert mittels Motion Tracking der Hnde durch eine Leap Motion Kamera sowie einer übersichtlich gestalteten Nutzeroberfläche auf einem Touchbildschirm. Ausgehend von dieser lassen sich Parameter wie Lautstärke, Besetzung, Raumeindrücke, tonale Stimmung (in Hz), Gefühlszustände ("Moods"), Positionsänderungen der Instrumentalisten, Artikulation jedes Spielers sowie das punktuelle Abspielen an einer gewünschten Position umsetzen. Die Leap Motion übersetzt Dirigatsgesten des Nutzers in Tempoänderungen, die in Echtzeit generiert werden. Mittels der Max/MSP-Library "Bach" wird die Notendarstellung auf der grafischen Benutzeroberfläche angezeigt. Die Klangerzeugung geschieht durch mehrere VST-Instrumente. Max/MSP Patches vernetzen die verschiedenen Elemente (Leap Motion Tracking, DMX Lichtsteuerung, Grafik und Bachscore) miteinander zu einem interaktiven Klangerlebnis.**

I. EINLEITUNG

Unter "Sonic Gestures" versteht man Bewegungen oder Mimiken des menschlichen Körpers, die Auswirkungen auf das Klangbild haben oder sogar Klänge erzeugen können. 'Sonic' bedeutet '(akustischer) Schall', während 'Gesture' auf die menschliche Komponente, die Körper- bzw. Ausdrucksbewegung hinweist [1]. Das grundsätzliche Prinzip von Sonic Gestures wird aktuell in Form von Klanginstallationen in vielen Museen aufgegriffen. Durch das in-Kontakt-treten mit den technischen Installationen steigt das Interesse des Museumsbesuchers und das subjektive Gefühl der Erlebbarkeit der Installation. Die Folge ist eine unterbewusste und doch geradlinige Ansprache zwischen Rezipient und wissenschaftlicher Forschungsquelle. Es kann abgeleitet werden, dass eine direkte Interaktion die Aufnahme- und Lernfähigkeit sowie das kognitive Gedächtnis fördert [2]. Von wissenschaftlicher Seite wird dieser Ansatz Sound and Music Computing (SMC) genannt. Dieses Versuchsfeld hat die Prämisse, die Gesamtheit der Klang- und Musikkommunikationskette von einem multidisziplinärem Standpunkt aus zu betrachten. Wissenschaftliche, technologische und künstlerische Methoden sollen zum Ziel des Verstehens, Generierens und Modellierens von akustischen Klängen durch computerbasierte Techniken beitragen [3].

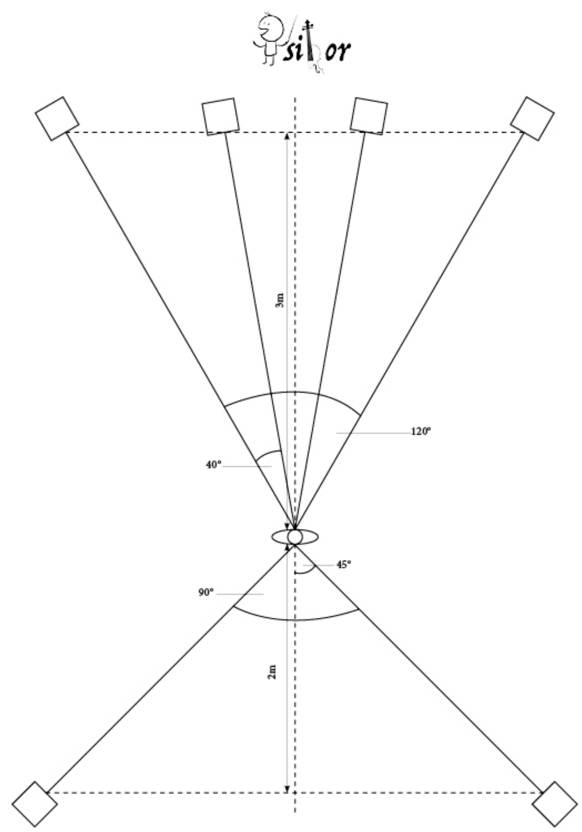


Abbildung 1. Isidor Setup

Mit Isidor soll dieser Ansatz des spielerischen Lernens aufgegriffen und erweitert werden. Die Erklärung des Namens unterstreicht den Grundgedanken der Forschung: 'Easy DAW' ("Digital Audio Workstation") verbalisiert 'Isidor'. Das Programm bietet einem breitgefächertem Publikum die Möglichkeit, direkten Einfluss auf die Darbietung eines Streichquartetts auszuüben. Die interaktive Installation setzt sich seitens der Hardware aus einem übersichtlichen User Interface auf einem Touchbildschirm, sechs Lautsprechern (vier Frontkanäle vor dem Rezipienten und zwei Rearkanäle hinter demselben zur Unterstützung der räumlichen Darstellung), einer Leap Motion Tracking Kamera (für die Tempoübersetzung des Dirigats) und via DMX gesteuerten Scheinwerfern (zur Unterstreichung der Tempoinformationen und Atmosphäre) zusammen.

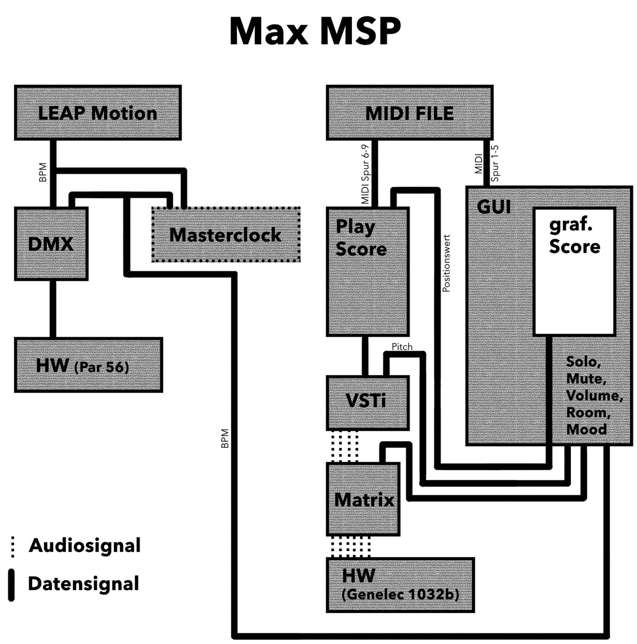


Abbildung 2. Technischer Aufbau Isidor

Softwareseitig wird eine vorprogrammierte MIDI-Datei mittels dem Max/MSP-Objekt Bachscore interpretiert, in Partiturdarstellung angezeigt und an VST Instrumente weitergeleitet, die Klänge generieren. Die Leap Motion Kamera erfasst dirigatsähnliche Handbewegungen des Nutzers, welche in einem separaten Max/MSP-Patch in Beats Per Minute (BPM) umgerechnet werden. An diesen Prozess ist auch die DMX-Steuerung angebunden, die durch farbiges Licht zusätzlich zum Hörerlebnis ein optisches Feedback über die aktuelle Geschwindigkeit des Stückes gibt und, in Folge dessen, die Stimmung intensiviert und das Nutzererlebnis um eine visuelle Ebene erweitert. Den Großteil der Steuerungen übernimmt der Rezipient über das Graphical User Interface (GUI). Ein zentraler Hauptbildschirm ohne Subscreens ermöglicht schnellen und intuitiven Zugriff mittels Touch-Steuerung auf folgende Funktionen: Transport (Play, Pause), Ausdruck bzw. Lautstärke jedes Instruments (gekoppelt an Note-On Expression), Mute-Funktion für jeden Spieler, Raumeindrücke (Studio/Dry, Small Hall, Big Hall), tonale Stimmung (in Hz), sogenannte "Moods" (bspw. wütend, verschlafen, vorsichtig), Positionsänderungen der Instrumentalisten sowie das punktuelle Abspielen ab einer gewünschten Position durch "Tap" mit dem Finger in die Partiturdarstellung.

II. State of the Art

Bereits 2001 wurde die Klanginstallation "Personal Orchestra" in Wien aufgestellt. Sie ist im 'Haus der Musik' stationiert, einem Zentrum der musikalischen Kulturbewahrung und -vermittlung. Die Herren Borchers, Samminger und Mühlhäuser behaupten mit 'Personal Orchestra' die erste Klanginstallation geschaffen zu haben, die es dem Nutzer ermöglicht, eine Audio-/Videoaufnahme in Echtzeit dirigieren zu können. Bei dem verwendeten Material handelt es sich nicht, wie bei Isidor, um verarbeitete MIDI-Daten, sondern um eine "echte" Tonaufnahme der Wiener Philharmoniker. Die Beeinflussung von Tempo, Lautstärke und Instrumentation ist durch Motion Tracking einer Infrarotkamera ohne hörbare Artefakte, die durch Time Stretching entstehen können, möglich. Hierbei sei jedoch angemerkt, dass die Tempovariation nur in einem eingeschränkten Rahmen umsetzbar ist. Wie bereits erwähnt, wurden alle Stücke aufgenommen und mittels Timestretching offline in verschiedenen Tempi gerendert. Bei einer Tempoänderung durch den Nutzer werden die passenden Audiodateien abgerufen und überblendet. Überschreitet ein Nutzer die Grenzen des geringsten oder höchsten vorgerenderten Tempos, hält die Wiedergabe an und der virtuelle Konzertmeister beschwert sich humoristisch über die Anstrengungen des Dirigenten. Ein weiterer Unterschied zu Isidor besteht darin, dass auch Videodaten abgespielt werden. Diese sind, wie das Audiomaterial, vorher aufgezeichnet und werden in Echtzeit überblendet [4].

Inspiration für die vorliegende Arbeit wurde weiters in der Klanginstallation "Effektorium" gefunden. Die seit 2014 im Leipziger Mendelssohn-Bartholdy Museum beheimatete Installation erlaubt es Nutzern ebenfalls, in die Rolle eines Dirigenten zu schlüpfen und ein virtuelles Orchester anzuführen. Säulenförmige Lautsprecher mit Display stehen stellvertretend für ausgewählte Musikergruppen im Raum. Die Leap Motion Kamera erfasst die Dirigatsgesten, deren ausgewertete Informationen mittels eines Max/MSP-Patches in Verbindung zu Ableton Live stehen. Die Audio-Workstation beheimatet Audiodateien, die in Echtzeit durch Timestretching auf das vorgegebene Tempo gedehnt bzw. gestaucht werden [5], [6].

Auch im Alltag geschieht in jüngerer Vergangenheit vermehrt Datengewinnung und -verarbeitung via Kinect oder Leap Motion-Technik. Kieran Flay widmete sich 2015 einer Gegenüberstellung dieser zwei Arbeitsmittel in Bezug auf die Verbindung von Gestikverarbeitung und digitalen Audiosystemen. Flay vergleicht die Technologien vor allem in Bezug auf die Parameter Durchführbarkeit, Nutzen und Funktionalität. Als weitere Option für gestikgesteuerte Systeme wird auf Smart TVs (Lautstärkesteuerung, Schnittstellenwechsel, Menü- bzw. Browsernavigation) oder Remidi Gloves (ein Handschuh mit Bluetooth-MIDI-Steuerung) verwiesen [7].

III. Methodik

Ziel des Projektes ist es, eine Installation zu gestalten, die dem Benutzer ein subjektives Kontrollgefühl verleiht und verständlich sowie intuitiv handhabbar ist. Um einen Einblick in den Stand der Forschung und ähnliche Projekte zu erhalten, wurde eine State Of The Art-Recherche durchgeführt. Die Ergebnisse gaben einen Rahmen vor, in dem sich das Projekt bewegen würde. Anschließend wählten die Autoren Technologien und Interfaces aus, welche die Umsetzung möglich machen. Aus dem weiteren Arbeitsfortschritt entstand ein Prototyp, der im Rahmen einer Ausstellung aufgebaut und von Besuchern ausprobiert werden konnte. Das verbale Feedback der Probanden wurde im Nachhinein eingearbeitet und das Installationsdesign dadurch optimiert.

IV. Signalfluss

Das vorliegende Projekt wurde zur Gänze in Max/MSP entwickelt. Herausgebracht von Cycling74 ist MAX eine graphische Entwicklungsoberfläche, mit der vornehmlich akustische und audiogenerative Projekte umgesetzt werden können. Zur grafischen Notendarstellung wurde die frei zugängliche, externe Max/MSP-Library “Bach” einbezogen[1]. BachScore ist ein Freeware Programm, dass von den Komponisten Andrea Agostini und Daniele Ghisi entwickelt wurde. Es handelt sich dabei um ein umfangreiches, traditionelles Notationssystem für Max/MSP. Da der Hauptfokus auf der klassischen Notation, Analyse und Komposition innerhalb der Programmierumgebung liegt, ist eine Anwendung als grafische Partitur nach Aussage der Entwickler eher unüblich und entsprechend (noch) nicht im vollem Maße implementiert.

Der strukturelle Aufbau (siehe Abb. 2) lässt sich am einfachsten anhand der vorprogrammierten MIDI-Datei verfolgen. Die ersten fünf Spuren des Files werden per Bach.Score grafisch im GUI dargestellt, während die Spuren sechs bis neun akustisch wiedergegeben werden. Jegliche Interaktion mit der graphisch angezeigten Partitur sendet wiederum Informationen an den wiedergebenden Score. Um akustische Informationen ausgeben zu können, werden die MIDI-Steuerdaten an einen Sampler (in diesem Falle Native Instruments’ KONTAKT-Factory Library) mit den vier Software Instrumenten (VSTi: Standard Soloinstrumente Violine, Viola und Cello) weitergegeben, wodurch vier Audiosignale generiert werden.

Diese VSTi können eine gemeinsame Änderung der akustischen Stimmung (Pitch: Standard = Kammerton a bei 440Hz) aus einem Wahlmen auf der graphischen Benutzeroberfläche heraus erfahren. Die vier Audiostreams durchlaufen eine Folge von Matrizen, um sie auf die sechs vorhandenen, physikalischen Ausgänge zu routen. Diese Matrizen können wiederum vom Nutzer durch die GUI beeinflusst werden (Mute-Funktion, Lautstärke jeder Stimme, Verräumlichung, etc.). Zum besseren Verständnis ist die dynamische Tempogestaltung separat zu betrachten. Ausgangspunkt hierfür ist das Motion Tracking der Leap-Motion-Kamera. Ihre ausgegebenen Werte werden auf BPM-Werte umgerechnet und dann auf einen Multiplikator für die Masterclock skaliert (genaue Besprechung an späterer Stelle).

Der BPM-Wert wird auch an die DMX-Steuerung weitergegeben. Dadurch wird die Farbwahl der Scheinwerfer der Installation bestimmt. Neben der Temposteuerung durch die Leap Motion Kamera können auch direkt über die GUI die sogenannten “Moods”, voreingestellte Kombinationen aus Tempo, Lautstärke und Artikulation, ausgewählt werden.

V. Hard- und Softwarespezifikationen zur Umsetzung

A. Audiosignalfluss

1) Direktschall: Die vier eigenständigen Inputs (V1, V2, Va, Vc) werden zuallererst an alle vier Frontkanäle gesendet. Dort gelangen sie in eine Matrix, die je nach Eingangswert (1-4) genau ein Signal ausgibt. Diese Eingangswerte sind voreingestellt und können dann in der GUI über ein abfragendes Array getriggert werden (genaue Besprechung folgt), wonach jeweils nur ein Wert an genau einer Position vorkommen kann. Die Stimmen werden also in ihrer Position direkt getauscht.

2) Diffusschall: Für jeden Frontkanal existiert eine eigene AUX-Schiene, die in ein Reverb-Plugin speist. Das Ausgangssignal jedes Frontkanals wird dort hinein gesendet und die Signale summiert, ehe sie verhallt werden. Dabei werden die Eingänge nach Position gewichtet (Bsp.: Linker Kanal im linken ReverbMix deutlich stärker vertreten als der Rechte Kanal; die Abstufung erfolgt in jeweils 3dB Schritten). Das generierte Summensignal wird abgegriffen und durch das in Max/MSP integrierte algorithmische Hall-PlugIn ‘Gigaverb’ für jeden Kanal einzeln prozessiert. Das Raumsignal wird abschließend mit dem trockenen Direktschall-Signal summiert (Diffusschall -6dB zu Direktschall) und an den physikalischen Output weitergeleitet.

In den rückwertigen Kanälen (“Rear”), die zur besseren Simulation eines räumlichen Hörerlebnisses benötigt werden, kommen die Ausgangssignale der Vorderkanäle, also die ursprünglichen Inputs, an. Diese werden miteinander summiert, wobei die beiden diagonal gegenüberliegenden Kanäle je um 3dB abgeschwächt werden. Die Verräumlichung erfolgt ebenso über Gigaverb und entspricht immer exakt demselben Preset der Verräumlichung der Frontkanäle.

Um der Ausbreitungsgeschwindigkeit von Schall im Raum Rechnung zu tragen, werden die Outputsignale der Rearkanäle zeitlich verzögert. Da die Verräumlichung durch Gigaverb algorithmisch geschieht, wird der Zeitfaktor des Delays der Rearkanäle an die jeweilige Nachhallzeit des künstlichen Raumes angepasst. Idealerweise sollte die Delayzeit dabei den Zeitraum, in dem beim Nutzer die reale Erstreflektion der Frontsignale an der rückwärtig raumabschließenden Wand eintrifft, nicht übersteigen.

[1] http://www.bachproject.net, 04.09.2017

Zu beachten bleibt außerdem, dass trocken aufgenommene VST-Instrumente das Klangerlebnis hinsichtlich Ortbarkeit und Klangqualität bedeutend verbessern.

3) Mute-Funktion: Wichtig für das Verständnis der Mute-Schaltung ist das Verhalten der wechselnden Positionierung der Eingangssignale (V1, V2 etc.), denen die statischen Kanäle 1-4 gegenübergestellt sind. Um dem Nutzer das Stummschalten einzelner Instrumente, egal an welcher Position, zu ermöglichen, müssen die Eingangssignale bereits gekappt werden, bevor sie einem Kanal zugewiesen werden. Dazu durchläuft jedes der vier Inputsignale vor dem Kanalrouting eine weitere Matrix. Diese hat jeweils zwei Inputkanäle, die sich zur einfachen Verständlichkeit als Hot und Cold bezeichnen lassen. So liegt an einem Input der Matrix das jeweilige Inputsignal (bspw. V1, V2) an, am Anderen kein Signal ("Stumm").

Nachdem die Verräumlichung aller AUX-Schienen durch die Ausgangssignale der Frontkanäle, also erst zu einem späteren Zeitpunkt im Routing, erfolgt, ist also sichergestellt, dass das Stummschalten eines jeweiligen Instrumentes (= Inputsignales) auch im Raumsignal, sowie den Rearkanälen stattfindet.

4) Tempo: Prinzipiell können in Bachscore Tempoänderungen implementiert werden. Problem dabei ist, dass diese Änderungen bereits im Vorfeld bzw. vor dem Abspielen der jeweiligen Position definiert werden müssen. Dabei kann das Tempo zu jedem Taktbeginn per Notenwert in BPM ("Viertelnoten in 120 BPM") definiert werden, wobei eine fließende Änderung des Tempos zum nächsten Takt per accelerando und ritardando möglich ist. Da das Konzept von Isidor jedoch verlangt, Tempoänderungen auch in Echtzeit zu ermöglichen, wird die Masterclock angegriffen, welche die Länge eines Taktes in Samples angibt und somit einen absoluten Zeitwert bietet, der moduliert werden kann.

Dabei muss die aktuelle Geschwindigkeit des Stückes durch die gewünschte (neue, durch das Dirigat bestimmte) BPM-Zahl dividiert werden, um den korrekten Multiplikator für die Masterclock zu erhalten (Bsp: 130 BPM Originaltempo zu 130 BPM anliegendes Tempo = Multiplikator 1).

5) Pitch: Das Streichquartett kann mittels einer Master-Kontrolle (d.h. gültig für alle vier Instrumente) in der Stimmung hinsichtlich des Referenztons variiert werden. Zur Auswahl stehen die internationale Standardstimmung auf 440Hz, die historische Stimmung auf 415Hz, die klassische Stimmung auf 430Hz und die Wiener Stimmung auf 466Hz. Die Änderung der Tonhöhe geschieht durch die Pitch Bend-Funktion des VST-Instrumentes. Problematisch dabei ist die Umrechnung von Hertz auf Cent auf linearer Basis, da sich die Tonhöhe logarithmisch verhält. Näherungsweise lässt sich für den relevanten Frequenzbereich um 440Hz ein Verhältnis von rund +/- 4Cent pro einem Hertz Verstimmung annehmen [8].

6) Darstellung vs. Klang in Midi: Beim Umgang mit Midinoten ist festzustellen, dass eine strikt notationsgetreue MIDI-Programmierung für die klassische Notendarstellung (Partitur) unumgänglich ist, die akustische Wiedergabe dieser Programmierung allerdings üblicherweise als sehr statisch und unrealistisch wahrgenommen wird (beispielsweise ergeben überlappende Midinoten einen flüssigeren Übergang und damit einen besseren Legato-Eindruck).

Aus diesem Grund wird ein Midifile verwendet, das aus insgesamt neun Einzelspuren besteht. Die ersten vier Spuren entsprechen einer exakten, notationsgetreuen Programmierung für die visuelle Darstellung jeder einzelnen Stimme. Die fünfte Spur ist für die Darstellungsbreite des Partiturfensters notwendig, worauf im Punkt "Darstellung der Partitur" nochmals konkret Bezug genommen wird. Die verbleibenden vier Spuren werden an den Sampler weitergeleitet und akustisch umgesetzt.

Darstellung der Partitur: Eine seitenweise Darstellung mittels Bach.Score ist zum aktuellen Zeitpunkt (noch) nicht möglich. Als Workaround werden daher in Isidor vier Bach.Score-Instanzen genutzt. Die Instanz, die tatsächlich abgespielt wird und Daten an den Sampler weitergibt, arbeitet versteckt im Hintergrund. Drei weitere Instanzen bilden untereinander positioniert die in der GUI sichtbare Partitur. Jede der drei Instanzen entspricht dabei einem Notensystem mit je 2 Takten. Entsprechend der eben diskutierten Umstände werden nur die ersten fünf Spuren grafisch wiedergegeben.

Prinzipiell folgt die Ansicht in Bach.Score immer dem Playmarker, also der aktuellen Abspielposition. Nachdem die grafischen Instanzen allerdings als passiv ("nicht abspielend") verstanden werden können, wird das händische Umblättern möglich.

Bei Bach.Score-Instanzen kann die Position des Playmarkers über die Message inscreenpos X.Y bestimmt werden, wobei X den Schlag angibt; der Taktbeginn entspricht dem X-Wert 0, während Y die Taktzahl angibt. Da insgesamt sechs Takte gleichzeitig zu sehen sind (zwei pro Instanz), werden bei jeder anliegenden grafischen Score-Instanz beim Blättern je 6 Takte zur aktuellen Position dazu addiert. Demnach springt die oberste Instanz von Takt 1 zu 7, von 7 zu 13, ... während die mittlere Instanz von 3 zu 9, von 9 zu 15, ... und das unterste System von Takt 5 zu 11, 11 zu 17, ... springt. Damit nicht über den Start- und Endpunkt hinaus geblättert werden kann, wurden für die Positionswerte per Clip-Objekt Grenzen definiert. Sind diese Werte erreicht, kann keine weitere Addition bzw. Subtraktion erfolgen.

Damit dieses Prinzip eines beständigen Sprunges der Ansichtsposition um sechs Takte überhaupt möglich ist, muss sichergestellt werden, dass pro Score-Instanz (also Notensystem) immer genau gleich viele Takte dargestellt werden und diese, auch aus grafisch-ästhetischen Aspekten, stets gleich groß sind. Die dargestellte Breite eines Taktes lässt sich in Bach per Default nicht einstellen. Dadurch erklärt sich die Notwendigkeit einer "Ansichtsspur" (Spur 5) in der programmierten MIDI-Datei. Diese Spur enthält einen - im Vergleich zum restlichen Notenmaterial - sehr geringen Notenwert, der konstant wiedergegeben wird. Im vorliegenden Fall entspricht dies einem Notenwert von zweiunddreißigstel Noten. Die benötigte Breite zur graphischen Darstellung dieser zweiunddreißig Einzelnoten, welche kontinuierlich ist, ermöglicht somit das Festlegen einer generellen Taktbreite. Da diese Spur einzig und allein zu diesem Zweck benötigt wird, ist sie im GUI durch ein weißes Viereck überdeckt, um dem User nur die vier hörbaren Instrumente im Notensystem anzuzeigen.

Das Springen zu einer Position durch Touch-Befehl auf den jeweiligen Takt basiert auf demselben Prinzip wie das Blättern. Wiederum werden “Inscreenposition”-Werte ausgegeben. Diese werden an der jeweiligen grafischen Score-Instanz ausgelesen und an die abspielende Instanz weitergegeben, die dadurch direkt zum entsprechenden Takt springt und die Wiedergabe an der gewählten Position startet.

B. Grafik und Design

Aus verschiedenen Entwürfen für die grafische Benutzeroberfläche wurde eine Lösung ausgewählt, die auf drei Spalten basiert: Kontrollen für alle Instrumente, Kontrollen für Einzelinstrumente und Kontrollen für die Partiturdarstellung.

Abbildung 3. Grafisches User Interface von Isidor

In den Kontrollen für Einzelinstrumente findet der Benutzer eine bildliche Darstellung des Instruments, dessen Namen, einen Fader für die Lautstärke des Instruments sowie einen Button um die Stimme stumm zu schalten. Auf der linken Seite der Instrumentenabbildung gibt es außerdem ein Wahlmen, in dem die Artikulation des Spielers (Legato, Staccato, Pizzicato, Tremolo) ausgesucht werden kann.

Hierbei sei angemerkt, dass der Volumefader nicht den Pegel des Ausgangssignals moduliert, sondern den MIDI-Control- Change Expression des VSTi ansteuert. Dementsprechend wird nicht nur beispielsweise die Lautstärke verringert, sondern auch die musikalische Artikulation angepasst (geringere Expression führt zu geringerem Druck des Bogens auf die Saiten, was in einem weicheren Klangbild resultiert), wodurch ein deutlich realistischeres und natürlicheres Klangbild erzeugt wird.

Von einer zuerst angedachten Lösung, über das Bild des Instruments in einen Subscreen mit erwähnten Kontrollen zu gelangen, wurde abgesehen, um die Kontrollelemente auf einem Blick ersichtlich zu behalten und eine Verschachtelung des Interfaces zu vermeiden. So soll es dem User auch einfacher gemacht werden, die Lautstärkenverhältnisse der Stimmen im Blick zu behalten und eine mögliche Quelle für Verwirrungen (bspw. versteckte Mute-Buttons) ausgeschlossen werden.

Eine weitere Entscheidungsgrundlage für die Anordnung ohne Subscreens war im Lauf der Entwicklung der Installation das Festlegen des verwendeten Touch-Bildschirmes. Hier sei erwähnt, dass viele dieser Entscheidungen bei der Umlegung auf andere Systeme und Kontrolleinheiten (bspw. Smartphone, Tablet) überdacht bzw. angepasst werden könnten.

In den Kontrollen für alle Instrumente (Room, Moods, Pitch) sind die Auswahlmenüs mit Toggle-Buttons ausgeführt. Interessanter ist die Positionsauswahl der Instrumente; hier gibt es eine Matrix, in der man für jede Stimme zwischen Links, Halblinks, Halbrechts und Rechts entscheiden kann. Beim Umsetzen einer Stimme tauscht diese den Platz mit dem Instrument an der jeweils neuen Position. Ursprünglich war angedacht, die Positionen mit 4 Monitoren, die jeweils unter den Front-Lautsprechern stehen, optisch darzustellen. Davon wurde abgesehen und stattdessen eine kleine, bildliche Darstellung jedes Instrumentes über der Positionsauswahl-Matrix eingefügt, die der Sitzordnung folgt. Hierbei war der Gedanke, dem Benutzer, wenn er nicht auf die Kontrolleinheit sieht, so wenig optische Reize wie möglich zu geben, damit der Fokus nicht vom Hörerlebnis abgelenkt wird.

Für die DMX-Steuerung wurde ein Interface der Firma “Enttec” (DMX USB Pro) verwendet, welches eine Kommunikation via USB ermöglicht. Dazugehörig wurde das Max/MSP-Objekt dmxusbpro der Firma Nullmedium[2] benutzt, um direkt aus Max/MSP Daten schicken zu können. Die Lichtsteuerung folgt, wie Eingangs erwähnt, dem aktuellen Abspieltempo des Stückes. Die BPM-Zahl wird abgegriffen und je nach deren Wert in Bereichen von 20 BPM einer Farbe zugeordnet. Die Grenzen liegen nach unten hin bei 30 BPM, nach oben hin werden alle Werte über 180 BPM derselben Farbe zugeordnet. Die Übergange zwischen den Farbbereichen geschehen fließend innerhalb von zwei Sekunden.

Wie bereits erwähnt, geschieht sämtliche grafische Umsetzung innerhalb von Max/MSP. Als Bildschirm und Kontrolloberfläche dient ein 24-Touch LCD der Firma “Ilyama” (ProLite B2483HS-B1).

C. Leap Motion

Durch den konstanten, technologischen Fortschritt entstehen laufend neue Wege zur Human-Computer-Interaction. Die auf diese Art entwickelten Schnittstellen eröffnen dem Nutzer viele Möglichkeiten. So verschieden die Technologien, so verschieden sind auch die Steuerungsdaten, welche für die Interaktion verwendet werden können. Zusätzlich zu haptischer

[2]http://www.nullmedium.de, 04.09.2017

und Sprachsteuerung können auch Kamerasysteme Daten ermitteln. Zwei Produkte, die oft für das Tracking von Körper- und Bewegungsdaten benutzt werden, sind die Xbox Kinect und die Leap Motion Kamera. Nach ausführlichen Vergleichen der Spezifikationen der zwei Kameras wurde aufgrund der besseren Einsatzfähigkeit im vorliegenden Kontext die Leap Motion Kamera verwendet.

Die Leap Motion Kamera kam 2013 mit dem Ziel auf den Markt, kleinste Bewegungen in sehr geringer Entfernung erfassen zu können. Während die Xbox Kinect Daten in einem sehr weiten Bereich von 0,5 - 3,5m abgreifen kann, ist der verwendbare Radius der Leap Motion mit 0,25 - 0,5m viel kleiner. Da dies jedoch für das Installationsdesign kein Problem darstellt, da nur das Tracking der Hände wichtig ist, fiel die Entscheidung zu Gunsten der Leap Motion Kamera. Auch die deutlich höhere Auflösung, welche Bewegungsänderungen in hundertstel Millimetern erfassen kann, galt als Entscheidungsgrundlage.

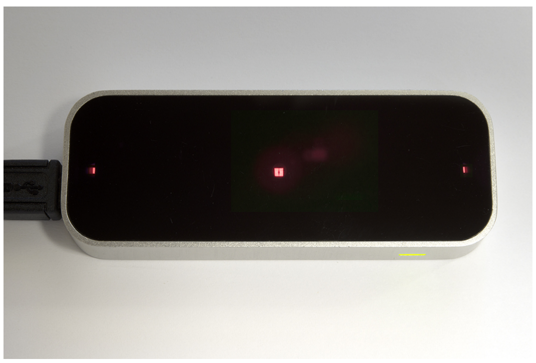
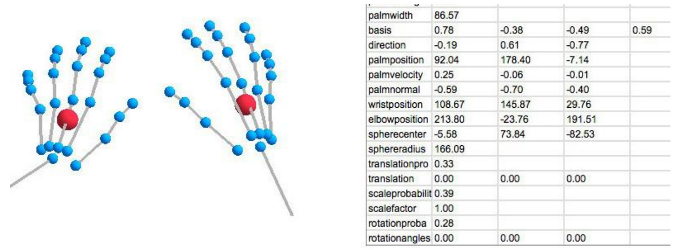


Abbildung 4. Leap Motion Kamera

Die Leap Motion Kamera funktioniert mittels drei LEDs, die das zu trackende Objekt beleuchten, welches durch zwei Infrarotkameras erfasst wird. Die Daten werden softwareseitig verarbeitet und in ein 3D-Modell der Hände umgewandelt. In der Software vorgegebene Bewegungsabläufe machen es möglich, verschiedene Bewegungen als Gesten zu erkennen (Swipe, Circle) oder das Tracking auf beliebige Weise einzuschränken (z.B. Tracking eines Fingers).

Die Tracking-Daten werden durch die eigene Software der Leap Motion Kamera erfasst. Mit Hilfe des Leap Motion Software Developer Kit und zweier Erweiterungen für Max/MSP konnten die Bewegungskoordinaten in Max abgegriffen werden. In Isidor wurde aufgrund besserer Isolationsmöglichkeiten die Leap Motion Erweiterung des "IRCAM" (Institut de Recherche et Coordination Acoustique/Musique) benutzt.

Nach einigen Versuchen stellte sich heraus, dass die beste Methode für die Umwandlung von Dirigatsgesten das Tracken der Palm-Position in der Höhe (X-Achse) ist. Um nur diese Daten zu erhalten, wird der Datenstrom, in dem alle Daten vorhanden sind, gefiltert und nur die extrahierten Daten weiterverarbeitet.



palmwidth	86.57			
basis	0.78	-0.38	-0.49	0.59
direction	-0.19	0.61	-0.77	
palmposition	92.04	178.40	-7.14	
palmvelocity	0.25	-0.06	-0.01	
palmnormal	-0.59	-0.70	-0.40	
wristposition	108.67	145.87	29.76	
elbowposition	213.80	-23.76	191.51	
spherecenter	-5.58	73.84	-82.53	
sphereradius	166.09			
translationpro	0.33			
translation	0.00	0.00	0.00	
scaleprobabilit	0.39			
scalefactor	1.00			
rotationproba	0.28			
rotationangles	0.00	0.00	0.00	

Abbildung 5. Leap Motion Gesten-Tracking

Mithilfe eines Wahlselektors kann in der GUI bestimmt werden, welche Hand getrackt werden soll. Die so erhaltenen Daten werden nun interpretiert und in ein Tempo (BPM) umgewandelt, das, wie bereits erwähnt, als Multiplikator für die Masterclock weiterverwendet wird.

Zuerst werden erfasste Bewegungsdaten der X-Achse, welche sich zwischen 0 und 400 befinden, auf einen kleineren Wert (0-1) skaliert. Da das Projekt ursprünglich für Mac OS angedacht war, wurde in der Ursprungsversion für das Feststellen einer Richtungsänderung der Palm Position (oben/unten, entsprechend der klassischen Dirigatsgesten) das J.Delta- Objekt aus der Jamoma Libary verwendet, das grundsätzlich die ersten drei Ableitungen einer Funktion berechnet. Durch den Umstieg auf ein Windows-Betriebssystem wurde dies jedoch aufgrund der fehlenden Implementierung unmöglich.

Um das J.Delta-Objekt zu ersetzen, werden die skalierten Daten in ein tfb- (Trigger-Float-Bang) Objekt geschickt, welches jede neue Koordinate speichert. Bekommt das Objekt einen neuen Input werden die beiden Werte subtrahiert und somit verglichen. Ist das Ergebnis negativ, findet eine Richtungsänderung statt. Jede negative Zahl wird an ein Change-Objekt weitergegeben, welches dafür sorgt, dass keine Koordinate doppelt verwendet wird. Jeder Negativwert löst nun einen Bang aus, der in ein Timer-Objekt geschickt wird, welches die vergangene Zeit zwischen zwei Bangs misst. Um von einem Zeitwert auf einen BPM-Wert zu kommen, wird das Ergebnis des Timer-Objekts durch 60000 dividiert. Um zu extreme Temposchwankungen zu verhindern, wird der mögliche BPM-Bereich auf 30-200 Schläge pro Minute festgelegt. Ein Slide-Objekt sorgt dafür, dass zwischen den Tempowerten interpoliert wird, was zu einem kontinuierlichen Tempowechsel führt. Um abschließend einen ganzzahligen BPM-Wert zu erhalten wird die errechnete Tempoinformation mit einem Round-Objekt gerundet.

Herausforderungen in der Programmierung stellte das Ersetzen des J.Delta-Objektes in der Windowsumgebung dar. Eine gute Übertragbarkeit der dirigierten Tempoinformation, um ein möglichst realistisches und steuerbares Erlebnis für den Nutzer zu ermöglichen, ist aufgrund des subjektiven Empfindens und dessen Ausdruck, anhand von Sensibilität des Trackings bzw. Feststellung des Zeitpunktes des tatsächlichen "Schlages", sehr schwer erreichbar.

VI. Testlauf

Um den Prototypen der Installation echten Rezipienten zum Test zur Verfügung zu stellen, wurde die Klanginstallation "Isidor" im Rahmen der "nextgeneration 7.0" im Zentrum für Kunst und Medien in Karlsruhe ausgestellt. Der Max/MSPPatch lief auf einem Macbook Pro (13z, Late 2011) auf dem Betriebssystem Windows 7. Als Audio-Interface wurde ein Saffire Liquid 56 benutzt, über dessen Outputs sechs Studiomonitore des Typs Genelec 1032B angesteuert wurden. Alle Filter an den Lautsprechern waren deaktiviert, die Gain-Regler waren alle auf den selben Wert eingestellt. Auch die Outputs des Saffire Liquid 56 spielten alle auf demselben Pegel aus; sämtliche Mischung geschah daher ausschließlich in Max/MSP. Als Lichter wurden zwei PAR56 LED-Scheinwerfer benutzt, die mit dem Laptop über das Enttec DMX USB Pro verbunden waren. Als Kontrolloberfläche diente der 24-Touchscreen der Firma Ilyama (Modell ProLite B2483HS-B1). Auf der Oberkante des Bildschirms lag die Leap Motion Kamera.

Beim Testaufbau wurde versucht, den Aufbaudimensionen (siehe Abb. 1) so weit wie möglich zu entsprechen. Aufgrund der Dimensionen des Raumes, in dem aufgebaut wurde, war dies jedoch nicht ausnahmslos möglich. Die beiden äußeren Frontlautsprecher standen in ca. 2 m Entfernung voneinander, der Zuhörer befand sich in ungefähr 2,5 m Entfernung von den Frontlautsprechern in mittiger Position. Die rückwertigen Lautsprecher standen innerhalb eines Meters zum Hörer, also sehr nahe im Vergleich zu den Idealdimensionen. Dieser Nachteil wurde durch zeitliche Verzögerung und Lautstärkenanpassung der Rear-Kanäle auszugleichen versucht.

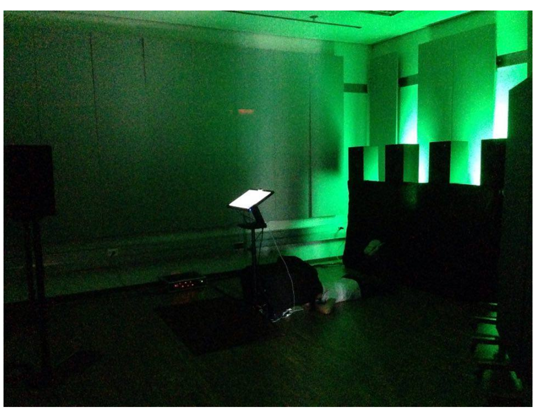


Abbildung 6. Aufbau im Rahmen der "nextgeneration 7.0" im Zentrum für Kunst und Medien (ZKM) in Karlsruhe

Bei dem Raum, in dem der Testlauf durchgeführt wurde, handelte es sich um den Aufnahmeraum eines Tonstudios des ZKM, dementsprechend waren die akustischen Begebenheiten sehr günstig und die Charakteristik des Raumklanges als trocken zu bezeichnen. Der Raum hatte eine Gesamtfläche von ungefähr 25m^2 bei einer Höhe von rund 3 Metern.

Die Installation war zwei Tage lang zugänglich und konnte von Besuchern getestet werden. Die Versuchspersonen waren gemischten Alters und Geschlechts und zu rund 70% Prozent Besucher der Ausstellung, während die restlichen 30% andere Ausstellende, Universitätsmitarbeiter oder Studenten des ZKM waren. Eine schriftliche Erklärung lag bereit; viele Besucher konnten dennoch anfängliche Berührungsängste mit der Installation erst nach einer persönlichen Einführung überwinden.

Die Reaktionen der rund 40 Probanden kann man als zum größten Teil positiv beschreiben. Viele Besucher schätzten vor allem die Möglichkeit der Temposteuerung durch das Dirigat, gaben aber auch an, dass sie sich noch nicht absolut in Kontrolle des Tempos fühlten. Alle Touch-gesteuerten Kontrollmöglichkeiten (Lautstärke, Position, Room, ...) wurden problemlos und mit Wohlwollen entgegengenommen. Auch das User Interface wurde als übersichtlich und optisch ansprechend empfunden. Die Autoren wurden von einem Probanden, der als Pädagoge arbeitet, auf das große Potenzial angesprochen, welches er in der Installation bezüglich Bildungsmöglichkeiten für Kinder sieht.

VII. Herausforderungen

Die wohl größte Herausforderung ist es, ein für den Nutzer nachvollziehbares, kontrollierbares und spannendes Erlebnis zu kreieren. Ziel der Installation ist es, sowohl für Laien als auch für Personen mit fachspezifischem Hintergrundwissen ansprechend gestaltet und verständlich zu sein. Dazu ist es notwendig, dass alle Elemente reibungslos miteinander kommunizieren.

Dies ist die grundlegende Herausforderung für die Entwicklung der Installation. Es gilt, ein System zu schaffen, welches mehrere Elemente und Schnittstellen miteinander verbindet und eine Kommunikationsbasis schafft. Die Auswahl der zu verwendenden Technologien spielte dabei eine wichtige Rolle. Von Anfang an stellte sich die korrekte Detektion der Tempodaten und die darauffolgende Tempomanipulation als schwierig heraus. Hier ist es notwendig, eine feine Abstimmung zwischen der Technik und der Wahrnehmung des Rezipienten herzustellen. Für ein bestmögliches Nutzerlebnis sollte jeder Benutzer das Gefühl haben, er wäre in vollkommener Kontrolle über die Installation.

VIII. Diskussion

Beschrieben wurde eine Installation zur Steuerung eines virtuellen Streichquartetts mittels MIDI-gesteuerten VST-Instrumenten und Tracking von Dirigatsgesten mittels Leap Motion Kamera. Nach der Konzeption und einem Testlauf sowie Nachbesserungen kann gesagt werden, dass die Installation für den Benutzer in Bezug auf subjektives Kontrollempfinden, Handhabung und Vergnügen eine zufriedenstellende Erfahrung ist. Vor allem das klangliche Ergebnis durch den Einsatz von VST-Instrumenten und DSP ist aufgrund seiner Variabilität hinsichtlich Tempo, Anschlagsdynamik, Raumklang, etc. sehr befriedigend und bietet im Vergleich zu ähnlichen Installationen, welche mit aufgenommenem Audiomaterial arbeiten, große Weiterentwicklungsmöglichkeiten. Der Benutzer kann mit einem leicht zu verstehenden Interface in viele Parameter eingreifen und bekommt umgehendes, audiovisuelles

Feedback. Entwicklungspotenzial gibt es nach wie vor in der Feinabstimmung, bspw. des Motion Trackings beim Dirigat. Da es sich bei der Installation um ein sehr reaktives System handelt, ist der Bedarf nach Feinjustierung naturgegeben sehr hoch, da jeder Eingriff viele verschiedene Parameter, vor allem hinsichtlich des Kontrollgefühls, beeinflusst.

IX. Ausblick

Möglichkeiten zur Weiterentwicklung des Projektes sehen die Autoren vor allem in der Leap Motion Technologie. Die Kamera erlaubt es dem Nutzer, beide Hände simultan zu erfassen und verschiedene Gesten zu erkennen. Die zweite Hand könnte somit dazu benutzt werden, andere Parameter anzusteuern - entsprechend dem Dirigat bspw. die Dynamik des Stückes. Wie bereits zuvor erwähnt, wurde von Probanden die Möglichkeit des Einsatzes der Installation für Bildungszwecke, speziell für Kinder, angemerkt. Letztlich bleibt auch die Weiterentwicklung des Bach-Objekts spannend für die Installation, da die Projektgruppe mit den Entwicklern von Bach in Verbindung trat und einige Anmerkungen und Anregungen möglicherweise in zukünftigen Versionen von Bach berücksichtigt werden könnten, was in der weiteren Entwicklung Wege ebnen würde. Um ein Beispiel zu nennen, würde die Möglichkeit, MIDI Control Changes in Bach verarbeiten zu können, große Verbesserungspotenziale für die realistische Klanggestaltung bieten.

Die Autoren hoffen, mit diesem Projekt Denkanstöße und Grundlagen für weiterführende Forschungen geschaffen zu haben.

Literatur

[1] A. Jylhä, “Sonic gestures as input in human-computer interaction: Towards a systematic approach,” in *Proceedings of the SMC2011-8th Sound and Music Computing Conference, Padova*, 2011.

[2] E. Hornecker and M. Stifter, “Learning from interactive museum installations about interaction design for public settings,” in *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. ACM, 2006, pp. 135–142.

[3] A. Camurri, G. De Poli, and D. Rocchesso, “A taxonomy for sound and music computing,” *Computer Music Journal*, vol. 19, no. 2, pp. 4–5, 1995.

[4] J. O. Borchers, W. Samminger, and M. Mühlhäuser, “Conducting a realistic electronic orchestra,” in *Proceedings of the 14th annual ACM symposium on User interface software and technology*. ACM, 2001, pp. 161–162.

[5] WhiteVOID, “Mendelssohn Effektorium Conducting a virtual orchestra,” Berlin, 2014, http://www.whitevoid.com/#/main/interactivestructures/mendelssohn-effektorium, letzter Zugriff: 04.09.2017.

[6] Aconica, “Effektorium-Interactive Museography/Interactive Sound and Light Installation,” Berlin, 2014, http://www.aconica.de/portfoliopage/effektorium-interactive-museography/, letzter Zugriff: 04.09.2017.

[7] K. Flay, “Gesture controlled musical conducting,” Semesterarbeit, School of Computer Sciences and Informatics Cardiff, Wales, 2015.

[8] E. Sengpiel, “Intervall-Umrechnung in Frequenzverhältnis nach cent und zurück,” n.d., http://www.sengpielaudio.com/Rechner-centfrequenz.htm, letzter Zugriff: 04.09.2017.

Evaluation of Data Transfer Methods for Block-based Realtime Audio Processing with CUDA

Christoph Kuhr*, Alexander Carôt† Department of Computer Sciences and Languages, Anhalt University of Applied Sciences
Köthen
Email: *christoph.kuhr@hs-anhalt.de, †alexander.carot@hs-anhalt.de

***Abstract*—Realtime audio production environments generally do not use GPUs, as long as they are not involved in 3D rendering or video production processes. Thus, the GPU is idle most of the time and can be utilized as an audio co-processor. The block-based streaming nature and floating point representation of computer audio hardware are very well suited for GPGPU programming techniques. In this paper we line out the data transfers as the most expensive part in the processing of realtime audio data and evaluate different data transfer methods and positively evaluate different data transfer methods with respect to future audio DSP applications.**

I. Introduction

Modern computer systems are equipped with a CPU and a GPU. CPUs control the peripheral hardware and perform calculations unrelated to 3D graphics or video decoding. A GPU in contrast is concerned with rendering 3D graphics or utilizing special hardware codecs to decode nowadays video codes like H264 [1].
If a computer system is used for any kind of audio production, that excludes 3D rendering and video decoding, the GPU is mostly idle. Additionally, GPUs are designed to handle multiple floating point operations at the same time in a threaded fashion.
These considerations promote the idea to use a GPU as an audio co-processor for signal processing purposes.
Computation intensive audio signal processing of realtime data has already been done, e.g. Wefers and Berg have used a GPU to process FIR and IIR filters [2], Jedrzejewski and Marasek have used the GPU to do impulse response computations for virtual room acoustics [3].

In this paper we will investigate the lower limit for the usage of a GPU for such signal processing tasks in a realtime audio production environment. The limit is given as the combination of channel count and sample buffer size in use. The bottlenecks in the communication between CPU and GPU are evaluated and discussed. Further, possible workarounds to increase the performance aspects under investigation are proposed and evaluated.

CUDA (Compute Unified Device Architecture) is a programming langauge designed for high-performance computing [4]. The idea is to make use of thousands of threads running in parallel, which is not possible with x86/x86_64 CPUs. Such parallel programs are called kernel in the CUDA domain.

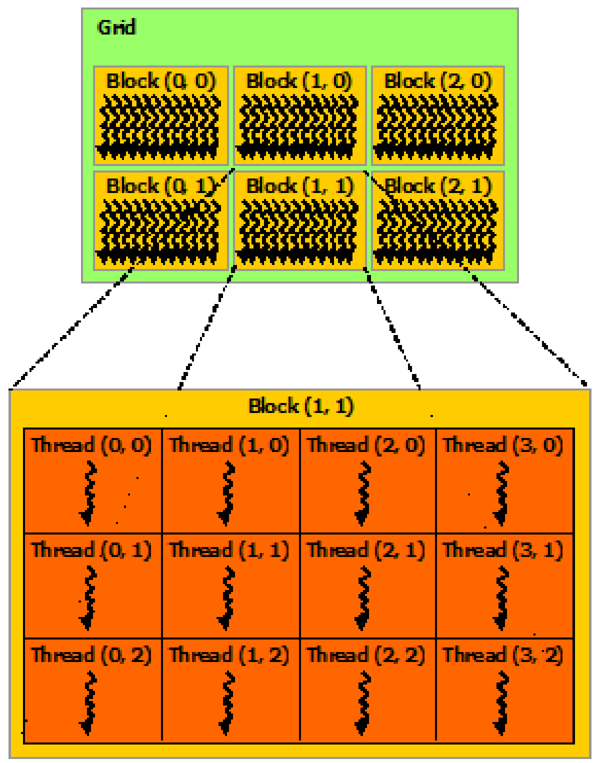


Figure 1: *CUDA Computing Grids [5]*

When a kernel is executed on the GPU, the kernel launches a grid of several blocks, the limit is depending on GPU features. Inside each block on the grid, multiple threads execute the actual computations at runtime. The same computation runs on each thread, but with different data. Threads can be handled in a synchronous or an asynchronous way. The latter requires the concept of streams for a destinct mapping of the data shared between the threads of one block. The structure of CUDA computing grids is shown in fig. 1.

The concept of CUDA streams [6] is very convenient for the problem at hand.

Different audio streams can be treated asynchronously, which is a better representation of their orthogonal nature then a matrix with an appropriate amount of rows and columns. This way the orthogonality may also be represented appropriately, but access to the matrix would be centralized and would experience possible racing conditions. Beyond, using a dimension (x, y or z) for the representation of the different audio channels, reduces the available dimensionality that is useable for calculations at runtime.

This paper is part of the research project fast-music [7]. The project has the goal to enable symphonic orchestras to rehearse via the public internet, by using the realtime communication software Soundjack [8] [9]. Research in the field of packet loss concealtment will use GPUs for complex signal processing based on machine learning algorithms.

II. Architecture

The work of Wefers and Berg [2] has also shown, that realtime processing of audio data with a GPU is possible.
The communication between CPU and GPU is realized via driver calls and shared memory, either DMA, GPU or CPU RAM. The CPU is also referred to as host and the GPU as device. Nowadays, system architectures where CPU and GPU share the same cache are used increasingly, albeit mainly in embedded systems. This architecture completely eleminates memory copies, since the memory is coherently accessible by the CPU and the GPU. In conventional systems which communicate via the PCIe bus, data has to be copied from CPU RAM to GPU RAM and back.
Since the API calls copying data between CPU and GPU have much overhead, it is more efficient to copy huge amounts of data. Thus, it is even more interesting to investigate the use case of small amounts of data, as generated and processed in the audio domain.

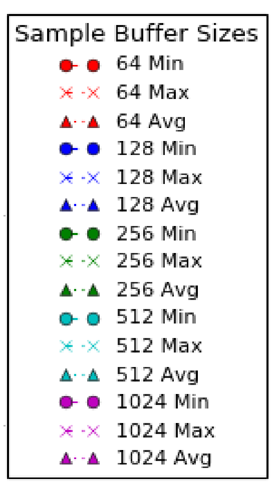


Figure 2: *Legend Data Transfer Method Measurements*

Realtime audio data is represented as a two dimensional vector field. At any sample point in time some analog digital converter process generates a sample, with typical bit depths of 16, 24 or 32 bits, either encoded as integer or floating point [10].
Computer audio hardware manages data by using buffers that consist of a predefined amount of samples. The audio driver repeatedly accesses the memory of the audio hardware and copies the sample buffers to the CPU RAM for further usage. The responsivness of such an audio system depends on the size of the sample buffers, while the response time reduces with an increasing sample buffer size. Typical sample buffer sizes are 64, 128, 256, 512, 1024 samples [11].

$$AudioDataBlock = SampleDepth \cdot SampleBufferSize \cdot ChannelCount$$

$$AudioDataBlock = 32bit \cdot \{64, 128, 512, 1024\} \frac{Samples}{s} \cdot \{2, 8, 16, 32, 64\}$$

Due to this block-based streaming nature, the data transfer and processing of audio data between CPU and GPU might reduce the impact of the data copying overhead, particularly if multiple audio channels are used.
The audio data, that we will transfer and process with the GPU, is provided by a professional audio driver and server combination called Jack Audio Connection Kit [12]. On top of a Linux ALSA [13] driver, Jack provides the means to interconnecting jack-aware audio software to the audio interface with 32 bit floating point precision. The floating point format requires the development of a prototype, because the Soundjack clients use an integer format instead of floating point and would require additional conversion.

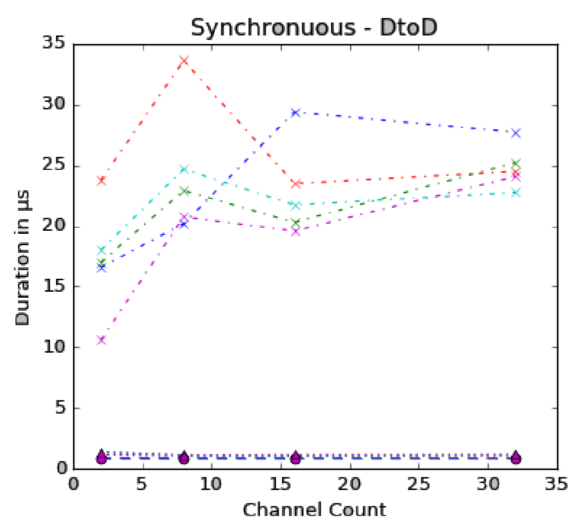


Figure 3: *Device to Device Copy Duration Synchronous Data Transfer Method*

We developed a most simple Jack client for testing purposes with varying channel counts and sample buffer sizes. The Jack client is linked against a shared library that provides the CUDA Kernel [4]. This way CUDA computations can be integrated in arbitrary C programs. The Jack Server configures the audio interface by utilizing the ALSA driver infrastructure. The most important configuration parameters for our investigations are the channel count and sample buffer size, called frame or period in the Jack domain. At runtime, the Jack Server requests our Jack client to process a frame with a callback function. If the callback function is not done with its computations in time, the Jack Server reports a buffer underrun, also called xrun in the Jack domain.

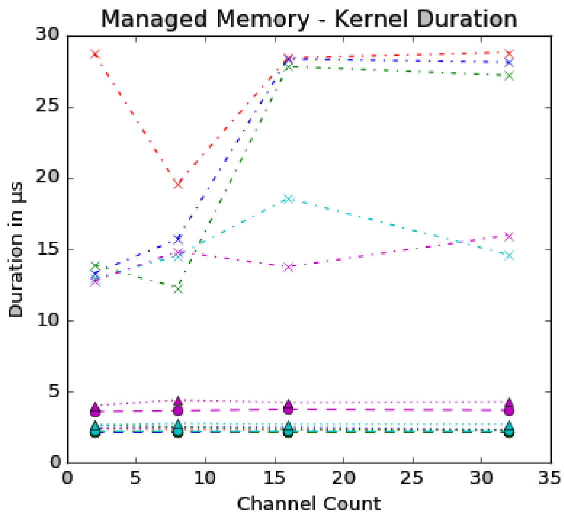


Figure 4: *Kernel Execution Duration Managed Memory Data Transfer Method*

A Nvidia Geforce GT940mx GPU with 2 GB of DDR3 RAM is connected to an Intel i7-6870 4-Core CPU with 16 GB DDR3 RAM via a PCIe x16 2.0 bus [14], in the system under test. Thus, the transfer rate between CPU and GPU is limited to the bus bandwidth of 8 GBps simplex.
The Nvidia Geforce GT940mx has a compute capability of 5.0 ($\geq$ 2.0), which allows it to use managed memory.

III. CUDA Memory Organization and Management

The data structure and data transfer between CPU and GPU are the bottlenecks for the entire signal processing. Three different data transfer methods can be used:

1) Synchronous data transfer

   A synchronous data transfer returns as soon as the memory operation on the GPU memory is done, with a success or failure result. For the GPU integration of synchronous data transfers, it is irrelevant whether the memory is pagable or pinned. Either type can be accessed. Pagable memory is memory from the virtual address space of CPU or the operating system.

2) Asynchronous data transfer

   An asynchronous data transfer returns immediately after invoking the data transfer, regardless of the result. The result of the operation has to be checked seperately. It requires the additional concept of streams for the integration on the GPU. Further, the host memory has to be pinned. Pinned memory addresses are allocated in the DMA address space of the host system.

3) Managed memory with coherent caches on CPU and GPU

   With managed memory, the requirement of memory copy operations is eliminated. The GPU driver allocates memory on the CPU and GPU respectively, manages any data access onto these memory segments implicitly and thus keeps the data in both memory locations coherent by small caching operations.

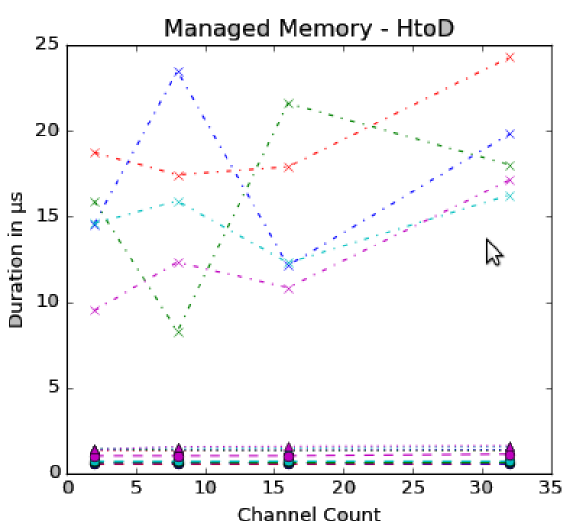


Figure 5: *Host to Device Transfer Duration Data Transfer Method*

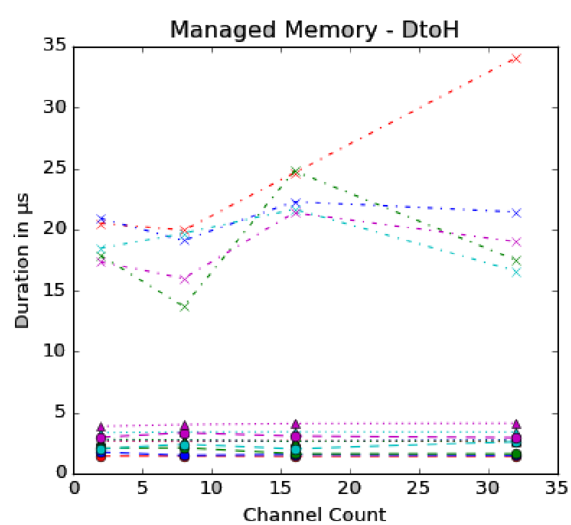


Figure 6: *Device to Host Transfer Duration Data Transfer Method*

The direction for data transfers is crucial as well. Three different directions are distinguished:

1) HostToDevice (HtoD or H2D)

   The *HostToDevice* mode utilizes the Direct Memory Access (DMA) memory of the host system. This enables the CPU to offload the data transfer operations to the GPU without waiting for the completion or result.

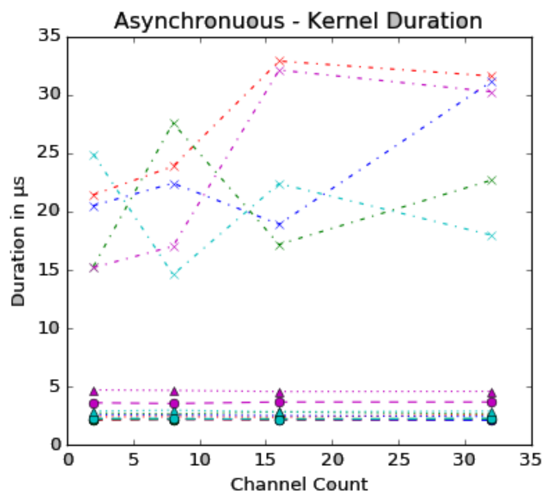


Figure 7: *Kernel Execution Duration Asynchronous Data Transfer Method*

2) DeviceToDevice (DtoD or D2D)

   Invoking CUDA memcpy between two GPUs uses memory copy operations between the RAM of both GPUs. If a D2D memory copy operation is issued on a single device however, the GPUs' internal cache is used for the data transfer.

3) DeviceToHost (DtoH or D2H)

   Although the *DeviceToHost* mode does not utilize the DMA memory it may also operate asynchronously, but slower since it is copied from GPU to CPU RAM.

IV. Experiments

We investigated the influence that the sample buffer size and channel count had on the data transfer rates. The audio channel count was varied between 2, 8, 16 and 32 channels, while each channel count was tested with each common sample buffer size of 64, 128, 256, 512 and 1024 samples per buffer. The samples were formatted as 32 bit floating point.
A simple CUDA kernel is provided for an exemplary computation. Each thread in a block handles exactly one sample, copies it from the input to the output buffer. This way 64 up to 1024 threads run in parallel in a single block.
The worstcase for the data transfer times, is given by the Jack servers buffersize and sample rate, which in this case is 48kHz (Sample Duration = $\frac{1}{48kHz} = 20.833\mu s$):

Sample Buffer Size	Worst Case Latency
64	1.334ms
128	2.667ms
256	5.334ms
512	10.667ms
1024	21.334ms

Table I: *Tolerable Worst Cast Latencies for Realtime Audio*

The profiling overhead of the NVidia Visual Profiler (NVVP) for 32 channels with 64 samples per buffer pushed the host machine to its limits. Thus, tests with 64 audio channels were omitted.

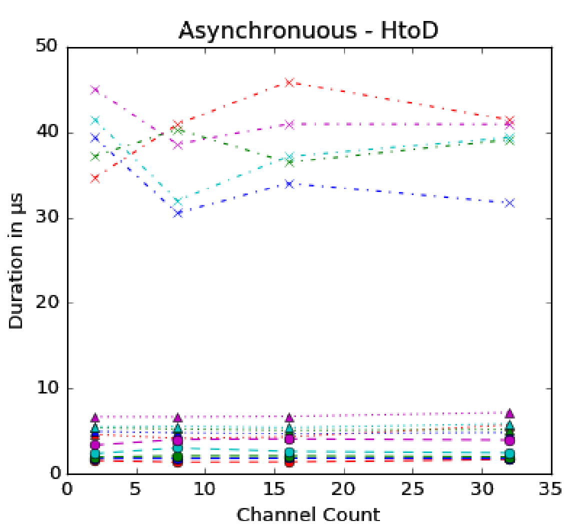


Figure 8: *Host to Device Transfer Duration Asynchronous Data Transfer Method*

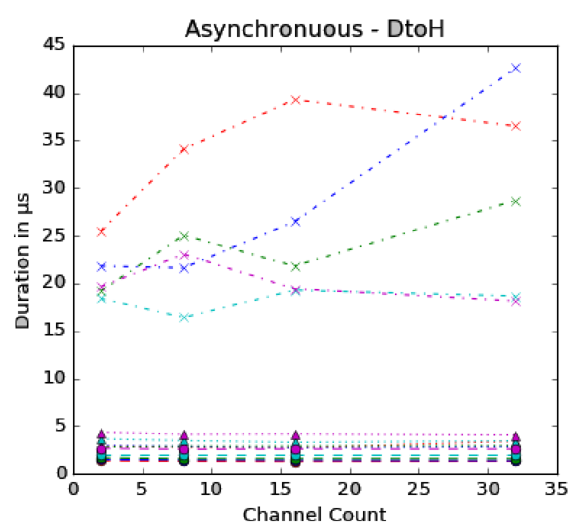


Figure 10: *Device to Host Transfer Duration Asynchronous Data Transfer Method*

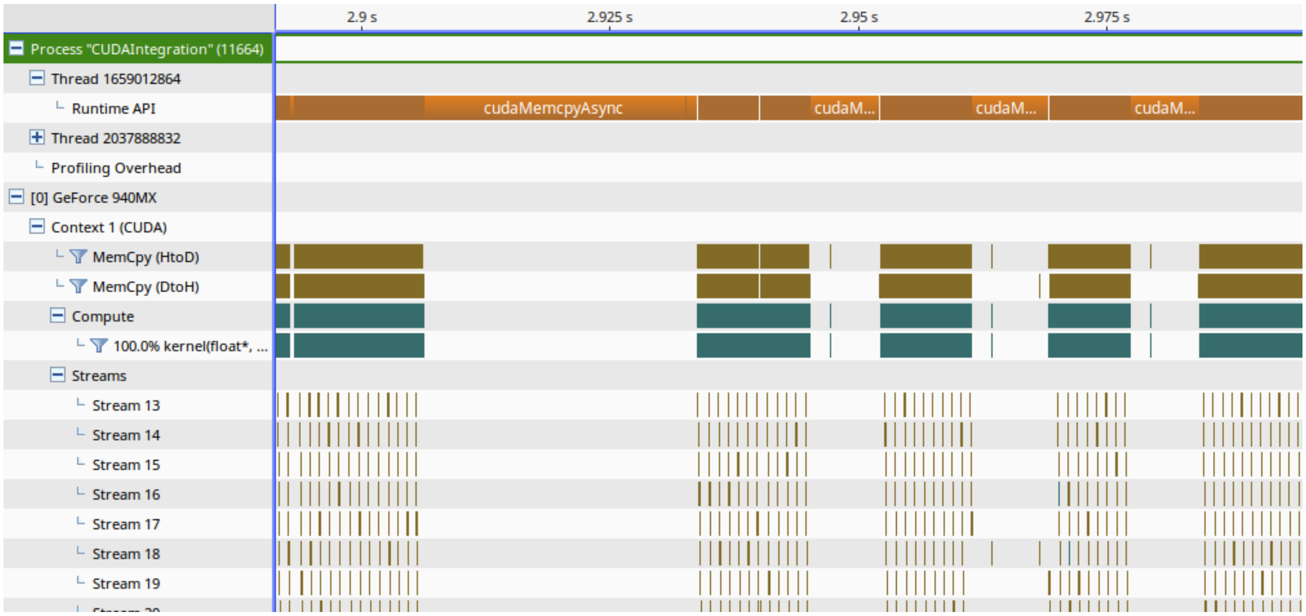


Figure 9: *NVVP Screenshot showing CUDA API Overhead*

V. Discussion

All combinations of transfer methods and modes, sample buffer sizes and channel counts take in average less then $10\mu s$ and show peaks of up to $46\mu s$, as visualized in fig. 4 to fig. 10. The visualized durations neglect the CUDA API and driver calls, they represent the execution on hardware only. The legend in fig. 2 is common for all figures.

A comparisson of fig. 5 and fig. 8 shows that the memory mapped H2D mode takes less time, at minimum, average and maximum then the asynchronous copy mode.
The kernel execution times for the two other transfer methods shown in fig. 4 and fig. 7, exhibit no significant difference. In fig. 3 only the device to device copy operation is shown, which does not involve any kernel launch. These findings suggest that the synchronous memory transfer method would also be suitable for the H2D copy mode. Since a kernel has to wait until all data is present in the GPU memory, it is of no consequence at this point, if the data is transfered synchronously or asynchronously. In contrast to the D2H mode, where a non blocking data transfer allows the processing chain to finish sooner. The magnitude of these savings is much lower then of the overhead introduced by the CUDA API and driver calls. This is observable in the rows below the CUDA Context in fig. 9, the three smaller gaps ($\approx 7ms$) on the right side and a larger gap ($\approx 28ms$) on the left side relate to the small chunks in the rows for the respective streams. These chunks are the hardware based memory operations as mentioned above and take only a few micorseconds in average.

All three memory organization modes exhibit a common problem of cyclic nature. At a given interval ($\approx 11s$ for pagable memory, $\approx 5s$ for pinned memory and $\approx 2.5s$ for managed memory) memory operations last approximately four times longer, resulting in the larger gap on the left side in fig. 9. These API and driver calls introduce jitter to the tested audio signal.

The turning point from where the CUDA API overhead is neglectable, can be quantified:

Channel Count	Sample Buffer Size
2	512
4	512
8	1024
16	1024
32	1024

Table II: *Channel Count and Sample Buffersize Limit for Realtime Audio Processing*

VI. Conclusions

All three memory transfer methods are able to operate on realtime audio data. Managed memory however is most convenient, because host and device pointers do not require any special handling and integrate smoothly into C code as well as CUDA code. For the usage with Jack however, two memory copy operations are still required, because Jack provides preallocated pointers to its buffer interface.
Low sample buffer sizes increase jitter, but no buffer underruns were detected. Although the duration of the CUDA API and driver calls suggest that underruns should occur with sample buffer sizes below 512 samples.

VII. Future Work

The evaluation of the data transfer method has been a feasability study for further goals. In the future, machine learning algorithms will be investigated in this environment as well as common signal processing algorithms, with respect to error concealment techniques and the generation of audio effects.

VIII. Acknowledgements

fast-music is part of the fast-project cluster (fast actuators sensors & transceivers), which is funded by the BMBF (Bundesministerium für Bildung und Forschung).

References

[1] *H.264: Advanced video coding for generic audiovisual services*, ITU-T Std. H.264, 2003.

[2] F. Wefers and J. Berg, "High-performance real-time fir-filtering using fast convolution on graphics hardware," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*. Graz, Austria: Institute of Technical Acoustics, RWTH Aachen University, Sep. 6–10, 2010, pp. DAFX–1 – DAFX–8.

[3] M. Jedrzejewski and K. Marasek, "Computation of room acoustics using programmable video hardware," in *Computer Vision and Graphics, Springer-Verlag Netherlands*. PJWSTK, 2006.

[4] *Getting Started with CUDA*, NVidia Corporation, 2008.

[5] *CUDA C PROGRAMMING GUIDE*, NVidia Corporation, 2013.

[6] *CUDA Streams, Best Practices and Common Pitfalls*, NVidia Corporation, Year unknown.

[7] (2017, Jun.) fast actuators, sensors and transceivers. [Online]. Available: https://de.fast-zwanzig20.de/

[8] (2017, Jun.) Soundjack - a realtime communication solution. [Online]. Available: http://http://www.soundjack.eu

[9] A. Carôt, "Musical telepresence - a comprehensive analysis towards new cognitive and technical approaches," Ph.D. dissertation, University of Lübeck, Germany, May 2009.

[10] A. V. Oppenheim and R. W. Schaefer, *Discrete-time signal processing*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, Inc., 1989.

[11] K. C. Pohlmann, *Principles of Digital Audio*, 5th ed. The Mcgraw-Hill Companies, 2005.

[12] (2017, Jun.) Jack audio connection kit. [Online]. Available: https://jackaudio.org

[13] (2017, Jun.) Advanced linux sound architecture. [Online]. Available: https://alsa-project.org/main/index.php/Main_Page/

[14] (2006, Dec.) Pci express base specification revision 2.0. PCI-SIG. [Online]. Available: https://members.pcisig.com/wg/PCI-SIG/document/download/8246

Session 3: Interactive Session - Special Track GLAMhack17

KuKoNö – KulturKontext Niederösterreich

Kerstin Blumenstein*, Barbara Margarethe Eggert†, Maria Grandl‡, Elisabeth Kasser-Höpfner§, Kathrin Kratzer†, Johannes A. Löcker-Herschkowitz¶, Georg Neubauer† und Florian Wiencek‖

*St. Pölten University of Applied Sciences, †Donau-Universität Krems, ‡Technische Universität Graz, §Amt der Niederösterreichischen Landesregierung, ¶Universität Wien, ‖Fluxguide

***Abstract*—Not least because of "Pokemon Go", everyone is talking about context-sensitive applications. These applications are also widely used in tourism and in cultural institutions, whose collections can not be shown to the public physically due to renovation or because of a lack of a permanent exhibition house. We use these experiences in the application KulturKontext Niederösterreich (KuKoNö) in order to make it possible to get to know and explore cultural objects from the State Collection of Lower Austria in the physical space. This paper presents the concept of the application KuKoNö and an example tour for St. Pölten.**

***Zusammenfassung*—Kontextsensitive Anwendungen sind nicht zuletzt nach dem Erfolg von "Pokemon Go" in aller Munde. Aber auch im Tourismus und in Kultureinrichtungen, deren Sammlungen durch Umbau oder durch ein fehlendes festes Ausstellungshaus physisch nur eingeschränkt der Öffentlichkeit gezeigt werden können, finden diese Anwendungen starke Verbreitung. Diese Erfahrungswerte greifen wir in der Applikation KulturKontext Niederösterreich (KuKoNö) auf, um das Kennenlernen und Erkunden von kulturellen Objekten aus den Landessammlungen Niederösterreich im physischen Raum zu ermöglichen. Dieses Paper präsentiert das Konzept der Applikation KuKoNö sowie eine Beispiel-Tour für St. Pölten.**

I. Einleitung

Kontextsensitive Anwendungen bieten den Nutzer*innen zur richtigen Zeit relevante Informationen. Insbesondere Location-based Services (LBS) stellen diesbezüglich einen attraktiven Ansatz für die Kulturvermittlung dar. Nicht zuletzt der Erfolg des Spiels "Pokemon Go" zeigt die Faszination von LBS auf Menschen unabhängig von Geschlecht, Alter oder Bildung [1]. Diese Begeisterung möchten wir mit der Applikation KulturKontext Niederösterreich (KuKoNö) auf das Kennenlernen und Erkunden von kulturellen Objekte aus den Landessammlungen Niederösterreich [2] übertragen.

Die Landessammlungen Niederösterreich sind defakto eine Sammlung ohne festen Ausstellungsort. Ausstellungsinstitutionen des Landes wie das Museum Niederösterreich (St. Pölten), das Karikaturmuseum Krems, bald auch die Landesgalerie Niederösterreich (Krems) und viele mehr können nur einen Teil der Landessammlungen als Exponate zeigen. Wie in so vielen Museen ist der große Rest der Sammlung den Kulturinteressierten verborgen.

Die grundlegende Zielstellung bei der Konzeption einer App war es deshalb, Objekte und Daten der Landessammlungen Niederösterreich an Orten in Niederösterreich erlebbar zu machen, zu denen sie einen direkten Bezug haben. Dies stellt nicht nur die Objekte selber in einen historischen Kontext, sondern bringt sie idealerweise zurück in das aktive kulturelle Gedächtnis der Niederösterreicher*innen und der Besucher*innen des Bundeslandes. Entstanden ist das Konzept der mobilen Applikation KuKoNö.

Mit der App lernen die Benutzer*innen "on the go" (spielerisch) die Kulturschätze Niederösterreichs kennen. Es ist geplant, die App KuKoNö zunächst für Objekte nutzbar zu machen, die einen Ortsbezug zu St. Pölten aufweisen. Hierdurch werden ausgewählte Orte im Stadtraum als Kulturraum erlebbar. Dafür nutzen wir unter anderem Augmented Reality, um historische Schichten sichtbar zu machen. Hierbei kann für die App KuKoNö u.a. auf aufbereitete Daten und Texte zurückgegriffen werden, die für das Landhausviertel durch den Sammlungsbereich Kunst im öffentlichen Raum erstellt wurden.

Nach erfolgreicher Implementierung der App für den Kontext St. Pölten kann diese auch für andere Regionen eingesetzt werden: Zum Beispiel für die Regionen Carnuntum und Wachau. In den Landessammlungen Niederösterreich befinden sich zahlreiche kunsthistorische, kulturgeschichtliche, archäologische und naturhistorische Objekte, die diesen Regionen zugeordnet werden können. Ferner könnte im Zuge von Landesausstellungen jeweils das Umland mittels der App KuKoNö medial angereichert werden. Für St. Pölten bietet die App sowohl einen Erstzugang zu Objekten aus dem Depot als auch zu Exponaten aus assoziierten Ausstellungshäusern. Sie soll generell das Interesse an materieller Kultur, auch an Alltagskultur, wecken. Darüber hinaus mag sie auch Nicht-Besucher*innen dazu anregen, erstmals ein Museum zu besuchen.

Im Folgenden wird das allgemeine Konzept der App erläutert und anhand einer Beispiel-Tour für St. Pölten konkretisiert.

II. Konzept

Durch die App KuKoNö werden Kulturschätze aus den Landessammlungen Niederösterreich virtuell mit relevanten Orten verknüpft. Diese Orte können bspw. der Entstehungsort des Objektes sein oder die Fundstätte. Zunächst erfolgt dies für den Kulturstandort St. Pölten. In der Landeshauptstadt ist bereits ein Kultur-Touristischen Leitsystems implementiert. Dessen Infostelen fassen einem Bilderrahmen gleich Blickachsen ein und weisen Ortsansässige sowie Tourist*innen auf

(a) Übersichtsansicht der Karte. Der graue Marker visualisiert die eigene Position.

(b) Beispielhafter Filter für Themen, Epoche und Art des Mediums.

(c) Karte im eingezoomten Zustand. Die Kunstobjekte werden mit dem KuKoNö-Icon markiert.

Abbildung 1. KuKoNö Konzept-Screens im Modus Entdecken.

Sehenswürdigkeiten und die damit verknüpfte Geschichte St. Pöltens hin.[1]

Anders bei KuKoNö: Ausgangspunkt ist hier der aktuelle Standort der Benutzer*innen. Dieser wird auf einer Karte auf dem eigenen mobilen Endgerät angezeigt. Mit Hilfe der Karte können die Kulturschätze gefunden und erlebbar gemacht werden. Dies lehnt sich an das Konzept der kartengestützten LBS aus dem Tourismus an, die praktische, kulturelle oder historische Informationen verorten, explorierbar, findbar und ortsbezogen erlebbar machen. Vergleichbare Apps, wie sie bereits für verschiedene Regionen implementiert wurden (siehe bspw. [4]), gibt es bislang für St. Pölten noch nicht.[2]

Eine Besonderheit der App besteht darin, dass ein virtueller Begleiter, das KuKoNö, die Benutzer*innen durch St. Pölten zu dessen Kulturschätzen führt. Das KuKoNö ist ein blasenförmiges Wesen in den Niederösterreichischen Landesfarben Blau und Gelb, das den Geist der Zeiten verkörpert. Seine Form ist von den tropfenförmigen Standortmarkierungen auf digitalen Landkarten inspiriert. Im Rahmen der Tour agiert das KuKoNö als virtueller Tourguide und stellt die Objekte aus den Landessammlungen an Orten vor, die mit diesen in Verbindung stehen. Es kann mit den Benutzer*innen der App sowie mit Personen und Tieren interagieren, die auf diesen Objekten dargestellt sind.

A. Modi

Die Benutzer*innen haben die Auswahl zwischen zwei verschiedenen Modi, um in Begleitung des Kulturgeists die Kulturschätze zu finden und zu erkunden: Entdecken und Tour.

1) Modus Entdecken: Im Modus *Entdecken* (Abb. 1) ist die Karte von Niederösterreich der Ausgangspunkt (Abb. 1a). Objekte können auf der Karte selbstständig gefunden werden. Die eigene Position wird auf der Karte durch einen grauen Marker dargestellt. Befinden sich mehrere Kulturobjekte nahe beieinander, werden diese zusammengefasst und geclustert als grüner Kreis mit der Anzahl der Objekte dargestellt. Beim Hineinzoomen werden die Cluster aufgelöst und so die genauen Positionen der Objekte sichtbar gemacht. Die Benutzer*innen können durch Betätigen des “In der Nähe”-Buttons den Fokus auf Objekte in ihrer unmittelbaren Umgebung legen. Es ist möglich, Objekte zu suchen oder die Objekte nach bestimmten Kategorien z.B. Epoche (Jahr), Medium (Objekttyp), Themen (Sammlungen z.B. Kunst, Landeskunde, Europäische Ur- und Frühgeschichte, Römerzeit) zu filtern (Abb. 1b). Die Benutzer*innen können sich die Objekte damit nach eigenen Interessen auswählen und so eine personalisierte Tour zusammenstellen. Der Stadtraum wird damit zur “Database City” [6], die Karte zum Koordinatensystem für Informationen und Geschichte(n) [7]. Zudem kann man sich von KuKoNö von Objekt zu Objekt navigieren lassen.

2) Modus Tour: Im Modus *Tour* (Abb. 2) haben die Benutzer*innen die Auswahl aus kuratierten Touren zu bestimmten Themen (z.B. zu Kunst im öffentlichen Raum, Wachau, Themenführung Wein, kindgerechte Familientour) oder Highlight-Touren (z.B. St. Pölten Highlights). Diese Touren haben eine vorgegebene Dauer. Die Benutzer*innen können aus einer Liste (Abb. 2a) die gewünschte Tour starten. Hierbei können einerseits verschiedene Interessenschwerpunkte berücksichtigt

[1] Eine vergleichbare ortsfixierte Infostation existiert z.B. in Berlin mit “Timescope”: Hier kann man durch Ferngläser historische Ansichten über den aktuellen Standpunkt blenden [3].

[2] Dr. Edith Blaschitz entwickelte gemeinsam mit dem Zentrum für angewandte Spieleforschung der Donau-Universität Krems eine Augmented Reality App für die ehemalige Tabakfabrik in Krems [5]

(a) Auswahl verschiedener Touren.

(b) St. Pölten Highlights – Übersicht des angewählten Objekts “Herrenplatz”.

(c) St. Pölten Highlights – Detailansicht Radnetz Y Nr. 1 mit AR. Um KuKoNö zu finden, wird das Smartphone über das Radnetz bewegt. © Landessammlungen Niederösterreich.

Abbildung 2. KuKoNö Konzept-Screens im Modus Tour.

werden, andererseits kann den besonderen Bedürfnissen von Einzelpersonen sowie Personen mit Kindern Rechnung getragen werden.

B. Objekte im Detail

Bei den Kulturschätzen angekommen, können die Benutzer*innen sich in der Detailansicht über die Objekte informieren. Aus den bereits existierenden Daten der Landessammlung stehen hier folgende Informationen zur Verfügung: Titel, Beschreibung (nicht immer vorhanden), Person(en) (z.B. Künstler*innen, Hersteller*innen), Bild(er), Entstehungsjahr, Sammlung, Objekttyp.

Die Daten, die von der Landessammlung zur Verfügung gestellt werden, dienen als Basis für die App. Diese werden, sofern sinnvoll, mit weiteren Daten angereichert. Es werden Daten der GND (Gemeinsame Normdatei) [8], VIAF (Virtual International Authority File) [9], WikiData [10] und der Wikipedia Verwendung finden. Die Daten werden in unterschiedlichen Formaten (MARC21-xml, RDFxml, RDF etc.) von den jeweiligen Anbietern angeboten und können in den meisten Fällen mittels API direkt in den Datenkorpus der Landessammlung eingebunden werden. Beispiel zur Verwendung von Daten anderer Quellen: Person(en) (z.B. Künstler*innen, Hersteller*innen): Durch die Verknüpfung dieses Eintrags zur Person mit Daten aus der GND (Geburtsort, Lebensdaten, Beruf etc.) und dem Import des Wikipedia-Eintrags werden den User*innen der App zusätzliche Informationen angeboten, welche eine intensivere Auseinandersetzung mit dem Thema ermöglichen. Des weiteren sind Verlinkungen zur Topothek [11] oder Wikimedia Commons [12] denkbar, damit interessierte Nutzer*innen die Möglichkeit erhalten, ihr Erlebnis zu vertiefen.

Bei ausgewählten Objekten kommt Augmented Reality (AR) zum Einsatz. So können die Benutzer*innen bspw. an der richtigen Position mit Blick in die richtige Richtung alte Fotografien sehen, die sich über die physische Realität der Jetztzeit legen. Historische Abbildung und körperliche Präsenz im Hier und Jetzt verschmelzen und versetzen die Nutzer*innen über die informationelle Erweiterung der Realität in die historische Zeit zurück.

Je nach Zielgruppe werden Gamification-Elemente von unterschiedlicher Komplexität in das Führungsprofil integriert, mittels derer real existierende Objekte vor Ort spielerisch erkundet werden können. In der Familientour können Kinder beispielsweise das KuKoNö aus einem Spinnennetz befreien (Abb. 2c). Quizfragen, die nur vor Ort gelöst werden können, sollen Jugendliche und Erwachsene ansprechen, zu intensiver Betrachtung und Auseinandersetzung anregen und auf weitere Objekte neugierig machen.

III. Tour St. Pölten Highlights

Das im folgenden vorgestellte Beispiel “Tour St. Pölten Highlights” steht exemplarisch für den oben beschriebenen Modus “Tour”, der kuratierte Führungen zu ausgewählten Themen bereit stellt.

Nachdem die App aktiviert ist, wird den Benutzer*innen der aktuelle Standort angezeigt und die Tour kann am nächstgelegenen Punkt beginnen. Gezeigt werden Objekte aus den Landessammlungen Niederösterreich (Kunst, Archäologie, Landeskunde und Naturkunde). Die gezeigten Sammlungsgegenstände stehen in Beziehung zu Objekten im Stadt-gebiet. Die Info besteht aus Bild, kurzer Objektbeschreibung (Sammlungsbereich, Datum, eventuell Technik, Beschreibung, Künstler*innen/Hersteller*innen), den aus anderen Quellen eingebundenen Daten (z.B. Wikipedia-Einträge zu Künstler*innen bzw. Institutionen) sowie weiterführenden Links (zu Gebäuden, weiteren Kunstwerken bzw. Objekten, Museen und Ausstellungen, zur Website publicart, Website Landessammlungen – in Zukunft auch auf die Onlinesammlung). Die Tour (Abb. 2b) wird durch sechs ausgearbeitete Beispiele präsentiert, welche in Folge durch weitere Objekte der Landessammlungen ergänzt werden können.

Bahnhof Aktie Mariazellerbahn
Sammlung Landeskunde, LK197/173
Beschreibung Aktie, Localbahn St. Pölten-Kirchberg a.d. Pielach-Mank, 200 Kronen, Wien, 1898 (heute Mariazellerbahn). Im Juli 1896 wurde die Konzession für die formell eigenständige Aktiengesellschaft der Lokalbahn St. Pölten – Kirchberg an der Pielach – Mank erteilt. Im selben Jahr wurde mit dem Bau durch das Niederösterreichische Landeseisenbahnamt begonnen. Die Stammstrecke von St. Pölten nach Kirchberg und die Zweigstrecke nach Mank wurden am 4. Juli 1898 eröffnet, Betriebsführer beider Strecken war das Landeseisenbahnamt selbst.

Rathausplatz Der Rathausplatz, Rudolf Pichler
Sammlung Kunst
Beschreibung Öl auf Leinen, 82 x 68 cm, 1969. Das Bild des niederösterreichischen Künstlers zeigt die Hauptfassade der ehem. Karmeliterkirche in St. Pölten. Sie befindet sich am südlichen Ende des Rathausplatzes neben dem Rathaus.

Englische Fräulein Besuch Franz Josef I (Abb. 3)
Sammlung Landeskunde, LK1714/a-b
Beschreibung Silbergelatine, 9 x 14 cm, 1910. Am 21. Juni 1910 besuchte Kaiser Franz Joseph die Stadt St. Pölten, um das neue Schützenhaus und die regulierte Traisen zu besichtigen. Von diesem Ereignis ist nicht nur eine Photoserie erhalten, auch ein Film wurde gedreht.

Abbildung 3. Tour St. Pölten Highlights – Detailansicht Kaiser Franz Josef. © Landessammlungen Niederösterreich.

Riemerplatz/Rathausgasse Radnetz Y Nr. 1, Walter Berger (Abb. 2c)
Sammlung Kunst im öffentlichen Raum, PA-279
Beschreibung Walter Bergers künstliches, doch exakt dem Bauplan einer Kreuzspinne folgendes Netz aus platiniertem Stahldraht verweist auf eine Naturbezogenheit, in der sich nicht nur romantisches Empfinden, sondern auch kosmische Ordnungsprinzipien widerspiegeln. Die Kunstsammlungen der Landessammlungen Niederösterreich besitzen auch noch die Entwurfszeichnungen zu diesem Projekt (KS-13582/1-6).

Herrenplatz Herrenplatz in St. Pölten, Reinhold Kukla
Sammlung Kunst, KS-9645
Beschreibung Aquarell 1926. Das Aquarell zeigt das Aussehen des Herrenplatzes Anfang des 20 Jahrhunderts. Mittig erhebt sich die Mariensäule und dahinter erkennt man den Übergang zum Domplatz.

Kulturbezirk Shedhalle St. Pölten, Margherita Spiluttini (Abb. 4)
Sammlung Kunst, KS-16787/3
Beschreibung C-Print, 40,5 x 49 cm, 1996-1997. 1997 wurde die Ausstellungshalle mit der Ausstellung “1000 Jahre Österreich” feierlich eröffnet. Das Foto zeigt das Aussehen der Shedhalle, wie sie wegen der spitz zulaufenden Oberlichtern genannt wird, vor der Erweiterung zum Landesmuseum (heute Museum Niederösterreich). Die Halle wurde von Hans Hollein geplant, das charakteristische, geschwungene Vordach wurde in die neue Planung miteinbezogen.

Abbildung 4. Tour St. Pölten Highlights – Detailansicht Shedhalle (Foto M. Spiluttini) mit AR. Wird das Smartphone an der richtigen Position in Richtung Museum Niederösterreich gehalten, wird das Kamerabild mit der Ansicht der Shedhalle von 1996 überlagert. © Landessammlungen Niederösterreich.

IV. Fazit

KuKoNö ist eine mobile Applikation für das spielerische und erlebnis-orientierte Erkunden von Kulturgütern in ihrem Wirkungsraum und fungiert als erweiternde Informationsebene zur physischen Realität. Durch das Erstellen und Generieren von Touren mit dem KuKoNö als virtuellem Guide ist es den Benutzer*innen möglich, eine Reise durch vergangene Kulturen Niederösterreichs zu unternehmen, Informationen über die dargestellten Kulturgüter zu erhalten und diese zu sammeln. Die Zielgruppen sind insbesondere durch die Gamification-Elemente der App breit gefächert und reichen von Kindern, Familien bis zu Schüler*innengruppen. Das Erlebnis der Sammlungsobjekte in objektspezifischem Kontext und die bewusste Interaktion und Auseinandersetzung stehen im Mittelpunkt des Konzepts.

Danksagung

Diese Arbeit wurde teilweise vom Bundesministerium für Wissenschaft, Forschung und Wirtschaft im Rahmen des FFG-COIN-Programms (MEETeUX Projekt, Nr. 7209770) sowie der Internet Privatstiftung Austria (IPA) durch ein Stipendium (Nr. 1587) unterstützt.

Literatur

[1] D. Takahashi, "Analytics firms show how pokémon go became a phenomenon," 2016, http://venturebeat.com/2016/07/14/analytics-firms-take-us-inside-the-pokemon-go-henomenon/, ausgehoben am 26.07.2016.

[2] Amt der Niederösterreichischen Landesregierung, Abteilung Kunst und Kultur, "Landessammlungen Niederösterreich," 2017, http://landessammlungen-noe.at/, ausgehoben am 26.10.2017.

[3] ART+COM Studios, "Timescope," 1996, https://artcom.de/project/timescope/, ausgehoben am 26.10.2017.

[4] Fluxguide Ausstellungssysteme GmbH, "Beispiele Fluxguide App-Projekte: Tourismus und Regionen," 2017, http://www.fluxguide.com/public/content/websites/fluxguide/upload/apptourismde1.pdf, ausgehoben am 26.10.2017.

[5] Donau-Universität Krems, "Rückschau und Zeitreise: Die Geschichte der Tabakfabrik Stein," 2017, https://www.donau-uni.ac.at/de/service/kultur/programm/archiv/25395/index.php, Ausgehoben am 26.10.2017.

[6] E. Gordon, *The Urban Spectator. American Concept-Cities from Kodak to Google.* Hanover, New Hampshire: Dartmouth College Press, 2010.

[7] E. Gordon and A. de Souza e Silva, *Net Locality – Why Location Matters in a Networked World.* West Sussex, UK: Wiley-Blackwell, 2011.

[8] Deutsche Nationalbibliothek, "Gemeinsame Normdatei (GND)," 2016, http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html, ausgehoben am 26.10.2017.

[9] OCLC, "VIAF: The Virtual International Authority File," 2017, http://www.viaf.org/, ausgehoben am 26.10.2017.

[10] "Wikidata," 2017, https://www.wikidata.org, ausgehoben am 26.10.2017.

[11] ICARus – Internationales Zentrum für Archivforschung, "topothek," 2017, http://www.topothek.at/de/, ausgehoben am 11.11.2017.

[12] Wikimedia, "Wikimedia commons," 2017, https://commons.wikimedia.org/wiki/Main_Page, ausgehoben am 11.11.2017.

Communities in biographischen Netzwerken

Ágoston Zénó Bernád, Maximilian Kaiser
Inst. f. Neuzeit- und Zeitgeschichtsforschung (INZ)
Österreichische Akademie der Wissenschaften

Sebastian M. Mair
Technische Universität Wien

Alexander Rind
Inst. f. Creative\Media/Technologies
Fachhochschule St. Pölten

***Abstract*—Biographical lexica are a rich data source for the Digital Humanities. For example, the connections between places can be studied based on the migrations of scholars. The work at hand resulted from the OpenGLAM.at Cultural Data Hackathon 2017 and describes the analysis of 151 biographies from the *Austrian Biographical Dictionary 1815–1950*. Community detection algorithms were applied to find groups of places that are densely connected internally and sparsely connected between groups. The resulting communities were examined in detail using network visualization.**

***Zusammenfassung*—Biographische Nachschlagewerke sind eine reichhaltige Quelle für Digital Humanities. Beispielsweise können Zusammenhänge zwischen Orten durch die Migration der dort wirkenden Personen erforscht werden. Vorliegender Beitrag entstand im Rahmen des OpenGLAM.at Kulturhackathons 2017 und beschreibt die Analyse von 151 Biographien aus dem *Österreichischen Biographischen Lexikon 1815–1950*. Mittels Communitiy Detection wurden stärker verknüpfte Bereiche des Migrationsnetzwerks ermittelt. In der Visualisierung wiederum konnten diese Communities dann näher erforscht werden.**

I. Einleitung

Die Erforschung historischer Personennetzwerke anhand digitalisierter biographischer Nachschlagewerke rückte in den letzten Jahren, wie dies u. a. die Projekte *Deutsche Biographie* [1], [2] und *BiographyNet* [3], [4] zeigen, immer mehr in den Fokus der Digital Humanities [5]. Zu diesen, nach dem Prinzip des Semantic Web konzipierten Vorhaben, gehört auch *Mapping historical networks: Building the new Austrian Prosopographical | Biographical Information System (APIS)*, das sich die semantische Anreicherung des *Österreichischen Biographischen Lexikons 1815–1950* (ÖBL) mittels Methoden der Computerlinguistik zum Ziel setzt [6], [7, S. 464–467].

Das seit 1954 erscheinende ÖBL stellt ein paradigmatisches Beispiel einer grenzüberschreitenden Nationalbiographie dar: Es umfasst nicht nur das Gebiet des heutigen Österreich, sondern den gesamten Raum des Kaisertums bzw. der österreichisch-ungarischen Monarchie und bietet somit ein Abbild zentraleuropäischer Kultur mit all ihren Wandlungen im 19. und 20. Jahrhundert. Die Printversion enthält in bis dato 14 Bänden (68 Lieferungen) über 18.000 Biographien im Buchstabenbereich A–Vo. Das Lexikon durchlief mehrere Etappen der Digitalisierung, die Online-Edition mit Biographien in einem Ad-hoc-XML-Format wurde 2009 umgesetzt [8], [9]. Die XMLs enthalten Name, Geburts- und Sterbedaten (einschließlich Orte), Berufs- sowie Werks- und Literaturangaben in strukturierter Form. Die meisten Informationen befinden sich jedoch im unstrukturierten Haupttext der Biographien.

Im Rahmen des OpenGLAM.at Kulturhackathons 2017 hat sich ein interdisziplinäres Team mit der Erforschung biographischer Daten beschäftigt und ist dabei vor allem folgenden zwei Forschungsfragen nachgegangen:

- Welche Städte hängen über die Wanderungsbewegungen von Personen zusammen (am Beispiel der beruflichen Tätigkeiten „Historiker“ und „Lehrer“)?
- Wie kann man diese Zusammenhänge anhand der extrahierten Named Entities[1] identifizieren und visualisieren?

II. APIS Datenset Humanities Scholars

Der untersuchte Datensatz [10] entstand im Frühjahr 2016 in Zusammenhang mit einer Projektpräsentation bei der Konferenz „Entangled Worlds – Vernetzte Welten“ am Institut für Mittelalterforschung der ÖAW [11]. Er basiert auf einer Auswahl von 151 Biographien aus dem ÖBL, in denen Orte und Institutionen annotiert wurden. Jede Biographie kann neben ihrer Zuordnung zu einer von 18 Hauptberufsgruppen („Literatur, Buch- und Zeitungswesen“, „Medizin“, „Religionen und Theologie“, „Sozial- und Wirtschaftswissenschaften“, „Politik“, „bildende und angewandte Kunst“ usw.), abhängig vom Wirkungsfeld der biographierten Person, auch noch bis zu 5 berufliche Tätigkeiten zugewiesen bekommen. Jede dieser Tätigkeiten („Altphilologe“, „Kunsthistoriker“, „Bibliothekar“, „Archivar“, „Volkskundler“ usw.) verweist wiederum auf eine der Hauptberufsgruppen. Durch die vielseitigen Lebens- und Karrierewege entstehen so Verflechtungen zwischen diesen Gruppen. Um dieses Phänomen der Interferenz näher zu untersuchen, fiel die Wahl aus thematischen Gründen auf Biographien von „Historikern“, die zu der Gruppe „Geisteswissenschaften“ zählen. Des Weiteren wurden die Daten durch das Kriterium der drei am häufigsten vorkommenden beruflichen Tätigkeiten aus der Berufsgruppe „Unterrichtswesen“ („Lehrer“, „Pädagoge“ und „Schulmann“) nochmals gefiltert.

Die Annotationen erfolgten in der am Austrian Centre for Digital Humanities (ÖAW | ACDH) entwickelten APIS Webapplikation [12], [13, S. 221–222]. Mittels eines Highlighter-Tools können so die Biographietexte unmittelbar über die Oberfläche bearbeitet, die darin enthaltenen Named Entities annotiert und mit Linked-Open-Data-Ressourcen verknüpft werden (Abbildung 1). Aus Datenbanken wie der Gemeinsamen Normdatei (GND) [14] und GeoNames [15] werden dabei je nach Quelle Informationen, wie z. B. Labels von Institutionen, Lebensdaten von Personen oder Geokoordinaten von

[1] Als Named Entities werden Satzelemente bezeichnet, die für Namen von Personen, Orten oder Institutionen stehen können.

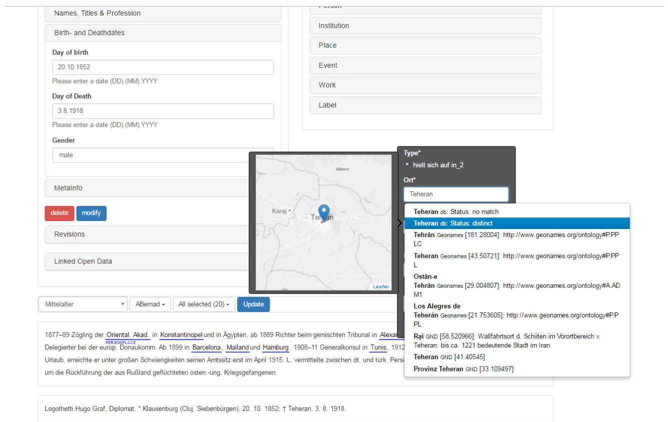


Abbildung 1. APIS Webapplikation

Orten, automatisch übernommen. Die APIS Webapplikation ist als virtuelle Forschungsumgebung konzipiert und bietet dem User neben Möglichkeiten der Datenerfassung, auch die der Datenvisualisierung (z. B. als Karten oder Netzwerkdarstellung) und des Datenexports in verschiedenen Formaten (JSON, GraphML, CSV, Excel).

III. Community Detection

Seit den Anfängen der sozialen Netzwerkanalyse steht die Untersuchung sozialer Strukturen und Teilgruppen von Graphen (engl. communities) im Fokus der Forschung. Als Thema lässt es sich über die Klassiker der Forschungsliteratur, beginnend in den 1930er-Jahren mit *Who shall survive?* von Moreno [16] bis zu der Untersuchung *Robust action and the rise of the Medici, 1400–1434* von Padgett und Ansell in den 1990er-Jahren [17] nachvollziehen. Im Kontext der historischen Netzwerkforschung findet dieses Thema in unterschiedlichen Bereichen wie z. B. der Mittelalterforschung [18], der Wirtschafts- und Sozialgeschichte [19] oder der Zeitgeschichte [20] Anwendung.

Vor dem Beginn einer derartigen Untersuchung war eine Vorbereitung der Daten erforderlich. Der APIS-Datensatz liegt als Netzwerk von Personen vor, die mit Orten verbunden sind. Diese Architektur wird als Bipartiter Graph bezeichnet. Zur weiteren Analyse wurde dieses Netzwerk auf einen einfacheren ungerichteten Graphen aus Orts-Knoten mit gewichteten Kanten projiziert. Das Gewicht der Kanten berechnet sich hierbei aus der Anzahl der Personen, welche mit zwei Orten verbunden sind. Daraus ergibt sich eine Nachbarschaftsmatrix A_{ij}. Jeder Knoten entspricht einer Zeile/Spalte. Falls zwei Knoten nicht über eine Kante verbunden sind, gilt $A_{ij} = 0$ andernfalls ist $A_{ij} = \omega_{ij}$ das Kantengewicht. Für diese Projektion wurde die Python Bibliothek networkx [21] eingesetzt und der resultierende Graph zur weiteren Analyse zurück in das GraphML Format übertragen.

Die zuvor formulierten Fragestellungen erfordern es zusätzlich zur Bestimmung der gängigen mikroskopischen und makroskopischen Eigenschaften wie Gradzentralität oder Netzwerkdurchmesser, vor allem die mesoskopischen Eigenschaften des konstruierten Netzwerks zu identifizieren. Welche Gruppen von Orten sind gleichermaßen mit mehreren Personen assoziiert. Übersetzt in Netzwerksprache lautet die Frage: Welche Bereiche innerhalb des Netzwerks sind untereinander stärker verknüpft als mit dem restlichen Netzwerk? Stärker verknüpfte Bereiche innerhalb eines größeren Netzwerks werden als Communities bezeichnet, wobei es zur Zeit keine präzisere, allgemein anerkannte Defintion einer Community gibt [22]. Basierend auf unterschiedlichen mathematischen Definitionen von Communities wurde eine Reihe von Community Detection Algorithmen entwickelt, wobei wir den Ansatz der Modularitätsmaximierung mit Grivan-Newman Null-Modell [23] mittels Louvain-Algorithmus [24] angewandt haben. Diese Konzepte sollen im folgenden Abschnitt näher erläutert werden.

Die Modularität Q misst die Wahrscheinlichkeit, dass eine Partitionierung $\Sigma = \{\sigma\}$ des Netzwerks in Gruppen von Knoten vom Zufall abweicht. Dies wird im Vergleich zu einem gegebenen Null-Modell p_{ij}, das unser Wissen ob der Konstruktion des Netzwerkes beinhaltet, berechnet.

$$Q = -\sum_{i \neq j} (A_{ij} - \gamma p_{ij}) \delta_{\sigma_i, \sigma_j} \qquad (1)$$

wobei A_{ij} die Nachbarschaftsmatrix ist. σ_i gibt die Community, welcher der Knoten i angehört, an. $\delta_{\sigma_i, \sigma_j}$ ist das Kronecker-Delta, das den Wert 1 annimmt falls $\sigma_i = \sigma_j$ andernfalls den Wert 0. Modularitätsmaximierung hat Probleme, kleine Communities zu finden, dieser Umstand kann durch geeignete Wahl der Auflösung γ kompensiert werden [25]. In Spezialfällen ist es möglich die optimale Auflösung algorithmisch abzuschätzen [26], darüber hinaus gibt es Verfahren plausible Wertebereiche zu finden [27]. In der Praxis wie auch in dieser Arbeit wird die Auflösung allerdings iterativ und mittels Domänenwissen im Rückschluss auf erzielte Netzwerkpartitionierungen angenähert. Das genutzte Girvan-Newman Null-Modell beschreibt, dass jeder Knoten eine gleich große Wahrscheinlichkeit hat, mit jedem anderen Knoten im Netzwerk verknüpft zu sein. So ist die Wahrscheinlichkeit zweier Knoten i und j mit k_i und k_j Verknüpfungen eine gemeinsame Kante zu haben

$$p_{ij} = \frac{k_i k_j}{2m} \qquad (2)$$

wobei m die Summe über alle Kantengewichte im Netzwerk ist. Die Maximierung dieser Modularitätsfunktion Q stellt ein NP-hartes Problem dar, der eingesetzte Louvain-Algorithmus ermittelt daher keine exakte, jedoch eine gute Lösung: Zuordnung der Knoten i zu Communities σ_i. In dieser Arbeit wurde die Community Detection über die graphische Oberfläche von Gephi [28] durchgeführt, und bei einer Auflösung von $\gamma = 9.5$ 49 Communities ermittelt.

IV. Visualisierung

Um die Communities von Orten zu untersuchen und abzuschätzen wie gut sich die Community Detection-Algorithmen für diese Daten eignen, wurden die Visualisierungsysteme

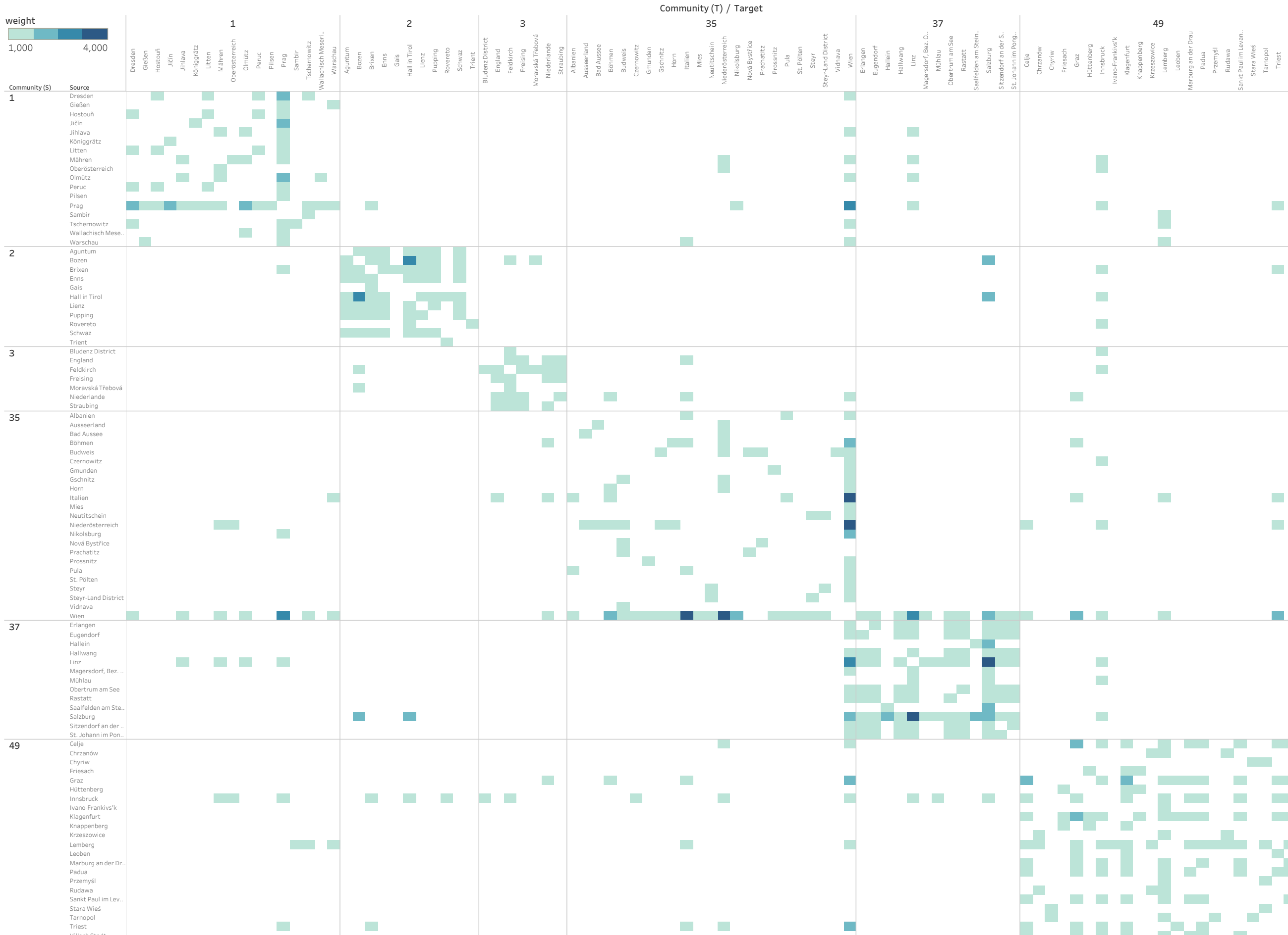


Abbildung 2. Visualisierung der Nachbarschaftsmatrix gruppiert nach Communities (Ausschnitt): Jede Zelle stellt die mögliche Verbindung zwischen zwei Orten aus Zeile und Spalte dar. Die Farbe der Zelle ist umso dunkler je mehr Personen bei beiden Orten erwähnt werden, also je höher das Kantengewicht ist. Ohne Kante bleibt die Zelle weiß. Die stärkere Vernetzung innerhalb der Communities ist hier deutlich sichtbar.

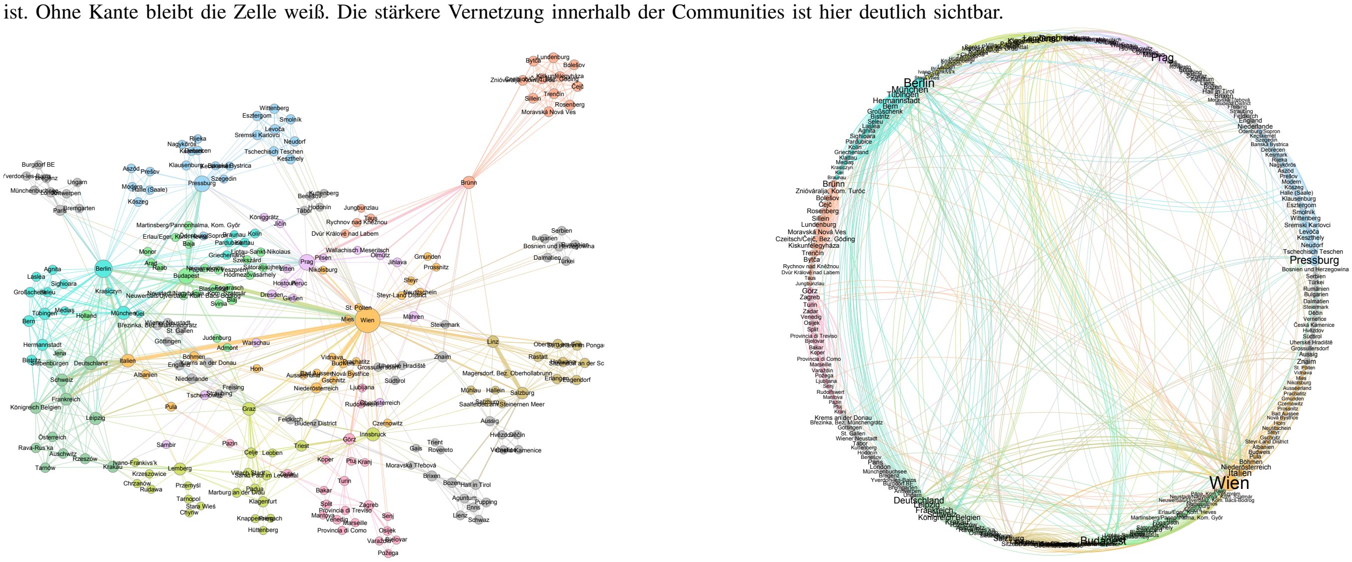


Abbildung 3. Visualisierung der Orte als Netzwerkknoten mit Verbindungen durch Wanderungsbewegungen. Die Farbe der Knoten entspricht der Community und die Größe dem Knotengrade. Links sind die Knoten auf Basis von Kräften positioniert und rechts in einem Kreis sortiert nach Community.

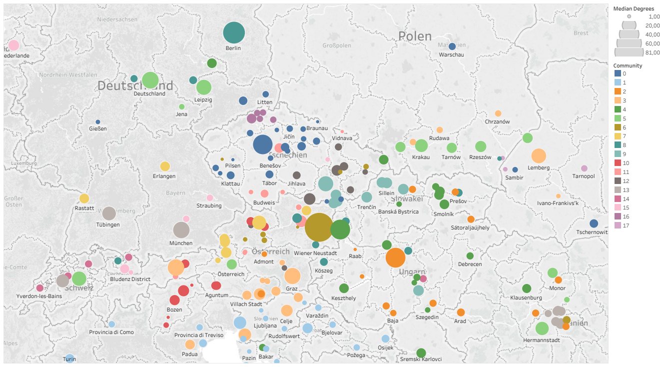


Abbildung 4. Visualisierung der Orte auf einer Landkarte. Die Farbe der Knoten entspricht der Community und die Größe dem Knotengrade.

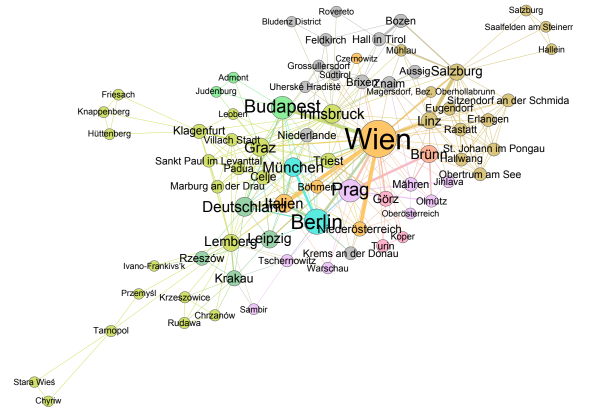


Abbildung 5. Filtern auf die Orte der Communities {Graz, Lemberg, etc.} und {Linz, Salzburg, etc.} und deren unmittelbarer Netzwerk-Nachbarschaft.

Gephi [28] und Tableau [29] eingesetzt. Derartige computerbasierte Visualisierungssyteme stellen interaktive visuelle Repräsentationen von Datensätzen zur Verfügung und unterstützen so Menschen dabei, ein mentales Modell der Datensätze zu entwickeln [30], [31]. Gerade im Bereich der Netzwerkvisualisierung gibt es ein umfangreiches Methodenrepertoire [32].

Prinzipiell bestehen die beiden Möglichkeiten Netzwerke als Matrix oder als Node/Link Diagramm darzustellen. Bei der Matrix-Darstellung wird jede Kante durch eine Tabellenzelle repräsentiert, deren Zeile und Spalte von Ursprungs- und Zielknoten abhängen. Diese Darstellung hat den Vorteil, dass auch größere Netzwerke übersichtlich dargestellt werden können. Wie in Abbildung 2 ersichtlich lassen sich bei geeigneter Sortierung der Knoten Communties gut überprüfen. Im Node/Link-Diagramm werden Knoten als Markierungen wie z. B. Kreise und Kanten als Linien zwischen diesen dargestellt. Diese Darstellung ist weitgehend bekannt und wird häufig verwendet, da es leichter möglich ist Pfaden über mehr als eine Kante hinweg zu folgen. Zur Anordnung der Knoten gibt es verschiedene Algorithmen – unter anderem ein Ausbalanzieren von an- und abstoßenden Kräften zwischen den Knoten oder die Anordnung der Knoten in einem Kreis (vgl. Abbildung 3).

Auch auf einer geographischen Karte können die Communities angezeigt werden (Abbildung 4). Dabei war es von Vorteil, dass die Orte auf Basis der verknüpften Linked-Open-Data mit Geokoordinaten versehen werden konnten. Alternativ wäre eine Set Visualisierung [33] wie KelpFusion [34] möglich gewesen.

Obwohl die Abbildungen nur das größte zusammenhängende Teilnetzwerk bestehend aus 243 Orten enthalten, ist die Visualisierung bereits überladen. Durch Filtern lassen sich Teile des Netzwerks im Detail erkunden. Beispielsweise zeigt Abbildung 5 Orte zweier Communities sowie Orte anderer Communities, die direkt mit jenen verbunden sind. Durch Verknüpfung mit dem Personen–Ort Netzwerk können auch die zugrundeliegenden Biographien in die visuelle Analyse einbezogen werden (Abbildung 6).

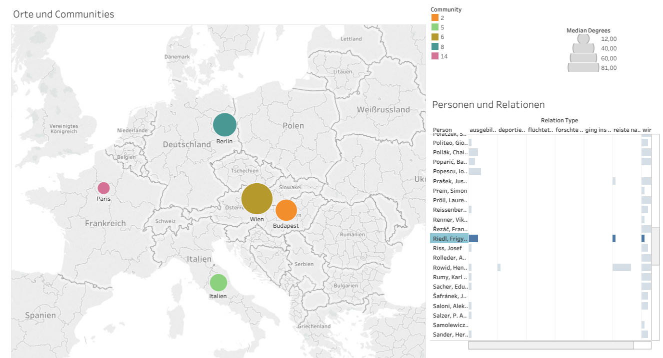


Abbildung 6. Interaktive Verknüpfung der Landkartenansicht von Orten mit den verknüpften Personen. Die gewählte Person hat eine außergewöhnliche Wanderbewegung durch Orte aus fünf Communities vollzogen.

V. Resümee

Anders als bei einer nach Zentralitätsmaßen wie Gradzentralität (engl. degree) oder Zwischenzentralität (engl. betweeness) geleiteten Darstellung, lassen sich über die Communities Gemeinsamkeiten in den Wanderungsbewegungen von Personen hervorheben und identifizieren. Wien war als Hauptstadt der Donaumonarchie bzw. Republik Österreich nicht nur politisches und administratives Zentrum, sondern auch für Wirtschaft, Gesellschaft und Kultur maßgebend. Die beruflichen Tätigkeiten der ausgewählten Personen wie „Lehrer“, „Historiker“ usw. beschränkten sich hingegen nicht ausschließlich auf die Metropolregionen. Mit der Auswahl des Samples einerseits und der Art von Analyse andererseits wird dadurch ein anderer Blickwinkel auf die Stadt und ihre Bedeutung ermöglicht. Es kann also zusammenfassend gesagt werden, dass es sich bei dieser Methode um einen vielversprechenden Ansatz zur Visualisierung und Analyse biographischer Daten handelt.

Danksagung

Dank aller Autoren gilt dem Organisationsteam des OpenGLAM.at Kulturhackathons sowie den TeamkollegInnen Christian Erlinger-Schiedlbauer, Janina Hoth und Barbara Kermer. Die Daten wurden im

Rahmen des durch den österreichischen Nationalfonds für Forschung, Technologie und Entwicklung geförderten Projekts *Mapping historical networks: Building the Austrian Prosopographical/Biographical Information System (APIS)* gewonnen und bereitgestellt. Die Analyse wurde teilweise durch Mittel des vorher genannten APIS-Projekts und des *KAVA-Time* Projekts (Austrian Science Fund (FWF): P25489-N23) gefördert.

LITERATUR

[1] „Deutsche Biographie," https://www.deutsche-biographie.de, Zugriff 24.10.2017.

[2] S. Stotz, V. Stuß, M. Reinert, und M. Schrott, „Interpersonal Relations in Biographical Dictionaries. A Case Study." in *BD2015. Biographical Data in a Digital World 2015. Proceedings of the 1st Conference on Biographical Data in a Digital World*, S. ter Braake, A. Fokkens, R. Sluijter, T. Declerck, und E. Wandl-Vogt, Hrsgg., Vol. 1399. CEUR-WS.org, 2015, S. 74–80.

[3] „BiographyNet," http://www.biographynet.nl, Zugriff 24.10.2017.

[4] S. ter Braake, A. Fokkens, und F. van Lieburg, „Mining Ministers (1572–1815). Using Semi-structured Data for Historical Research," in *Social Informatics: SocInfo 2014 International Workshops, Barcelona, Spain, November 11, 2014, Revised Selected Papers*, L. M. Aiello und D. McFarland, Hrsgg. Cham: Springer, 2015, S. 279–283.

[5] F. Jannidis, H. Kohle, und M. Rehbein, Hrsgg., *Digital Humanities: Eine Einführung*. Stuttgart: J.B. Metzler, 2017.

[6] Österreichische Akademie der Wissenschaften, Hrsg., *Österreichisches Biographisches Lexikon 1815–1950*, Wien, 1954ff.

[7] Á. Z. Bernád, „Das Österreichische Biographische Lexikon 1815–1950. Ein Nachschlagewerk im Wandel der Zeit," in *Hortus Amicorum. Köszöntőkötet Egyed Emese tiszteletére*, K. A. Bartha, A. Biró, Z. Demeter, und G. N. Tar, Hrsgg., Kolozsvár, 2017, S. 457–467.

[8] „Österreichisches Biographisches Lexikon 1815–1950 Online Edition," http://www.biographien.ac.at, Zugriff 24.10.2017.

[9] C. Gruber und R. Feigl, „Von der Karteikarte zum biografischen Informationsmanagementsystem. Neue Wege am Institut Österreichisches Biographisches Lexikon und biographische Dokumentation," in *Biografische Lexika im Internet. Internationale Tagung der „Sächsischen Biografie" in Dresden, 30.–31.05.2008*, M. Schattkowsky und F. Metasch, Hrsgg., 2009, S. 55–75.

[10] Á. Z. Bernád, M. Kaiser, M. Schlögl, und K. Lejtovicz, „APIS Dataset humanities scholars," Sep. 2017, https://doi.org/10.5281/zenodo.896281, Zugriff 24.10.2017.

[11] Á. Z. Bernád, M. Kaiser, K. Lejtovicz, und M. Schlögl, „Mapping historical networks. Working with biographical data," in *Entangled Worlds – Vernetzte Welten*, J. Preiser-Kapeller, Hrsg., Wien, im Erscheinen.

[12] „Testinstanz der Webapplikation APIS," https://apisdev.eos.arz.oeaw.ac.at, Zugriff 24.10.2017.

[13] K. Lejtovicz und M. Schlögl, „APIS – Eine Linked Open Data basierte Datamining-Webapplikation für das Auswerten biographischer Daten," in *DHd 2017. Digitale Nachhaltigkeit, Konferenzabstracts, 13.–18.02.2017*, Bern, 2017, S. 220–223.

[14] „Gemeinsame Normdatei (GND)," http://dnb.de/gnd, Zugriff 24.10.2017.

[15] „GeoNames," http://www.geonames.org/, Zugriff 06.11.2017.

[16] J. L. Moreno, *Who shall survive?: A new approach to the problem of human interrelations*. Washington: Nervous and Mental Disease Publishing Co, 1934.

[17] J. F. Padgett und C. K. Ansell, „Robust Action and the Rise of the Medici, 1400–1434," *American Journal of Sociology*, Vol. 98, Nr. 6, S. 1259–1319, 1993.

[18] R. Gramsch, *Das Reich als Netzwerk der Fürsten: Politische Strukturen unter dem Doppelkönigtum Friedrichs II. und Heinrichs (VII.), 1225–1235*. Ostfildern: Thorbecke, 2013.

[19] M. Stark, „Soziale Einbettung eines ländlichen Kreditmarktes im 19. Jahrhundert," Univ. Dissertation, Trier, 2012.

[20] M. Düring, „Hilfe für Verfolgte während des Nationalsozialismus," in *Vom Papier zum Laptop. Perspektiven elektronischer Tools zur partizipativen Visualisierung und Analyse sozialer Netzwerke*, M. Schönhut, M. Gamper, und M. Stark, Hrsgg. Bielefeld: transcript Verlag, 2011.

[21] A. A. Hagberg, D. A. Schult, und P. J. Swart, „Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Pasadena, CA USA, Aug. 2008, S. 11–15.

[22] S. Fortunato und D. Hric, „Community detection in networks: A user guide," *Physics Reports*, Vol. 659, S. 1–44, 2016.

[23] M. E. Newman und M. Girvan, „Finding and evaluating community structure in networks," *Physical review E*, Vol. 69, Nr. 2, S. 026113, 2004.

[24] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, und E. Lefebvre, „Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, Vol. 2008, Nr. 10, S. P10008, 2008.

[25] S. Fortunato und M. Barthélemy, „Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, Vol. 104, Nr. 1, S. 36–41, 2007.

[26] M. Newman, „Community detection in networks: Modularity optimization and maximum likelihood are equivalent," *arXiv preprint arXiv:1606.02319*, 2016.

[27] V. A. Traag, G. Krings, und P. Van Dooren, „Significant scales in community structure," *Scientific reports*, Vol. 3, 2013.

[28] M. Bastian, S. Heymann, und M. Jacomy, „Gephi: An Open Source Software for Exploring and Manipulating Networks," in *International AAAI Conference on Weblogs and Social Media*, 2009, http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154, Zugriff 25.10.2017.

[29] „Tableau Desktop," https://www.tableau.com/products/desktop, Zugriff 25.10.2017.

[30] S. K. Card, J. D. Mackinlay, und B. Shneiderman, Hrsgg., *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann, 1999.

[31] H. Schumann und W. Müller, *Visualisierung: Grundlagen und Allgemeine Methoden*. Berlin: Springer, 2000.

[32] S. Hadlak, H. Schumann, und H.-J. Schulz, „A Survey of Multi-faceted Graph Visualization," in *Proc. Eurographics Conference on Visualization – State of The Art Report, EuroVis STAR*, R. Borgo, F. Ganovelli, und I. Viola, Hrsgg. Eurographics, 2015, S. 1–20.

[33] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, und P. Rodgers, „The State-of-the-Art of Set Visualization," *Computer Graphics Forum*, Vol. 35, Nr. 1, S. 234–260, 2016.

[34] W. Meulemans, N. Henry Riche, B. Speckmann, B. Alper, und T. Dwyer, „KelpFusion: A Hybrid Set Visualization Technique," *IEEE Trans. Visualization and Computer Graphics*, Vol. 19, Nr. 11, S. 1846–1858, Nov. 2013.

Session 4: Visualization

Interaction Concepts for Collaborative Visual Analysis of Scatterplots on Large Vertically-Mounted High-Resolution Multi-Touch Displays

Mohammad Chegini*, Lin Shao*, Dirk J. Lehmann†, Keith Andrews‡ and Tobias Schreck*
*Institute of Computer Graphics and Knowledge Visualisation, Graz University of Technology, Austria
Email: {m.chegini, l.shao, t.schreck}@cgv.tugraz.at
†University of Magdeburg, Germany
Email: dirk@isg.cs.uni-magdeburg.de
‡Institute of Interactive Systems and Data Science, Graz University of Technology, Austria
Email: kandrews@tugraz.at

***Abstract*—Large vertically-mounted high-resolution multi-touch displays are becoming increasingly available for interactive data visualisation. Such devices are well-suited to small-team collaborative visual analysis. In particular, the visual analysis of large high-dimensional datasets can benefit from high-resolution displays capable of showing multiple coordinated views.**

This paper identifies some of the advantages of using large, high-resolution displays for visual analytics in general, and introduces a set of interactions to explore high-dimensional datasets on large vertically-mounted high-resolution multi-touch displays using scatterplots. A set of touch interactions for collaborative visual analysis of scatterplots have been implemented and are presented. Finally, three perception-based level of detail techniques are introduced for such displays as a concept for further implementation.

I. Introduction

Large high-resolution displays are becoming an affordable option for the visualisation of data [1]. Large displays have proved to be effective for tasks such as comparative genomics analysis [2], graph topology exploration [3], and sensemaking [4]. Large vertically-mounted (landscape-orientation) high-resolution multi-touch displays are particularly effective for collaborative analysis by small teams. However, previous research has often focused on horizontally-mounted tabletop surfaces or vertically-mounted displays with more distant interaction [5]. In this paper, a set of user interactions to support scatterplot matrices analysis on vertically-mounted displays are introduced. These techniques help analysts to efficiently select a scatterplot from scatterplot matrices and explore it collaboratively.

Some physical and virtual interactions with large displays were described in the previous literature. Modalities range from natural interactions like speech, body tracking, gaze, and gestures to the use of secondary control devices like mobile phones, tablets, or Wii controllers [6]. Of these, multi-touch interactions provide a fluid and intuitive interface suitable for up-close interaction in front of the display by small groups. Although there are studies about collaborative interaction with large displays (e.g. [7], [8]), they usually focus on single-user interaction [9]. Since typical multi-touch interactions do not support collaboration, more research needs to be done on cooperative gestures, modalities and the dynamics of group work around these devices. Cooperative gestures are known to enhance the sense of teamwork and increase the participation of team members [10].

Screen size and resolution are particularly important for information visualisation of multivariate datasets. Having a large display allows multiple, linked views, such as scatterplot matrices and parallel coordinates [11] to be provided simultaneously. If the screen is not high-resolution, the user experience of near distance interaction decreases significantly. For instance, on screens with less than sixty pixels per inch, the user is not able to read from the screen up-close [12]. Furthermore, users can make more observations with less effort using physical navigation (e.g., walking) rather than virtual [1]. More screen space can be used to either provide a better overview of a dataset or to provide more details of a portion of it. For example, users can see both an entire scatterplot matrix, specific scatterplots, and parallel coordinates plots at the same time. As a result, users may have the opportunity to gain more insight into large datasets.

Previous studies [5] suggest that vertically-mounted displays are more suited to parallel tasks within a group, due to reduced visual distraction and the possibility to share information through physical navigation like turning the head or walking. On tabletop displays, if users are not on the same side of the table, the shared view often needs to be reoriented.

This paper addresses the design gap between standard interaction techniques for large, multi-touch displays and advanced interaction techniques and visual feedback for collaborative scatterplot and scatterplot matrix analysis. Design concepts for such interaction techniques have been implemented as a proof of concept and are presented. The techniques include

Fig. 1. Two users collaboratively analyse a dataset on a large vertically-mounted multi-touch screen. User A on the left drags a Regression Lens, while user B on the right adapts the degree of the regression model using the floating toolbox. The display is an Eyevis 84-inch 4K/Ultra-HD 60Hz multi-touch LCD monitor with a resolution of 3840×2160.

scatterplot selection from scatterplot matrices, collaborative regression model analysis, and an extension of the Regression Lens [13] to include a floating toolbox. As a proof of concept, the techniques are developed on a large display.

The paper is structured as follows: Section II discusses related work. Several novel interaction designs for collaborative visual analysis of scatterplots on large displays are introduced in Section III. The use case and current implementation of the proposed interaction techniques are described in Section IV. Section VI introduces the concept of perception-based level of visual detail. The paper concludes with a discussion of open problems and future work in Section VII.

II. Related Work

At a high level, information visualisation systems consist of two components: visual representation and interaction. Visual representation concerns the mapping from data to display [14]. The interaction starts with a user's intent to perform a task, followed by a user action. The system then reacts and feedback is given to the user [15]. It is essential to consider both visual representation and interaction when designing an application for information visualisation.

A. Visualisation on Large Displays

Researchers in various fields are increasingly confronted with the challenge of visualising and exploring high-dimensional datasets [13], [16]. Keim argues that although many traditional techniques exist to represent data, they are often not scalable to high-dimensional datasets without suitable analytical or interaction design [16].

With the current size and resolution of typical computer displays, it is challenging to represent entire datasets on one screen using techniques like scatterplot matrices or parallel coordinates. The user is often forced to resort to panning and zooming, leading to frustration and longer task completion times. Ruddle et al. [17] conducted an experiment in which participants searched maps on three different displays for densely or sparsely distributed targets. They concluded that since the whole dataset fits on a larger display, sparse targets can be found faster.

Multiple linked views are often used to gain a better understanding of a high-dimensional dataset. Such views are usually connected by techniques such as brushing or combined navigation [18]. Every view occupies space on display. If more space is available, additional views can be shown simultaneously. Allowing the user to access multiple windows increases performance and satisfaction [19]. Isenberg et al. [20] present hybrid-image visualisation for data analysis, where two images are blended to achieve distance-dependent perception. This concept might be especially helpful for collaborative visual analysis tasks on vertically-mounted displays, where users observe data from various distances.

B. Visual Data Analysis and Multi-Touch Interaction

Previous researchers proposed various interaction techniques for large displays and multi-dimensional dataset interaction on multi-touch displays. Ardito et al. [18] proposed a classification of large display interaction having five dimensions: visualisation technology, display setup, interaction modality, application purpose, and location. Khan presented a survey of interaction techniques and devices for large, high-resolution displays [6]. The survey categorises modalities of interaction into speech, tracking, gestures, mobile phones, haptic and other technologies such as gaze and facial expression.

Tsandilas et al. presented SketchSliders [21], a tool that provides a mobile sketching interface to create sliders which interact with multi-dimensional datasets on a wall display. In comparison, in this paper, interaction is performed directly on the display rather than using a secondary touch device. Zhai et al. [22] introduced gesture interaction for wall displays based on the distance of the user from the screen. The gestures can be performed in far or near mode. Unlike the techniques described in this paper, the proposed interaction gestures are not directly related to visual analytics tasks. Heilig et al. [23] developed multi-touch scatterplot visualisation on a tabletop display. Sadana and Stasko [24] proposed advanced techniques for scatterplot data selection on smaller touch-based devices, such as tablets and smartphones, whereas this paper focuses on large multi-touch displays.

MultiLens supports various gestures for fluid multi-touch exploration of graphs [25]. The Regression Lens [13] allows the user to interactively explore local areas of interest in scatterplots by showing the best fitting regression models inside the lens. The idea of visualising local regression models is also studied by Matković et al. [26]. Rzeszotarski et al. [27] introduced Kinetica, a tool for exploring multivariate data by physical interactions on multi-touch screens. Kister et al. [25] presented BodyLenses, a promising set of magic lenses for wall displays, which are mostly controlled by body interaction and therefore suitable for interacting with wall displays from a distance.

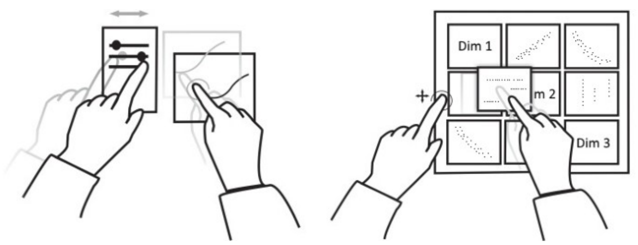


Fig. 2. On the left, a user is drags a Regression Lens with the right hand while adjusting the lens with the left hand. On the right, a user drags a scatterplot with the right hand while panning through the scatterplot matrix with the left hand.

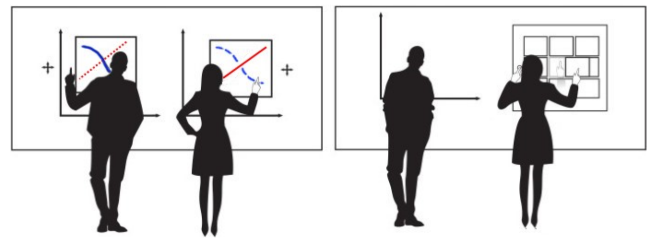


Fig. 3. On the left, two users collaboratively analyse a scatterplot. Both users create a regression model for a subset of selected data. The created models are displayed in their partner's respective lens as well, supporting comparison of local data models. On the right, one user analyses a scatterplot, while their partner selects interesting plots in the scatterplot matrix and passes them over by holding the background and swiping the right hand.

In comparison to this work, the aforementioned studies either focus on a different type of interaction and medium or are not designed for collaborative visual analytics tasks.

C. *Collaborative Visualisation*

Large displays are well-suited to collaboration [28], [29]. Jakobsen and Hornbæk [5] conducted an exploratory study to understand group work with high-resolution multi-touch wall displays. The study suggests that using this kind of display helps users to work more efficiently as a group and fluidly change between parallel and joint work. A large display benefits group working on a shared task, since users can operate on one common physical medium and share information on it.

Morris et al. [10] formalised the concept of cooperative gestures as a set of gestures performed by multiple users and interpreted as a single task by the system. Liu et al. developed CoReach [9], a set of gestures for collaboration between two users over large multi-touch displays. Comparing the use of a large vertically-mounted display against two ordinary desktop displays, Prouzeau et al. [30] concluded that groups obtain better results and communicate better on large, vertically-mounted displays.

An experiment by Pedersen and Hornbæk [31] showed that users prefer horizontal surfaces over vertically-mounted displays, but this result was limited to simple single-user tasks and not collaborative tasks with different dynamics. Vertically-mounted displays allow users to obtain an overview of their data by stepping back from the display and make it possible to interact from afar as well as up close. Badam et al. [32] proposed a system for collaborative analysis on large displays by controlling individual lenses through explicit mid-air gestures.

Although these studies are not directly related to collaborative scatterplot analysis on large multi-touch displays, they do provide valuable insights into the design process of such systems.

III. Proposed Interaction Techniques

Current standard multi-touch interaction techniques are not designed for collaboration on vertically-mounted high-resolution displays [9]. Here, both single-user and collaborative interactions are proposed for the analysis of scatterplots and scatterplot matrices on such devices. Some of the interaction techniques are based on the concept of the Regression Lens [13], which supports real-time regression analysis of subsets of a scatterplot through lens selection and manipulation. With Regression Lens, a user can select a local area in a scatterplot and observe the regression model of selected points [13]. Shao et al. proposed operations to adjust and manipulate the regression model shown in the Regression Lens, such as changing the degree of the regression model or inverting its axes. Figure 1 illustrates some of the suggested collaborative gestures on an 84-inch 4K/ULTRA-HD@60HZ multi-touch LCD monitor produced by Eyevis [33]. The user on the left finds interesting scatterplots and passes them to the user on the right. The user on the right analyses the plots using the Regression Lens [13]. In the rest of this section, four interaction designs for both collaborative and single scatterplot analysis are introduced. Later in Section IV, an implementation of these techniques is demonstrated.

A. *Lens and Floating Toolbox*

Magic lens techniques like DragMagics [34] and BodyLens [35] are used to explore local regions in a visualisation. An extended version of the basic lens concept provides for more fluid interaction with large multi-touch displays. For instance, as shown in Figure 2, after a region of interest has been selected in a scatterplot using the dominant hand (here the right hand), a toolbox appears next to the other side of the lens (near the non-dominant hand), where the user can use sliders and touch buttons to adjust the lens. For example, the user can change the degree of the regression model. The lens can be dragged with one hand, while being adjusted with the second hand, thus potentially speeding up performance.

B. *Two-Handed Interaction with Scatterplot Matrices*

A scatterplot matrix consists of pairwise scatterplots arranged in a matrix, with dimensions typically labelled in the diagonal cells. Since the number of dimensions is usually high, panning and zooming within the scatterplot matrix is almost inevitable. With common multi-touch interactions, the scatterplot or dimension label is dragged to the corner of

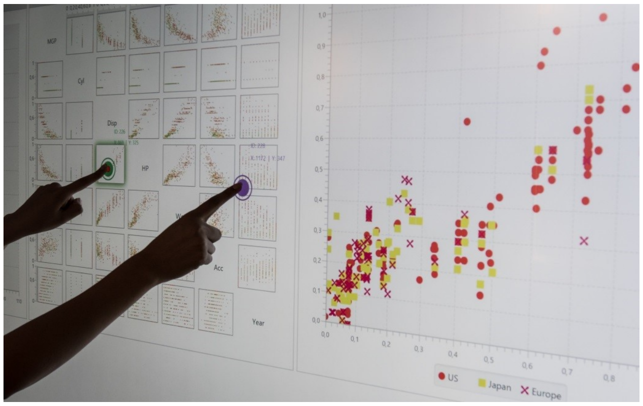


Fig. 4. A user selects a scatterplot of interest from a scatterplot matrix by touching and holding the left hand on the scatterplot. Swiping with the right hand then passes the selected scatterplot to the right hand side of the display for more detailed analysis.

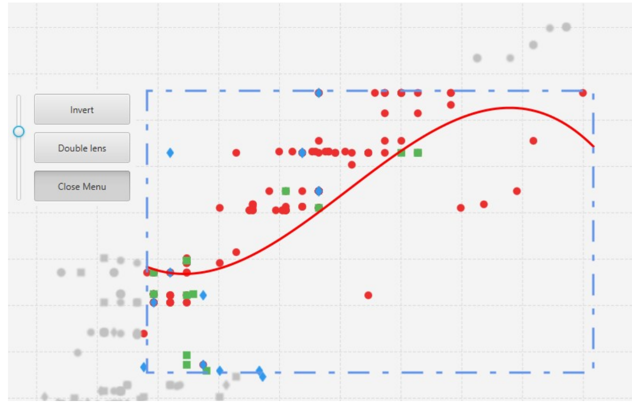


Fig. 5. A Regression Lens containing a cubic regression model is shown. At the left side of the Regression Lens, a floating toolbox with various options is visible.

the scatterplot matrix for panning. It is not feasible to zoom into or out of a scatterplot matrix while dragging another object. Based on two-handed interaction on tablets [36], a two-handed technique is proposed whereby the dominant hand is responsible for dragging items, while the non-dominant hand performs common operations. As shown on the left side of Figure 2, the user drags a scatterplot around to reorder the plots in the scatterplot matrix. Panning is performed by the non-dominant hand. With this two-handed technique, the interactions needed to reorder scatterplots in a scatterplot matrix can be reduced.

C. Collaboration using Gestures

On large vertically-mounted collaborative displays, it is not always desirable to move from one side of the screen to the other to perform a task. Instead, collaborative gestures can be used to pass objects. Based on the ideas of Liu et al. [9], collaborative gestures on scatterplots are proposed. In the right-hand side of Figure 3, the user on the left is analyses a scatterplot, while the user on the right selects another scatterplot of interest. By holding the background of the scatterplot matrix with one hand, and swiping with the other hand, the scatterplot is passed over to the partner. The partner can then decide whether or not to load the scatterplot for comparison. This technique can also be used for other tasks. For example, in Figure 4, the user selects a scatterplot of interest from a scatterplot matrix by touching and holding it with one hand (here, the left hand) and swipes the other hand in the direction of the analysis panel to load that scatterplot for more detailed analysis.

D. Collaborative Lens

In collaborative analysis, visual feedback plays an essential role. When two analysts work on a vertically-mounted display without proper visual feedback, they need to communicate more and turn their heads more often. A collaborative lens can help ameliorate this issue. As illustrated on the left side of Figure 3, the user on the left side of the screen creates a regression lens and regression model in blue. Meanwhile, the user on the right side of the screen creates their regression lens and regression model in red. Both users can see the other user's regression model reflected in their own regression lens.

IV. Implementation

Proof-of-concept interaction techniques for single-user and collaborative analysis of scatterplots and scatterplot matrices have been implemented on a vertically-mounted Eyevis 84-inch multi-touch display with a resolution of 3840 × 2160 pixels and a frame rate of 60 Hz. Figure 1 demonstrates a typical setup of the implemented application with two users working on the screen.

The prototype application is written in Java, using JavaFX for the user interface and the TUIO [37] and the TUIOFX library [38] for multi-touch interaction. To enable multiple users to work on the same screen with different widgets and user interface elements at the same time, a concept called focusArea from the TUIOFX library is used [39]. The application follows the widely-used Model-View-Controller (MVC) architecture.

V. Use Case

The use case for the prototype application is to improve interaction with the Regression Lens on multi-touch screens. The developed interaction techniques were tested with the well-known car dataset from the UCI Machine Learning Repository [40].

For the interaction technique shown in Figure 1, user A (on the left) and user B (on the right) select two different plots from the shared central area containing the scatterplot matrix. For this technique, the user holds and touches a scatterplot with one hand and swipes to the right or left with the other hand to maximise it. This technique is elaborated in detail in Section III-C. After that, users A and B select an area in the scatterplot separately and toggle the Collaborative Lens option in the Floating Toolbox. As described in Section III-D, each

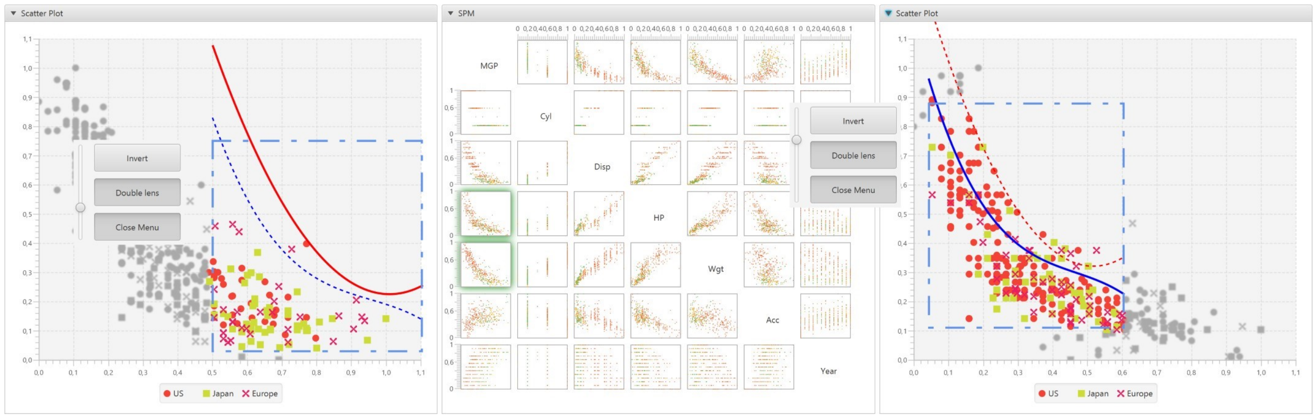


Fig. 6. The left and right panels are scatterplots for User A (left) and B (right) respectively. The central area of the screen contains a shared scatterplot matrix. User A on the left draws an arbitrary rectangle and is interested in the quadratic regression model of the selected records, shown in red. User B on the right chooses to observe the cubic regression model of the selected area, shown in blue. User A can see the cubic regression model of the right panel in dashed blue and user B can see the left panel regression model in dashed red. Selected scatterplots are highlighted in green in the scatterplot matrix.

user is now able to observe the regression model of the other user in their regression lens. Figure 1 shows two users working side by side on a large vertically-mounted multi-touch display, after creating two separate Regression Lenses and toggling to the Double Lens option. The exact state of the screen is shown in Figure 6. A single Regression Lens with a floating toolbox is visible in Figure 5.

VI. Perception-based Level of Visual Detail Concepts for Scatterplots

Users of large vertically-mounted high-resolution displays may take up positions at varying distances from the display, and hence may perceive more or less detail in the display. At greater distances from a large high-resolution display, less detail is perceived. Here, *perceived pixel density* (PPD) is defined as the number of pixels mapped to a single cell on the retina of the user's eye. PPD increases quadratically as distance to the screen increases. The human perceptual system tends to average out too large PPD w.r.t. colour, brightness, and contrast [41], for example a red pixel and a green pixel is perceived as brown.

The perceptual effect of *averaging* is well known, for instance in the perception of secondary colours as a mixture of two primary colours or in the phenomena of metamerism. More related effects include simultaneous contrast [42], after-images [43], and the Chubb effect [44]. Without delving too deeply into perception psychology, note that a sophisticated theory for averaging effects are already available and well described. For the purpose of this discussion with respect to large high-resolution displays, it is sufficient to state that the effect of averaging a set of pixels is already exploited in practice by techniques such as image mosaics [41] and halftone techniques [45], as illustrated in Figure 7.

Since PPD and related averaging effects are a function of distance from the display, screen distance can be seen as an interactive parameter which can be exploited for visual data analysis. Three techniques are proposed to apply a perception-based level of detail to scatterplots on large vertically-mounted high-resolution displays.

Firstly, the concept of *superpixels* is similar to image mosaics. A superpixel consists of a set of pixels in a small rectangular area of the screen, for example a regular grid of say 50×50 pixels. The average colour, brightness, and contrast properties of superpixels can be used to visualise data for users farther from the screen. At the same time, the individual colouring of pixels comprising a superpixel can be used to visualise more detailed information for users who are closer to the screen.

Secondly, the concept of a *Screen Progressive Visual Glyph* (SPVG) utilises the colour, brightness, and contrast values of a glyph to encode different secondary information for closer users. In Figure 8, the scatterplot on the left visually encodes two different classes (brown and cyan) in the data. This is easily perceivable by a distant user. On the right, a user who is closer can make out an additional level of detail: the dots of the scatterplot in fact contain an additional histogram representing the distribution of the related class in the data. In this case, the circles representing the mapped data points are SPVGs. The difference between SPVGs and superpixels is that SPVGs encode different visual details of the same data at different distances. In this way, they could be understood as a data filter concept as well. SPVGs can be placed on the screen on demand and are not restricted to a regular grid, providing greater flexibility.

Thirdly, *variational textures* are related to halftone techniques. Structural variations of an underlying texture can be used to visually encode fine data details for users who are very close to the screen, while these details will immediately disappear when the user goes further away.

These proposed approaches for level of visual detail align well with Shneiderman's mantra for information visualisation [46]: "Overview first, zoom and filter, details on demand". In this case, distance from the screen is an additional degree of

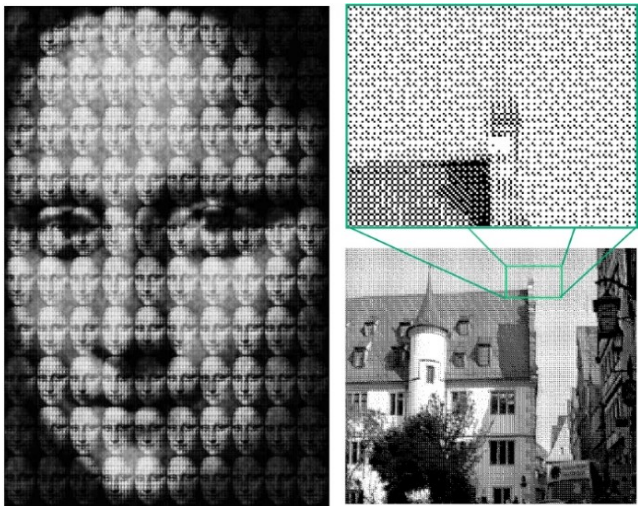


Fig. 7. On the left, a multi-image mosaic of the Mona Lisa [41]. On the right, an example of halftone dot sampling [45].

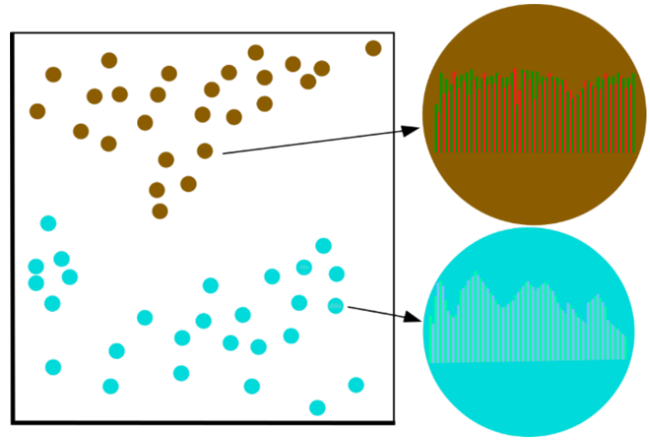


Fig. 8. Screen Progressive Visual Glyphs (SPVGs): On the left, dots on a scatterplot representing items belonging to two classes (brown and cyan) are seen by distant users as simple dots. On the right, users who are closer to the screen can perceive an additional histogram showing the distribution of items.

freedom, controlled by each user individually as they move closer to or further away from the display. The approaches are discussed as a concept and not implemented yet.

VII. Discussion and Future Work

The concepts described in this paper are first designs of appropriate touch interaction for the visual interactive analysis of scatterplot data on large vertically-mounted high-resolution multi-touch displays. The interactions support small-group collaborative analysis, by exchanging patterns or settings from one user's view to the others. The interaction design is currently based on user selections, but is generalisable to other basic techniques. The interaction techniques have been implemented as a proof of concept. They still need to be evaluated with real users and real tasks as part of future work. Mapping out the design space for this combination of visualisation and display device may well yield further interesting interaction designs.

The idea of exploiting perception-based level of detail for the visualisation of scatterplots on large displays is new. Detailed information can be rendered inside the marks of the plot, becoming perceivable once users are closer to the screen. Again, this is a proof of concept and requires further development and evaluation.

While large high-resolution displays can improve the exploration of large scatterplot spaces, further data analysis support is needed to scale up with the number of data points and dimensions. Traditional techniques like cluster analysis and aggregation can help with scalability. Another relevant line of improvement is to adjust the view to the user's need and situation. In [47], the authors propose using eye tracking to infer user interest and using this information to recommend additional relevant but previously unseen views for exploration. While that work was developed as a desktop application, it might be interesting to incorporate eye-tracking support to recommend views for small collaborative team work on a large display. Moreover, adding group activity recognition and therefore pro-active interaction, can support collaboration by preventing information overload [48].

VIII. Concluding Remarks

This paper presented challenges and solutions for collaborative and single-task multi-touch interaction on large vertically-mounted high-resolution displays. The techniques presented are well-suited for collaborative analysis tasks with scatterplots and scatterplot matrices. They are potentially generalisable for other data exploration and visual analytics practices but require further implementation and evaluation. Also, perception-based visualisation of scatterplots is introduced as a possible direction for further research.

References

[1] K. Reda, A. E. Johnson, M. E. Papka, and J. Leigh, "Effects of display size and resolution on user behavior and insight acquisition in visual exploration," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2759–2768.

[2] R. A. Ruddle, W. Fateen, D. Treanor, P. Sondergeld, and P. Ouirke, "Leveraging wall-sized high-resolution displays for comparative genomics analyses of copy number variation," in *Biological Data Visualization (BioVis), 2013 IEEE Symposium on*. IEEE, 2013, pp. 89–96.

[3] A. Prouzeau, A. Bezerianos, and O. Chapuis, "Evaluating multi-user selection for exploring graph topology on wall-displays," *IEEE Transactions on Visualization and Computer Graphics*, 2016.

[4] C. Andrews, A. Endert, and C. North, "Space to think: large high-resolution displays for sensemaking," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2010, pp. 55–64.

[5] M. R. Jakobsen and K. Hornbæk, "Up close and personal: Collaborative work on a high-resolution multitouch wall display," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 21, no. 2, p. 11, 2014.

[6] T. K. Khan, "A survey of interaction techniques and devices for large high resolution displays," in *OASIcs-OpenAccess Series in Informatics*, vol. 19. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2011.

[7] K. Vogt, L. Bradel, C. Andrews, C. North, A. Endert, and D. Hutchings, "Co-located collaborative sensemaking on a large high-resolution display with multiple input devices," *Human-Computer Interaction–INTERACT 2011*, pp. 589–604, 2011.

[8] P. Isenberg, S. Carpendale, A. Bezerianos, N. Henry, and J.-D. Fekete, "Coconuttrix: Collaborative retrofitting for information visualization," *IEEE Computer Graphics and Applications*, vol. 29, no. 5, pp. 44–57, 2009.

[9] C. Liu, O. Chapuis, M. Beaudouin-Lafon, and E. Lecolinet, "Coreach: Cooperative gestures for data manipulation on wall-sized displays," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 6730–6741.

[10] M. R. Morris, A. Huang, A. Paepcke, and T. Winograd, "Cooperative gestures: multi-user gestural interactions for co-located groupware," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 1201–1210.

[11] A. Inselberg, "The plane with parallel coordinates," *The visual computer*, vol. 1, no. 2, pp. 69–91, 1985.

[12] M. Ashdown, P. Tuddenham, and P. Robinson, "High-Resolution Interactive Displays," in *Tabletops - Horizontal Interactive Displays*, 2010, pp. 71–100.

[13] L. Shao, A. Mahajan, T. Schreck, and D. J. Lehmann, "Interactive regression lens for exploring scatter plots," in *Computer Graphics Forum*, vol. 36, no. 3. Wiley Online Library, 2017, pp. 157–166.

[14] J. S. Yi, Y. A. Kang, J. Stasko, and J. Jacko, "Toward a deeper understanding of the role of interaction in information visualization." *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, pp. 1224–31, 2007.

[15] B. Lee, P. Isenberg, N. H. Riche, and S. Carpendale, "Beyond mouse and keyboard: Expanding design considerations for information visualization interactions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2689–2698, 2012.

[16] D. Keim, "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1–8, 2002.

[17] R. A. Ruddle, R. G. Thomas, R. S. Randell, P. Quirke, and D. Treanor, "Performance and interaction behaviour during visual search on large, high-resolution displays," *Information Visualization*, vol. 14, no. 2, pp. 137–147, 2015.

[18] J. C. Roberts, "Exploratory Visualization with Multiple Linked Views," in *Exploring Geovisualization*. Amsterdam: Elseviers, 2005, pp. 159–180.

[19] M. Czerwinski, G. Smith, T. Regan, B. Meyers, G. G. Robertson, and G. Starkweather, "Toward characterizing the productivity benefits of very large displays." in *Interact*, vol. 3, 2003, pp. 9–16.

[20] P. Isenberg, P. Dragicevic, W. Willett, A. Bezerianos, and J.-D. Fekete, "Hybrid-image visualization for large viewing environments," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2346–2355, 2013.

[21] T. Tsandilas, A. Bezerianos, and T. Jacob, "Sketchsliders: Sketching widgets for visual exploration on wall displays," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 3255–3264.

[22] Y. Zhai, G. Zhao, T. Alatalo, J. Heikkilä, T. Ojala, and X. Huang, "Gesture interaction for wall-sized touchscreen display," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 2013, pp. 175–178.

[23] M. Heilig, S. Huber, M. Demarmels, and H. Reiterer, "Scattertouch: a multi touch rubber sheet scatter plot visualization for co-located data exploration," in *ACM International Conference on Interactive Tabletops and Surfaces*. ACM, 2010, pp. 263–264.

[24] R. Sadana and J. Stasko, "Expanding selection for information visualization systems on tablet devices," in *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 2016, pp. 149–158.

[25] U. Kister, P. Reipschläger, and R. Dachselt, "Multilens: Fluent interaction with multi-functional multi-touch lenses for information visualization," in *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 2016, pp. 139–148.

[26] K. Matković, H. Abraham, M. Jelović, and H. Hauser, "Quantitative externalization of visual data analysis results using local regression models," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 2017, pp. 199–218.

[27] J. M. Rzeszotarski and A. Kittur, "Kinetica: naturalistic multi-touch data visualization," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2014, pp. 897–906.

[28] C. Andrews, A. Endert, B. Yost, and C. North, "Information visualization on large, high-resolution displays: Issues, challenges, and opportunities," *Information Visualization*, vol. 10, no. 4, pp. 341–355, 2011.

[29] P. Isenberg, T. Isenberg, T. Hesselmann, B. Lee, U. Von Zadow, and A. Tang, "Data visualization on interactive surfaces: A research agenda," *IEEE Computer Graphics and Applications*, vol. 33, no. 2, pp. 16–24, 2013.

[30] A. Prouzeau, A. Bezerianos, and O. Chapuis, "Trade-offs between a vertical shared display and two desktops in a collaborative path-finding task," in *Proceedings of Graphics Interface 2017*, 2017.

[31] E. W. Pedersen and K. Hornbæk, "An experimental comparison of touch interaction on vertical and horizontal surfaces," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. ACM, 2012, pp. 370–379.

[32] S. K. Badam, F. Amini, N. Elmqvist, and P. Irani, "Supporting visual exploration for multiple users in large display environments," in *Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on*. IEEE, 2016, pp. 1–10.

[33] "Eyevis display 84-inch," 2017. [Online]. Available: http://www.eyevis.de/en/products/lcd-solutions/4k-ultra-hd-lcd-monitors/84-inch-4k-uhd-lcd.html

[34] A. Prouzeau, A. Bezerianos, and O. Chapuis, "Towards road traffic management with forecasting on wall displays," in *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 2016, pp. 119–128.

[35] U. Kister, P. Reipschläger, F. Matulic, and R. Dachselt, "Bodylenses: Embodied magic lenses and personal territories for wall displays," in *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. ACM, 2015, pp. 117–126.

[36] K.-P. Yee, "Two-handed interaction on a tablet display," in *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2004, pp. 1493–1496.

[37] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza, "Tuio: A protocol for table-top tangible user interfaces," in *Proc. of the The 6th Intl Workshop on Gesture in Human-Computer Interaction and Simulation*, 2005, pp. 1–5.

[38] M. Fetter and D. Bimamisa, "Tuiofxtoolkit support for the development of javafx applications for interactive tabletops," in *Human-Computer Interaction*. Springer, 2015, pp. 486–489.

[39] M. Fetter, D. Bimamisa, and T. Gross, "Tuiofx: A javafx toolkit for shared interactive surfaces," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. 1, p. 10, 2017.

[40] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[41] A. Finkelstein and M. Range, "Image Mosaics," Princeton University, Computer Science Department, Technical Report TR-574-98, Mar. 1998.

[42] P. Burgh, "Peripheral viewing and simultaneous contrast," *The Quarterly Journal of Experimental Psychology 16(3)*, pp. 257–263, 1964.

[43] S. Anstis, B. Rogers, and J. Henry, "Interactions between simultaneous contrast and colored afterimages," *Vision Research*, pp. 899–911, 1978.

[44] C. Chubb, G. Sperling, and J. Solomon, "Texture interactions determine perceived contrast," *Proc Natl Acad Sci, 86*, pp. 9631–9635, 1989.

[45] R. Steinbrecher, "Bildverarbeitung in der praxis," *RST-Verlag. München-Wien-Oldenburg*, 2005.

[46] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *IEEE Visual Languages*, 1996, pp. 336–343.

[47] L. Shao, N. Silva, E. Eggeling, and T. Schreck, "Visual exploration of large scatter plot matrices by pattern recommendation based on eye tracking," in *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*. ACM, 2017, pp. 9–16.

[48] D. Gordon, J.-H. Hanne, M. Berchtold, A. A. N. Shirehjini, and M. Beigl, "Towards collaborative group activity recognition using mobile devices," *Mobile Networks and Applications*, vol. 18, no. 3, pp. 326–340, 2013.

Evaluation of the User Experience of Interactive Infographics in Online Newspapers

Julia Langer
Information Technology and Information Management
University of Applied Sciences Burgenland
Eisenstadt, Austria
1410639004@fh-burgenland.at

Michael Zeiller
Information Technology and Information Management
University of Applied Sciences Burgenland
Eisenstadt, Austria
michael.zeiller@fh-burgenland.at

***Abstract*— Information graphics are a powerful tool to communicate complex information. Adding interactive elements to infographics that are published in online media enables journalists to tell even more complex and exiting stories. However, the usability of such complex data presentations is crucial for their acceptance among readers of online newspapers. The results of a usability study of interactive infographics published in online newspapers reveal weaknesses and success factors for designing interactive infographics to ensure an improved user experience.**

Keywords—interactive infographics; information graphics; information visualization; interaction; usability; user experience; data-driven journalism; online journalism; online newspaper.

I. Introduction

Information visualization describes the use of visual representations of abstract data to amplify cognition [1] [2]. The visual representation of information enables users to effectively and efficiently perceive, recognize and interpret information. Especially information graphics (short: infographics), that combine graphics, image and text, are an efficient means to communicate complex information, data or knowledge [3]. Static representations can be enhanced with interaction (e.g., filtering, selection, input of data, navigation) to provide users with different ways of controlling how and which kind of information is presented [4]. Since Shneiderman [5] proposed his *Visual Information-Seeking Mantra: Overview first, zoom and filter, then details-on-demand,* interaction has been a key principle for the success of information visualization.

Due to their capability to communicate complex data, information and knowledge efficiently, infographics are often applied in data-driven journalism. In data-driven journalism large amounts of data are collected, evaluated, interpreted and presented to readers [6] [7]. Lorenz [8] defines data-driven journalism as a workflow where data is the basis for analysis, visualization, and storytelling. Based on large amounts of data, data journalists explain new insights and tell complex stories that are enhanced by (interactive) visual representations [6] [9].

Information visualization supports data journalists in multiple ways. In the reporting phase information visualization helps them to identify themes and questions, to identify outliers, or to find typical examples [6]. When journalists publish stories based on their investigations, information visualization is an appropriate medium of communication for storytelling – apart from simply attracting attention due to beautiful graphics. In the publishing phase (interactive) visualizations can play multiple roles: they help to illustrate new insights in a more compelling way, they can remove unnecessarily technical information from prose, or they offer a new perspective since they can show changes over time, show connections, or compare values much more efficiently than text [6].

Stories published in online media can take advantage of narratives including complex graphics and especially interactive infographics. Especially in online newspapers we find an emerging number of stories including interactive infographics. Due to interactive elements readers can explore the data and can control by themselves which and how much information shall be displayed. Adding interactivity introduces an additional level of required skills to users (i.e., data literacy) to control and navigate within the interactive graphics. Additionally, inadequate user experience, flaws in the infographics' usability and simple mistakes in the interactive presentation can lead to wrong conclusions and force readers to stop exploring the infographics [10].

Although interactive infographics are increasingly used in online media, readers face the challenge of finding and getting access to the interactive infographics because they are not marked properly and not all control elements for interaction can be identified [11]. Since previous studies (e.g., [11]) reveal a significant lack of convenience during the utilization of interactive infographics in online newspapers, this paper focusses on the user experience of interactive infographics that have been published in German-speaking online newspapers. The results of a usability test, that has been applied to several interactive infographics, are presented. The usability test reveals some weaknesses, but also success factors that can help journalists and designers of interactive infographics to improve the user experience for readers of online news stories.

Section II gives a short introduction to interactive infographics which is followed by a brief overview on related work on usability of (interactive) infographics in section III. In section IV we introduce the usability test method based on *Thinking Aloud* and a questionnaire. Sample results on selected infographics and selected ergonomic principles are presented in section V. We end up with some remarks on future work in section VI and conclusions in section VII.

II. Interactive Infographics

A. Information Graphics

Information graphics are visual representations of information or data, e.g. as a chart or diagram, telling a specific story [3] [12]. They combine graphics, image, text and numbers to communicate information, data or knowledge efficiently [13]. Infographics can be used to communicate complex topics and draw the attention of percipients to them. They provide the percipient with new insights and a quick overview on complex facts on subjects like politics, science, technology, and nature that are hard to understand just using text-based information. However, despite obvious advantages there is an ongoing debate on visual embellishment.

While trying to create appealing infographics designers have to prevent from adding unnecessary visual embellishment – chart junk. They should adhere to a reduced approach using plain and simple charts, e.g., by following the data-ink ratio for non-interactive infographics proposed by Tufte [14] to reduce chart junk.

B. Infographics and Interaction

Most infographics published in books, newspapers, magazines, on TV, or online media provide static representations. However, an increasing number of infographics published in online media can be manipulated by the user interactively. Interaction is the ability to change in reaction to the user and enhances all types of static information visualization [4]. In the context of infographics there exist several methods of interaction to manipulate a visual representation, like scrolling, overview plus detail, focus plus context, filtering, or data reordering [15].

Weber and Wenzel [13] define interactive infographics as a visual representation of information that integrates several modes (at least two) – e.g., image/video, spoken or written text, audio, layout, etc. (the image mode is constitutive) – to a coherent ensemble that offers at least one option of control to the user. Interactive infographics can be controlled by, e.g., Start or Stop button, forward or backward button, menu item to select, timeline or time controller, filter, data request or input box [13]

C. Types of Interactive Infographics

Following Weber [16] we can classify interactive infographics according to five distinctive features that cover interaction as well as narrative issues: degree of interactivity, activity model, communicative intent, topic, and the classic questions *What*, *Where*, *When*, *How*, etc. [17]. Additionally, features like genre or visual narrative can be applied, too [18]. Most important for the usability of interactive infographics are the *degree of interactivity* and the *activity model* [11].

The degree of interactivity of interactive infographics is made up of three levels [13]: *Low* interactivity, *medium* interactivity, and *high* interactivity. While a low level of interactivity allows a user to manipulate interactive infographics without changing the graphics itself (e.g., zooming, mouseover effects for showing details, Next or Start buttons), on a medium level a user can manipulate the graphics (e.g., using a timeline slider or menu items) by applying changes and comparing information. In contrast, a high level of interactivity enables the user to fully explore the infographics. He/she can interact with information by input of data, retrieving data, or filtering, thus changing the content.

The activity model identifies the way users can interact with the infographics and distinguishes between:

- Linear
- Nonlinear
- Linear-nonlinear

The linear type restricts the user to move forward or backward through a predetermined linear sequence [19]. The step-by-step course is predefined by the author, i.e., this is an author-driven style of interaction [18]. The user can only follow a strict path using navigation tools like Start, Stop, Forward, Backward, or Next and cannot explore the visualization by himself. [16]

In contrast, a nonlinear visualization does not provide a prescribed ordering. This type offers the user many ways to explore and query the visualization, including free exploration without predefined navigation paths. Thus, its narrative is reader-driven [18]. Navigation tools for nonlinear infographics include input box, data query, filter, or brushing. [16]

The third type called linear-nonlinear is a combination of the other approaches. This type enables the author to communicate his message using a predefined path, but additionally it allows the user a limit amount of selection, for example using interactive timelines, time controller, or an integrated menu for navigation. [16]

III. Related Work

Interactive infographics shall communicate complex topics fast, easy, in an easily understandable way to a broad audience. To achieve this goal they have to be user-friendly, i.e., the usability has to be well designed. However, designing and creating interactive infographics is a challenging task [10]: After identifying and structuring the topic and deriving an appropriate type of representation the multimedia elements – written text, spoken text (audio), images (photos, diagrams, graphics), videos (video, animation) – have to be combined in a meaningful way. Interactive elements and hypertext elements have to be defined and embedded within a concept of navigation [10]. Since online newspapers are read by a broad audience and not by experts, only a limited knowledge on the linguistic knowledge (e.g., technical terms), structural knowledge (structure of the information service), application knowledge (e.g., utilization of interactive elements like buttons of sliders), and functional knowledge (e.g., filtering of data) can be assumed [20]. There are a few "standards" for designing static information graphics (for print and web), but for interactive applications in the web no standards exist. Burmester, Wenzel and Tille [10] provide some recommendations for designing interactive infographics they derive from a user study on 23 interactive infographics.

The utilization of interactive infographics has been analyzed in some studies. Some authors take a general and global view on interactive infographics [21] and some authors focus on the utilization by journalists and publishing houses [9] [11] [22]. Only a few studies have been published that analyze the utilization by readers and focus on usability issues, like Schumacher [23], Burmester, Mast, Wenzel and Tille [24], and Zwinger, Langer and Zeiller [25]. Since a study by Zwinger and Zeiller [11] revealed a significant lack of convenience and usability during the utilization of interactive infographics published in online newspapers of Austria, Germany and Switzerland, the usability of such infographics will be analyzed in detail.

IV. Usability of Interactive InfoGraphics

A. Research Question

This study examines the usability of interactive information graphics that have been published in online newspapers. It focusses on the utilization of interactive infographics by readers of those online news. We analyze how readers perceive, interpret and interact with interactive infographics. We investigate whether background knowledge (structural, application, functional knowledge) and previous experience is required to provide a sufficient user experience. The requirements of users related to a user-friendly design (based on the international standard ISO 9241 on the ergonomics of human-system interaction) of interactive infographics for online news are identified due to an analysis of usage problems and identifying weaknesses. Success factors and potential areas of improvement will be shown.

Therefore, we focus on the following question: Which success factors improve the usability of interactive information graphics in online journalism?

B. Usability and User Experience

Usability – in particular web usability – can be defined in various ways. One of the pioneers Jakob Nielsen [26] defines usability as “a quality attribute that assesses how easy user interfaces are to use”. Consequently, he characterizes usability by five quality attributes: learnability, efficiency, memorability, errors, and satisfaction [26]. The international standard ISO 9241 on the ergonomics of human-system interaction defines in part 11 usability as the *extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use* [27]. Information systems with high usability have to be user friendly, easy to use, ease to learn, and its interface has to be ergonomically designed [28].

The term “user experience” not only considers the actual use of an information system, but includes the anticipated usage (before using the system) and the processing of the use situation (identification or distancing; after using the system) [28]. According to Norman and Nielson [29] user experience encompasses “all aspects of the end-user's interaction with the company, its services, and its products”. ISO 9241-210 defines user experience as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" [30].

C. Method

The user experience and in particular the usability of an interactive system can be measured in various ways [31]. Eyetracking would be a method of first choice on evaluating the usability of an online services including infographics [31] [32]. However, most eyetracking tools cannot cope with the dynamic behavior of interactive infographics and do not track the changes on the screen, e.g., caused by mouseover, interactive timelines, time controllers, and integrated navigation. Thus, a combined method based on a questionnaire and the well-known qualitative method *Thinking Aloud* (Think Aloud) is applied. By combining both methods the weaknesses in the presentation of infographics and problems of the users while interacting with them can be identified and the course of action can be reconstructed [27].

1) Thinking Aloud

Thinking Aloud is a cheap, flexible, robust and easy to learn usability test [33]. Since it was first introduced to interface design by Lewis [34] in 1982 it became one of the most popular methods for usability testing. While testing the usability of an interactive infographics the test persons are encouraged to continuously comment their actions and their thoughts. Ideally the test person describes all paths of action taken and all of his/her impressions [35]. The verbalized thoughts of the test persons are recorded (audio and/or video recording). We used the recording feature of QuickTime Player to record the spoken word of the test persons plus the movement of the mouse on the screen and mouse clicks.

This method provides immediate response that enables the test moderator to draw conclusions on the actions taken and emotions already during the evaluation. A few test persons are sufficient to derive qualitative feedback of good quality [28].

2) Questionnaire

After the test phase where the Thinking Aloud protocol had been applied the evaluation was continued by a questionnaire the test persons had to fill in. A usability evaluation can apply the questionnaire ISO 9241/110-S [28]. It follows the ergonomic principles of the ISO standard 9241-110: *suitability for the task*, *suitability for learning*, *suitability for individualization*, *conformity with user expectations*, *self-descriptiveness*, *controllability*, and *error tolerance*. Each of the seven principles is evaluated by five items on a seven-point Likert scale (“very negative” to “very positive”). Since this is a standard questionnaire, some items may be inappropriate for a specific usability evaluation.

To perform the usability test on interactive infographics the ISO 9241/110-S questionnaire was modified and adapted to the special needs [36] of this specific test. The number of items in each section (i.e., ergonomic principle) had been adapted: *conformity with user expectations* used the five original items; *suitability for the task*, *self-descriptiveness*, and *controllability* were reduced to four items; *suitability for individualization* and *error tolerance* were reduced to three items; and *suitability for learning* was reduced to two items. The Likert scale for assessing each item was reduced from seven to four levels: ‘very negative’ “- -“ | ‘negative’ “-“ | ‘positive’ “+” | ‘very positive’ “+ +” [36].

To illustrate which items had been included in the adapted questionnaire the items for the most significant and meaningful ergonomic principles will be listed in detail: *suitability for the task*, *conformity with user expectations*, *self-descriptiveness*, and *controllability*.

Example 1: The principle "suitability for the task" consists of the following items: The interactive infographics …

- is complicated and confusing / is straightforward and clearly structured.
- is boring and unimaginative / is exciting and creatively designed (i.e., motivates to interact).
- includes unnecessary elements for control and interaction / includes an appropriate number of elements for control and interaction.
- offers too much and unnecessary information to keep me informed / offers exactly fitting and necessary information to keep me informed.

Example 2: The principle "conformity with user expectations" is made up of the following items: The interactive infographics …

- complicates orientation and use due to a non-uniform design of interaction elements / makes orientation and use easy due to a uniform design of interaction elements.
- contains text that is hard to read / contains easily readable text.
- complicates orientation and use due to bad color-coding / makes orientation and use easy due to good color-coding.
- reacts slowly and with unpredictable turnaround times and reaction times / reacts fast and with predictable turnaround times and reaction times.
- includes interactive elements that contradict my expectations and habits / includes interactive elements that correspond to my expectations and habits.

Example 3: The principle "self-descriptiveness" is made up of the following items: The interactive infographics …

- offers no overview of interactive elements / offers a good overview of interactive elements.
- uses vague and unclear terms and abbreviations / uses terms and abbreviations that are easily understood.
- uses ambiguous and unclear symbols and icons / symbols and icons that can be easily understood.
- includes unnecessary comments and explanations / includes helpful comments and explanations.

Example 4: The principle "controllability" consists of the following items: The interactive infographics …

- allows for a cumbersome adoption of navigation tools / allows for easy adoption of navigation tools.
- offers difficult actions and changes using buttons / easy offers actions and changes using buttons.
- allows to undo single steps in a complicated way / allows to undo single steps easily.
- provides complicated and insufficient sorting, filtering and selection of information / provides simple and sufficient sorting, filtering and selection of information.

V. Usability Study

In the presented usability test six interactive infographics have been evaluated by eight test persons [36]. These infographics had been published in German-speaking newspapers from 2012 till 2016: two infographics had been published by “Kurier” from Austria, two infographics by “Spiegel” from Germany, one had been published by “Berliner Morgenpost” from Germany, and the sixth infographic had been published by “20min” from Switzerland. The infographics have been selected according to the activity model (section II.C). To ensure a balanced usability test two infographics have been chosen for each type: linear, nonlinear, and linear-nonlinear.

Although this is a quite small sample, we included examples from different kinds of media in all three countries, different topics, all three types of the activity model, and consequently different levels of complexity. Obviously the small number of infographics tested cannot represent all characteristics of possible occurrences. Furthermore, the “quality” of the specific implementation (e.g., design, depth of content, adherence to usability guidelines) has a great influence on the result – including a risk of introducing a bias to the results. Thus, the results are only partially representative – see also section *VI*.

TABLE I. SELECTED INTERACTIVE INFOGRAPHICS

Type	Publisher	Title
Linear	Kurier	Vegan auch bei Fleischessern beliebt [37]
Linear	Spiegel	So sank die „Titanic“ [38]
Nonlinear	Berliner Morgenpost	WM 2014 [39]
Nonlinear	Kurier	Interaktive Formel-1-Grafik: Fahrer, Strecken, Rekorde [40]
Linear-nonlinear	20min	Eishockey [41]
Linear-nonlinear	Spiegel	Chronologie der Katastrophe in Fukushima [42]

The usability test has been carried out in January 2017. Eight persons at the age of 18 to 30 evaluated the six infographics.. Four test persons had already been using interactive infographics before. The other four test persons did not have previous experiences in using interactive infographics. However, all test persons were familiar with interactive systems and experienced users of the World Wide Web and its applications. Four persons were female, four persons were male. [36]

The test started by presenting the first interactive infographic to the test person. The test person explored the infographic individually, but had to perform a small number of specific tasks. According to the Thinking Aloud protocol he/she had to speak out loud, so the moderator could follow his/her comments and verbalized thoughts (and record them). Immediately after finishing all tasks the questionnaire had to be filled in by the test person on his/her own. This cycle was repeated for all six interactive infographics. A test session had an average duration of 70 minutes. [36]

In the following subsections some sample results of the usability test for selected infographics are presented. We have chosen one representative for each type of the activity model: ‘So sank die „Titanic“’ (linear, IV.A), ‘Interaktive Formel-1-Grafik: Fahrer, Strecken, Rekorde’ (nonlinear, IV.B), Eishockey (linear-nonlinear, IV.C). We will present selected results of the questionnaire for the most significant and meaningful ergonomic principles introduced in section III.C.2: *suitability for the task*, *conformity with user expectations*, *self-descriptiveness*, and *controllability*.

A. Results on Case 1: So sank die "Titanic"

The interactive infographic "So sank die „Titanic“" has been published by Spiegel Online in 2012 (Fig. 1) [38]. This infographic of linear type provides a Forward button and a Backward button. Users can step forward and backwards step-by-step among 11 individual images. Some images are animated, but the animation cannot be controlled by the user.

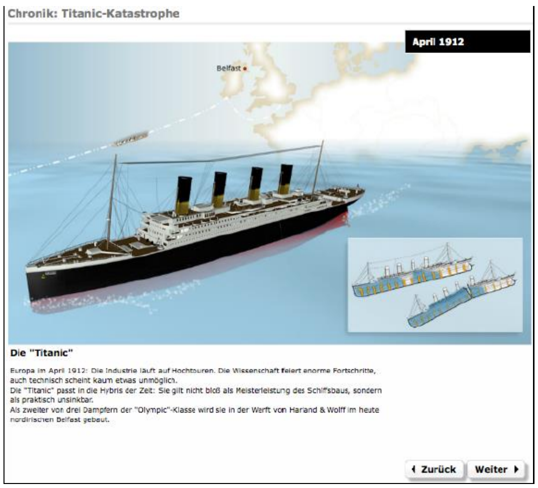


Fig. 1. Example linear type: *So sank die "Titanic"* [38]

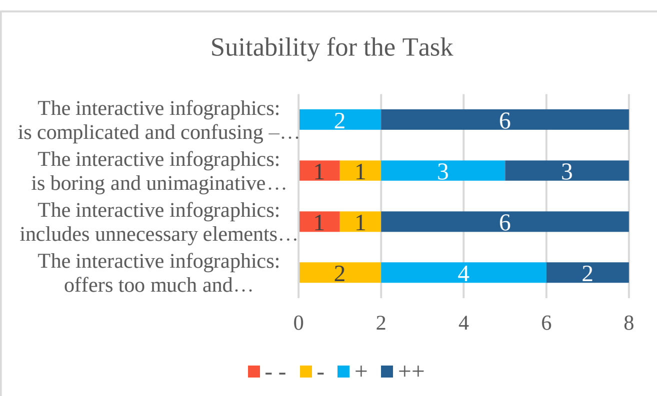


Fig. 2. Results for case 1: Suitability for the task

The results on the principle *suitability for the task* (Fig. 2) show that this infographic is reviewed as straightforward and clearly structured by all test persons (i.e., rated + or ++). Six persons had been motivated to interact and experienced its design exciting and creative. Furthermore, also six persons agreed that the number of elements for control and interaction are appropriate. Six persons indicated that this infographics offers exactly fitting and necessary information while two persons disagreed. The response during the *Thinking Aloud* protocol confirmed these results and provided more detailed qualitative feedback on those issues.

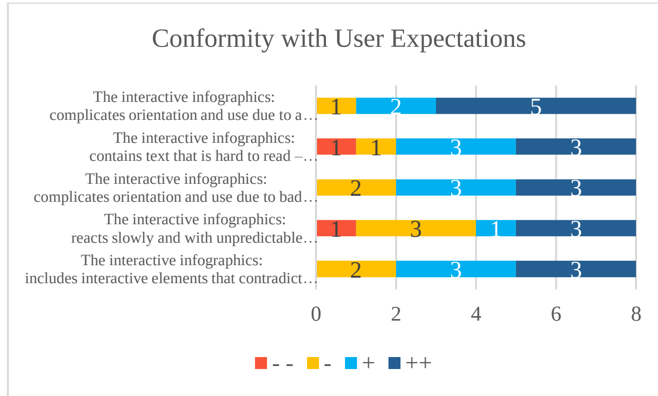


Fig. 3. Results for case 1: Conformity with user expectations

The response on the principle *conformity with user expectations* (Fig. 3) show that seven test persons agreed that this infographics makes orientation and use easy due to a uniform design of interaction elements. The majority of users (6 out of 8) could read the text elements easily. The other two users complained about the readability in animated images. Six persons could navigate easily within the infographic due to good color-coding. Two persons were irritated by the inconsistent use of the color *red*. The turnaround times and reaction times have been experienced differently. Four persons evaluated them positively, while the other four persons had been annoyed by slow reaction times. The majority (again 6 out of 8) persons indicated that the interactive elements correspond to their expectations and habits.

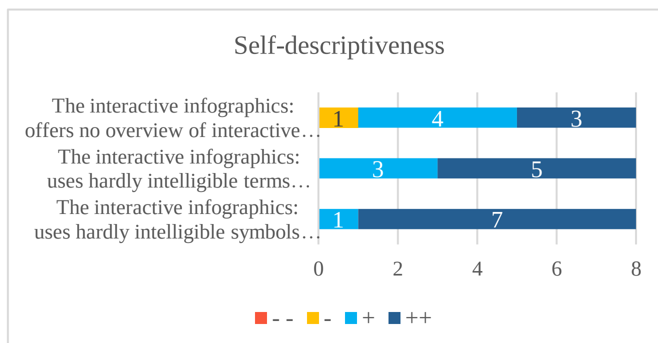


Fig. 4. Results for case 1: Self-descriptiveness

The ergonomic principle *self-descriptiveness* has been experienced by all test persons in a very positive way (Fig. 4). Seven persons could easily identify the interactive elements. Terms and abbreviations, as well as symbols and icons could easily be understood and interpreted by all test persons.

B. Results on Case 2: Interaktive Formel-1-Grafik

The interactive infographic "Interaktive Formel-1-Grafik: Fahrer, Strecken, Rekorde" has been published by the Austrian newspaper Kurier on its website in 2016 (Fig. 5) [40]. It is an infographic of the nonlinear type that enables users to fully explore the infographic. Users can select a Formula-1 driver with a drop-down menu which provides them with information on the team, the Formula-1 debut, the results of the last season, etc.. Additionally, users can navigate through all Grand-Prix tracks by clicking on flags. The users can go forward and backwards using buttons and reset the infographic.

Fig. 5. Example nonlinear type: *Interaktive Formel-1-Grafik: Fahrer, Strecken, Rekorde* [40]

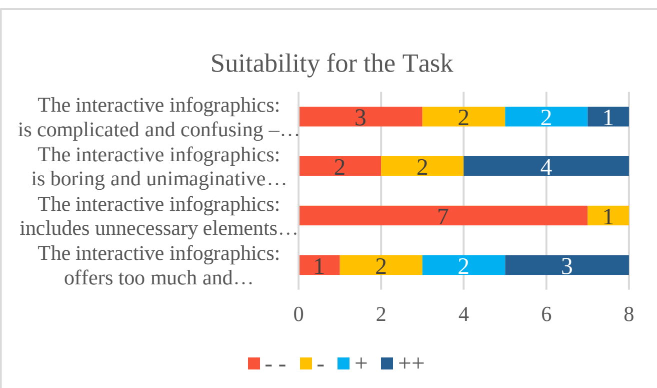


Fig. 6. Results for case 2: Suitability for the task

The answers in the questionnaire concerning the principle *suitability for the task* (Fig. 6) show that the majority of test persons experienced this infographics as being complicated and confusing (5 out of 8). This negative experience might be caused by the fact that all test persons identified unnecessary elements for control and interaction. All users had been confused by the fact that additional functions ("maintain", "exclude") were provided when clicking on some interactive elements – with unclear functionality. Half of the users found this infographics as boring and unimaginative while the other half found it exciting. Three persons mentioned that this infographic does not provide the right amount of information.

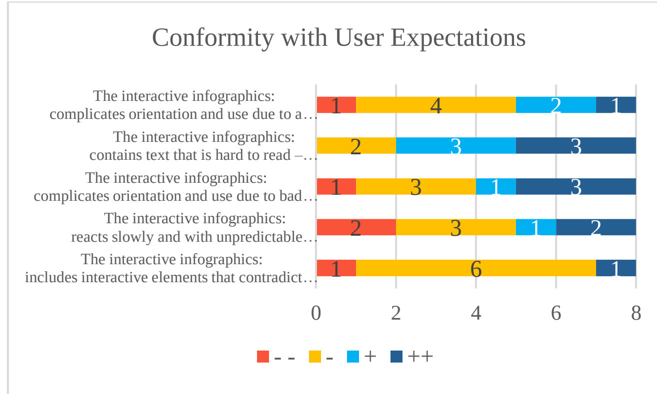


Fig. 7. Results for case 2: Conformity with user expectations

While evaluating the principle *conformity with user expectations* (Fig. 7) we found that five test persons mentioned that this infographic complicates orientation and use due to a non-uniform design of interaction elements. Some users simply could not find out where to click to induce a reaction by the infographic. The readability of text is good (for six persons). Opinions diverged on the ease of orientation and ease of use due to color-coding. While four people were satisfied, four people were irritated, especially by the usage of the same color for interactive and non-interactive elements. The majority (five out of eight) experienced the turnaround times and reaction times as to long and unpredictable. A large majority of the test persons (seven out of eight) was disappointed because several interactive elements contradicted their expectations and habits.

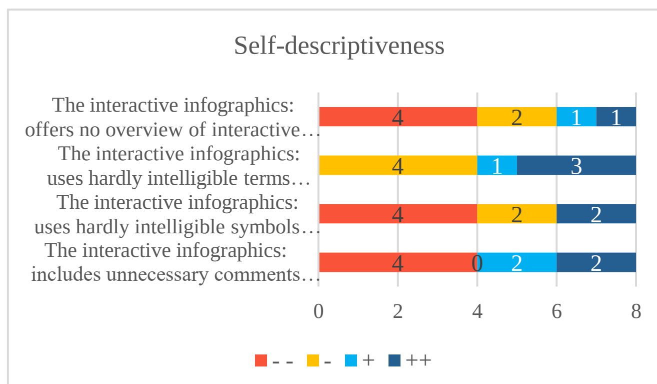


Fig. 8. Results for case 2: Self-descriptiveness

The results on the ergonomic principle *self-descriptiveness* (Fig. 8) show that six test persons expressed their opinion that this infographic does not offer a good overview of interactive elements. Half of the test persons was satisfied with the terms and abbreviations being used, but the other half was not. Only two test persons were satisfied with the use of symbols and icons. The others were dissatisfied with the use of symbols (especially the usage of flags). Although the infographic provides some comments and advices, four test persons felt the comments and explanations as being not helpful.

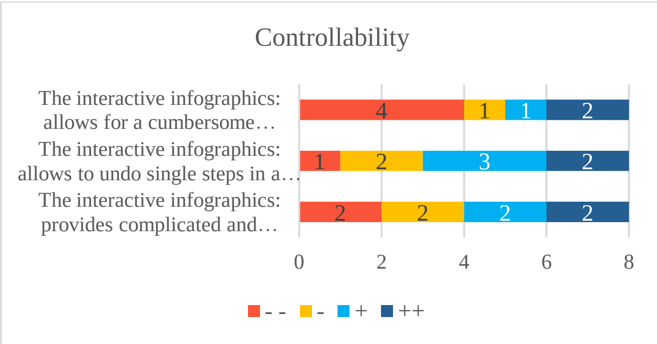


Fig. 9. Results for case 2: Controllability

The answers in the questionnaire regarding the principle *controllability* (Fig. 9) show that the majority of test persons (five out of eight) experienced the adoption of navigation tools being cumbersome. The *Thinking Aloud* protocol revealed that some persons had been confused by the fact that detailed information on the racing drivers as well as information on the tracks can be retrieved – but they do not influence each other. Users can undo their actions and use forward and backward buttons, but not all of them could find these buttons.

C. Results on Case 3: Eishockey

The third case that will be presented is an interactive infographic that has been published by the Swiss online news platform 20min on ice hockey in 2015 (Fig. 10) [41]. The type of the activity model is linear-nonlinear which combines the other two approaches. The infographic consists of a start page and graphics describing different issues on ice hockey. The navigation bar on the top enables users to freely move within the infographic. A linear progress is supported by forward and backward buttons. Red circle icons provide the users with additional information.

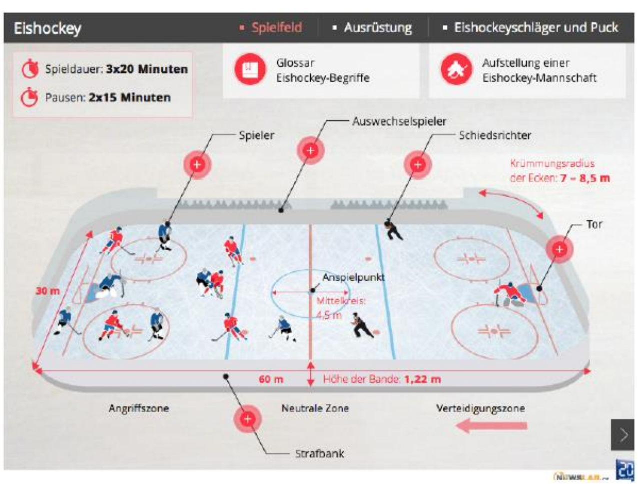


Fig. 10. Example linear-nonlinear type: *Eishockey* [41]

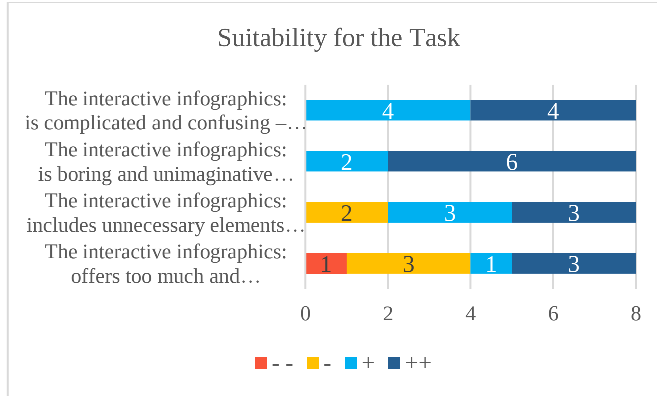


Fig. 11. Results for case 3: Suitability for the task

Again we start with the results on the ergonomic principle *suitability for the task* (Fig. 11). The test persons agree that this infographic is straightforward and clearly structured. The majority found the number of elements for control and interaction appropriate. However, two persons mentioned that the infographic also includes unnecessary elements for control and interaction. Although half of the test persons was overwhelmed by too much information, the other half was quite satisfied with the amount of information provided.

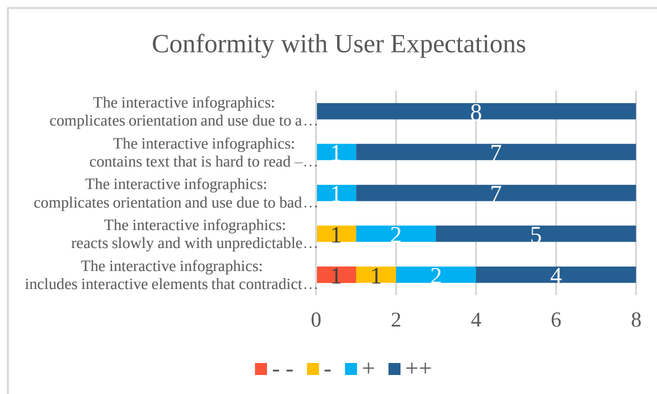


Fig. 12. Results for case 3: Conformity with user expectations

The feedback concerning the principle *conformity with user expectations* (Fig. 12) was quite biased. All test persons were very satisfied. Only slight criticism could be found on the use of interactive elements because they contradicted the expectations and habits of two test persons.

The test persons had also been very satisfied with this infographic when responding to the questionnaire on the ergonomic principle *self-descriptiveness* (Fig. 13). Only one person did not identify the forward and backward buttons right from the beginning and rated the item “overview of interactive elements” negative. Most test persons commented the issues of self-descriptiveness very positive during the *Thinking Aloud* protocol.

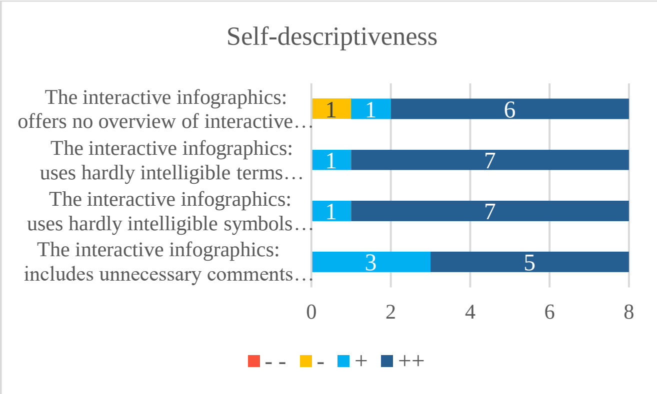


Fig. 13. Results for case 3: Self-descriptiveness

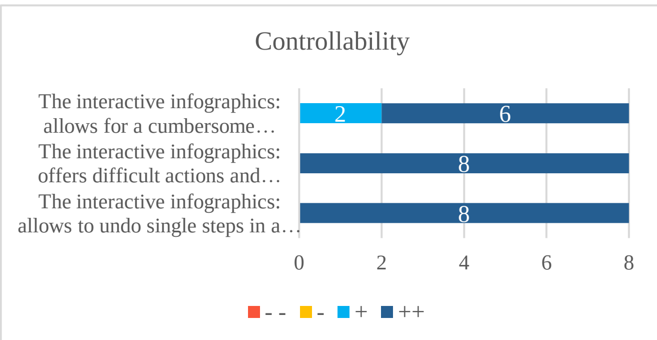


Fig. 14. Results for case 3: Controllability

The interactive infographic on ice hockey performed also very well concerning the principle *controllability* (Fig. 14). It was very easy for all test persons to adopt to the navigation tools. Only two test persons needed more time than the others to adopt because they did not identify the navigation bar at the top immediately. Nevertheless, all test persons could perform changes and actions using the provided buttons without difficulties, including undoing operations.

D. Summary of Results

Summarizing the results of the usability test of all six infographics (including the other three examples not presented in detail here, but in [36]) reveals that users of infographics of the linear type do not necessarily need previous experience. The test persons perceived this type as straightforward and simple. The linear type provides a step-by-step experience and there are no additional, unnecessary elements of interaction which has been confirmed by the test results. Exploring the infographics corresponds to the expectations and habits of users (i.e., high conformity with user expectations). Both analyzed examples made orientation easy due to a uniform design and color-coding of the interaction elements. However, this type of infographics has a major problem in user experience: Since this kind of interaction is very simple, it risks being boring and uninspired because users would like to have more means of interaction

In contrary, nonlinear interactive infographics are found to be exiting and creative because they offer a large variety of possibilities to fully explore the infographics in a very individual way. However, this type risks to be perceived as being complicated and confusing. The test persons noticed that both examples that have been analyzed offer unnecessary control elements for interaction and non-essential information. They needed much more time to get familiar with the interaction elements and control tools. Nevertheless, a uniform design of the interactive elements can increase the user experience significantly. Users appreciate that they can move around and navigate within the infographics – as long as the infographics remains user-friendly.

Whereas the nonlinear model often requires previous experience – depending on the usability of the implementation – both examples of the linear-nonlinear type have shown that they can be easily used. The test persons found both examples of this type to be exciting, but at the same time evaluated them as being straightforward and having a clear design. An easy and fast adoption is supported by a combination of structured information delivery (linear activity) and individual exploration (nonlinear activity). Additionally, this effect is assisted by good usability, e.g., by uniform design and color-coding that facilitate orientation.

The most compelling success factors for appealing and usable interactive infographics are a clear and straightforward structure, an appropriate number of elements for control and interaction, a uniform design of interaction elements that have to correspond to the user expectations and habits, good color-coding, terms and abbreviations as well as symbols and icons that can be easily understood, and navigation tools that can be easily adopted.

VI. Future Work

The presented results are based on the analysis of a quite small sample of six interactive infographics. A continuative study will analyze a much larger number of infographics. Again, a balanced usability test will have to analyze infographics of the linear, nonlinear, and linear-nonlinear type (activity model). To avoid biased results a widespread selection of infographics on various topics, published in different media (online newspapers and online magazines) will be tested to cover different style and quality of the implementation. Thus, there are a number of criteria that have to be considered while selecting the infographics, e.g., type of activity model, type of media, design, depth of content, publishing date, technology, and topic.

Furthermore, a larger number of test persons will be involved: persons of different age, different internet skills and familiarity with interactive designs, and different levels of experience of using (interactive) infographics. Since a test session should not last much longer than in this test (70 minutes), we will have to assign a number of infographics (e.g., six to a maximum of eight) at random while still ensuring a proper overall distribution of the criteria mentioned above.

Although the test method using *Thinking Aloud* and the adapted questionnaire based on ISO 9241/110-S was quite suitable, the continuative study will use eye tracking and/or another approach that is capable to track interactions (e.g., based on time stamps). An additional short questionnaire two weeks after the first questionnaire might help to analyze the influence of the usability and other parameters on the information recall.

VII. Conclusions

The user experience – notably the usability – of interactive information graphics in interactive online media was evaluated by performing a usability test on six interactive infographics. Eight test persons evaluated those infographics that had been published in online newspapers in Austria, Germany and Switzerland according to the Thinking Aloud protocol and a questionnaire following a modified version of the ISO 9241/110-S questionnaire. Based on the results of the usability test a number of success factors have been identified.

Data journalists and designers that use interactive infographics for storytelling in online newspapers and magazines – daily news or infographics for scrollytelling – may apply those success factors when designing new, compelling infographics. Although there are some fundamental influencing factors like the degree of interactivity (low, high, medium) and the activity model (linear, nonlinear, linear-nonlinear) the specific design and implementation improving the usability will boost the acceptance of the infographics – and the stories being told – among readers enormously.

References

[1] C. Ware, Information Visualization: Perception for Design, 3rd ed. Waltham: Elsevier, 2012.

[2] C. Chen, Information Visualization: Beyond the Horizon, 2nd ed. London: Springer, 2006.

[3] N. Yau, Visualize This: The FlowingData Guide to Design, Visualization, and Statistics. Weinheim: Wiley Publishing, 2011.

[4] R. Spence, Information Visualization: An Introduction, 3rd ed. Heidelberg: Springer, 2014.

[5] B. Shneiderman, "The Eyes Have It: a task by data type taxonomy for information visualization" Proceedings of the 1996 IEEE Symposium on Visual Languages, 1996, pp.336-343.

[6] J. Gray, L. Bounegru, and L. Chambers (Eds.), Data Journalism Handbook. European Journalism Centre and the Open Knowledge Foundation, 2012. http://datajournalismhandbook.org/

[7] L. Matzat, Datenjournalismus: Methode einer digitalen Welt. Konstanz, München, UVK Verlagsgesellschaft, 2016.

[8] M. Lorenz, "Status and outlook for data-driven journalism" In Data-driven journalism: What is there to learn? Amsterdam, European Journalism Centre, 2010, pp. 8-17.

[9] W. Weber, H. Rall, "“We Are Journalists.” Production Practices, Attitudes and a Case Study of the New York Times Newsroom" In W. Weber, M. Burmester, R. Tille (Eds.), Interaktive Infografiken. Berlin, Heidelberg: Springer Vieweg, 2013, pp. 161-172.

[10] M. Burmester, A. Wenzel, and R. Tille, "Navigation – Interaktion – Informationspräsentation. Rezeptions- und Nutzungsprobleme bei interaktiven Infografiken" In W. Weber, M. Burmester, R. Tille (Eds.), Interaktive Infografiken. Berlin, Heidelberg: Springer Vieweg, 2013, pp. 105-134.

[11] S. Zwinger, and M. Zeiller, "Interactive Infographics in German Online Newspapers", In W. Aigner, G. Schmiedl, K. Blumenstein, M. Zeppelzauer, and M. Iber (Eds.), Forum Media Technology 2016, CEUR Workshop Proceedings, vol. 1734, pp. 54-64. urn:nbn:de:0074-1734-7.

[12] A. Cairo, The functional art: an introduction to information graphics and visualization. Berkeley, CA: New Riders, 2013.

[13] W. Weber, and A. Wenzel, "Interaktive Infografiken: Standortbestimmung und Definition" In W. Weber, M. Burmester, R. Tille (Eds.), Interaktive Infografiken. Berlin, Heidelberg: Springer Vieweg, 2013, pp. 3-23.

[14] E. Tufte, The Visual Display of Quantitative Information, 2nd ed. Cheshire, CT: Graphics Press, 2001.

[15] R. Mazza, Introduction to Information Visualization. London: Springer, 2009.

[16] W. Weber, "Typen, Muster und hybride Formen: Ein Typologisierungsmodell für interaktive Infografiken" In W. Weber, M. Burmester, R. Tille (Eds.), Interaktive Infografiken. Berlin, Heidelberg: Springer Vieweg, 2013, pp. 25-37.

[17] D. Roam, The Back of the Napkin: Solving problems and selling ideas with pictures. London: Marshall Cavendish, 2008, pp. 138-139.

[18] E. Segel, J. Heer, "Narrative Visualization: Telling Stories with Data" In IEEE Transactions on Visualization and Computer Graphics, vol. 16, no. 6, 2010, pp. 1139-1148.

[19] R. Sims, „Interactivity: A Forgotten Art?“ In Computers in Human Behavior, vol. 13, no. 2, 1997, pp. 157-180.

[20] H-J. Bucher, "Online-Interaktivität - ein hybrider Begriff für eine hybride Kommunikationsform. Begriffliche Klärungen und empirische Rezeptionsbefunde" In: C. Bieber, C- Leggewie, Interaktivität. Ein transdisziplinärer Schlüsselbegriff. Frankfurt am Main: Campus Verlag, 2004, pp. 132-167.

[21] W. Weber, M. Burmester, R. Tille (Eds.), Interaktive Infografiken. Berlin, Heidelberg: Springer Vieweg, 2013.

[22] J. George-Palilonis, and M. Spillman, "Storytelling with Interactive Graphics: An Analysis of Editors' Attitudes and Practices" In Visual Communication Quarterly, vol. 20, no. 1, 2013, pp. 20-27.

[23] P. Schumacher, Rezeption als Interaktion: Wahrnehmung und Nutzung multimodaler Darstellungsformen im Online-Journalismus. Baden-Baden: Nomos, 2010.

[24] M. Burmester, M. Mast, R. Tille, and W. Weber, "How users perceive and use interactive information graphics: an exploratory study" In Proceedings International Conference Information Visualization 2010, July 26-29, London, 2010, pp. 361-368.

[25] S. Zwinger, J. Langer, M. Zeiller, "Acceptance and Usability of Interactive Infographics in Online Newspapers" In Proceedings International Conference Information Visualisation 2017, July 11-14, London, 2017, pp. 176-181.

[26] J. Nielsen, Usability 101: Introduction to Usability. Nielsen Norman Group, 2012. https://www.nngroup.com/articles/usability-101-introduction-to-usability/

[27] International Organization for Standardization, Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts. ISO/DIS 9241-11:2015.

[28] F. Sarodnik, and H. Brau, Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendungen, 2nd ed. Bern: Hans Huber, 2011.

[29] D. Norman, J. Nielsen, The Definition of User Experience (UX). Nielsen Norman Group, n.d. https://www.nngroup.com/articles/definition-user-experience/

[30] International Organization for Standardization, Ergonomics of human system interaction - Part 210: Human-centered design for interactive systems. ISO/DIS 9241-210:2009, 2009.

[31] T. Tullis, B. Albert, Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics, 2nd ed. Waltham, MA: Morgan Kaufmann, 2013.

[32] J. Nielsen, and K. Pernice, Eyetracking Web Usability. Berkeley, CA: New Riders, 2010.

[33] J. Nielsen, Thinking Aloud: The #1 Usability Tool. Nielsen Norman Group, 2012. https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

[34] C.H. Lewis, Using the "Thinking Aloud" Method In Cognitive Interface Design (Technical report). IBM. RC-9265, 1982.

[35] T. Stapelkamp, Informationsvisualisierung: Web-Print-Signaletik. Berlin: Springer Vieweg, 2013.

[36] J. Langer, Analyse der Usability von interaktiven Infografiken im deutschsprachigen Online-Journalismus. Bachelor thesis, Eisenstadt: University of Applied Sciences Burgenland, 2017.

[37] Kurier. Vegan auch bei Fleischessern beliebt, 2015. Retrieved December 15, 2016 from https://kurier.at/genuss/vegan-auch-beifleischessern-beliebt (offline).

[38] Spiegel. So sank die Titanic, 2012. Retrieved September 15, 2017 from http://www.spiegel.de/flash/flash-28319.html.

[39] Berliner Morgenpost. WM 2014, 2014. Retrieved September 15, 2017 from http://interaktiv.morgenpost.de/wmspieler. Checked September 15, 2017.

[40] Kurier. Interaktive Formel-1-Grafik: Fahrer, Strecken, Rekorde, 2016. Retrieved September 15, 2017 from https://kurier.at/sport/motorsport/interaktive-formel-1-grafik-fahrer-streckenrekorde-zum-durchklicken/187.685.483.

[41] 20min. Eishockey, 2015. Retrieved September 15, 2017 from http://www.20min.ch/interaktiv/vizualne/2015_03_Eishockey/index_CHE.html.

[42] Spiegel. Chronologie der Katastrophe in Fukushima I (Daiichi), n.d.. Retrieved September 15, 2017 from http://www.spiegel.de/flash/flash-25484.html.

A Bigram Supported Generic Knowledge-Assisted Malware Analysis System: BiG2-KAMAS

Niklas Thür[1], Markus Wagner[1], Johannes Schick[1], Christina Niederer[1], Jürgen Eckel[3], Robert Luh[2], Wolfgang Aigner[1]
[1]Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences, Austria
[2]Josef Ressel Center for Unified Threat Intelligence on Targeted Attacks, Austria
[3]IKARUS Security Software GmbH, Austria
Email: [1,2]first.last@fhstp.ac.at, [3]eckel.j@ikarus.at

Abstract—**Malicious software, short "malware", refers to software programs that are designed to cause damage or to perform unwanted actions on the infected computer system. Behavior-based analysis of malware typically utilizes tools that produce lengthy traces of observed events, which have to be analyzed manually or by means of individual scripts. Due to the growing amount of data extracted from malware samples, analysts are in need of an interactive tool that supports them in their exploration efforts. In this respect, the use of visual analytics methods and stored expert knowledge helps the user to speed up the exploration process and, furthermore, to improve the quality of the outcome. In this paper, the previously developed KAMAS prototype is extended with additional features such as the integration of a bi-gram based valuation approach to cover further malware analysts' needs. The result is a new prototype which was evaluated by two domain experts in a detailed user study.**

I. Introduction

Malicious software, or short malware, is one of the biggest threats to computer systems these days [1]. 'Malware' refers to software programs, which are designed to cause damage or perform other unwanted actions on a computer or network. Therefore malware plays a big part in most computer intrusions and security incidents. Malware includes inter alia: viruses, trojan horses, worms, rootkits, scareware, and spyware [1]. By now there are millions of malicious programs and the number is increasing every day.

"Malware analysis is the art of dissecting malware to understand how it works, how to identify it, and how to defeat or eliminate it" [1]. In malware analysis, there are two basic approaches to examine a malware program: the static and the dynamic approach. Often the malware analyst only has the potentially malicious executable, which includes the machine code but is not human-readable. Therefore, static malware analysis involves the investigation of the malware executable as well as certain reverse-engineering tasks to recover the sample's source code. On the other hand, dynamic analysis requires the execution of the malicious software on e.g. a virtualized host machine to detect the malware's runtime behavior [1]. To cover all of the malware analyst's needs, Wagner et al. [2] performed a problem characterization and abstraction elaborating the analysts needs in relation to behavior-based malware analysis. In the article by Wagner et al. [3] a design study for a behavior-based knowledge-assisted malware analysis system (referred to as KAMAS) is described. The malware analyst's workflow involves the tasks of examining potentially malicious behavior patterns, selecting them, categorizing them, and storing the found rules in the knowledge database (KDB) [3]. We developed an interactive prototype to extend the KAMAS design study [3] with a new feature of **Bi-G**ram supported **G**eneric **K**nowledge-**A**ssisted **M**alware **A**nalysis **S**ystem (BiG2-KAMAS) [4]. A focus group meeting with members of an Austrian IT security company, the Information security department of St. Pölten UAS and the developers of the initial KAMAS prototype was conducted to identify the tasks and needs for additional features requested by the IT security company to extend the KAMAS design study [3]. Based on this feature list, the paper at hand contributes the following:

1) Integrating a generic data loading process enabling KAMAS to load any kind of data, based on a given structure;
2) Storing benign rules and their highlighting when loading new cluster files, thereby supporting the analyst;
3) Identifying malicious or benign call sequences by including a bi-gram based valuation;
4) Presenting in detail two user studies validating the new features.

This paper is structured as follows: Sect. II provides background knowledge about the work of our collaborators and related work in the field of malware analysis. In Sect. III we describe the prototype's design, visualization methods and implementation. Furthermore, Sect. IV defines the integration of additional knowledge in the prototype's knowledge database. Sect. V shows the prototype's evaluation method, while results are discussed in Sect. VI.

II. Related Work

Shiravi et al. [5] published a survey related to network security visualization, comparing the data sources and visualization techniques of thirty-eight different systems. Furthermore, Egele et al. [6] presented a general literature for malware analysis techniques and tools. In their work they surveyed different approaches for dynamic automated malware analysis and compared them based on their analysis techniques.

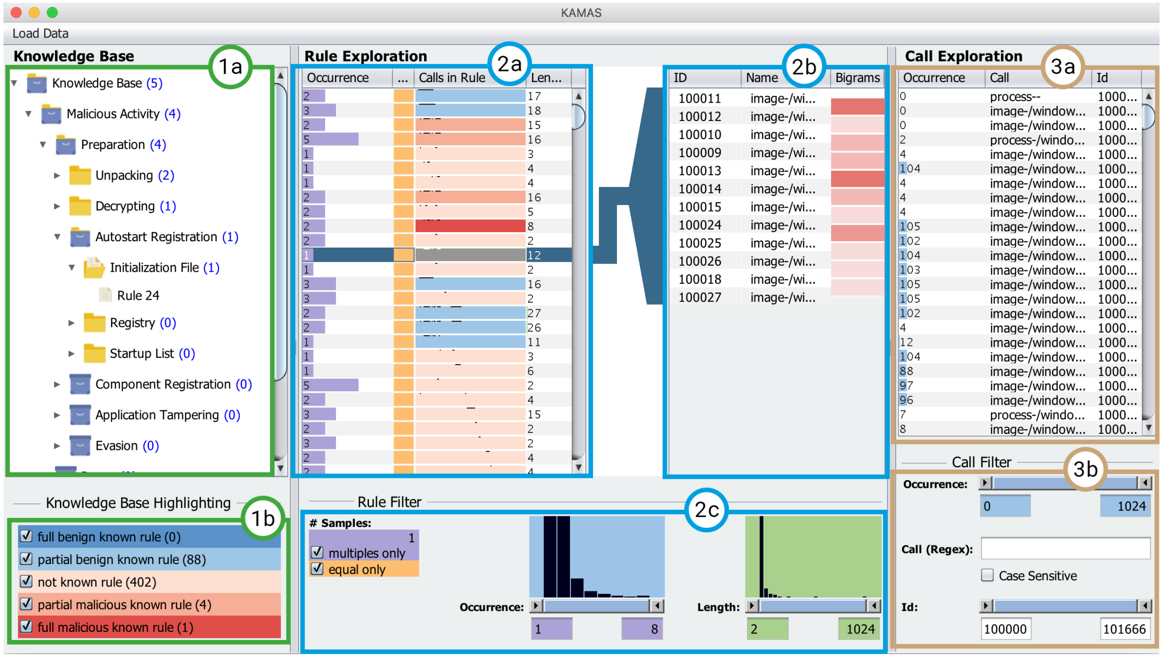


Fig. 1. The BiG2-KAMAS prototype and it's three sections: Section 1 shows the knowledge base including the KDB (1a) with it's new category for benign activity. Beneath the knowledge base highlighting filters are displayed (1b). Section 2 shows the rule exploration area including the bigram visualization (2b) and new color highlighting for benign rules (2a). Finally, section 3 shows the call exploration area.

Likewise, Bazrafshan et al. [7] surveyed various heuristic malware detection techniques as well as malware obfuscation techniques. Additionally, Wagner et al. [8] published a survey of 25 different visualization systems for malware analysis. The objective of their work was the comparison and categorization of the malware systems visualization methods and features and categorizing them along their novel 'Malware Visualization Taxonomy'. Furthermore, McNabb and Laramee [9] published a survey of surveys: Mapping The Landscape of Survey Papers in Information Visualization.

In 2017, Wagner et al. [3] published a paper on a Knowledge-Assisted Malware Analysis System, referred to as KAMAS. In their user study, they found out that the experts are not only interested in visualizing patterns. A supportive valuation approach was implemented by Luh et al. [10], [11], calculating the degree of maliciousness based on system and API call bi-grams. Somarriba et al. [12] presented another malware detector system for Android Malware Behavior. Besides, Marschalek et al. [13] published a system for threat detection using a real-time monitoring agent to gather all or only selected system events and visualize these using event propagation trees. Xiaofang et al. [14] published a paper of a malware variant detection approach using Similarity Search "by processing malware as content fingerprint" [14]. Jain et al. [15] presented a visual exploration approach of android binary files. Their approach is based on the visualization of android .dex files to analyze and compare malicious android executables. David et al. [16] presented "a novel deep learning based method for automatic malware signature generation and classification" [16]. Wrench and Irwin [17] published an approach in which they identify and classify Remote Access Trojans (RATs) and other malicious software based on the programming language PHP.

III. Prototype Concept

This section describes the new features of the '**Bi-Gram** supported **G**eneric **K**nowledge-**A**ssisted **M**alware **A**nalysis **S**ystem (BiG2-KAMAS), conceptually grounded on the KAMAS prototype [3].

A. Data

In its current iteration, BiG2-KAMAS bases its visualization on sequential traces of Windows kernel operations amounting to benign and malicious application behavior in the context of OS and user-initiated processes. These *events* are typically abstractions of raw system and API calls that yield information about the general behavior of an unknown application sample or resident process [8]. Raw calls may include wrapper functions (e.g. `CreateFile`) that offer a simple interface to the application programmer, or native system calls (e.g. `NtCreateFile`) that represent the underlying OS or kernel support functions. In the context of BiG-KAMAS and its data providers, events are collected directly from the Windows kernel. We employ a driver-based monitoring agent [13] designed to collect and forward a number of events to a database server. This gives us unimpeded access to events depicting operations related to process and thread control, image loads, file management, registry modification, network socket interaction, and more. For example, a shell event that creates a new binary file on a system may be simply denoted as a triple `explorer.exe,file-create,sample.exe`. Additional information captured in the background includes various process and thread ID information required to uniquely identify an event within a system session and to link individual events to a full sequence (trace) needed for further processing stages. Based on aforementioned traces, BiG2-KAMAS uses two distinct mechanisms to further process arbitrary kernel event sequences:

Pattern inference: Our introduced framework has been developed in concert with an event extraction system called SEQUIN [11]. SEQUIN uses grammar inference extended with statistical evaluation to automatically identify and crop relevant sequences (rules) from traces of kernel-level behavioral data for further processing and visualization. Generally speaking, grammar inference is the process of computationally assembling a formal ruleset by examining the sentences of an unknown language [18]. In the information security domain, grammar inference is primarily used for pattern recognition, computational biology, natural language processing, language design programming, data mining, and machine learning. Grammar inference has also been proven to be a feasible approach to anomaly detection, since "algorithmic incompressibility is a necessary and sufficient condition for randomness" [19]. We use grammar inference as key component in the process of 'compressing' a sequential trace for extracting relevant behavioral patterns.

To achieve inference by compression in a computationally feasible way, we selected an algorithm that losslessly produces (without changes to order and immutability) a context-free grammar (CFG) in unsupervised operation. As opposed to context-sensitive grammars, languages created by a CFG can be recognized in $O(n^3)$ time, which is a relevant distinction for all future parsing efforts. The choice ultimately fell on Sequitur [20]. Sequitur is a greedy compression algorithm that creates a hierarchical structure (CFG) from a sequence of discrete symbols by recursively replacing repeated phrases with a grammatical rule. The output is a compressed representation of the original sequence. The algorithm creates this representation through the application of two base properties: *rule utility* and *bi-gram uniqueness*. Rule utility checks if a rule occurs at least twice in the grammar, while bi-gram uniqueness observes if two adjacent symbols occur only once. Assuming we have a string `abcdbcabcd`, where every character represents an event, the first bi-gram of that trace would be `ab`, followed by a second bi-gram `bc`, and so forth. See Table I for a complete example of the process.

Sequitur is linear in space and time. In terms of data compression, the algorithm can outperform other designs that achieve data reduction by factoring out repetition. It is almost as performant as designs that compress data based on probabilistic predictions [20].

Bi-gram extraction and scoring: In addition to rule inference, BiG2-KAMAS uses precomputed maliciousness scores of event bi-grams separately explored using a sentiment-like extraction system based on the log likelihood ratio (LLR) test [10]. An LLR test is a statistical method used test model assumptions, namely the quality of fit of a reference (null) and an alternative model. When determining the occurrence of rarely observed events – which are often at the core of malicious traces – likelihood ratio tests show significantly better results than alternatives such as x^2 or z-score tests [21].

TABLE I
OPERATION OF SEQUITUR AFTER [20]. PROPERTY APPLICATION IS *italicized*.

Symbol	String	Grammar	Remarks
1	a	S → a	
2	ab	S → ab	
3	abc	S → abc	
4	abcd	S → abcd	
5	abcdb	S → abcdb	
6	abcdbc	S → abcdbc	bc appears 2x
		S → aAdA	*bigram uniqness*
		A → bc	
7	abcdbca	S → aAdAa	
		A → bc	
8	abcdbcab	S → aAdAab	
		A → bc	
9	abcdbcabc	S → aAdAabc	bc reappears
		A → bc	
		S → aAdAaA	*bigram uniqness*
		A → bc	aA appears 2x
		S → BdAB	*bigram uniqness*
		A → bc	
		B → aA	
10	abcdbcabcd	S → BdABd	Bd appears 2x
		A → bc	
		B → aA	
		S → CAC	*bigram uniqness*
		A → bc	B used only 1x
		B → aA	
		C → Bd	
		S → CAC	*rule utility*
		A → bc	
		C → aAd	

In preparation for sentiment-assisted visualization, we use the LLR method to learn likely benign and malicious event sequences in big corpora of recorded kernel operations (traces). The resulting sentiment dictionary can be used to accurately and effectively determine if an investigated event bi-gram is contextually suspicious. Specifically, we compute the LLR score for each bi-gram to highlight collocations characteristic to sequences of malicious and benign system events [10].

The resulting occurrence counts (shown in Table II) are the basis for this calculation: Following the approach by [21], we define the number of times both event tokens occur in combination (k_{11}), the number of times each token has been observed independently from the other (k_{12} and k_{21}, depending on the relative position in the bi-gram), and the number of times the token was not present at all (k_{22}).

TABLE II
EVENT OCCURRENCE MATRIX [10]

	A	!A
B	k_{11}=$k(AB)$	k_{12}=$k(!AB)$
!B	k_{21}=$k(A!B)$	k_{22}=$k(!A!B)$

The same process is later applied to the pattern's general occurrence in a labeled benign versus malicious corpus. The final result is a normalized sentiment rating ranging from $+1.0$ (benign) to -1.0 (malicious). Unknown bi-grams are ultimately scored against the resulting dictionary, the outcome of which is at the core of the bi-gram evaluation feature in the new BiG2-KAMAS prototype.

B. Visualization Design

Structure: Wagner et al. [3] describe in their article that since IT-security experts are commonly familiar with programming IDEs, they used the design concept of IDEs like Eclipse or Netbeans for their prototype. The updates to the new prototype also follow this design concept approach. In contrast to the previous prototype, the new one has an additional view. In this initial view the KDB is situated on the left side, which can be compared to the project view in Eclipse. On the right side only the file load buttons are displayed, which can be compared to the initial view of Eclipse, where no project has been opened yet.

Coloring: For the rule highlighting as well as the Bi-Gram visualization we selected a sequential color scheme from red to blue. Red indicates that the rule or bi-gram is malicious and a blue one stands for a benign rule or bi-gram. To avoid problems with red and green hues for colorblind people [22, p. 124], we used blue instead of green and select colorblind-safe qualitative colors from Colorbrewer[1].

Layout: The prototype is structured into three parts: knowledge base, rule exploration area and call exploration area (see Figure 1). On the left side the knowledge base is visualized with it's 'Knowledge Database (KDB)' (see Figure 1:1a) and the KDB's color highlighting filters (see Figure 1:1b). The KDB is displayed as a tree, in which each category of the database can have several subcategories. Each category with subcategories is shown with a box icon (see Figure 1:1a) and the ones without subcategories are displayed with folder icons. Each rule, which is stored in the database, is displayed with a paper icon. Beneath the KDB the 'Knowledge Base Highlighting' filters are displayed (see Figure 1:1b). Each filter can be activated or disabled with its checkbox and updates the result of the prototypes filter pipeline and visualization of the 'Rule Overview Table' (see Figure 1:2a).

After loading and translating the input file, the system updates the 'Graphical User Interface' (GUI) and visualizes new elements. In the middle the 'Rule Exploration' area (see Figure 1:2) is visualized, while the right side contains the 'Call Exploration' area (see Figure 1:3).

In the 'Call Exploration' area all the included system or API calls of the loaded input file are represented in the call table (see Figure 1:2b) as described by Wagner et al. [3]. The rules included in the input file are visualized in the rule overview table located in the 'Rule Exploration' area (see Figure 1:2a). If the user loads several trace files, each trace file will be displayed as one rule.

[1]http://colorbrewer2.org

The background of the third column of the 'Rule Overview Table' indicates whether a rule is fully benign, partially benign, not known, partially malicious or fully malicious. The background of the malicious rules will be painted in red and the background of the benign rules in blue. The fully known rules will be displayed in a dark red/blue while the partially known rules are highlighted in a light red/blue (see Figure 1:1b). The red color highlighting for malicious activity is adopted of the KAMAS prototype [3]. If a rule is fully known and, therefore, highlighted in dark red, the rule is included as-is in the KDB. A partially known rule is only a part of one rule in the KDB. This kind of rule has at least one additional call at the beginning or at the end of a fully known rule [3]. If an input file was loaded, the system automatically calculates the knowledge state of each rule. For this purpose, the system compares each rule of the input file with each rule of the KDB. After the calculation process the system highlights the rules in the corresponding colors in the rule overview table.

Bi-Gram Visualization: The rule detail table is located next to the rule overview table (see Figure 1:2b). The rule detail table automatically updates its content when clicking on a rule in the rule overview table and represents all system and API calls included in the selected rule. From left to right, the table displays the unique id as well as the name of the call. The last column visualizes the new bi-gram based valuation approach for the corresponding calls. As mentioned before, the prototype uses the bi-gram approach of Luh et al. [10]. A bi-gram is an n-gram where the length of $n = 2$. An n-gram, in turn, is a coherent sequence of n elements. In this approach the elements are system or API calls. Each bi-gram has a score in the range [-1, 1], which indicates whether this pair of calls is malicious or benign. For bi-gram based valuation, two different visualization approaches were implemented following a semantic zooming approach: First, if the width of the bi-gram column is bigger than 75px, the prototype visualizes the bi-gram values as bar charts (see Figure 2:a), whereby each bar starts in the middle of the bi-gram column. If the bi-gram score is between 0 and -1, the bi-gram is malicious. Therefore, the red color bar chart unfurls from the middle towards the left side. If the bi-gram score is between 0 and 1 the bi-gram is benign and the bar chart is visualized from the middle to the right side in a blue color. The colors correspond to the KDB highlighting. The visualization approach was chosen to give the user a quick but still precise overview of the bi-gram based scores.

If the width of the bi-gram column is smaller than 75px and therefore the bar charts are hardly recognizable, the system switches to the second visualization. Here, the bi-gram values are visualized as a color-filled rectangle (see Figure 2:b). As before, a red colored rectangle indicates that the bi-gram is malicious and a blue one stands for a benign bi-gram. To visualize the value of the malicious or benign bi-gram, the system changes the alpha value of the displayed color. Therefore, the darker the color, the higher the respective value. Since the difference of an alpha value between 255 and 240 is

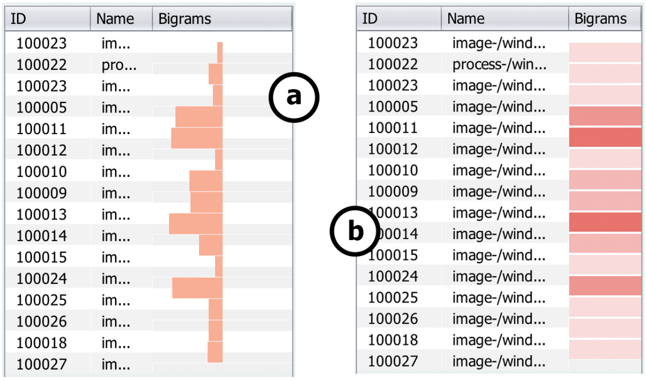


Fig. 2. The two different visualisations methods of the call bi-grams. The first method visualises the bi-grams as bar charts (a), whereas the second visualisation uses the alpha channel to show the severity of the bi-gram (b).

not easy to recognize and every value below 100 is generally difficult to see, we decided to implement only four graduation steps for the alpha value. The visualization with the alpha value is less precise than the visualization with the bar charts but, at the same time, significantly easier to interpret. Table III shows the different graduation steps and their value ranges.

TABLE III
COLOUR GRADUATION STEPS FOR THE ALPHA VALUE BI-GRAM VISUALISATION.

Colour	Alpha value	Value ranges
	200	>= 0.75
	150	>= 0.5 && <0.75
	100	>= 0.25 && <0.5
	50	>= 0 && <0.25
	50	<0 && <= -0.25
	100	<-0.25 && <= -0.5
	150	<-0.5 && <= -0.75
	200	<-0.75

C. Interaction

Like the KAMAS prototype of Wagner et al. [3], the BiG2-KAMAS's functionality will be described in accordance to the four steps of the visual information seeking mantra of Shneiderman et al. [23], namely overview, rearrange and filter, details-on-demand and, extract.

Overview: The BiG2-Kamas prototype has an additional initial view where the user can decide whether to load a Sequitur input file or several raw trace files. When the analyst loads a Sequitur file, the rule and call tables will be filled with the rule and call data included in the input file. Each entry in the rule overview table represents one rule of the loaded cluster. Furthermore, the histograms in the rule exploration area give a quick impression of the distribution in the rule occurrence and length [3]. When the user loads one or more trace files the rule and call tables will also be filled with the data of these files. Contrary to a loaded Sequitur file, each entry of the rule overview table represents an entire trace file. Thus, if the user loads three traces the rule overview table will have only three rows. Furthermore, due to the fact that the user analyses several independent trace files the histogram for the rule occurrence is insignificant. Therefore, only one histogram for the trace length will be displayed in the rule filter area.

Rearrange: If the rule overview table and the call overview table are loaded with data, the user can rearrange their content by clicking on a table's column. This will re-sort the included data and update the visualization [3]. The content of the rule detail table cannot be rearranged since the calls are shown in their sequential order and should therefore not be changeable.

Filter: In the next step the user can reduce the number of rules or trace files by using the rule/trace and call filters [3]. No matter which files were loaded, the user always has the opportunity to filter the rules or traces by the included calls (events). The user can rearrange the call filters or select a specific call in the call overview table to reduce the number of shown rules [3]. Furthermore, the analyst can filter the rules or specific traces by using the filters in the rule exploration area. If loading a Sequitur file, the analyst can filter the rules by their occurrence, length, whether they are equally distributed in the input file or if they match, partially match, or don't match the stored rules in the KDB [3]. By changing the filter settings, the included rules in the rule overview table automatically update immediately. If one or more trace files were loaded, the analyst can only filter the shown traces in the rule overview table by their length. In addition, the highlighting and filtering of the KDB is switched off.

Details-on-Demand: If the user wants to analyze a rule or trace, he/she can open the rule/trace in the rule detail table by selecting it in the rule overview table. This will display all the included calls in the rule detail table in their sequential order [3]. The bi-grams provide information whether a combination of two calls is malicious or benign. This should support the user in finding interesting call sequences more quickly.

Extract: Independent of the loaded files the analyst can add a new rule to the database using two different ways. One method is to simply select one rule or trace in the rule overview table and simply drag and drop it in one leaf category of the KDB. This will add the entire rule or trace file to the database [3]. Alternatively, the analyst can select several calls of interest in the call overview table and add these by dragging and dropping them to the KDB. When adding a new rule to the KDB, a popup window will show up where the analyst can assign the rule a specific name. If the user has loaded a Sequitur file, the system will now update the knowledge state for all rules as well as the highlighting in the rule overview table for further analysis.

D. Implementation

Since the BiG2-KAMAS prototype is based on the prototype of Wagner et al. [3], it also uses a data-oriented design concept [24]. To increase the performance of the prototype, the system only works with integer comparisons. Therefore, the input data only includes the call ids. It is only possible to translate a call id to the actual call value with an additional translation file. This translation file is also used for the bi-grams. The original bi-gram file has several columns in which only the string values of the system or API calls are stored. To increase the performance and to reduce memory usage, the BiG2-KAMAS prototype generates its own bi-gram file. When starting the prototype the system checks with md5 hash values to determine whether the translation file or the original bi-gram file has changed. If so, the system converts the original bi-gram file to the translated bi-gram file in which also the integer values of the system calls are stored. Like the prototype of Wagner et al. [3] the new prototype is using the action pipeline for filter options. This enables dynamic query environments and real-time data operations.

To evaluate the robustness and performance of the BiG2-KAMAS prototype three different Sequitur cluster-grammar files containing between 10 and 500 rules were used. The file with 500 different rules contained a total amount of 30,000 system and API calls. To test the bi-gram functionality, a bi-gram file with nearly 117,500 bi-gram entries was loaded. On a machine with an 2.1GHZ Dual-Core processor and 12GB of memory it took the system about four minutes to translate the original bi-gram file to the translated bi-gram file. The malware and bi-gram samples were collected by collaborators in the Josef Ressel Center TARGET of St. Pölten UAS.

IV. Externalized Knowledge Integration

As Wagner et al., [3] described in their article, we integrated a knowledge database to support the user during their analysis tasks. The KDB is based on the malware behavior schema of Dornhackl et al., [25]. The KDB is located at the left side of the prototype and is implemented in a hierarchical structure (tree structure). In the BiG2-KAMAS prototype the KDB was extended by one additional category to store the benign rule data, namely benign activity. In the current version of the prototype there is only one category to store benign rule data. Each category is displayed with either a box or a folder icon, the category description and the number of included rules in the integrated subfolders. The analyst can add new rules by drag & drop. When adding a new rule, the KDB automatically unfolds closed categories. Additionally, a popup window opens in which the analyst can enter a rule name. To investigate a rule stored in the KDB, the user can open a context menu by right clicking on the chosen rule. The context menu will show two different menu items, namely ‘Information’ and ‘Delete’. The information menu item opens a popup window in which the analyst is presented the following information:

- **Assigned Concept:** This information tells the analyst in which schema category (concept) the rule is currently categorized. The assigned concept is implemented as a selection list to give the user the opportunity to change the assigned concept. For that purpose, the analyst must select a different concept in the list and press the save button at the bottom of the pop up window.
- **Rule Name:** Here, the actual rule name is displayed. The rule name is implemented as a text field to quickly change it if necessary.
- **Included Calls:** Finally, the calls included in the stored rule are displayed in a table. Thus, the calls are visualized in their sequential order and each call will be shown with its unique call id which corresponds to the call id of the translation file and the actual call value. In the current version of the prototype it is only possible to investigate the included calls in their sequential order, but not to delete specific calls which are listed in the table.

The second menu item is the “delete’ item, which allows the analyst to delete the currently selected rule. Furthermore, when selecting a concept instead of a rule, the BiG2-Kamas prototype will show a context menu with which the user can disable a category and all its integrated subcategories. Thus, the analyst can disable the entire KDB or only specific categories. If the user disables a category all the included rules will no longer be considered in the knowledge base highlighting and filtering.

When the user clicks the right mouse button to open the corresponding context menu before selecting a rule or category, the system automatically selects the rule/category at the actual mouse position.

Searching: If the user searches for interesting rules or specific calls or call groups he/she can use the call filter options to reduce the data to be analyzed. In the call exploration area, the user can search for a specific call by entering its name or use regular expressions to find an entire call group. Beneath the search text field the user can enable case sensitive search with the corresponding checkbox ’Case Sensitive’. Filtering or searching the calls affects the data shown in the call overview and rule overview table. Additionally, to find rules of interest the analyst can use the rule exploration filters or the knowledge base filters.

V. Prototype Evaluation

This section describes the procedure of the performed user studies, the specific results, as well as further feature requests. For the prototype validation, a user study with two domain experts was conducted. The domain experts validated the functionality as well as the visual design interface.

Participants: Both participants work at St. Pölten UAS and have more than five years of experience in the field of malware analysis. The first participant is between 30 and 39 years of age, male and holds a masters degree. The second participant is between 60 and 69 years of age, male, and holds a PhD. Generally, both participants are well experienced in this field and can be categorized as experts.

Design and Procedure: Each participant was interviewed individually and had already tested the previous version of the prototype at least once. First, the participants received a short introduction to the new features of BiG2-KAMAS and also a quick reminder of the basic features and workflow. The participants were asked to mention additional missing functionalities and to criticize all potential usability issues.

Both participants took part in the same two scenarios: First, the participants had to load a Sequitur file, investigate the loaded rules and filter specific call sequences. At the end they had to store a rule in the KDB and name it. In the second scenario, the participants had to load three trace files. They were asked if they perceived any differences when loading trace files instead of a Sequitur file. At the end they had to investigate a rule stored in the KDB and move it to a different category.

Equipment and Materials: The latest version of the BiG2-KAMAS prototype was used in the evaluation. For the first user scenario, the participants had to load a Sequitur file with about 500 rules and 30,000 system and API calls. In the second scenario, three trace files with a length between ten and fifteen calls were used. The bi-gram file had a total number of about 117,000 bi-grams. The translated bi-gram file had already been generated so that the participants did not have to wait until the system finished the translation process. As evaluation equipment, two different setups were used. Both participants worked on a 13 inch Macbook Pro with a Retina display (screen resolution of 2560x1600) and a mouse for navigation. Participant #1 worked with an additional 20 inch Monitor with a full HD screen resolution and an external keyboard. Each user test was conducted with the same version of the BiG2-KAMAS prototype and was documented on paper.

A. *Results*

The following section discusses the results of both scenarios. Both the results of ‘Scenario 1’ (Sequitur file) and ‘Scenario 2’ (trace files) will be presented. Both participants had no problem loading the different files for the user scenarios.

Scenario 1: *Loading and Analyzing a Sequitur file.* Both participants quickly recognized the additional color scheme for the new benign category. The colors for the knowledge base highlighting were assessed as easily understandable and the additional rule counter next to the knowledge base filters were mentioned as being very useful. Participant 1 mentioned that if a rule in the rule overview table is highlighted, it would be useful to know which rule or rules of the KDB match this rule in the table. Therefore, a tooltip would be helpful which tells the user the names of the matching rules of the KDB. Furthermore, participant 2 suggested to always show the rule counter of the KDB’s categories. If there are currently no rules in a category, the counter should be zero.

When participant 2 first saw the bar chart bi-gram visualization, he assumed it visualizes the occurrence of the combined call sequence. In contrast, the alpha color visualization was immediately recognized as an indicator for maliciousness or benignity. Participant 1 also mentioned that the alpha color visualization is easier and faster to recognize. Furthermore, both participants mentioned that the color visualization is not as precise as the bar chart visualization and therefore would only be useful for initial malware classification. Participant 1 suggested an additional tooltip to display the accurate bi-gram value. Participant 2 remarked that it would be more useful if the calls in the call overview table only showed the beginning and the end of the call’s value. This would simplify finding a specific call in a group of similar calls. Additionally, he recommended a search button for the regular expression call filter. This could help some users, since currently it is only possible to search by pressing the enter key. Adding a new rule to the KDB was no challenge for either participant and both valued the ability to give the rule a specific name.

Scenario 2: *Loading and analyzing three trace files.* Both participants had no difficulties with loading the three trace files. They also recognized quickly that each entry in the rule overview table now represents one trace. Neither of them realized that the knowledge base filters and highlighting were disabled. Participant 1 suggested to gray out the knowledge base filters to make it clear that these are disabled. Participant 2 proposed to change the headings for the trace file analysis view in order to avoid confusion. He remarked that it could be misleading if the headings say e.g. ‘Rule Overview Table’ when analyzing a trace file. Furthermore, both participants recommended to change the occurrence column in the rule overview table to the file names of the traces. As the last task, the participants had to change the corresponding category of a random rule. Even if both participants solved this task easily, both remarked that it would be useful if the user could move a rule from one category to another per drag & drop.

B. *Result Analysis*

This section gives an overview of the issues which were mentioned during the expert reviews. Like Wagner et al. [3] each issue was rated based on Nielsen’s [26] severity ratings. Table IV shows the potential new features noted by the test persons and includes three columns: ‘feature requests’ (FR), ‘severities’ (SE) and the effort it would take to implement these changes [3]. The features mentioned in the table include small cosmetic changes as well as real usability improvements. The only feature mentioned by all participants is an additional tooltip which shows the actual bi-gram values.

TABLE IV

LIST OF REMARKED FEATURE REQUEST AND SEVERITIES AND THE EFFORT IT WOULD TAKE TO IMPLEMENT THEM IN THE PROTOTYPE. (FR: 1 = NICE TO HAVE, 2 = GOOD FEATURE, 3 = ENHANCES USABILITY; SE: 1 = MINOR, 2 = BIG, 3 = DISASTER; EFFORT: 1 = MIN, 2 = AVERAGE, 3 = MAX) [3].

Description	FR	SE	Effort
KDB: Move a rule to another category by using drag & drop.	2	1	1
KDB: Show the rule counter even if zero rules are included.	1	-	1
KDB: Gray out the knowledge base filters if they are disabled.	2	1	1
Tables: Highlighted rules in the rule overview table should show the KDB’s corresponding rules.	3	2	3
Tables: Change the occurrence column to the trace file names.	2	1	2
Tables: Show only the begin and the end of the calls in the call overview table.	3	2	2
Tables: Implement a search button for the call regex search.	1	-	1
Bigram: Tooltip to show the bi-gram values.	3	-	1
Headings: Change the headings when loading trace files.	2	-	1

VI. Discussion & Reflection

The performed user studies described in Section V confirmed that the four feature requests, which are determined in Section I are fulfilled by the BiG2-KAMAS prototype:

1) *Generic data loading:* The BiG2-KAMAS prototype is structured to enable the generic loading of data sequences. To make this possible the input data as well as the prototype's database are based on unique identifiers (id) instead of the actual values. Thus, all system-internal comparisons are based on integer values instead of string values. Only with the corresponding translation table, the system can translate the ids to the actual values. Thus, it is possible to load data sequences independent of their actual values as long as there is a translation table through which the prototype can translate the data. Furthermore, the system was adopted to also offer the opportunity to load raw system or API call based traces. In this state the KDB highlighting and filtering is disabled but the user can explore the loaded trace files and add new rules to the KDB. The prototype can't only load Sequitur call sequences, but also independent data sequences as long as the the data sequence has the given structure and a translation file.

2) *Extend the KDB with benign rules:* To fulfill this requirement the KDB was extended with an additional category for benign activity. In this category, all rules which are identified as benign can be stored. Additionally, the KDB's highlighting and filter pipelines were extended to identify and filter partially and fully benign rules. Rules with a partially or fully benign knowledge state are highlighted in blue in order to avoid the combination of the colors red and green.

3) *Implementation of bi-gram based valuation:* To support the bi-gram approach of Luh et al, [10] the prototype's rule detail table was adopted. Since many domain experts mentioned [3] that the arc-diagram visualization is not very helpful, it was replaced by the bi-gram visualization. Bi-gram based valuation is implemented with two different approaches. If the width of the bi-gram column is bigger than 75px the valuation is visualized with bar charts and colored in red (malicious) or blue (benign). If the width is less than 75px the bi-gram visualization uses the alpha channel to show the severity of the bi-gram (see Table III).

4) *User studies to validate the new features:* The results of the user studies show further feature requests which could be implemented in a future project. However, both participants mentioned that the bi-gram visualization is very helpful for identifying potentially malicious or benign call sequences and, therefore, helps to decide whether a rule is malicious or not.

Future Work: For the behavior-based malware analysis process, it could be valuable to implement a rule creation process where the analyst can build their own rules based on the known system and API calls [27]. Furthermore, it could be beneficial to edit the stored rules in the KDB or to build new rules based on existing patterns. Further avenues for future work are to include possibilities to hide, shrink an expand areas to provide the user with more flexibility. Moreover, to update the occurrence column of the Call Exploration area (see 1:3a) to show the relation to the total number of occurrences included in the loaded file. Additionally, normalizing the occurrence dataset and visualization to this total could be beneficial.

Categorization of BiG2-KAMAS: Like the KAMAS prototype [3] the BiG2-KAMAS prototype can be categorized as a Malware Forensic as well as a Malware Classification tool in the Malware Visualization Taxonomy of Wagner et al. [8]. However, due to the bi-gram based valuation the BiG2-KAMAS prototype offers the malware analyst an additional assistance for the Individual Malware Analysis.

VII. Conclusion

In this work, we presented a design study for a Bi-gram Supported Generic Knowledge-Assisted Malware Analysis System (BiG2-KAMAS). The prototype is based on the KAMAS prototype [3] and extended by additional features such as generic data loading, an extension of the KDB to enable the analysis of benign rules, and the implementation of a bi-gram based valuation approach. The requirements were discussed in a focus group meeting and then implemented as part of a functional prototype. After implementing the new features, two user studies were conducted to evaluate the design and the functionality of the new BiG2-KAMAS prototype.

Acknowledgments

The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged.

This work was supported by the Austrian Science Fund (FWF) via the “KAVA-Time” project (P25489-N23) and by the Austrian Federal Ministry of Science, Research and Economy under the FFG Innovationscheck (no. 856429). We would also like to thank all focus group members and test participants who have agreed to volunteer in this project.

References

[1] M. Sikorski and A. Honig, *Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software*, 1st ed. No Starch Press, 2012.

[2] M. Wagner, W. Aigner, A. Rind, H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato, “Problem characterization and abstraction for visual analytics in behavior-based malware pattern analysis,” in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec ’14. ACM, 2014.

[3] M. Wagner, A. Rind, N. Thür, and W. Aigner, “A knowledge-assisted visual malware analysis system: Design, validation, and reflection of KAMAS,” *Computers & Security*, vol. 67, pp. 1–15, 2017.

[4] N. Thür, M. Wagner, J. Schick, C. Niederer, J. Eckel, R. Luh, and W. Aigner, “Big2-kamas: Supporting knowledge-assisted malware analysis with bi-gram based valuation,” in *Poster of the 14th Workshop on Visualization for Cyber Security (VizSec)*, Phoenix, Arizona, USA, 2017.

[5] H. Shiravi, A. Shiravi, and A. Ghorbani, “A survey of visualization systems for network security,” vol. 18, no. 8, pp. 1313–1329, 2012.

[6] M. Egele, T. Scholte, E. Kirda, and C. Kruegel, “A survey on automated dynamic malware-analysis techniques and tools,” vol. 44, no. 2, pp. 6:1–6:42, 2008.

[7] Z. Bazrafshan, H. Hashemi, S. Fard, and A. Hamzeh, “A survey on heuristic malware detection techniques,” 2013, pp. 113–120.

[8] M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner, “A survey of visualization systems for malware analysis,” in *Eurographics Conference on Visualization (EuroVis) - STARs*. The Eurographics Association, 2015.

[9] L. McNabb and R. S. Laramee, "Survey of surveys sos - mapping the landscape of survey papers in information visualization," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 589–617, Jun. 2017. [Online]. Available: https://doi.org/10.1111/cgf.13212

[10] R. Luh, S. Schrittwieser, and S. Marschalek, "LLR-based Sentiment Analysis for Kernel Event Sequences." IEEE, 2017.

[11] R. Luh, G. Schramm, M. Wagner, and S. Schrittwieser, "Sequitur-based Inference and Analysis Framework for Malicious System Behavior," 2017.

[12] O. Somarriba, U. Zurutuza, R. Uribeetxeberria, L. Delosières, and S. Nadjm-Tehrani, "Detection and visualization of android malware behavior," vol. 2016, p. e8034967, 2016.

[13] S. Marschalek, R. Luh, M. Kaiser, and S. Schrittwieser, "Classifying malicious system behavior using event propagation trees." ACM Press, 2015, pp. 1–10.

[14] B. Xiaofang, C. Li, H. Weihua, and W. Qu, "Malware variant detection using similarity search over content fingerprint." IEEE, 2014, pp. 5334–5339.

[15] A. Jain, H. Gonzalez, and N. Stakhanova, "Enriching reverse engineering through visual exploration of android binaries," in *Proceedings of the 5th Program Protection and Reverse Engineering Workshop*, ser. PPREW-5. ACM, 2015, pp. 9:1–9:9.

[16] O. E. David and N. S. Netanyahu, "DeepSign: Deep learning for automatic malware signature generation and classification." IEEE, 2015, pp. 1–8.

[17] P. M. Wrench and B. V. W. Irwin, "Towards a PHP webshell taxonomy using deobfuscation-assisted similarity analysis." IEEE, 2015, pp. 1–8.

[18] A. Stevenson and J. R. Cordy, "A survey of grammatical inference in software engineering," *Science of Computer Programming*, vol. 96, pp. 444–459, 2014.

[19] L. Ming and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*. Springer Heidelberg, 1997.

[20] C. G. Nevill-Manning and I. H. Witten, "Identifying hierarchical structure in sequences: A linear-time algorithm," *J. Artif. Intell. Res. (JAIR)*, vol. 7, pp. 67–82, 1997.

[21] T. Dunning, "Accurate methods for the statistics of surprise and coincidence," *Computational linguistics*, pp. 61–74, 1993.

[22] C. Ware, *Information Visualization: Perception for Design*. Elsevier, 2012.

[23] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *Proc. of VL*, 1996, pp. 336–343.

[24] R. Fabian, "Data-Oriented Design," 2013, accessed on Nov. 11, 2015. [Online]. Available: http://www.dataorienteddesign.com/dodmain/dodmain.html

[25] H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato, "Malicious behavior patterns," in *SOSE*. IEEE, 2014, pp. 384–389.

[26] J. Nielsen, *Usability engineering*. Boston: Academic Press, 1993.

[27] M. Wagner, A. Rind, G. Rottermanner, C. Niederer, and W. Aigner, "Knowledge-assisted rule building for malware analysis," in *Proceedings of the 10th Forschungsforum der österreichischen Fachhochschulen*, FH des BFI Wien. Vienna, Austria: FH des BFI Wien, 2016.

Rule Creation in a Knowledge-assisted Visual Analytics Prototype for Malware Analysis

Johannes Schick[1], Markus Wagner[2], Niklas Thür[2], Christina Niederer[2], Gernot Rottermanner[2], Paul Tavolato[2], Wolfgang Aigner[2]
St. Pölten University of Applied Sciences, Austria
Email: [1]dm171566@fhstp.ac.at, [2]first.last@fhstp.ac.at

***Abstract*—The increasing number of malicious software (malware) requires domain experts to shift their analysis process towards more individualized approaches to acquire more information about unknown malware samples. KAMAS is a knowledge-assisted visual analytics prototype for behavioral malware analysis. It allows IT-security experts to categorize and store potentially harmful system call sequences (rules) in a knowledge database. To meet the increasing demand for individualization of analysis processes, analysts should be able to create individual rules. This paper is a visualization design study, which describes the design and implementation of a Rule Creation Area (RCA) into KAMAS and its evaluation by domain experts. It became clear that continuous integration of experts in interaction processes improves the knowledge generation mechanism of KAMAS. Additionally, the outcome of the evaluation revealed that there is a demand for adjustment and re-usage of already stored rules in the RCA.**

I. Introduction

Nowadays, domain experts have to deal with an ever increasing number of malicious software (malware) which in addition is becoming more targeted, persistent and unknown. Malwares are able to disturb computer operations and gather personal information of the system's owner without raising attention [1] [2]. When it comes to analyzing methods, there are two approaches for the identification of malware. On the one hand, software can be analyzed without actually executing it, which is called static analysis. Obfuscation techniques used by malware developers can render this task virtually impossible [3]. Dynamic analysis observes actions performed by potential malware while it is being executed in a protected environment. More precisely, analysts observe execution traces of programs; for the sake of simplicity malware analysts often reduce these traces to function calls, neglecting all other simple machine instructions. Therefore, dynamic analysis is also known as behavior-based analysis [3].

In behavior-based analysis malware analysts have to deal with large amounts of data, which can lead to a very complex analysis process: a trace of a malware sample may often comprise thousands of system calls and analysts have to find similar system call patterns within thousands of such traces. In order to simplify this process, analysts need automated approaches for finding such patterns and categorizing them as potentially harmful or harmless. However, such identification of patterns relies heavily on the analysts knowledge, which makes it impossible to automate this process completely [4]. These patterns of behaviors can be defined as a formal language using formal grammars (syntactic pattern recognition [5], [6] or for more details [7]). The task of the analyst is the development of a set of grammar rules incorporating their knowledge about (malicious) behaviors of malware samples. In this context, visual analytics (VA) is needed to support the analysts in integrating their knowledge. VA plays an essential role in supporting data analysis, since it combines data processing capabilities of computer systems with the knowledge and experience of users [8].

According to Keim et al. [9], VA also connects automated analysis techniques with interactive visualizations in order to combine different types of information and obtain understanding from complex data sets. To make reasoning out of this massive amount of data, it is necessary to include ”implicit” [10] or ”tacit” [11] knowledge in the analysis process. By externalizing the implicit/tacit knowledge of domain experts, it is possible to provide explicit knowledge in form of data, which is independent from the current user of the system. This extracted knowledge can subsequently be connected through interactive visualization tools [11]. In addition to these findings, Lee et al. [12] stated that visualization is necessary to analyze potential malware more effectively.

This paper provides a design study [13] dealing with the implementation and evaluation of a separate Rule Creation Area (RCA) [14] into a Knowledge-Assisted Visual Malware Analaysis System (KAMAS) [15]. In order to meet quality standards, this paper follows a problem-oriented research approach. In conjunction with this prototype, this means that the visualization and implementation of the system is performed under consideration of specific real-world problems defined by domain experts [13]. Thus, the main objectives of this research are:

- Clarify, why the implementation of a separate area for rule creation is necessary and how malware analysts can benefit from it.
- Presenting the design and implementation of the RCA into the KAMAS system with a detailed description of all involved components and functionalities.
- Conducting an evaluation of the implemented system in cooperation with malware analysis experts to proof the effectiveness of the deployed functionalities.
- Reflection of the implemented functionalities under consideration of their evaluation by real world users and the resulting future research.

II. Related Work and Background

Since there were no interactive visualization tools available which cover all requirements for malware analysts, Wagner et al. [15] developed the KAMAS prototype. With KAMAS, analysts are able to categorize function call traces in terms of their potential harmfulness and store them into a knowledge database (KDB). The KDB assists them in further observation and simplifies the analyzing process. In order to expand the range of functionalities and subsequently improve the effectiveness of KAMAS, Wagner et al. [16] suggested an interface design for the RCA [14]. The RCA allows the construction of completely new rules by using single system and API calls in the same structure as generated by the sequitur algorithm [17]. These rules can subsequently be stored in the KDB.

Knowledge Generation in VA: Thomas and Cook [18] define VA as approach to gain knowledge from massive, dynamic, ambiguous, and often conflicting data. Based on the research by Sacha et al. [19] following findings can be determined: VA uses data to draw conclusions on a specific application field and gain insights into the problem domain. On the one hand, there is the combination of perceptive skills following the capability of drawing conclusions by humans. On the other hand, there is the computing and data storage capability of machines. Both of these aspects can be combined in visual representations. Interactions with VA tools provide a possibility for analysts to detect patterns in their data, thus assisting them in verifying or falsifying their initial hypothesis. By clustering and classifying the found patterns, the outcome of their exploration can be visualized.

According to Chen et al. [10], the aim of knowledge-assisted visualization is to automate reasoning about abstracted information from a set of data. Additionally, they also mentioned that the field of knowledge-assisted visualization is still in development, since the growing amount of data requires systems to continuously adapt to these challenges. As stated by Wang et al. [11], the nature of tacit knowledge can be defined as intimate and specialized. However, by using interactive visualization it can be connected with explicit knowledge, which is processable by computers or can be stored in a database [11]. Recent research has focused even more on the role of humans in this process. In order to deal with the increasingly ambitious challenges in the field of VA, the philosophy has to shift from a ”human in the loop” philosophy to a ”human is the loop” viewpoint [20]. This new approach focuses on recognizing the workflow of analysts and consequently adapting interaction processes to the needs of the analyst.

Appliance of VA Techniques to Malware Analysis: According to Alazab et al. [21], all executable programs have the aim to perform actions using API calls. The process of malware analysis involves the observation of system call sequence patterns and the actions they cause. Both Alazab et al. [21] and Mohaisen et al. [22] emphasize classification and clustering of patterns in terms of their maliciousness or benignity as a main task for malware analysts. AMAL, a behavior-based malware analysis system by Mohaisen et al. [22] is an example for a program, which is capable of fulfilling this task. Just like KAMAS, it tries to tackle shortcomings of existing systems by combining methodologies of static and behavior-based approaches. By running malware samples in a virtualized environment, the system collects data which is subsequently used for automated classifying and clustering of samples into different malware families. However, AMAL does not provide an interactive user interface, nor does it provide the possibility to integrate externalized expert knowledge into the analysis process. Another project with similar approach to KAMAS is the visualization tool SEEM [23], which enables analysts to compare large sets of malware and their associated attributes.

As mentioned in Section I, supportive visualization is needed in order to provide a more efficient approach of analyzing potential malware samples [12]. In their state of the art report on visualization systems in the field of malware analysis, Wagner et al. [24] concluded that future systems should provide a compound of classification overviews for comparison and detail views for individual analysis.

III. Method

In general, this paper is a design study following the design principles/ideas proposed in [13], which is described as problem-orientated research approach. This includes a problem definition, the design and implementation of a visualization system which solves the problem, the evaluation of the prototype as well as a reflection about lessons learned and possible improvements [13]. The problem was defined by Wagner et al. [16] in their design study, which addresses the need for the implementation of a separate area for rule creation in the KAMAS prototype. All scientific publications directly related to KAMAS [4] [16] [15] [24] served as a basis for the general understanding of the prototype and its background.

Requirements & Features: The functionalities and interface design of the RCA were designed according to the rule building screen prototype ’CallNet’ presented in [16]. Furthermore, ’CallNet’ and its desired functionalities were already reviewed by usability experts. In order to ensure compatibility with the sequitur algorithm [17], the extracted knowledge has to be stored in a rule-based interface and structure. Based on the task definitions and the outcome of the design study, following key requirements (R) for the RCA can be defined:

- **R1 Consistency:** To ensure an effective usage of the user interface, it is necessary to provide consistent interaction techniques throughout the whole system. In this specific case, the input data for the RCA originates from another interface section of the program. Therefore, the interaction visualization should be related to the movement of data, e.g. ’Drag & Drop’ operations.
- **R2 Creation Support:** The amount of data offered by this system is particularly high. Thus, additional support in the process of rule creation is important. By giving the analyst additional interaction possibilities, e.g. automatically validated suggestions for single calls, the rule creation process can be accelerated.

Moreover the interface has to provide the possibility to switch the highlighting of these calls based on higher or lower frequency to support creating rules with individual preferences.

R3 Editing Options: The process of rule creation requires the system to allow the editing of rules at any time and to offer a quick way to restart the process. As a consequence, the analyst has to be able to reorder and delete single calls of the dropped call sequence and to reset the whole RCA to its default state.

R4 Knowledge Extraction and Extension: Finally, the newly created rule should be used to extend the spectrum of computerized knowledge in the system. Therefore, it is necessary to offer the possibility of moving rules from the RCA to the KDB. By implementing this functionality, the knowledge generation loop (see Figure 2) can be expanded, which should subsequently improve the effectiveness of the analysis process.

The features of the RCA were implemented according to the defined key requirements. In general, the design and implementation followed a user-centered design process [25]. During the development process, continuous exchange with researchers/developers of the KAMAS prototype was performed. Thus, it was possible to adjust requirements and discuss alternative solutions.

Evaluation: In order to evaluate the implemented features of the prototype, two malware analysis experts reviewed the system in the course of a semi-structured, qualitative user test. During this test, both experts had to solve different tasks, which occur in the rule creation process. The results were documented by written notes and afterwards categorized based on their importance. Afterwards, the results of the evaluation were summarized and rated in a list inspired by Nielsen's severity rating procedure [26]. With these ratings, it was possible to provide a clear overview of the most important findings as well as potentially negligible aspects.

IV. Design and Implementation

The design and functionalities of the RCA are based on the 'CallNet' prototype [16], which allows users to create rules from scratch with system and API calls. The KAMAS prototype and its implemented functionalities developed by Wagner et al. [15] served as a basis to expand the prototype's spectrum of features. The implementation resulted in the realization of the RCA (see Figure 1), which was achieved using the programming language Java.

Call Exploration: The 'Call Exploration' table (see Figure 1.3) provides a list of all system and API calls of the loaded file showing their occurrence in the file, the name and the ID of the call. In addition to the already available functionalities from [15], the possibility to drag single calls from the table to the RCA was implemented.

KDB: The KDB (see Figure 1.1) offers the possibility to save and organize rules in tree structure based concepts visualized as folder structure. Furthermore, the analyst can access information of already stored rules like the name, the assigned concept and the calls it consists of.

RCA in General: The RCA (see Figure 1.2) generally consists of three main areas. First, the analyst can drop single calls, which he previously selected and dragged from the 'Call Exploration' table into the Rule Creation Table (RCT) (see Figure 1.2.b). Secondly, above and below the RCT, the interface provides suggestions for single calls which occur either before (see Figure 1.2.a) or after (see Figure 1.2.c) the dropped system call sequence. At last, on the bottom of the RCA the analyst has the possibility to reset the whole RCA to its default state (see Figure 1.2.d) and to switch the highlighting of the call suggestions (see Figure 1.2.e).

Rule Creation Table in the RCA: After adding the first call from the 'Call Exploration' table to the RCT, an additional row gets added on the top of the table. This row makes it possible to drag the newly created rule (which contains all single calls inside the RCT) and add it to the KDB. Furthermore, the number in the second column of the RCT represents the occurrence of the newly created rule in the loaded file. If there is a need to reorder calls inside the RCT, this can be achieved by simply dragging a single call and move it to the desired position. The original call from the desired position then switches position with the dragged call. Also, single calls can be deleted from the RCT by right clicking on the desired call and using the 'Delete' pop-up. It must also be pointed out that every interaction performed in the RCT affects the occurrence column and call suggestions, since these components depend on the values inside the RCT. Even though a rule usually can contain 1 to n calls, the maximum number of calls inside the RCT was limited to eight calls in order to provide enough space for the other areas in the RCA.

Call Suggestions in the RCA: Above and below the RCT, the interface offers suggestions for calls, which can be dragged and dropped into the RCT. The suggestions above (see Figure 1.2.a) represent calls from the loaded file which occur before system call sequences with the same structure as the one inside the RCT, whereas the calls below (see Figure 1.2.c) represent calls which occur after the currently dropped system call sequence. Moreover, the font size of the call suggestions varies depending on their occurrence. By default, more frequent single calls are displayed with a bigger font. If a single call appears in multiple system call sequences of the loaded file, the font size increases by one for every found similar single call. Thus, every call suggestion displayed in the user interface is unique and the analyst gets a better overview of which single calls are more or less frequent.

Control Buttons in the RCA: Provided that the RCT contains at least one single call, a 'Reset' button (see Figure 1.2.d) is available at the bottom of the RCA. This button offers the possibility to set the whole area back to its default state. The second button (see Figure 1.2.e) is responsible for handling the highlighting of the call suggestions and is only visible when the currently dropped system call sequence offers suggestions. With the use of this button, the analyst can switch between highlighting more or less frequent call suggestions.

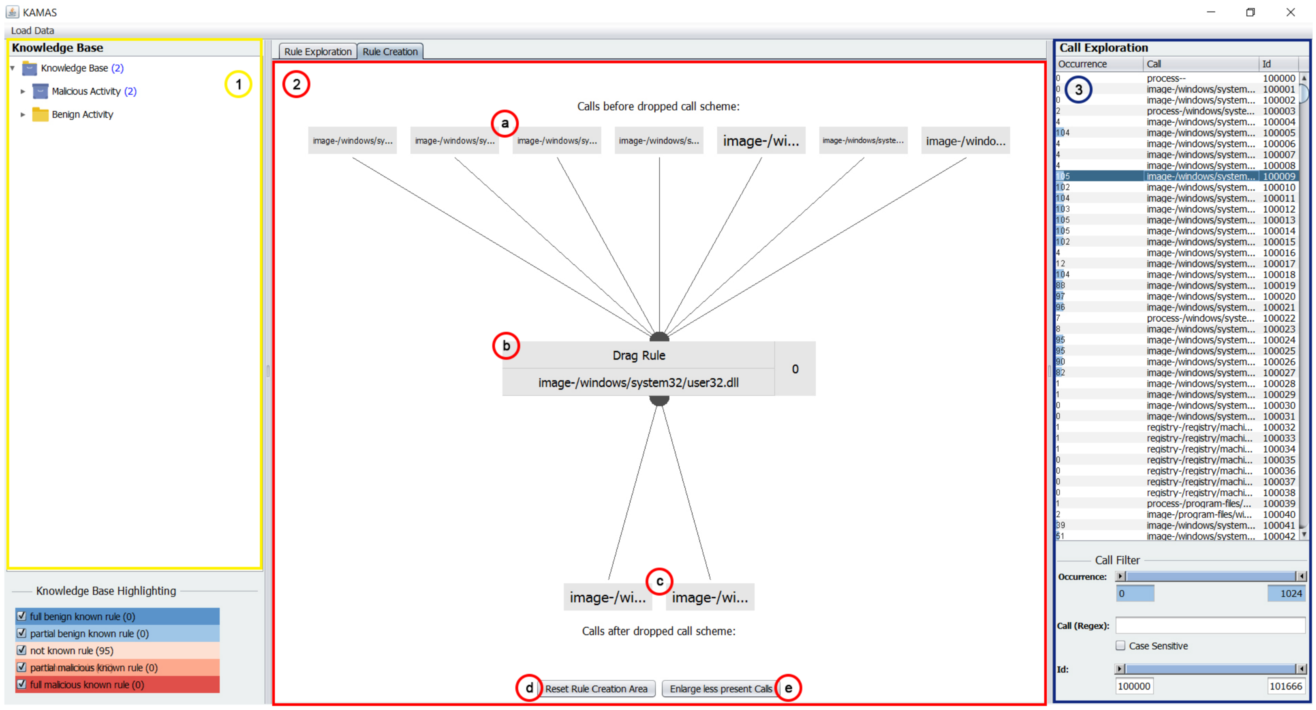


Fig. 1. User interface of the KAMAS prototype with activated RCA. 1) The KDB where newly created rules can be stored. 2) The RCA including the call suggestions before (2.a) and after (2.c) the currently dropped call sequence, the rule creation table (2.b), the button to reset the whole RCA (2.d) and the button to change the call suggestion size according to their occurrence (2.e). 3) The 'Call Exploration' table with a list of all single calls included in the currently loaded file.

Filter Pipelines for Call Suggestions: Depending on the currently dropped system call sequence, the call suggestions get validated through one general and two separated filter action pipelines. The general pipeline loops through all rules of the loaded file and eliminates every rule which does not include the exact same system call sequence as in the RCT. In the next step, the remaining rules serve as input data for the call suggestion validation which is finally displayed above and below the RCT. The first pipeline loops through every remaining rule and checks if there is another single call before the first one of the sequence. If so, this call is displayed as call suggestion above the RCT in the user interface. The same applies to the second pipeline, except that it extracts the single call after the last one of the sequence which is subsequently displayed below the RCT.

Usage Scenario: First, the analyst loads a new file into the system and KAMAS automatically provides an overview of all included single calls in the 'Call Explorer' (see Figure 1.3) as well as all preprocessed rules in the 'Rule Explorer'. The 'Rule Explorer' serves as a graphical summary and exploration area with colored highlighting of all included rules depending on the current knowledge state of the KDB. If the analyst wants to store one of these rules in KDB, this can either be achieved by selecting the full rule or just specific parts of this rule and drag and dropping it to the KDB. Further functionalities linked to the 'Rule Explorer' are described by Wagner et al. [15]. In the initial version of the KAMAS prototype, the rule storing process was limited to use either preprocessed rules or their included single calls. Thus, the analyst was not able to change patterns like e.g. the order of included single calls inside a rule. Following the implementation of the RCA, the analyst can now switch to the 'Rule Creation' screen and create own rules from scratch. In the beginning, the analyst can explore and select specific calls from the single call table and drag them into the RCA. After the desired calls were added to RCT, the interface offers suggestions for calls which occur before and after the currently dropped call sequence. These calls can also be used in the further creation process by dragging them into the RCT. Additionally, it is possible to highlight either more or less frequent call suggestions by increasing their font size. This can be achieved by clicking the 'Enlarge less/more frequent calls' button. The number next to the calls inside the RCT represents the occurrence of the currently dropped rule in the analysis file. During the whole process, the analyst can adjust the created rule by reordering calls inside the RCT via drag and drop or deleting unnecessary calls via right clicking on the desired call and using the 'Delete' pop-up. Finally, the rule can be dragged at the top of the RCT and moved to the KDB. Afterwards the RCA can be reset to its default state by clicking the 'Reset Rule Creation Area' button. The analyst can now return to the 'Rule Exploration' screen and continue the analysis with an updated KDB containing the newly created rule.

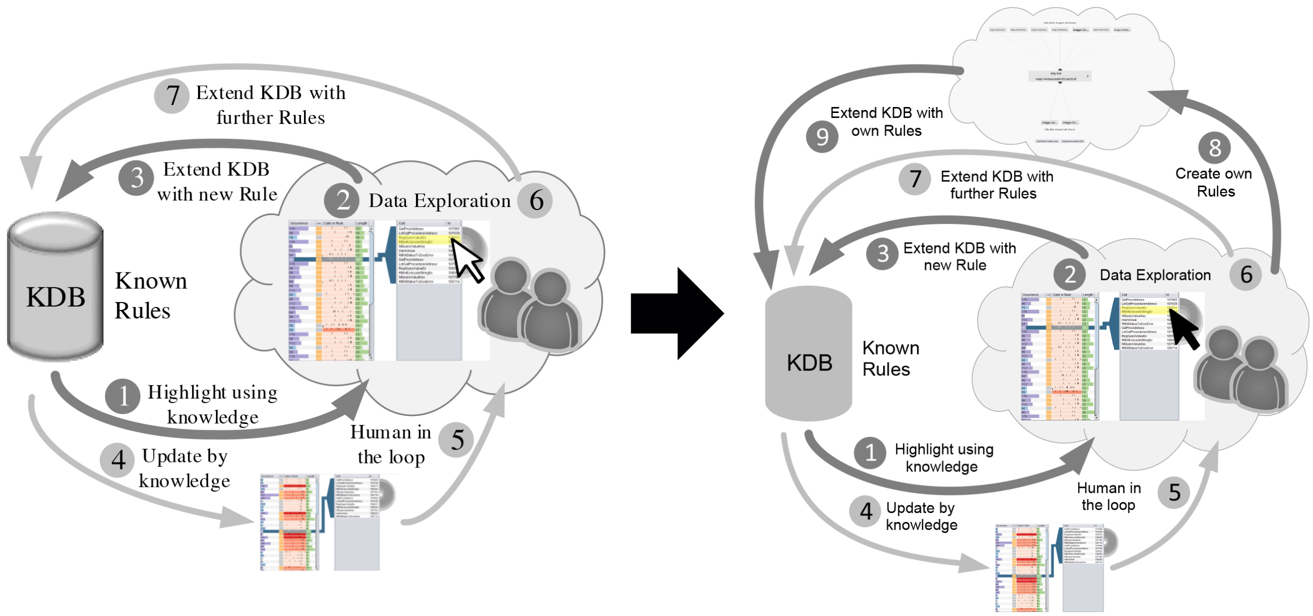


Fig. 2. Comparison of the Knowledge Generation Loop before and after the implementation of the RCA. **Left**: Knowledge Generation Loop of the initial KAMAS prototype visualized by Wagner et al. [15]. **Right**: Knowledge Generation Loop including rule creation process. The range of possibilities for the analyst has increased.

V. The Knowledge Generation Loop

Initial Knowledge Generation Loop: Wagner et al. [15] provided an overview of the knowledge generation process in the KAMAS system (see Figure 2 on the left). It highlights the KDB, which contains all known rules and the data exploration through the analysts as central elements in the knowledge generation loop. By extending the KDB with new rules, the system automatically revalidates the highlighting in the 'Rule Exploration' area depending on the new knowledge state. The rules used for the extension of the KDB are automatically generated and displayed in the 'Rule Exploration' table.

Extension of Knowledge Generation Loop: With the implementation of the RCA, the possibilities for knowledge generation have expanded. The usage scenario now includes the following process (see Figure 2 on the right): The analyst can load a new file, which is going to be checked automatically against the stored data in the KDB (see Figure 2.1). Afterwards, the system provides a visualization of the loaded rules in comparison to its current knowledge state in the system (see Figure 2.2). In contrast to the initial KAMAS prototype, the analyst can now choose between two different options to expand the KDB. The analyst can either use automatically generated rules from the 'Rule Exploration' table (see Figure 2.3 and Figure 2.7), or create own rules, which can include presently unknown sequences based on the system's current state (see Figure 2.8 for the creation and Figure 2.9 for the knowledge generation). As a result, the KAMAS prototype is not limited to rely on automatically generated rules anymore, but rather offers more flexibility by providing an opportunity to create rules from scratch based on individual needs.

VI. Evaluation

As third step in this design study [13], it was necessary to evaluate the newly implemented functionalities with real world users. Therefore, a formative usability evaluation including a usability test with malware analysis experts was conducted and finally compressed to the most important findings.

A. Method

TABLE I
Overview of the domain experts who participated in the user test. (E = expert)

Person	Age	Gender	In field	Experience	Education
E1	30-39	male	5 years	expert	MSc
E2	60-69	male	6 years	advanced	PhD

Participants: For this user test, two malware analysis experts (see Table I) were invited to test and comment all functionalities of the RCA. Both experts had more than five years of experience in this field and were also part of previous KAMAS case studies. Therefore, both of them were familiar with the general appearance and functionality of the user interface.

Design and Procedure: In the beginning, the experts received a brief overview about the main functionalities of the RCA. In order to provide a realistic testing environment, an execution trace sample was provided and loaded into the system. Then, the experts were asked to test each possible feature and to speak out their thoughts on the user interface.

Apparatus and Materials: Both case studies were performed in a seminar room. To guarantee the testing of all possible features, a questionnaire based on the functionalities of the

TABLE II
OVERVIEW OF EVALUATED FEATURE REQUESTS, SEVERITIES AND EFFORT (FR: 1: = NICE TO HAVE, 2: = GOOD FEATURE, 3: = ENHANCES USABILITY; SE: 0: = NO PROBLEM, 1: = COSMETIC PROBLEM, 2: = MINOR PROBLEM, 3: = MAJOR PROBLEM, 4: = CATASTROPHE; EFFORT: 1: = MIN, 2: = AVERAGE, 3: = MAX).

Description	Feature Request (FR)	Severity (SE)	Effort
Call Exploration: Change selection mode to single selection	2	3	1
Call Exploration: Display only the last part of single call names	2	3	1
RCT: Provide a clear visualization of the drop location	3	3	2
RCT: Add arrow symbols to visualize the possibility of reordering	1	1	1
RCT: Display only the last part of single call names	2	3	1
Enlarge-Button: Change ’Enlarge’ to ’Highlight’	1	1	1
Call Suggestions: Display current highlighting status in separate label	1	1	1
Call Suggestions: Change ’before/after’ to ’which appear before/after’	1	1	1
Call Suggestions: Change ’scheme’ to ’sequence’	1	1	1
KDB: Implement possibility to use known rules in the RCA	3	3	3
RCA: Display name of rule if it is already stored in the KDB	2	3	2
RCA: Implement a separate save button	1	2	2
Connection lines: Use logical elements	1	0	2

system was provided. The comments stated by the experts were documented by notes on the printed questionnaire.

B. Results

Moving Single Calls to the RCT: E2 mentioned that it was difficult to find the desired calls, since the names of the displayed single calls were not fully readable due to lack of space in the ’Call Exploration’ table. He suggested to display only the last part of the names because this part mostly differs from other single call names. Both of them showed uncertainty regarding whether if it is possible to move more than one call at the same time or not. Consequently, they were not sure if their desired calls were correctly dropped into the RCT. E2 suggested to change the selection mode of the ’Call Exploration’ table to single selection when the RCA is activated in order to avoid misconceptions.
Moving Call Suggestions to the RCT: While they tried to move the calls to the RCT, both of them were confused by the drop location inside the RCT. They stated that they were not able to recognize a drop scheme, which led to the assumption that the calls get randomly added to the table.
Reordering of Calls Inside the RCT: E1 did not recognize the possibility of reordering in the first place and suggested to add arrow symbols near the table cells of the RCT in order to make it more clear. Furthermore, E2 mentioned that long call names could get cropped off inside the table cells. Thus, the reordering of multiple calls with the same structure could be difficult to recognize.
Change Highlighting of Call Suggestions: Both experts had problems understanding the naming convention of the button. Since the word ’Enlarge’ is present in both versions of this button, E2 expected the call suggestions displayed above and below the RCT to extend to the follow-up call. Furthermore, he suggested to change the description linked to the call suggestions to ’Calls which appear before/after dropped call sequence’ to clarify the meaning. E1 added to show the current highlighting state in a separate label to provide a more clear appearance for the user.
Deletion of Single Calls and Reset of RCA: Both experts were able to delete single calls from the RCT and reset of the whole RCA to its default state without any uncertainty.
Adding Created Rule to KDB: Both experts expressed their wish for a possibility to drag rules from the KDB into the RCA. Additionally, the RCA should display the name in a label when editing an already created rule. In order to provide an alternative for the drag and drop approach, E1 suggested to implement a separate save button for the storing of rules into the KDB.
General Exploration: Both experts were pleased with the general appearance of the user interface. They found the functionalities to be valuable and the interface easy to understand. Furthermore, the simplicity of the user interface was rated positively.

C. Rating

Based on the experts’ comments, the exploration results were combined and rated in a list of the most important issues (see Table II). The rating procedure in this list is inspired by Nielsen’s severity ratings [26]. It includes a description of the issue, feature requests (FR), severities (SE) as well as the associated effort for the solution of the issue. The conducted rating is illustrated in Table II.
Summary: The conducted evaluation showed that the implemented functionalities were well received by the domain experts, although there are still certain improvements to consider for the future. By rating found issues and suggested improvements (see Table II), it was possible to determine major areas for further development of the current prototype.

VII. LIMITATIONS

Following the evaluation by malware analysis experts, certain limitations in the scope of functionalities for this prototype can be determined:
Adjustment of Stored Rules: This prototype does not provide a possibility to drag already stored rules from the KDB into the RCA. However, the workflow of malware analysts also includes the manual adaption of already found rules [4]. With

the implementation of this feature, it would be possible to cover all essential needs of malware analysts and subsequently improve the analysis process even more.

Displaying of Rule Names: The process of rule creation can also lead to a situation, where analysts are constructing rules, which are already stored in the KDB. However, the RCA is not able to recognize already known rules and consequently does not provide the rule name in the user interface. By enabling the RCA to check the currently constructed rule against the KDB and subsequently recognize known rules, the workflow of analysts can be enhanced by e.g. preventing the storage of duplicates in the KDB.

Creating Rules with More than Eight Single Calls: The RCA offers the possibility to create rules with up to eight single calls. Nevertheless, rules can contain much more single calls in reality. Since the RCA also provides call suggestions above and below the RCT, the capacity of space in the RCA is rather limited. To overcome this, for example Focus+Context and/or aggregation techniques could be applied.

Drop Location Visualization: As mentioned in Section VI, the RCA does not provide a visual preview of the currently dragged single call in the RCT. Since both experts were struggling with this issue, the implementation of a visual preview of the dragged single call would have enormous potential for improving the quality of the user interface.

VIII. Reflection and Conclusion

In order to complete the methodology of Sedlmair et al. [13], this section focuses on the reflection of the combined results emerging from the design and implementation of the prototype and its evaluation by real world users. The requirements (R1 - R4) described in Section III were omnipresent during all steps in this design study and serve as point of reference for the following reflection.

R1 Consistency: In order to stick to the defined requirements, drag and drop operations served as the major interaction technique in this prototype. This involves the addition of single calls and call suggestions to the RCT, the reordering of calls inside the RCT as well as the storing of the created rule in the KDB. Both analysts were comfortable with the handling of the given interaction possibilities. However, the evaluation showed that additional visualization is needed to make the outcome of drag and drop operations fully transparent.

R2 Creation Support: As mentioned in the beginning, analysts have to deal with a large amount of data during the exploration process. Therefore, the implemented prototype provides call suggestions to accelerate and simplify the rule creation process. Based on the currently dropped call sequence, the previously described filter pipeline (see Section IV) validates the displayed call suggestions automatically. Additionally, the prototype offers a possibility to highlight more or less frequent call suggestions, which assists analysts in their decision making process.

R3 Editing Options: To ensure editability during the rule creation process, the prototype provides possibilities to delete and reorder single calls in the RCT as well as a button to restart the whole process from scratch. A particularly interesting outcome of the evaluation was that both experts expressed their wish for reusing/adjusting already stored rules in the RCA. This aspect was not taken into account during the development of the current prototype version. After the evaluation, it can be considered as highly recommendable to implement this feature.

R4 Knowledge Extraction and Extension: The possibility to drag the newly created rule and store it in the KDB was also well received by the experts. As mentioned in Section IV, the implementation of this feature expands the knowledge generation loop (see Figure 2). Analysts are now more flexible when it comes to the extension of the KDB. By providing the possibility to create individual rules based on the experts current state of knowledge, the prototype shifts towards the in the beginning mentioned ”human is the loop” philosophy [20]. As a result, the overall knowledge generation process is getting more individualized and the following analysis process can draw upon different expertises.

Lessons Learned: In the course of this design study, it became clear that the continuous integration of domain experts in interaction processes enhances the efficiency of the analysis procedure. As the number of malware families is growing, higher importance has to be attached to the integration of expert knowledge [24]. At the same time, VA techniques have to adapt to the need for more human integration in the analysis process [20]. With the implementation of the RCA based on the interface design prototype by Wagner et al. [16], both previously mentioned challenges were tackled. In cooperation with malware analysis experts, the implemented prototype was proven to enhance the knowledge generation process and to handle the need for increasing focus on human interactions in VA. However, the evaluation also revealed that interaction visualization is a key factor for providing a satisfying solution. Additionally, it showed that there are still possibilities to improve the knowledge generation process. Subsequently, humans could be even more integrated into the previously mentioned knowledge generation loop. In this system, the knowledge is stored based on the same rule structure as they are generated by sequitur [17]. But the storage of knowledge depends on the structure of the underlying data. Thus, also value ranges or process structures can be used.

Future Work: The usage of already stored rules for rule creation can be seen as the next logical step for further development of the presented prototype. Additionally, the enhancement of interaction visualization should round off the overall appearance and usability of the user interface. In general, further exchange with malware analysis experts should be taken into account in order to stay on track with the developments in the scene.

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) via the “KAVA-Time” project (P25489-N23). We would also like to thank all focus group members and test participants who have agreed to volunteer in this project.

References

[1] T. Micro, "Addressing big data security challenges: The right tools for smart protection," *US: Trend Micro*, 2012.

[2] E. Gandotra, D. Bansal, and S. Sofat, "Malware Analysis and Classification: A Survey," *Journal of Information Security*, vol. 05, no. 02, p. 56, 2014.

[3] M. Egele, T. Scholte, E. Kirda, and C. Kruegel, "A Survey on Automated Dynamic Malware-analysis Techniques and Tools," *ACM Computing Surveys*, vol. 44, no. 2, pp. 6:1–6:42, 2008.

[4] M. Wagner, W. Aigner, A. Rind, H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato, "Problem Characterization and Abstraction for Visual Analytics in Behavior-based Malware Pattern Analysis," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*. ACM, 2014, pp. 9–16.

[5] K. Fu, *Syntactic pattern recognition and applications*, ser. Prentice-Hall advanced reference series: Computer science. Prentice-Hall, 1982.

[6] R. Gonzalez and M. Thomason, *Syntactic pattern recognition: an introduction*. Addison-Wesley Publishing Company,Reading, MA, 1978.

[7] H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato, "Malicious Behavior Patterns," in *IEEE International Symposium on Service Oriented System Engineering*, 2014, pp. 384–389.

[8] E. Kandogan and U. Engelke, "Agile Visual Analytics in Data Science Systems," in *IEEE International Conference on High Performance Computing and Communications; IEEE International Conference on Smart City; IEEE International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2016, pp. 1512–1519.

[9] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, *Mastering the Information Age Solving Problems with Visual Analytics*. Eurographics Association, 2010.

[10] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. v. Liere, K. L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver, "Data, Information, and Knowledge in Visualization," *IEEE Computer Graphics and Applications*, vol. 29, no. 1, pp. 12–19, 2009.

[11] X. Wang, D. H. Jeong, W. Dou, S.-W. Lee, W. Ribarsky, and R. Chang, "Defining and applying knowledge conversion processes to a visual analytics system," *Computers & Graphics*, vol. 33, no. 5, pp. 616–623, 2009.

[12] D. Lee, I. S. Song, K. J. Kim, and J. h. Jeong, "A Study on Malicious Codes Pattern Analysis Using Visualization," in *International Conference on Information Science and Applications*, 2011, pp. 1–5.

[13] M. Sedlmair, M. Meyer, and T. Munzner, "Design Study Methodology: Reflections from the Trenches and the Stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

[14] J. Schick, M. Wagner, N. Thür, C. Niederer, G. Rottermanner, P. Tavolato, and W. Aigner, "Supporting knowledge-assisted rule creation in a behavior-based malware analysis prototype," in *Poster of the 14th Workshop on Visualization for Cyber Security (VizSec)*, Phoenix, Arizona, USA, 2017.

[15] M. Wagner, A. Rind, N. Thür, and W. Aigner, "A knowledge-assisted visual malware analysis system: Design, validation, and reflection of kamas," *Computers & Security*, vol. 67, pp. 1–15, 2017.

[16] M. Wagner, A. Rind, G. Rottermanner, C. Niederer, and W. Aigner, "Knowledge-assisted rule building for malware analysis," in *Proceedings of the 10th Forschungsforum der österreichischen Fachhochschulen*, FH des BFI Wien. Vienna, Austria: FH des BFI Wien, 2016.

[17] R. Luh, G. Schramm, M. Wagner, and S. Schrittwieser, "Sequitur-based inference and analysis framework for malicious system behavior," in *Workshop for Formal Methods in Software Engineering (ForSE), 3rd International Conference on Information Systems Security and Privacy (ICISSP)*, SCITEPRESS Digital Library. Porto, Portugal: SCITEPRESS Digital Library, 2017, pp. 632–643.

[18] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Computer Graphics and Applications*, vol. 26, no. 1, pp. 10–13, 2006.

[19] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim, "Knowledge Generation Model for Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1604–1613, 2014.

[20] A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, "The human is the loop: new directions for visual analytics," *Journal of Intelligent Information Systems*, vol. 43, no. 3, pp. 411–435, 2014.

[21] M. Alazab, S. Venkataraman, and P. Watters, "Towards Understanding Malware Behaviour by the Extraction of API Calls," in *2010 Second Cybercrime and Trustworthy Computing Workshop*, 2010, pp. 52–59.

[22] A. Mohaisen, O. Alrawi, and M. Mohaisen, "AMAL: High-fidelity, behavior-based automated malware analysis and classification," *Computers & Security*, vol. 52, pp. 251–266, 2015.

[23] R. Gove, J. Saxe, S. Gold, A. Long, and G. Bergamo, "SEEM: A Scalable Visualization for Comparing Multiple Large Sets of Attributes for Malware Analysis," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 72–79.

[24] M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner, "A survey of visualization systems for malware analysis," in *Eurographics Conference on Visualization (EuroVis) - STARs*, R. Borgo, F. Ganovelli, and I. Viola, Eds. Cagliari (Sardinia / Italy): The Eurographics Association, 2015, pp. 105–125.

[25] H. Sharp, Y. Rogers, and J. Preece, *Interaction Design: Beyond Human Computer Interaction*. John Wiley & Sons, 2007.

[26] J. Nielsen, *Usability Engineering*. Morgan Kaufmann Publishers Inc., 1993.

Session 5: HCI

A Variable Low-cost Platform for Conducting Work Design Experiments

Hendrik Stern
Faculty of Production Engineering
University of Bremen,
&
BIBA - Bremer Institut für Produktion und Logistik GmbH
at the University of Bremen
Hochschulring 20, 28359 Bremen, Germany
Email: ste@biba.uni-bremen.de

Till Becker
Faculty of Production Engineering
University of Bremen,
&
BIBA - Bremer Institut für Produktion und Logistik GmbH
at the University of Bremen
Hochschulring 20, 28359 Bremen, Germany
Email: tbe@biba.uni-bremen.de

***Abstract*—Due to the change of manufacturing work caused by the introduction of cyber-physical production systems (CPPS) further work design research is necessary. We propose to conduct work design experiments in order to design the future work places for the requirements of highly computational and cognitive tasks. Thus, we developed a low-cost experiment platform for an easy setup of experiments. First, we present a brief overview of work design research methods and recent experiments. Second, we present the experiment platform in detail and show how to setup work design experiments on it.**

I. Motivation

Many production resources and processes experience a change towards cyber-physical production systems. This means a combination of physical entities with computational elements in order to make them intelligent [1]. It leads to new products and processes, such as autonomous driving or smart homes, and also to smart production systems [2] [3]. These smart factories will change the way of working as well. There will be new work areas and a new task allocation between humans and machines. Besides, that development goes along with a highly-increased computerisation. Hence, for example, workers will have to deal with smart glasses, wearables, tablets, exo-skeletons, and more. To sum up, the human work will get more cognitive, more digitalised and less physical [4].

In order to facilitate and support this development, we need new work area design principles in order to enable a human-oriented work in the future factories. Therefore, our research deals with the primary research question: How do the work areas for human workers in the future factories have to be designed?

A suitable method to answer this research question is to conduct work design experiments [5]. By experiments, work design researchers are able to test different work design setups regarding their effects on key figures of interest, such as work performance, percentage of errors, work load perception, or motivation. In practice, when preparing experiments, a lot of preconditions and side effects have to be considered. First, implementing a work design experiment into the actual manufacturing process requires us to keep the production process ongoing. This leads to preconditions for the experiment setup which may influence the experiment outcome and subsequently biases the results. Second, running experiments in the actual manufacturing process are cost-intensive due to its disturbing effects on the previous production process and the incalculable outputs.

Therefore, a different way for conducting work design experiments is needed. We propose to use a variable, low-cost experiment platform to easily (pre)-test work design ideas without the mentioned harmful effects on the production key figures. The experiment platform shall be applicable to a wide range of work design starting points. Due to its separation from the production process and its re-usability, researchers are enabled to gain insights on the effects of particular work design ideas in an easier way.

In this paper, we present such an experiment platform from a technical point of view. Alongside, we focus on work design research experiments and how to run them on the experiment platform.

II. Work Design Research

What are the standard, classic approaches to design human work? From a human-oriented point of view work design deals with the creation of jobs, which enable a safe and neither physical nor mental exhaustive way of working [6]. For example, the tasks should be feasible, reasonable or satisfying [7]. Besides, work area design is concerned with the creation of varying and manifold tasks. Therefore, tasks should be as complete as possible. That means that tasks, for example, should have a clear objective, allow an autonomous decision about the tools to be used, and should provide a result review [6] [8].

But what about work area design for the future factories? We consider the standard work area design ideas as still being important but not fully sufficient anymore. Therefore, we suggest to add new work area design ideas to the previous ones. These new criteria focus on the design of the interface between the humans and the machines, such as use of assistance systems, illustration, robustness, or feedback [4]. In order to gain insights on their influence on work performance and perception further work design research is needed.

A. Methods

Work design research is mostly carried out by empirical methods. Two of the main options are observation and questioning. Observation can be distinguished by several criteria [5] [9]:

1) Open or hidden observation: Open refers to a situation where the observed persons are aware of being observed (due to the presence of an observer or a visible camera). Hidden refers to a situation, where the observed persons are not aware of being observed.
2) Participating or non-participating observation: In case of a participating observation the researcher is working with the test persons cooperatively. In case of a non-participating observation, the researcher stays passive.
3) Systematic or non-systematic observation: A systematic observation is performed following a fixed and standardized scheme and stays constant when repeated. A non-systematic observation is explorative and can vary if re-executed.
4) Artificial or natural situation: In case of an artificial situation, the investigated work design setup has been created for research purposes only. In case of a natural situation the investigation takes place on the job directly.
5) Self- or external observation: A self-observation is present, if the test person is observing him- or herself. In an external situation, the researcher observes the test person.

In human factors and work design research the mainly used method is an open, non-participating external observation [5]. For our experiment platform we therefore decided to stay with this proven setup. Further, we chose to perform the observations in a systematic way, which increases the reliability and usability of the results [5]. Finally, the observations shall take place within an artificial situation. As outlined earlier, that way we can separate the experiments from the ongoing manufacturing process.

Besides, we combined the observation part with the other main research method, the questioning. The experiment platform offers the possibility of including one or more questionnaires into the experiments at any time.

B. Experiments

Conducting experiments as a way of applying observations and questioning is a common research method in human factors and ergonomics science. Their topics and research goals cover numerous different aspects. In the following, we provide a brief overview of topics of interest and experiments conducted recently:

Jeske et al. did a study on the influence of different task descriptions on the learning process of workers. They showed a relationship between the design of task descriptions and work performance and mental work reception [10]. Another study dealt with the workers acceptance of head-mounted displays. The authors described a relationship between technology acceptance and wearing comfort or view restrictions [11]. An experimental investigation by Ganßauge was concerned with the light conditions for surveillance tasks. They showed the impact of different light conditions on human vigilance [12]. More studies further investigated topics around the mental stress related to cognitive tasks [13], on trust issues towards autonomous systems [14], or on the examination of mental stress in factories [15].

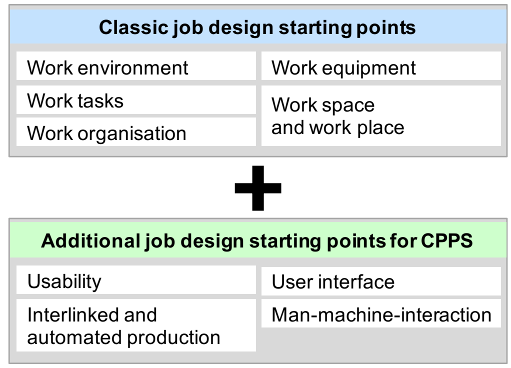


Fig. 1. Work design starting point in CPPS

The topics of these studies show some work design starting points (i.e. aspects of work design, which are necessary for human-oriented work design). As also discussed in [4], most contributions in work design research have been made prior to the rise of modern, cyber-physical production systems. Therefore, they are mainly dealing with partially obsolete understandings of manufacturing work. For example, highly physical-related work design actions such as the consideration of required brawn, which is necessary for executing specific tasks, are mentioned. However, since the majority of physical work tasks will be automated in cyber-physical systems, this topic might not be as important for the major part of future work places as it was before. Thus, additional work design actions, which fit the new situation of cyber-physical production systems, have to be considered. Figure 1 shows a summary on work design actions for future production systems.

III. Experiment platform

A. Technical and functional description

The experiment platform is mainly based on a Raspberry Pi 2 B microcontroller in combination which a 7 inch touch display. The experiment software is a self-developed Python program, running on a regular Linux operating system for Raspberry Pi. The test person is able to communicate with the system via the touch display. Prior to the experiment, the investigator sets up the work task or the work setup to be investigated. During the experiment, the system automatically collects data about the test person's performance (observation) and records answers in the questionnaires (questioning). After

Fig. 2. Experiment platform

the experiment, a results file is provided to the investigator. Figures 2 and 3 show the system and its functional diagram.

The Raspberry Pi (1) is mounted to the touch display (2) and to a display case (3). Further, that component is mounted on a cubical box (4). In order to fit the budget, we used a plastic lottery box as a basis and a case for the experiment platform. It came with a prefabricated horizontal slot (which is meant to be used as an opening for lots or sheets), which we use for cable feedthrough. Inside of the cubical box a battery, an XBee unit,

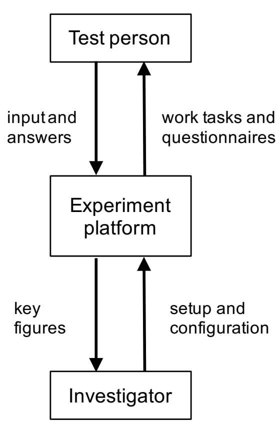


Fig. 3. Functional diagram of the system

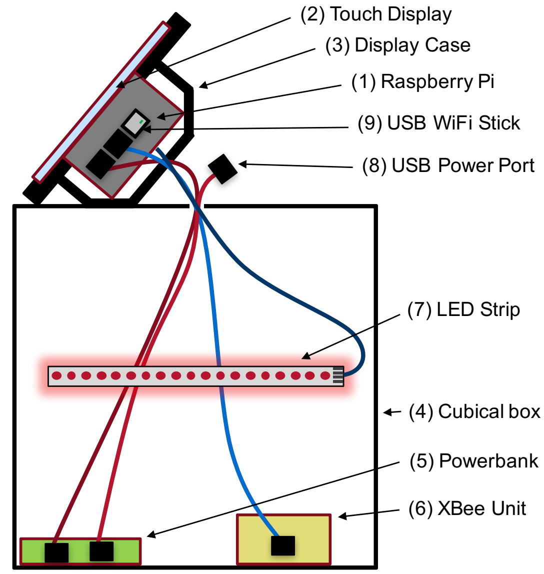


Fig. 4. Cross section of the experiment platform

an LED strip, and wires are stored. We use a standard USB powerbank (5) for power supply of the whole system. The XBee unit (6) is a radio module for a reliable data transfer and meant to be used for an optional communication between two or more experiment systems. Additionally, the LED strip (7) is used for illuminating the experiment system and can be used as a part of the experiments, e.g. as a supporting visible effect. All components can be controlled by the Python program. Besides, we installed an USB power port (8) outside of the box to have an easily reachable charging option. A USB WiFi stick (9) is used for setting up the experiment and for data exchange with the investigator's computer. Figure 4 shows a cross section of the system.

Due to the use of popular electronic components, such as the raspberry pi or a USB powerbank, the total price (234 euros) of the experiment platform is very affordable. It is within the range to standard tablet computers, which could be seen as an alternative solution. But, based on its modular design, the Raspberry pi based platform can be modified and extended

TABLE I
MATERIAL LIST

Component	**approx. Price in €**
Raspberry Pi 2 B	40
7 Inch Touch Display	75
SD Card	12
WiFi USB Stick	8
Display case	22
Cubical box	30
USB Powerbank	20
USB cables	10
Jumper cable	2
LED Strip	15
Total	**234**

more easily. Therefore it increases the fit of the system to the desired experiment. Table 1 gives an overview on the used parts and estimated retail prices.

B. Setup of experiments

One of the main goals while creating the experiment platform was to include the option to easily change the experiment setup. This way, the subject of the experiment can be varied in order to test the effects of these variations. Further, main parts of the software can be re-used for another investigation with a different subject of the experiment.

Therefore, an experiment process follows a sequence of pages, which are shown on the touch display. A sequence consists of questionnaire pages, text and information pages, and task pages. They can be arranged in any order. The test person faces these pages step by step and can move back and forth along these pages (with restrictions). Depending on the page type, the test person receives instructions, is asked to answer questions, or is asked to fulfil a task or solve a problem. These answers, results, and solutions are recorded by the experiment system. Further, several key figures such as the time spent on every page or the number of touches on every page are measured. All data is stored in a results file.

Exemplary, an experimental investigation on the influence of work design elements in cyber-physical production systems shall be presented. The investigation will be conducted using the experiment platform. First, the test persons fill out a general skills questionnaire. This information is used for a general classification of the test persons. Second, test persons get to the task description and the task illustration. Further they are asked to perform the task execution, i.e. to solve a given scheduling problem. Third, the test persons answer a second questionnaire. The questions deal with their reception of the tasks regarding motivation, task complexity, or task difficulty. Fourth, the test persons get to a results page. Here they see a comparison of their solution and the optimal solution. Via the experiment setup, particular work design elements can be switched on or off in order to test their impact. Figure 5 illustrates this process. Finally, figure 6 shows a situation during the experiment execution.

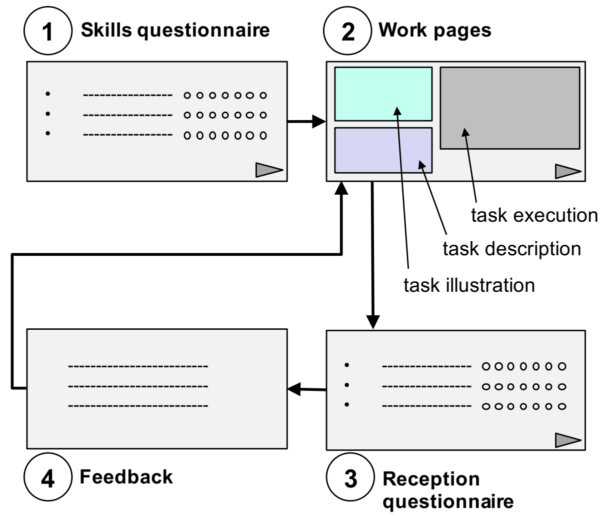


Fig. 5. Exemplary sequence of pages

Fig. 6. Experiment platforms in use

IV. Possibilities and Outlook

Besides the exemplary experiment setup presented earlier, the experiment platform can be used for other experiments or purposes as well. Here, both the work design elements of interest and the tasks to work on can be varied. For example, alternatives to the scheduling tasks could be the creation of batches. Via the touch display the test person could be asked to pool orders or production resources in order to optimize the material flow and subsequently the logistical key figures. Moreover, one or more experiment platforms could be used to model picking tasks. Then, for instance, a touch display represents a shelf compartment. In this case, the test person is asked to mark the requested items of a bill of materials.

Further, the experiment platform is suitable for non-investigative purposes such as training of workers. Here, the several experiment platform can be combined by using the radio module function and arranged as a group work exercise. Thus, skills in collaborative work can be enhanced. Besides, job-related training can be carried out by using the experiment platform. Instead of introducing changes in the manufacturing work on the job, the platform enables a decoupled test environment.

As already outlined earlier, the experiment platform contains a xBee unit to enable communication among two or more platforms. This component has not been integrated into the software yet. We plan to include this function in the next research steps in order to make experiments and training with multiple test persons or platforms available. Additionally, after finalization of experiment platform, we plan to provide the software under an open source license for interested researchers and practitioners.

Acknowledgment

The work of Hendrik Stern and Till Becker has been supported by the Institutional Strategy of the University

References

[1] R. Baheti and H. Gill, "Cyber-physical Systems," *The Impact of Control Technology*, no. 1, pp. 161—-166, 2011.

[2] V.-G. M. und Automatisierungstechnik (GMA), "Cyber-Physical Systems: Chancen und Nutzen aus der Sicht der Automation," 2013.

[3] E. Geisberger and B. Manfed., *agendaCPS*, 2012, vol. 1. [Online]. Available: http://link.springer.com/10.1007/978-3-642-29099-2

[4] H. Stern and T. Becker, "Development of a Model for the Integration of Human Factors in Cyber-physical Production Systems," *Procedia Manufacturing*, vol. 9, pp. 151 – 158, 2017, 7th Conference on Learning Factories, CLF 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2351978917301488

[5] C. M. Schlick, R. Bruder, and H. Luczak, *Arbeitswissenschaft*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. [Online]. Available: http://link.springer.com/10.1007/978-3-540-78333-6

[6] J. Wendsche, "Arbeitsgestaltung," *Lehrbuch Organisationspsychologie / Heinz Schuler, Klaus Moser*, pp. 643–695, 2014.

[7] W. Rohmert, "Aufgaben und Inhalt der Arbeitswissenschaft," *Die berufsbildende Schule*, no. 24, pp. 3–14, 1972.

[8] E. Ulich, *Arbeitspsychologie*, 7th ed. Zurich [u.a.]: vdf, Hochschulverl. an der ETH [u.a.], 2011, xIV, 891 S. ; 240 mm x 170 mm : graph. Darst.

[9] J. Friedrichs, *Methoden empirischer Sozialforschung*, 19th ed., ser. Rororo-Studium ; 28, Sozialwissenschaft. Reinbek bei Hamburg: Rowohlt, 1975, 429 S.

[10] T. Jeske, F. Meyer, and C. M. Schlick, "Einfluss der Gestaltung von Arbeitsplänen auf die Anlernzeit sensumotorischer Tätigkeiten," *Zeitschrift für Arbeitswissenschaft*, vol. 68, no. 1, pp. 1–6, Mar 2014. [Online]. Available: https://doi.org/10.1007/BF03374416

[11] B. M. Grauel, J. N. Terhoeven, S. Wischniewski, and A. Kluge, "Erfassung akzeptanzrelevanter Merkmale von Datenbrillen mittels Repertory Grid Technik," *Zeitschrift für Arbeitswissenschaft*, vol. 68, no. 4, pp. 250–256, Dec 2014. [Online]. Available: https://doi.org/10.1007/BF03373926

[12] R. Ganßauge and A. Hoppe, "Arbeitsplatzgestaltung bei Überwachungstätigkeiten – Grundlagen für zukünftiges Handeln durch Vigilanzforschung," *Zeitschrift für Arbeitswissenschaft*, vol. 70, no. 2, pp. 108–114, Aug 2016. [Online]. Available: https://doi.org/10.1007/s41449-016-0017-8

[13] P. Jeschke, B. Lafrenz, and S. Wischniewski, "Vergleich subjektiver und objektiver Beanspruchungsmessung bei dual-2-back-Aufgaben," *Zeitschrift für Arbeitswissenschaft*, vol. 70, no. 4, pp. 211–219, Dec 2016. [Online]. Available: https://doi.org/10.1007/s41449-016-0033-8

[14] G. Pöhler, T. Heine, and B. Deml, "Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen," *Zeitschrift für Arbeitswissenschaft*, vol. 70, no. 3, pp. 151–160, Nov 2016. [Online]. Available: https://doi.org/10.1007/s41449-016-0024-9

[15] U. Lenhardt, "Psychische Belastung in der betrieblichen Praxis," *Zeitschrift für Arbeitswissenschaft*, vol. 71, no. 1, pp. 6–13, Mar 2017. [Online]. Available: https://doi.org/10.1007/s41449-017-0045-z

Comic Experience: Narrative & Collaborative Drawing on a Multi-Touch Table in an Art Museum

Christina Niederer, Stefanie Größbacher, Wolfgang Aigner, Peter Judmaier and Markus Seidl
Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences, Austria
first.lastname@fhstp.ac.at

***Abstract*—Most art museums provide audio guides or, more recently, multi-media guides, with static context such as background information to enrich their exhibits with an extra layer of content. Usually, there is no actual interaction with the museum's exhibit possible, no hands-on experience that fosters a deeper cognitive engagement. The integration of multi-touch tables has a great potential for collaborative experiences. We designed a touch table application that allows for collaborative and active drawing experiences and conducted two usability studies, one in a laboratory setting and one in the field. The design study was structured in three phases: domain and problem analysis, user experience and interface design, and evaluation. The results show that the collaborative aspect – drawing on one picture simultaneously in different personal areas - was accepted and praised by the visitors. The study indicates that museums with mostly passive viewable artefacts can profit from interacitve and collaborative content, which enhances the general experience in their exhibitions.**

I. Introduction

In art museums the exhibition design is limited, because their focus is on displaying collections of objects such as paintings, sculptures, multimedia works, and installations. Usually, there are hardly any opportunities for visitors to interact with artefacts or other visitors, other than discussing exhibited objects. Most art museums try to increase their visitors' interactivity by handing out handheld devices providing static content like audio tours [1], [2]. Multi-touch technology, in combination with appropriate interaction design concepts, allows for true interactivity between visitors and the exhibition objects. The presented research examines the emerging role of interactivity with exhibition objects by developing a collaborative drawing and viewing application running on a multi-touch table and a web application for smartphones. The collaborative drawing and viewing application adds interactive elements in accordance to the visitors' desire for self-expression. The interactive comic experience specifically developed for the Karikaturmuseum Krems makes drawing for visitors easier, actively engaging them with drawing styles of exhibited artists, and allows collaboration with other visitors, even outside of the context of the museum.

Based on a user-centered design approach we conducted a design study to investigate whether using the collaborative drawing application introduces novel user behaviors or social interactions. Moreover, we studied how digital brushes have to be designed and implemented for strokes performed by human fingers in order to work well on the touch surface, independently of the target group and its drawing skills. To answer these questions, we applied an ensemble of research methods: First, we identified user's needs and created personas and scenarios. The needs were then taken into account while developing paper prototypes and the interactive application. Furthermore, two usability studies (one in a lab environment, one in the field) were conducted to evaluate the multi-touch application in general, and the user interface in particular.

In the next section we discuss *Related Work* dealing with multi-touch and multi-user approaches in museums, participatory projects and drawing applications. The section *Design Study* describes the research methods used during the development of the application and the application's features. In the section *Evaluation* we give details of the procedure, participants and test results of the conducted studies. In the section *Discussion* we summarize and interpret our findings of the two usability studies. Finally, we discuss possible directions for future research in the section *Conclusion*.

II. Related Work

Large-scale table-top devices have already demonstrated their great potential in public use of interactivity and collaboration in the past. In 2002, the project *SmartSkin* [3] investigated a new sensor architecture for making interactive surfaces sensitive to human hand and finger gestures. Besides technical achievements, the study of Rekimoto [3] also reported new insights into interaction techniques using multiple fingers. One year later, a study with *Diamond Touch* was conducted by Dietz and Leigh [4]. They proposed a touch-sensitive input device which allows multiple, simultaneous users to interact in an intuitive fashion. Nowadays, multi-touch table-top devices can be found in various locations such as airports, information centers, retail stores, and museums [5], [6]. To provide an overview of work related to our problem domain, we focused on multi-touch and multi-user table-top applications, participatory projects, as well as drawing applications.

A. Multi-user Table-Top Applications

The Museum of Science and Technology in Islam [7] demonstrates 1500 years of history of Muslims on a large multi-touch table. Visitors can simultaneously interact with the application and create a social learning experience. Furthermore, Horn et al. [8] conducted a survey at the *Harvard Museum of Natural History*, showing that visitors collaborate effectively and engage in on-topic discussions of the exhibition. They presented a design and evaluation of a tabletop multi-user game to help visitors learn more about evolution. The multi-touch table installation of Hornecker [9] in the *Berlin Museum of Natural History* demonstrates that information-browsing applications may be inappropriate for a museum's context, as it was not used much and hardly provided discussion topics. The potential of interactive tabletops was not exploited satisfactorily.

Multi-user scenarios can also be found in other areas besides a museum's context. Blumenstein et al. [10] have described *inter alia* general requirements and challenges for multi-user and multi-device scenarios from the perspective of interactive data visualization.

B. Participatory Projects

Ideum [11] developed a photo kiosk for the *Crystal Bridges Museum of American Art* in conjunction with the exhibition *Warhol's Nature and Jamie Wyeth*. The participatory aspect of the project was that visitors were able to capture their own photograph and then choose different style elements to apply to their photograph based on the works of the two American artists. After styling their photograph, users could send it via email to either themselves or others.

Moreover, the *Indianapolis Museum of Art (IMA)* developed a number of participatory projects [12]–[14] that allow visitors to contribute to the museum experience by creating their own content and sharing it with the public. In 2013 the IMA launched a drawing competition with the *Matisse, Life in Color* exhibition encouraging visitors to create drawings inspired by the works of the French artist [12]. This concept based on an app available on a number of iPads was installed in the exhibition entrance. The created drawings could then be submitted via the app to a provided competition website, where people could view submissions and rate and comment the drawings. IMA stated that this participatory project worked well, because visitors could see themselves and/or their works represented within the network.

C. Drawing Applications

There are a number of drawing applications on the World Wide Web where users can draw on their own device and then share it with others or draw collaboratively over the web. *Awwapp* [15] and *sketchpad* [16] are two well known examples. *Awwapp* offers collaborative drawing by connecting through the Internet. The available functions are very basic but effective. Basic functions that most of the applications include are a pencil with different sizes and color, an eraser, texts, and sometimes images that can be placed. Deleting the whole image, as well as saving and sharing it, are additional functions. Drawing applications on multi-touch tables can be found in domains like design, in the form of a brainstorming tool [17] or in educational organizations [18], [19]. Partarakis et al. [20] presented a painting game for pupils, introducing physical objects to a large touchscreen. The aim of this installation was to teach drawing techniques to pre-schoolers.

Beside the work of Partarakis et al. [20], the usage of drawing applications on multi-touch tables in a museum's context has not been investigated. Apart from these studies, there is a lack of research on the integration of drawing applications on multi-touch tables in a museum's context, especially for art museums with the focus on collaboration and participation.

III. Design Study

Our design study is divided into three parts: gathering information to deduce requirements, conceptual design for an easy-to-use interactive comic experience, and evaluating the application to identify problems.

A. Requirements Research

The first step in the process of defining requirements was to collect qualitative data about the potential users of the museum. In this research phase real world observations and interviews were conducted. After collection, the information was modeled in form of personas. In the final stage scenarios were developed to define the requirements.

1) Observations & Interviews: Firstly, we physically visited the museum to gain insights by interviewing the museum's employees and conducting observations such as: what does the exhibition area look like, what is the average exhibition period or who are the visitors. Four employees of different functions were interviewed: the director of the museum, a cashier, and two museum warders. The conversations took roughly 15 minutes. The museum provided all their data they had already collected about their visitors over the years. During the observations we also analyzed published advertising materials, the gift shop and the guestbook. The document analysis showed that the guestbook is full of sketches and little cartoons, showing the visitors' desire to express themselves not only graphically, but also by relating their drawings to the context of the museum's exhibition.

2) Personas & Scenarios: Based on the interviews and observations three personas [21] were created: an older married couple, a class of high school juniors and a young guy in his twenties. These personas became the main characters of the developed scenarios. The scenarios describe their visits to the museum, why they go there in the first place, how they act in the museum and how they react to and interact with the table. Storyboards have been created to illustrate the scenarios (Figure 1). The output of this process was the requirement definition.

3) Results for Requirements: The target group of the museum includes nearly every age group (young children as well as retirees), and different social groups (tourists, students, regulars). Regarding the touch table the target group is reduced to people interested in technology. The list below presents user requirements for the target domain:

- **Expressing themselves graphically:** The paper guestbook shows that visitors express themselves by drawing funny sketches based on the exhibition topic.
- **Collaborative work:** Sketches in the paper guestbook are often drawn by more than one person.

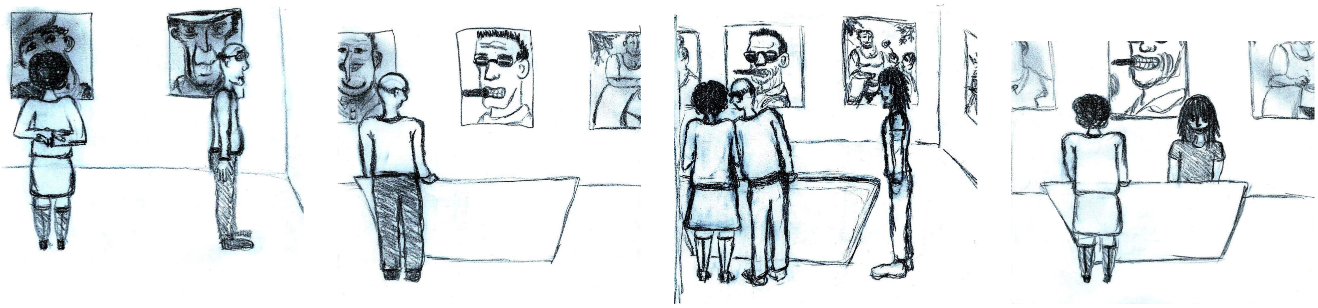


Fig. 1: Storyboard image from one of the scenarios in the museum exhibition: An elderly married couple visits the Karikaturmuseum Krems and comes across a touch table. The husband is interested in new techniques and shows his wife how to read comics on the table. Tom joins them at the table and starts the drawing application. The wife is all eyes and copies his interactions. They draw a panel together.

- **Self-representation:** The drawings in the paper guestbook, which any visitor of the museum can flip through, are nearly always signed.

B. Design

Paper prototypes showing the concept of the drawing application and the first design of the look-and-feel of the application were developed. In the next section we describe the final interactive prototype running on a multi-touch table.

The core concept of the drawing experience is the narrative aspect to it. In case of the Karikaturmuseum museum, the visitors get the chance to become storytellers by drawing panels for a collaboratively created comic. The collaboration is not limited to the museum context because user-generated drawings are exhibited both on the interactive table and the web application, once the comic is finished. Visitors who do not wish to draw in public and have their work presented publicly can interact with the table by flipping through completed stories. This way, the multi-touch table caters not only to the needs of visitors who wish to be actively involved in the exhibition, but also to those who prefer to passively take in the art presented.

Start Screen: As proposed and evaluated by Klinkhammer et al. [22], we divided the whole screen into four personal working areas seen in Figure 2, where the user can interact with the tool. The main screen contains elements for drawing comics (pencils) and one element showing already finished artwork (book).

These two features are included to cover the needs of different forms of participation in museums identified by Simon [23]. On the one hand, the visitors, who are "creators", can produce content by drawing panels for a comic and, on the other hand, the so called "spectators" read and discover finished comics.

To start the application, the user has to drag one element into their personal working area. The selected element then pops up in the chosen area and the user can then either start drawing a panel for one of the provided stories, or look at completed comics (Figure 2). Figure 3 shows the provided

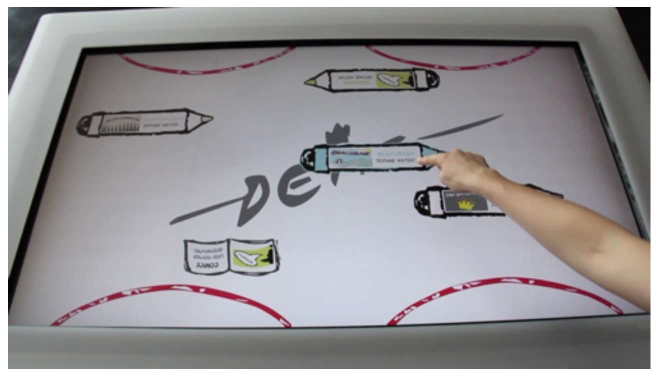


Fig. 2: Interactive element "start screen": drag and drop a pencil in one of the four semicircles to start drawing.

Fig. 3: Story options to choose when drawing a comic: starting illustrations are caricatures by Austrian artist Manfred Deix.

stories, which are based on famous drawings by Deix, a well-known Austrian cartoonist. Furthermore, the user chooses a story they want to contribute to.

Collaboration Concept: After choosing a story the users have to decide if they want to draw on their own or collaboratively in a group (Figure 4). To work collaboratively, the system provides the possibility to draw individually on two

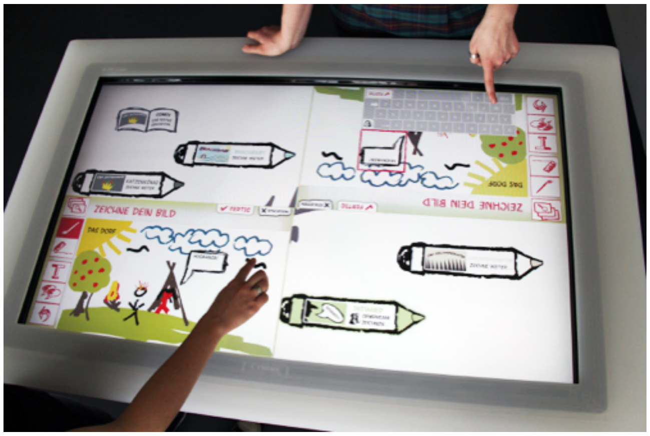


Fig. 4: Divided Screen – Personal Working Areas: The users are able to work on four areas on their own or collaboratively. This means that users work on their personal working area, but when cooperating with another visitor they draw in their respective drawing areas simultaneously.

different working areas. Each user sees what the other draws in their own working area. To do so, the user needs to share their story before starting to draw. The chosen story then reappears on the main screen, giving other visitors the chance to join this work. If users choose to work on a comic, they get to see the last three pictures that have been drawn for the selected story in the form of a carousel. This way, users get a glimpse of how the story developed so far, without telling everything that has been going on up to this point. By not knowing the whole story the comics should turn out more compelling. This concept was used to encourage the creativity of every user and to get interesting stories.

Sketching: The sketching part of the system provides a drawing application with various tools: brushes, balloons, text areas, an eraser and the functionality of undo (Figure 5).

The basis for the design of the brush implemented on the multi-touch table was the analogue drawing behaviour with a pen. To make it easy to use for the broad visitor audience we integrated one type of brush. The line style of the brush is comparable to a felt pen. To vary the type of brush, a thickness slider with a preview area and a colour palette was implemented. The colours depend on the story the user has chosen.

The interaction concept for adding text elements and bubbles is based on known concepts of graphic applications such as Adobe InDesign or Photoshop. The text box appears on the surface and users are able to drag and drop the box into the place of the picture where they want it to be.

Related to the text input methods, we decided to integrate a soft keyboard based on the QWERTY approach [24]. The physical keyboard elements are mapped to the on-screen keyboard. The touch elements have a squared shape and the size of the touch elements was adapted to a finger-friendly size.

Fig. 5: Drawing Interface

Following the Story: After finishing a drawing, users can sign the comic panel by filling out a form with a name, residence and an email address. Then users can see their picture lined up with the previous panels. This allows the visitors to see how the drawing just finished integrates into the whole comic strip.

By scanning the provided QR-Code on the multi-touch table, visitors can take the story home with their personal smartphone seen on Figure 6. The QR-Code leads to a mobile web application, which links to the comic the users took part in the museum. So, visitors stay in contact with the exhibition and the collaborative aspect does not end when leaving the museum. The integrated QR-Code does not provide extra information about the exhibited arts in the museum [25], but rather complements the mobile website.

Fig. 6: Interface Design for scanning the QR-Code

Reading Finished Comics: The application on the touch table also provides the possibility to look through finished comic strips from other visitors. Thus, the visitors can get an idea of the stories and inspiration for their own sketching work. The interface is arranged similar to the drawing area seen in Figure 7. On the left, there is a tool bar showing the different stories. In the main area, different versions of one selected story are listed.

Fig. 7: Reading finished comics in the exhibition. On the left, the sidebar shows the main stories. The grey area on the right contains all versions of a finished story in the museum.

C. Prototype

The prototype was developed for a 40 inch framed high definition (1920 x 1080 px) table-top, including infrared tracking to discern the touch points. Up to four museum visitors may use the application simultaneously. The application was developed for a multi-touch and multi-user approach and combines a touch table with mobile devices (Figure 8). The system is an interactive installation where visitors can do creative, graphical and collaborative comic storytelling. On the one hand, the users may sketch a drawing and become part of a bigger story and, on the other hand, they can look through already completed artworks by other users.

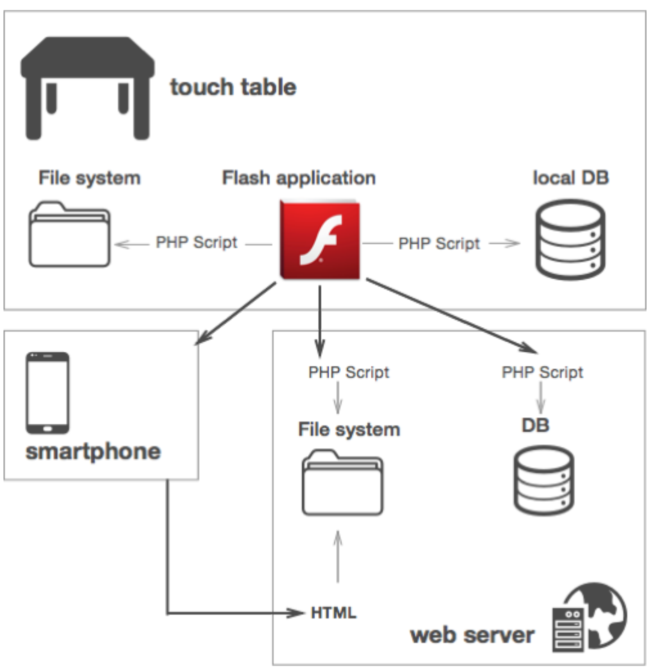


Fig. 8: Component diagram: The architecture of the interactive installation. It contains a touch table, smartphones and a web server.

The research about different technologies showed that Flash (Actionscript 3.0) together with the framework Open Exhibits (http://openexhibits.org/) for recognizing gestures is the system most suitable for us on the touch table. At the time of prototype implementation, Flash had a large community and is well-documented. Furthermore, we chose this platform for our comic experience application because of the experiences with Flash on multi-touch tables in previous projects concerning the stability and easy installation on Windows PCs.

Our application supports up to four simultaneous users and the process of drawing needs sensitive reactions by the system. The gesture framework Open Exhibits provides the advantages of predefined touch-gestures and the support of simultaneous touch events, which are needed to develop collaborative applications. The first step in the implementation phase was to build a clear object- and action structure, defining which data-objects should be used and which actions would be performed on those objects. A database contains all the data objects and their relations. This database is also used for our website, where the users may open their drawn images from home or on their smartphones.

IV. Evaluation

Two user tests were performed: one was conducted in a laboratory setting (at an early conceptual stage) and one in the field (with a completed first release prototype). The study design and results are presented in the next sections.

A. User Study in Laboratory Setting

1) Prototype: The interactive prototype running on the touch table at this stage of the design study already included these functions: opening the drawing application via drag and drop interaction; selecting the brush and setting its width; drawing on the comic panel; erasing the lines; adding text and bubble elements and typing text into it. Based on a user-based usability test we evaluated the prototype to investigate (1) how effective the drawing application is and (2) how satisfying it is for the users to draw with their fingers on a multi-touch table. (3) Furthermore, the text-input method, in our case a soft keyboard (Figure 8), was part of the analysis. The aim was to find out how easy it is for visitors to type on a touch-based keyboard.

2) Procedure and Participants: 13 high school pupils (11 female and 2 male) at the age of 14 to 15 years participated in the user test. At this point of development the main functions of the application were fully developed and implemented on the touch table. The test equipment consisted of two 40 inch framed table-tops including infrared to discern the touch points and two DSLR cameras recording the interactions with the system and the users' feedback. During the observations handwritten notes were taken. The participants were supposed to complete a set of seven predefined tasks and were divided into two groups: one group consisted of single students working through the tasks and the second group were four students working together on four individual areas. Both groups faced the same tasks to complete and did so simultaneously in two separate rooms.

3) Study Design: The tasks the students had to complete included: (1) describing what they see, (2) drawing a cat, (3) letting the cat talk, and (4) changing what the cat is saying. Due to the qualitative character of the study, the subjects were asked to fill out a questionnaire on how well they were able to handle the application and how much they enjoyed doing so at the end of the usability test.

Furthermore, a focus group discussion with all 13 students regarding questions such as: Did they like the application? Would they improve certain functions? Do they have general recommendations? was initiated to get a broad range of viewpoints and insights. During the test, the thinking aloud approach was followed [26].

4) Test Results: The results suggest that drawing with fingers on a multi-touch table is very effective and easy to use, though some of the students struggled to draw as accurate as they wanted to.

The suggestions of the questionnaire indicated that the drawing part is very satisfying for the participants. All students ranked the application between 1 and 3 (grades 1 to 5, 1 indicating the highest satisfactory level). They commented that they would try drawing on the table in the museum, as well. Some students also revealed that they like to be creative. Regarding the brush design and variety of colours provided, they expressed the wish for a thinner brush and more colour combinations.

The text-input via the keyboard (Figure 9) revealed some room for improvement: during the test it could be observed that participants had problems with typing on the keyboard. The touch areas were too small, causing the keyboard to close itself when they hit the drawing area instead. The subjects also called for a cursor.

Fig. 9: The improved keyboard design as a result of the laboratory test.

B. Field Study

The second user test was conducted on-site in the museum, testing the application in real world circumstances in the field. At this point, the development of the prototype was basically finished. Based on the previous prototype for the laboratory setting this prototype was improved and extended. The improvements included: one brush with more colours to choose from (8 main colours); collaborative drawing functionality; a keyboard adapted by resizing the keys. The application was extended by the functionality of reading comics.

This field study examined (1) how satisfying it is for the user to draw, (2) how effective the concept of collaborative drawing on a multi-touch table is, and (3) whether stories were being developed.

To record user behaviour and interactions remotely, we installed a webcam beside the touch table and used the Software *iSpy* [27] to adapt the recording time of the camera to the opening hours of the museum. We also implemented instrumentation functionality to log usage data while using the application (such as which tools were used, which stories were chosen, or how long drawing sessions took).

1) Procedure and Participants: The table was set up in one of the rooms of the museum (called Deix room) over a period of one week. Posters on the sides of the table explained that it was a university project, that visitors were invited to draw comics and that users would be filmed when using the table. These were the only explanations museum visitors received regarding the use of the table. A camera filmed the interactions of the visitors when using the table. When closing the application, a pop-up with a voluntary questionnaire appeared. In the background, we logged data to get more insights about the interaction behaviours of the visitors.

The test participants were a random group of visitors, regardless of age and media literacy, who attended the exhibition at the Karikaturmuseum museum in the time period of one week. The exact number of participants is unknown, as the camera that recorded the users was positioned in a way that guaranteed their anonymity. Overall, 185 sessions were captured.

2) Study Design: The camera was arranged at the side of the table, filming from high above. So, the whole table was in focus while the angle provided the anonymity of the visitors.

The questionnaire was structured in two parts: After giving their approximate age, gender and their reason for visiting the museum (or deciding not to answer), they could choose between different smileys (laughing, neutral, sad) to state whether or not they found navigating through the application easy, they instantly knew how to use the drawing tools, and if they liked drawing on the table.

User behaviour was also logged in the background. The following research questions were the basis for the logging functionality: (1) Which of the four comics based on a Deix drawing ("king of the cats", "women on the beach", "playing indians", "hedgehog") was chosen most often? (2) Did users close the drawing app before finishing their picture? (3) How many pictures were finished in total? (4) Which comics did they like to read? (5) During which times of the day was the table used? (6) On which day of the week was it used most often? (7) Do users prefer working alone over working in groups? These are some of the questions the log contributed answers to.

3) Test Results: Collaboration and Stories - Findings showed that users are more likely to work on a comic alone instead of in a group. Their favourite story was a story about the "king of cats", but when drawing alone the story about a "woman on the beach" was chosen most often. Concerning working in groups, we could observe that people help each other and work together rather than destroying the work of the other drawers. Even when people work separately on their own sketches, they stop to help users with problems in using the application.

Interestingly, the developed panels/drawings exhibited less elements of a comic, such as text boxes or speech bubbles. A few participants used a thin brush to write texts to complete their panel instead of using the text tool (Figure 10). Overall, within 185 sessions, visitors used text elements only every fourth session and every second session they placed speech bubbles on their drawings.

Reading Finished Comics: In the days of the field study, no comic was finished completely. Since only completed stories can be read on the table, the visitors were not able to flip through stories. Overall, the video recording showed that the interface design and interaction was clear and easy-to-use, though.

Participants: In the testing period visitors between the ages of 11 to 25 and 36 to 50 years attended the exhibition. The application was used by more women than men.

Questionnaire: The overall response to the drawing application was very positive. 48 of 60 visitors rated the drawing application as positive. Respondents were asked to indicate whether the tools (brush, text and bubble, undo) are immediately clear. 41 of 60 participants categorized the tools as very easy to understand and easy-to-use.

V. Discussion

The findings of the two usability and user-experience studies can give some implications and experiences for the design of collaborative drawing applications with the focus on storytelling on multi-touch tables in art museums. The results seem to be consistent with other research as shown in the following sections.

Concept & Interaction Design: The interface of the drawing application was designed in analogy to well-known applications such as *Adobe Photoshop* or *Illustrator*. The tool palette is positioned on the left side of the interface, with the drawing area next to it. By tapping a tool, a menu opens and provides the different choices the tool offers. By using drag and drop or tapping on the selection of the text elements, for example, they appear on the drawing area. The video observations of the field study and the personal observation in the laboratory show that only few users prefer the possibility to drag and drop elements over tapping on an element and have it appear on the drawing area. The comic reading section is structured the same way; on the left, one can choose between the four different stories and next to the sidebar the presentation area is positioned. Based on the statements of the conducted questionnaires, and interaction behaviors of visitors as seen on the videos of both studies, we can state that this structure of the interface works well for a wide range of visitors.

The user studies showed that one brush is efficient enough as long as the thickness can be adjusted properly enough.

Regarding the text-input method we can only interpret the results of the study. 185 sessions were detected during the field study in the museum, but only 57 of the drawings were signed by the visitors and only few comic panels exhibit text elements or speech bubbles. This may result from the input method for text, using a keyboard known from the smartphone applications. So, we can confirm the results of Wigdor et al. [28] stating that text input on large multi-touch tables can be problematic and that more research besides Hinrich et al.'s [24] study to investigate new methods for textual input ought to be done.

The concept of collaborative storytelling in this specific context of the museum Karikaturmuseum Krems works well. We found some storytelling aspects in one comic, but there were no completed comics. As there are four different stories to choose from, it takes a while until one comic is completed. The number of panels resulting in a comic was defined as too high. We recommended a number of panels for a story of approximately 10 panels. A way to get visitors to complete comics quicker, and thus be able to offer the application's full functionality, could be to start off with only one story and have visitors unlock the other three stories by completing one story after the other, until all four stories are available. This would also force users to work collaboratively, which could then lead to have museum visitors interact more easily. In that way, we can avoid the problem that there are no comics to read on the multitouch table, as well. It is important to show at least a message saying that this area is empty until the predefined number of panels for one story has been finished.

Collaboration: Our findings show that users are more likely to work on a comic alone instead of in a group. So, we can confirm the results of Block et al. [29] and express the need to provide a meaningful single-user experience. But surprisingly, this phenomenon can be noticed differently depending on the topic of the story. Some topics are more likely to be drawn collaboratively than others. To ascertain why there is a noticeable imbalance in the stories with regards to visitors working alone or collaboratively, more research would be necessary.

Besides the known behavior of social learning and peripheral interest identified by Hinrichs and Carpendale [30], we could determine the fact that users help each other by giving hints or performing the interaction for their partners on their personal area.

Research Methods: The conducted research methods, the user study in the laboratory setting on the one hand, the user study in the field on the other hand, can be categorized as very useful. The early on user-centered research in the laboratory gave important insights on the problems of the interface and user behaviors while interacting with the drawing application on the multi-touch table. Here, we were able to discern that the concept of collaborative storytelling while drawing is effective and satisfying for users even early in the design process.

The second usability study in the field with video record-

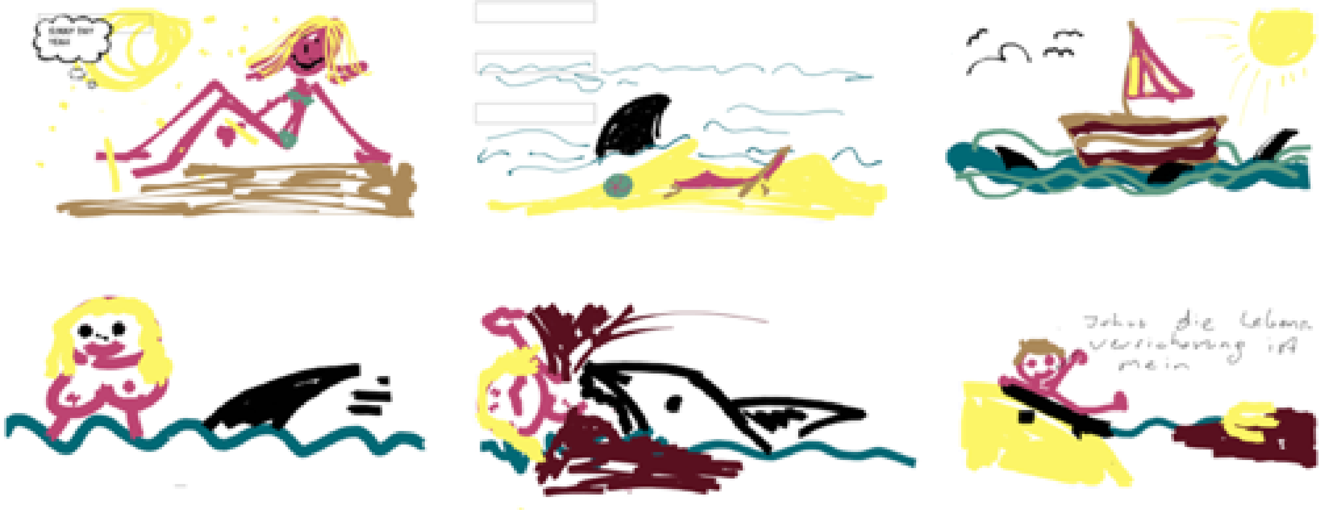


Fig. 10: Example narrative created during the field study.

ings, logging and a questionnaire provided different insights into the usage of the drawing application. The video recordings gave us the chance to identify the overall usage of the drawing application and insights into social interactions in groups or alone. To get more information of the visitors and their interactions the logging of the data gave us knowledge about the exact number of sessions, used stories and tools as well as the age of the visitors. So, the combination of video recordings and the data logging can be recommended as it is very useful. In upcoming studies we would integrate a personal questionnaire again, asking the visitors about their interaction with the interface to get more insights into the needs and wishes of the target group.

VI. Conclusion & Future Work

The presented study was designed to determine the effects of the integration of an application on a multi-touch table in the context of the Karikaturmuseum Krems. Thus, a drawing application based on a collaborative concept was developed and tested in the field and in a laboratory environment. The results show that there is large potential for introducing such kind of digital technology in a museum's context. The development of systems such as mentioned for collaborative drawing applications with a storytelling aspect to them for art museums introduces some challenges, such as:

Interplay Between Table-top and Smart Device: Concepts for multi-display scenarios that incorporate both large displays and small personal mobile devices have to be explored in depth in further studies. We approached this subject by giving visitors the chance to take elements of the museum's exhibition home, thus keeping them connected to the development of the stories, as well as the museum in general. Calling up a website on their personal smartphone is a step toward the multi-display trend defined in 2010 by Isenberg et al. [31].

Collaboration: Museum studies have found that people often visit exhibitions in groups [9], [32]. Yet, many museums offer elements where visitors work on individual tasks sequentially or parallel, but never collaboratively [31]. So, systems should provide aspects of collaborative work not only in the form of integrating large tabletops but also interaction concepts and game concepts for working in groups on one task. With our tool, we can introduce a storytelling approach for a drawing application with the focus on collaborative drawing, by allowing museum visitors to draw on their personal working area, but simultaneously draw in collaboration with other users.

In future research we plan to focus on identifying which aspects of the application work well in any museum, which are specific to a certain type of museum, and which only cover the particular needs of the Karikaturmuseum Krems.

Acknowledgments

We want to thank Gottfried Gusenbauer, the director of the Karikaturmuseum Krems, who contributed to the project in many fruitful discussions. This work was supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) under the ICT of the future program via the VALiD project (no. 845598) and by the Austrian Federal Ministry of Science, Research and Economy under the FFG COIN program (MEETeUX project, no. 7209770).

References

[1] A. M. GmbH. HAUS des MEERES - aqua terra zoo - audioguide. 2017. Retrieved August 07, 2017 from https://www.haus-des-meeres.at/de/Besucherinfo/Audioguide.htm.

[2] Kunsthistorisches museum wien. 2017. Retrieved August 07, 2017 from https://www.khm.at/erfahren/kunstvermittlung/audioguide/.

[3] J. Rekimoto, "Smartskin: An infrastructure for freehand manipulation on interactive surfaces," in *Proc. of the CHI '02*. ACM, 2002, pp. 113–120. [Online]. Available: http://doi.acm.org/10.1145/503376.503397

[4] P. Dietz and D. Leigh, "DiamondTouch: A multi-user touch technology," in *Proc. of the UIST '01*. ACM, 2001, pp. 219–226. [Online]. Available: http://doi.acm.org/10.1145/502348.502389

[5] M. I. Berkman and A. Karahoca, "A direct touch table-top display as a multi-user information kiosk: Comparing the usability of a single display groupware either by a single user or people cooperating as a group," *Interact. Comput.*, vol. 24, no. 5, pp. 423–437, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.intcom.2012.07.002

[6] A. N. Antle, A. Bevans, J. Tanenbaum, K. Seaborn, and S. Wang, "Futura: Design for collaborative learning and game play on a multi-touch digital tabletop," in *Proc. of the TEI '11*. ACM, 2011, pp. 93–100. [Online]. Available: http://doi.acm.org/10.1145/1935701.1935721

[7] Interactive timeline table mosti - Museum of science and technology in islam. 2014. Retrieved April 05, 2016 from XXXX.

[8] M. Horn, Z. Atrash Leong, F. Block, J. Diamond, E. M. Evans, B. Phillips, and C. Shen, "Of BATs and APEs: An interactive tabletop game for natural history museums," in *Proc. of the CHI '12*. ACM, 2012, pp. 2059–2068. [Online]. Available: http://doi.acm.org/10.1145/2207676.2208355

[9] E. Hornecker, " I don't understand it either, but is is cool - visitor interaction with a multi-touch table in a museum," in *Proc. of the TABLETOP '08*. IEEE, 2008, pp. 113–120. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4660193

[10] K. Blumenstein, M. Wagner, and W. Aigner, "Cross-Platform InfoVis Frameworks for Multiple Users, Screens and Devices: Requirements and Challenges," in *DEXiS WS on Data Expl. for Interactive Surfaces. In conj. wirth ACM ITS*, 2015.

[11] Warhol / wyeth interactive photo kiosk. 2015. Retrieved April 05, 2016 from XXX.

[12] F. Fantoni, K. Jaebker, and T. Leason, "Participatory experiences in art museums: Lessons from two years of practice," in *Selected Papers from Two International Conferences*. Museums and the Web LLC, 2014, pp. 191–200. [Online]. Available: http://mw2014.museumsandtheweb.com/paper/participatory-experiences-in-art-museums-lessons-from-two-years-of-practice/

[13] Curate your own exhibit - crystal bridges. 2016. Retrieved August 08, 2017 from http://ideum.com/work/curate-your-own-exhibit.

[14] Textile viewer and creator exhibit - IAIA. 2017. Retrieved August 07, 2017 from http://www.ideum.com/work/textile-viewer-creator.

[15] awwapp - A web whiteboard. 2016. Retrieved April 05, 2016 from https://awwapp.com/.

[16] sketchpad - Draw, create, share. 2015. Retrieved April 05, 2016 from https://sketch.io/sketchpad/.

[17] P. Clifton, A. Mazalek, J. Sanford, C. Rbola, S. Lee, and N. Powell, "SketchTop: design collaboration on a multi-touch tabletop." ACM Press, 2011, p. 333. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1935701.1935778

[18] K. A. Nováková, "FOR STUDENTS: Sketching apart together - a collaborative seminar focused on architecture studio education," in *Proc. of the eCAADe Conference*, 2012, pp. pp. 77–85.

[19] E. M. Mercier and S. E. Higgins, "Collaborative learning with multi-touch technology: Developing adaptive expertise," *Learning and Instruction*, vol. 25, pp. 13 – 23, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959475212000850

[20] N. Partarakis, M. Antona, and C. Stephanidis, "Paint-It: A Children's Habit Revised," in *Second International Conference, DAPI 2014, Held as Part of HCI Interational 2014 Proceedings*, N. Streitz and P. Markopoulos, Eds. Springer International Publishing, 2014, pp. 164–171.

[21] A. Cooper, R. Reimann, and D. Cronin, *About Face: The Essentials of Interaction Design*, 3rd ed. Wiley, 2014.

[22] D. Klinkhammer, M. Nitsche, M. Specht, and H. Reiterer, "Adaptive personal territories for co-located tabletop interaction in a museum setting," in *Proc. of the ITS '11*. ACM, 2011, pp. 107–110. [Online]. Available: http://doi.acm.org/10.1145/2076354.2076375

[23] Simon, Nina, *The Participatory Museum*. Museum 2.0, 2010.

[24] U. Hinrichs, M. Hancock, S. Carpendale, and C. Collins, "Examination of Text-Entry Methods for Tabletop Displays," in *Workshop on TABLETOP '07*, 2007, pp. 105–112.

[25] S. Medić and N. Pavlovi, "Mobile technologies in museum exhibitions," in *TURIZAM - International Scientific Journal*, volume 18, issue 4 ed. Faculty of Science, Department of Geography, Tourism & Hotel Management, 2014, pp. 166–174.

[26] M. W. Van Someren, Y. F. Barnard, J. A. Sandberg, and others, *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press London, 1994, vol. 2.

[27] iSpy - open source camera security software. 2015. Retrieved April 05, 2016 from http://www.ispyconnect.com.

[28] D. Wigdor, G. Perm, K. Ryall, A. Esenther, and C. Shen, "Living with a tabletop: Analysis and observations of long term office use of a multi-touch table," in *Workshop on TABLETOP '07*, 2007, pp. 60–67.

[29] F. Block, J. Hammerman, M. Horn, A. Spiegel, J. Christiansen, B. Phillips, J. Diamond, E. M. Evans, and C. Shen, "Fluid grouping: Quantifying group engagement around interactive tabletop exhibits in the wild," in *Proc. of the CHI '15*. ACM, 2015, pp. 867–876. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702231

[30] U. Hinrichs and S. Carpendale, "Gestures in the wild: Studying multi-touch gesture sequences on interactive tabletop exhibits," in *Proc. of the CHI '11*. ACM, 2011, pp. 3023–3032. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979391

[31] P. Isenberg, U. Hinrichs, M. Hancock, and S. Carpendale, *Tabletops - Horizontal Interactive Displays*. Springer London, 2010, ch. Digital Tables for Collaborative Information Exploration, pp. 387–405.

[32] U. Hinrichs, H. Schmidt, and S. Carpendale, "EMDialog: Bringing Information Visualization into the Museum," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1181–1188, Nov 2008.

HoloKeys - An Augmented Reality Application for Learning the Piano

Dominik Hackl
University of Applied Sciences
Upper Austria
4232 Hagenberg/Austria
Email: dominikhackl@gmx.at

Christoph Anthes
University of Applied Sciences
Upper Austria
4232 Hagenberg/Austria
Email: christoph.anthes@fh-hagenberg.at

***Abstract*—This paper describes the design and the implementation approach of a piano training application. HoloKeys is an Augmented Reality tool which is capable to superimpose the keys to be played on a real piano. Musical pieces are loaded as MIDI files, interpreted and can be displayed in two different ways. This prototype provides many possibilities for extension which can make it a powerful teaching tool.**

I. Introduction

Augmented Reality (AR), described by Azuma as a technology where the user sees 'the real world, with virtual objects superimposed upon or composited with the real world' [1], has become a hot topic in the recent years. The application areas are wide spread and range far beyond simple advertisements and virtual manuals from advanced training to sophisticated remote collaboration scenarios. Using AR to train musical instruments has a long tradition in the field but because the rapid development in AR Head-Mounted-Displays (HMDs) this application area has gained new attention.

We present HoloKeys, a prototypical implementation of an AR training tool for learning the piano. HoloKeys runs on an HMD which the user is wearing while sitting in front of a physical piano. The application indicates notes that are supposed to be played by displaying virtual keys superimposing the physical keyboard with two different approaches. Acquiring the musical data dynamically by loading and processing MIDI (Musical Instrument Digital Interface) files, the application is fully agnostic considering the musical pieces to be trained. To achieve the required precision for the augmentations on the piano, the application was implemented using fiducial marker tracking. Since this application is a prototype, an extensive collection of possible enhancements and prospects for the future is given.

A. *Outline*

The remainder of this paper is structured as follows: The next chapter provides an overview of the related work in music teaching applications. Chapter III will introduce the conceptual design of the application describing the architecture and the user interface. Implementation details are provided in Chapter IV. Finally conclusions are drawn and an outlook into the future work is given.

II. Related Work

Music education has a long tradition in the field of AR. In an early approach Cheng and Robinson provided a visual sheet music overlay displayed planar in the visual field of the user. The display of the augmentations is triggered when he looks at the hands. The type of sheet is depending on which hand he looks. The augmentation is not registered (meaning it is not directly spatially interconnected) to a real object opposed to the approach presented in this publication. An HMD is used for display [2]. Cakmakci et al. augmented the information which string to pull on a guitar with the intention to reduce cognitive discontinuities compared to the traditional way of learning an instrument. They were the first to provide information on the interaction to be taken in an immediate way on an instrument [3]. The registration of the guitar and the virtual hand is implemented with the help of fiducial markers.

In order to avoid the use of fiducial markers on the piano Huang et al. use their knowledge on the application domain and track the keys of the piano for pose estimation with the help of natural feature recognition [4]. Unfortunatly they provide no details on the diplay used, but the frame-rate of 15 frames per second, implies that it has not been developed for a head-tracked system.

Chow et al. focus on the educational level of AR piano teaching showing that with the help of augmentations and gamification components the motivation and interest in learning the piano could be increased. They provided a system illustrating the notes to be played by lines approaching the keys. Their findings also indicate that notation literacy does not increase using their system of illustration [5]. We use a similar approach for the augmentations of the notes to be played but rely on a optical see-through HMD instead of a video-based HMD.

Opposed to this visualisation approach Torres-Fernandez et al. introduce a virtual character which illustrates how well the piano player has performed. To interpret the played music they compare the input from a MIDI keyboard with an initially loaded MIDI file [6]. A similar analysis was suggested and implemented earlier by Barakonyi and Schmalstieg [7]. They make use of fiducials for tracking and a desktop AR system equipped with a webcam and a traditional screen.

In terms of visualisation Weing et al. demonstrate a system in the area of Spatial Augmented Reality where they project the keys to be pressed directly on the piano. Different modes show for example the current and the next keys to be pressed. If a wrong key is pressed it is highlighted in red to provide feedback to the user [8].

Zhang et al. use a completely virtual keyboard and track the hand of the user with fiducial markers and the finger positions with a self-developed data glove. Their approach targets the rehabilitation of the motor function of stroke survivors rather than teaching the piano [9].

Compared to these existing and presented approaches our system is unique in terms of used display technology.

III. Conceptual Design

The following chapter gives an overview of the application's hardware and software components and explains how the individual parts interact with each other.

A. Architecture Overview

The application's setup is illustrated in Fig. 1 and consists of the following two hardware components.

1) The Piano: The core component is a physical piano which is used for the actual playing. Underneath the piano keyboard which is usually made of 88 keys a fiducial marker is placed which is used by the application for tracking. The keys of a regular piano are standardized in size which makes the application fully independent considering the type of piano. In case a keyboard is used the key width can be adjusted.

2) The Head-Mounted-Display: The user sits in front of the piano and wears an HMD on which the application runs. Through the HMD the user sees augmentations in the form of highlighted keys on top of the real keyboard. The HMD also handles tracking by recognizing the image marker with the help of computer vision algorithms. The HMD therefore keeps track of the player's position and displays the augmentations accordingly. Additionally, the HMD is responsible for sound output of the music to be played. This gives the user an impression on how the piece is supposed to sound and makes it easier to play along with it.

B. Interface

In order to manage different settings and control the playback, a simple user interface was implemented. The originally two-dimensional UI is placed inside the 3D scene using world-stabilized coordinates. Considering the usually static setup of the application with the user sitting in front of the piano, the world-stabilized menu is a reasonable approach [10]. User input works through gaze-based interaction combined with gestures.

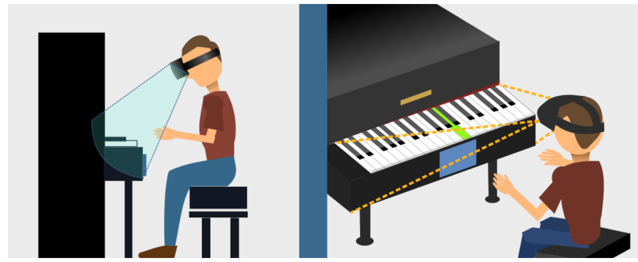


Fig. 1. Illustration of the conceptual design. The user, sitting in front of the piano and wearing an HMD, looks down at the keyboard. When there are notes to be played the respective key is highlighted. Underneath the keyboard there is an image marker which is used for tracking.

1) The Main Menu: The initial scene of the application is the main menu. There the user can select the musical piece to play as well as the desired playback speed. By pressing the start button the application will switch to playback mode and begin visualizing and playing the musical piece.

2) Playback Mode: In playback mode the user sees the augmentations of the keys to be played superimposing the physical keyboard. Additionally a timeline shows the current playback position and gives the user the option to jump to different positions inside the piece. With the pause button the user is able to interrupt the playback or return to the main menu.

3) Calibration Mode: In calibration mode the application displays an augmentation of only one key, the middle C. The user can adjust the position of the marker until the virtual key perfectly fits the real one. This is useful to setup the optimal position of the marker on the piano. Additionally the user can also adjust the pitch of the virtual piano sound in calibration mode because this does not necessarily match with the real piano. Playback volume can be adjusted in the HMD.

C. Display of Augmentations

Generally the HMD displays an augmentation of a bright green key to indicate that the actual key on that position has to be pressed. Two different approaches as seen in Fig. 2 were tested and both have their advantages and disadvantages concerning predictability and Field Of View (FOV) limitations.

1) The Instant Approach: The moment a key is supposed to be pressed it becomes highlighted. Once it is supposed to be released it switches back to normal. This way the user can more or less observe the playing of the piece in real-time, comparable to watch the fingers of an actual pianist. While this approach can be useful for advanced players, it is hardly possible to learn a new piece or even to play along with it, because the player has no way of predicting the next notes. Still, observing this looks great and could be used for showcase purposes (self-playing piano), as the limited FOV is also less of a problem there.

2) The Beatmania Approach: Note objects are created far in the distance and from there start moving towards the particular keys. As soon as the virtual object reaches the real key, the note should be played. With this approach, which became

Fig. 2. Comparing the two tested approaches. Left: The Instant Approach. Right: The Beatmania Approach.

popular with the game 'Beatmania' [11] and is still used in many music rhythm games today, the user can anticipate the upcoming notes and prepare accordingly. When learning a piano piece the musician's brain utilizes its 'muscle memory' and fine motor skills rather than memorizing each individual note [12]. Therefore learning a piece with the Beatmania approach should be equally efficient than learning it from sheet music, especially for beginners.

IV. Implementation

This chapter goes into detail regarding the concrete implementation of HoloKeys. It starts with a brief overview of used hardware and software tools followed by an in-depth description of the two main development tasks, visualization and MIDI processing.

A. Used Technologies

The application was developed for tablet devices as well as the HoloLens. The tablet approach is mainly used for demonstration purposes, rather than actual training.

1) Hardware:

- HoloLens[1]
  The HoloLens as a current AR HMD provides good sensory support as well as spatial audio and stereoscopic display capabilities. Its main disadvantage the limited FOV poses an issue to the applicability of this use case.

2) Software: To allow cross-platform and cross-device development the following set of tools and libraries was used.

- Unity[2]
  Unity is traditionally a game engine which has found wide adoption in the whole domain of Mixed Reality [13]. It allows scene setup and provides scripting capabilities. The applications developed with Unity can easily be deployed on a multitude of target platforms including iOS and Android devices as well as UWP (Universal Windows Platform) devices.
- Vuforia[3]
  The Augmented Reality part of the project is based on Vuforia, an AR tracking library which perfectly integrates with Unity. Vuforia supports several different tracking methods ranging from recognizing plain images to complex objects. With a specific setup, Vuforia can also be used on the HoloLens.
- C# Synth Project and MIDI Support[4]
  The C# Synth Project is an open-source library which is used for processing MIDI data and synthesizing it to audio data. MIDI is an industry standard for interconnection between musical instruments and digital devices. Its file format represents musical information like notes values, volume and tempo. Although MIDI is a complex format, it is still the most popular and commonly used format to store musical data. For piano pieces the format is usually sufficient because only one channel is required to store a series of notes and tempo changes.

B. Visualization and Tracking

The application's visuals consist of a Unity 3D scene which renders the virtual keys, combined with Vuforia's tracking abilities to provide the information on where to render the keys.

1) Vuforia's image target: For this application tracking via fiducial marker and image target was used. The image target in Unity is a planar object in 3D space which is associated with a set of 2D images. These images represent the markers that are placed somewhere in the real world. Once the camera recognizes a marker the application can trace back the position of the HMD and can therefore project all augmented objects accordingly.

2) Tracking setup: Marker images and other tracking settings can be configured in Vuforia's web interface. This configuration with all related assets is then compiled into a Unity package that can be imported into Unity after that. In Unity two components of Vuforia, ARCamera and ImageTarget, are used. Subordinate objects of the ImageTarget become affected by the marker-related projection.

3) Generating the keyboard: In order to display the currently played keys, first an entire virtual keyboard is displayed half-transparently superimposing the real one. A script takes care of automatically generating all 88 key objects. One base key object is placed in the scene and aligned at around 90 degrees relative to the ImageTarget. This registration has to match with the real world relation between marker and piano keyboard. All other keys are then generated as duplicates of the base object with respective offset and color (black or white).

C. Audio and MIDI Playback

The two core components of the C# Synth Project library are the MidiSequencer which handles loading and processing MIDI data and the MidiStreamSynthesizer which handles the actual audio playback.

[1] https://www.microsoft.com/en-us/hololens
[2] https://unity3d.com/
[3] https://www.vuforia.com/
[4] https://csharpsynthproject.codeplex.com/

1) Handling key actions: During playback the MidiSequencer fires two events that are relevant for this application: MidiNoteOn and MidiNoteOff. These two events are respectively fired when the playback of a note is triggered or terminated and therefore indicate exactly the time when a key is pressed and released. In the implementations of these two event handlers the MIDI code of the affected note is passed as a parameter. The only operation is to map this MIDI code to our according key object and set its material color to either green (in NoteOn) or the default color (in NoteOff).

2) Combining the audio sources: The MidiStreamSynthesizer creates actual audio data based on the sequencer's input. To make sure that this audio data is actually redirected to Unity's audio source, the special method OnAudioFilterRead has to be implemented. This method supports direct writing into the audio buffer and therefore redirect the contents of the StreamSynthesizer to Unity's audio source.

V. Conclusion

As a prototype the application serves well, but due to the limited FOV, which will most likely increase in the next years with the following generations of AR hardware, its real world usage could be doubted. Furthermore, an evaluation of the different augmentation methods would be useful. Especially when trying out a few more possible approaches, a user test could find out which of the methods are most likely to work in a real-world scenario. A more in-depth study of musical augmentation methods would also be useful for teaching other instruments or even in completely different areas of music.

A. Future Work - The Virtual Piano Teacher

A long-term vision could be the creation of a full-featured virtual piano teacher using AR. Especially early-stage piano learning contains many tasks that could be implemented with AR technologies like the one explained in this paper combined with gamification elements.

1) Use Cases:

- Learning notes and the piano keyboard
  Simple exercises or games to recognize the note names and match it with the proper keys could really increase the early-stage learning rate. For beginners the note names could be augmented on top of every key until they become familiar with it.
- Learning easy to intermediate musical pieces
  Especially for smaller pieces the AR learning approach could surpass traditional learning by music sheets. Beginners who are not used to reading music yet, would still be able to learn pieces quickly on their own. Additionally a lot more useful information like fingering, expression and dynamics could be displayed during playback.
- Technical exercises
  The importance of regular technical exercises for piano students is huge but generally underestimated and disliked. With the introduction of AR and gamification, a whole lot of enjoyable and still pianistically valuable exercises could be realized. By adding some sort of level system, the student would be even more aware of his progress and more likely to remain motivated.
- Dictionary of chords, scales etc.
  A very useful utility not only for beginners but also for advanced pianists would be a piano dictionary. The player could look up all possible chords and scales and would be able to see them highlighted right on top of his keyboard. Especially for jazz piano where complex chords and scales are common, this technology would be of great service.

2) Further Improvements:

- Using music sheets as markers
  The use of music sheets, perhaps in the form of a special music book, as fiducial markers could eliminate the need for additional markers placed on the piano. It could not only automatically detect the musical piece to be played but also indicate, when to turn the sheets or even highlight musical attributes on the sheets.
- Checking the learning performance
  Real-time feedback of the user's playing could greatly contribute to the learning experience. This could be achieved on the one hand by using MIDI keyboards to directly receive the MIDI input of pressed keys or on the other hand by recording and deconstructing the audio data. The first approach would be technologically straight-forward but would limit the application to electronic keyboard instruments while the second approach would be more flexible but complicated to implement and perhaps inaccurate [14].

The possibilities of the virtual piano teacher are enormous but all are based on the core concept of the technique explained in this paper. As soon as there are improvements in AR hardware, especially concerning FOV, virtual piano teachers can be implemented and actually start to become a helpful tool.

References

[1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, August 1997.

[2] L.-T. Cheng and J. Robinson, "Personal contextual awareness through visual focus," *IEEE Intelligent Systems*, vol. 16, no. 3, pp. 16–20, 2001.

[3] O. Cakmakci, F. Brard, and J. Coutaz, "An augmented reality based learning assistant for electric bass guitar," in *10th International Conference on Human-Computer Interaction*, 2003.

[4] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano AR: A markerless augmented reality based piano teaching system," in *Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2011.

[5] J. Chow, H. Feng, R. Amor, and B. C. Wunsche, "Music education using augmented reality with a head mounted display," in *Fourteenth Australasian User Interface Conference (AUIC2013)*. Melbourne, Australia: ACM, Jan. 2013, pp. 73–79.

[6] C. A. T. Fernandez, P. Paliyawan, and C. C. Yin, "Piano learning application with feedback provided by an ar virtual character," in *5th Global Conference on Consumer Electronics*. Kyoto, Japan: IEEE, Oct. 2016.

[7] I. Barakonyi and D. Schmalstieg, "Augmented reality agents in the development pipeline of computer entertainment," in *4th international conference on Entertainment Computing (ICEC'05)*. Sanda, Japan: Springer, Sep. 2005, pp. 345–356.

[8] M. Weing, A. Rhlig, K. Rogers, J. Gugenheimer, F. Schaub, B. Knings, E. Rukzio, and M. Weber, "P.i.a.n.o.: Enhancing instrument learning via interactive projected augmentation," in *Conference on Pervasive and ubiquitous computing adjunct publication (UbiComp13)*. Zurich, Switzerland: ACM, Sep. 2013, pp. 75–78.

[9] D. Zhang, Y. Shen, S. Ong, and A. Nee, "An affordable augmented reality based rehabilitation system for hand motions," in *International Conference on Cyberworlds (CW '10)*. Singapore, Singapore: IEEE, Oct. 2010.

[10] M. Billinghurst and H. Kato, "Collaborative mixed reality," in *International Symposium on Mixed Reality (ISMR '99)*. Springer, 1999, pp. 261–284.

[11] S. Steinberg, *Music Games Rock*. P3: Power Play Publishing, 2011. [Online]. Available: http://www.musicgamesrock.com/

[12] R. Shusterman, "Muscle memory and the somaesthetic pathologies of everyday life," *Human Movement*, vol. 12, no. 1, pp. 4–15, 2011.

[13] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," *Presence: Telemanipulator and Telepresence Technologies*, vol. 2351, pp. 282–292, 1994.

[14] S. Dixon, "On the computer recognition of solo piano music," in *Proceedings of Australasian computer music conference*, 2000, pp. 31–37.

Browser Application for Virtual Audio Walkthrough

Thomas Deppisch
Student, Graz University of Technology
and University of Music and Performing Arts
Email: thomas.deppisch@student.tugraz.at

Alois Sontacchi
University of Music and Performing Arts
Institute of Electronic Music and Acoustics
Inffeldgasse 10, 8010 Graz, Austria
Email: sontacchi@iem.at

***Abstract*—We present an application allowing an interactive virtualization of auditory scenes. It enables the user to navigate through the virtual scene inside a web browser. Audio signals are spatialized for headphone playback using a binaural Ambisonics approach. A mixture of cues is used to activate and enhance distance perception. Customized scenes are created using a simple text file which contains meta data regarding properties of the virtual room and the audio objects. In order to scale the audio reproduction quality corresponding to available computational power, parameters like Ambisonics order and image source order are used to adjust the virtualization during runtime.**
The source code is provided online[1].

I. Introduction

Hitherto in conventional and classical audio recordings the acoustic perspective within the recording has been defined by the tonmeister. However, new developments [1] provide the possibility to follow new practices in media/audio immersion: Listeners can navigate throughout a production visiting any favored position of interest. The addressed invention [1] relates to an audio production, processing, and playback apparatus to convey a multichannel interactive audio experience, allowing the listener to traverse an entire sound scene. Hereinafter, we present a web based implementation of this approach.
Before going into implementation details, the following introduction states how direction and distance of acoustic sources are perceived and reproduced. Basic concepts of the Web Audio Application Programming Interface (API) for audio processing in a browser environment are shown as well.

A. Perception of direction

Cues for the perception of an acoustic source direction are classified into monaural and binaural cues [2]. Binaural cues utilize information from differences in both ear signals while monaural cues utilize equivalent parts of both ear signals to determine the direction of a sound source [2]. Binaural cues can be further divided into interaural level differences (ILDs) and interaural time differences (ITDs). ILDs arise due to head shadowing effects for signals with small wave lengths compared to the diameter of the head. Hence, lateral sources produce higher levels on the ipsilateral ear than on the contralateral ear [2]. The delayed arrival of a sound signal at the contralateral ear in comparison to the ipsilateral ear results in an interaural time difference. Such a delay is evaluated using the phase difference in both ear signals. For wave lengths smaller than the diameter of the head these phase differences do not contain useful information. Therefore, ITDs are predominantly used for localization of signals with low frequency content [2]. Still, evaluation of the signal envelope allows localization based on ITDs for higher frequency signal components [3].
Monaural cues are manifested in direction dependent spectral changes of the ear signals' frequency responses. These spectral changes emerge due to reflections on pinna and torso, resulting in constructive and destructive interferences. Spectral localization cues are predominantly important for localization of elevated sources in sagittal planes, to prevent confusions and ambiguities [2].

B. Head-related transfer function

Both, monaural and binaural cues are incorporated in the head-related transfer function (HRTF) and its time domain representative, the head-related impulse response (HRIR) [2]. The HRIR can be obtained by placing microphone probes inside the ear channels of a test person or dummy head and measuring the impulse response for a number of source directions [4]. The HRIR is generally direction-dependent and hence can be used to simulate direction of a source in binaural synthesis. For distances smaller than 1 m the HRTF also shows distance-dependent spectral variations. For non-static sources or when head movements are incorporated, interpolation of a finite number of measured HRTFs is essential [5]. The anthropometric differences between human individuals result in individual spectral differences in HRTFs which can lead to an impairment of the binaural experience when using non-individualized HRTFs.

C. Perception of distance

Distance perception for acoustic sources is generally less accurate than the perception of direction [6]. There are several acoustic cues which allow a distance estimation for sound sources but also non-acoustic cues that play a big role in overall distance perception. The most prominent acoustic distance cue is the inverse distance law for sound pressure which states a 6 dB reduction of sound pressure level when doubling the source distance in free field conditions [6].
Another acoustic distance cue is the direct-to-reverberant energy ratio in reflective environments. Here, close sources provide a greater amount of direct energy in comparison to reverberant energy [6]. For sources further away than 15 m, air

[1] https://git.iem.at/thomasdeppisch/walkthrough

absorption results in high frequency attenuation and therefore in spectral distance cues [2]. Furthermore, for sources closer than 1 m an increase in low frequency ILDs has a strong impact on distance perception for close sources [7].

D. The Web Audio API

The Web Audio API[2] (WAA) allows modular audio processing in a web browser environment. Audio signals thereby are sent through an audio routing graph consisting of audio nodes which can be connected arbitrarily. A source node such as the *MediaElementAudioSourceNode* allows the integration of audio files into the routing graph. Several predefined audio nodes such as *BiquadFilterNode*, *DelayNode*, *GainNode* and *ConvolverNode* provide the possibility of realtime audio processing. The *AudioDestinationNode* connects the audio routing graph to the audio hardware. The WAA also allows basic spatialization by providing a *SpatialListenerNode* and a *SpatialPannerNode*. Customization of settings like HRTF set, distance function and directivity function are currently not possible [8].

II. Related work

So far, traversing a sound scene in reproduction could be realized by audio spatialization based on isolated recordings combined with additional spatial recordings or rendering of reverberation (object-based). Although the listener is meant to be located at a central position, by changing the arrangement of the virtual sources the playback perspective at the reproduction side can be adapted. There are several products allowing the use of this approach, e.g. Fraunhofer Spatial Sound Wave[3], or the Ambix Plugin Suite [9].

Moreover, Pihlajamäki and Pulkki [10] presented a different approach based on the DirAC [11] method. There the sound field is decomposed into a non-diffuse and diffuse part. Then the non-diffuse part gets resynthesized by assigning a direction to each frequency band. Transformations of the direction vectors, gain control and diffuseness control are used to simulate translations of the listener.

A method for sound field navigation using Ambisonics was presented by Allen and Kleijn [12]. After the directional decomposition of a signal, an adjustment for the translated origin is performed by filtering. Re-encoding is done in respect to the new angles based on the translation vector.

BogJS is a JavaScript framework for object-based audio rendering in browsers[4]. A demo[5] shows the use case of auditory scene virtualization in a web browser. As the spatialization is done solely with Web Audio API functionalities, the possibilities of personalization (e.g. change of the HRTF set) and flexible adjustments (e.g. of the distance gain function) are restricted.

[2]https://www.w3.org/TR/webaudio/

[3]https://www.idmt.fraunhofer.de/en/institute/projects_products/q_t/spatialsound_wave.html

[4]https://github.com/IRT-Open-Source/bogJS

[5]https://lab.irt.de/demos/object-based-audio/interactive/

III. Recording an auditory scene for virtualization

For recording of auditory scenes with the goal of later virtualization two approaches are feasible: Virtualization of sound objects recorded through spot microphones or virtualization of the scene recorded by multichannel microphone arrays (cf. figure 1). In the first case every microphone signal represents an acoustic object in the virtual space, e.g. a musical instrument. In the second case the signals of one microphone array represent a part of the sound field spatially sampled at one point in the room. Hence, the overall sound intensity of the multichannel microphone arrays needs to be normalized, so a higher density of microphone arrays in one part of the room does not result in a higher intensity. A hybrid approach combining spot microphones and multichannel microphone arrays is also feasible. During playback every microphone capsule is interpreted as a virtual speaker object which then gets placed in the room according to its original position.

Fig. 1. Recording an auditory scene using multichannel microphone arrays.

IV. Binaural synthesis using a virtual Ambisonics approach

A. Ambisonics

By solving the Helmholtz equation a spherical harmonics transform is obtained. Applying the spherical harmonics transformation to a point source leads to Ambisonics encoding (cf. eq. (1)) and decoding equations (cf. eq. (2)) [13].

$$\vec{\chi}_N(t) = \vec{y}_N(\vec{\theta}_0)s(t) \tag{1}$$

$$\vec{s}_{ls}(t) = \mathbf{D}\ \mathrm{diag}\{\vec{a}_N\}\ \vec{\chi}_N(t) \tag{2}$$

Multiplication of the signal $s(t)$ with the spherical harmonics evaluated at the desired source position $\vec{\theta}_0$ contained in $\vec{y}_N$, yields the Ambisonics encoded signals $\vec{\chi}_N(t)$ (eq. (1)). The order at which the evaluation of spherical harmonics is truncated is called Ambisonics order N. The encoded signals are decoded to speaker signals $\vec{s}_{ls}(t)$ by multiplication with a suitable decoder matrix $\mathbf{D}$ (eq. (2)). The decoder matrix can be obtained in several ways such as mode-matching, sampling or AllRAD [14]. The vector $\vec{a}_N$ can contain psychoacoustically motivated optimization factors, e.g. for *max* $\vec{r}_E$ optimization. *Max* $\vec{r}_E$ optimization reduces sidelobes and therefore leads to

a more distinct source localization (cf. [14], [15]).
Apart from full periphonic (3D) Ambisonics, circular harmonics can be employed to obtain planar (2D) Ambisonics. Further, mixed-order schemes are used to encode horizontal source information in higher order than vertical information [16]. Rotation of a sound field is done efficiently in the Ambisonics domain by matrix multiplication as described in [17].

B. Virtual Ambisonics approach

For binaural synthesis a virtual Ambisonics approach is used [18]. A regular distribution of virtual speakers is placed around the virtual listener. Decoding of the encoded Ambisonics signals $\vec{\chi}_N(t)$ at the virtual speaker positions $\vec{\theta}_q$ is achieved by multiplication with the decoding matrix $\mathbf{D}_{vls}$. The binaural signals for the left and right ear (eq. (3), (4)) are obtained by convolving the resulting virtual loudspeaker signals with their corresponding HRIRs and summing them up for each ear.

$$s_l(t) = \sum_{q=1}^{m} HRIR_{l,q}(\vec{\theta}_q) * (\vec{e}_q^{\,T}\ \mathbf{D}_{vls}\ \vec{\chi}_N(t)) \qquad (3)$$

$$s_r(t) = \sum_{q=1}^{m} HRIR_{r,q}(\vec{\theta}_q) * (\vec{e}_q^{\,T}\ \mathbf{D}_{vls}\ \vec{\chi}_N(t)) \qquad (4)$$

This approach, in contrast to HRTF interpolation methods [5], allows a rotation of the encoded sound field in the Ambisonics domain [17] instead of interpolation of HRTFs for every sound object. Therefore, the number of needed HRTFs is only depending on the number of virtual speakers and not on the number of virtual sound objects. This can reduce the amount of convolutions needed and hence reduce the computational effort.

V. Implementation

A. General functionality

The application uses an interaction of JavaScript code and Web Audio API (WAA) audio nodes based on C++ implementations. Background signal processing such as convolutions, filtering and gain adjustments are accomplished by WAA audio nodes. The calculations to retrieve the values for spatialization are done in JavaScript code. For Ambisonics processing, classes of the open source JavaScript library JSAmbisonics[6] [19] were adapted to provide periphonic as well as planar Ambisonics processing. As JSAmbisonics is built on top of the WAA as well, a seamless integration is possible.
In the following the construction of auditory scenes allowing a virtual walkthrough is explained step by step.

B. Scene File

To construct the virtual scene, meta data needs to be provided in a simple text file, the scene file (cf. figure 2). A valid scene file needs to follow the JSON[7] (JavaScript Object Notation) standard.

[6]https://github.com/polarch/JSAmbisonics
[7]http://json.org/

```
[{
        "type": "room",
        "width": 4.5,
        "length": 5.5,
        "height": 4,
        "listenerStart": {"x":2,"y":1}
},{
        "type": "mono",
        "name": "Noise",
        "position": {"x":1,"y":1,"z":1},
        "gain": 0.8,
        "NFC" : 1,
        "orientation": {"azim":90,"elev":-45},
        "distGain": {"a": 1.4,"g0": 1},
        "file": "sounds/noise.wav"
},{
        "type": "fourChannelArray",
        "name": "Oktava",
        "center": {"x":4,"y":4},
        "centerDistance": 0.5,
        "directivity": 0.5,
        "file": "sounds/oktava1.ogg",
        "channelMapping": {"speaker1":1,"speaker2":2,"
             speaker3":3,"speaker4":4}
}]
```

Fig. 2. Example for a valid scene file containing scene meta data.

In the first section the scene file provides information of the room as well as coordinates for the starting point of the virtual listener. Below the room data an arbitrary number of audio objects can be defined. Objects of type *mono* are based on a mono audio track, e.g. a spot microphone recording. Objects of type *fourChannelArray* represent a spatially sampled part of the sound field recorded by a microphone array consisting of four capsules. Each defined audio object has parameters like position, gain, orientation, distance gain function, directivity and reference to a sound file. Optionally, near field compensation filters (NFC) which approximate the filters given in [20] can be activated for mono objects. Objects of type *fourChannelArray* are defined by a center position and a center distance for each of the four corresponding virtual speakers. The sound file of a *fourChannelArray* object contains four separate mono channels. To map these four mono channels to the corresponding virtual speaker object, a channel mapping parameter is provided.
By using the directivity parameter a virtual speaker radiation directivity can be controlled. The directivity gain follows equation (5) and hence enables interpolation between omnidirectional ($\gamma = 1$), cardioid ($\gamma = 0.5$) and figure of eight ($\gamma = 0$). These directivity patterns are also valid for the three dimensional space as the angle φ is calculated as the angle between the vector pointing from the virtual speaker to the listener and the vector pointing in the same direction as the speaker.

$$g_{dir} = \gamma + (1-\gamma)cos(\varphi) \qquad (5)$$

The distance gain function follows equation (6) and can be adjusted by using the parameters α and g_0. The resulting distance gain equals 1 for a distance $r = 1$, linearly interpolates to g_0 for distances $r < 1$ and decreases by $1/r^{\alpha}$ for distances $r > 1$.

$$g_{dist} = \begin{cases} g_0 + (1-g_0)r & ,\text{if } r \le 1 \\ \frac{1}{r^{\alpha}} & ,\text{if } r > 1 \end{cases} \qquad (6)$$

Figure 3 shows the default distance gain functions for objects of type *mono* and type *fourChannelArray*. For far distances the default distance gain decreases by $1/r^{1.4}$. This is an overproportional decrease compared to the inverse distance law and responds to the fact that physical distance is generally rather underestimated by humans when sound intensity is the primary cue [6]. For close distances the default gain depends on the object type: For *fourChannelArray* objects which do not represent an actual audio source but a part of the sound field, the distance gain decreases for close distances, so a single virtual speaker does not get too prominent. For *mono* objects the gain remains constant at $g_{dist} = 1$.

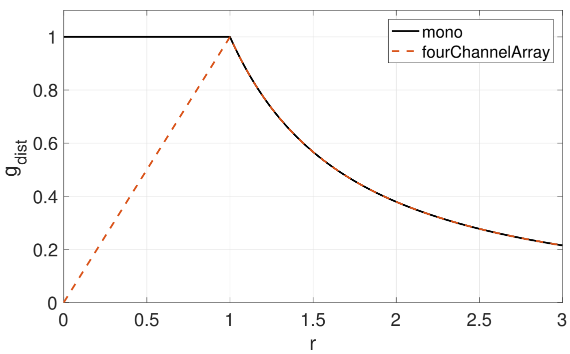


Fig. 3. Default distance gain as a function of distance r for objects of type *mono* (solid black) and type *fourChannelArray* (dashed red).

C. Virtualizing the scene

In the next step, when activated, mirror image sources for each audio object are built. These additional copies of sources follow the concept of simulating room reflections by mirroring sources along the room boundaries as explained in [21]. Image sources of first and second order are provided in the application and can be activated during runtime. Activation of image sources in big auditory scenes can lead to performance impairments due to the fact that the number of sources and hence all calculations for spatialization are multiplied.
From this point onwards the signal processing steps are displayed in a block diagram (cf. figure 4). Relating to the position of the virtual listener, angle and distance to each virtual speaker are calculated dynamically. From this data directivity gain and distance gain as described in eqations (5), (6), as well as a dynamic delay line (equation (7)) are adjusted.

$$\Delta t = \frac{r}{343\frac{m}{s}} \tag{7}$$

As the delay is adjusted dynamically to fit the distance r between listener and speaker, it is able to reproduce the Doppler shift.
Image sources are then lowpass filtered simulating a high frequency loss caused by absorption during reflections on room boundaries. In the last step before Ambisonics encoding, loudspeaker objects corresponding to *fourChannelArray*

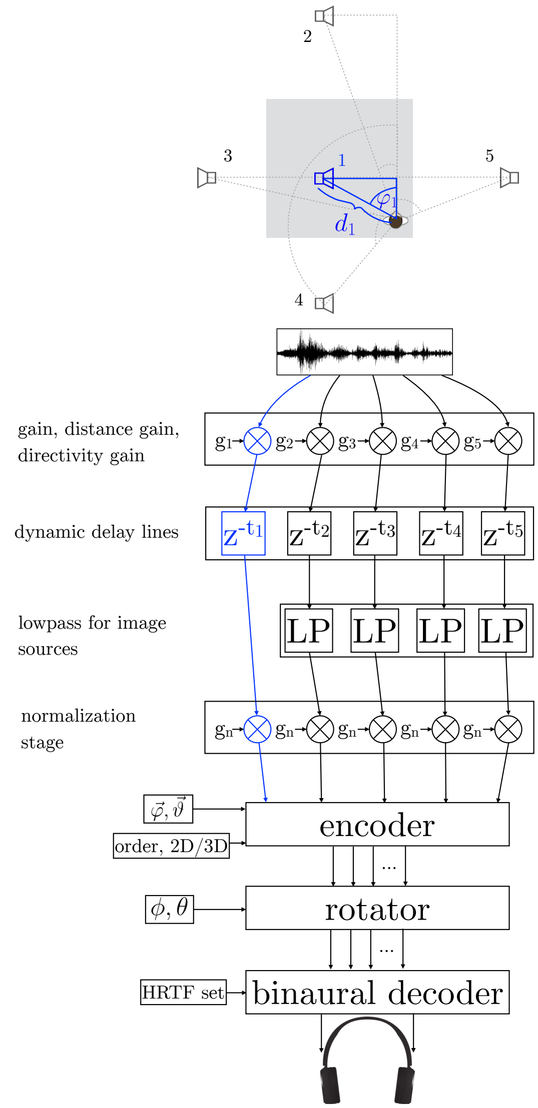


Fig. 4. Signal processing block diagram.

sources are intensity normalized as described in section III.
Before binaural headphone signals are obtained, Ambisonics encoding, rotation and decoding takes place. The encoder evaluates spherical or circular harmonics at the speaker directions relative to the virtual listener. *Mono* as well as *fourChannelArray* sources are encoded in an adjustable Ambisonics order N. If *fourChannelArray* sources (first order Ambisonics microhones) are encoded in a higher order than first order, the sound field does not get reproduced accurately. Yet, due to the superposition of several sound field sample points, audio information from the four room directions get reproduced more sharply when higher order encoding is enforced. For close distances all Ambisonics channels but the W-Channel (contains omnidirectional information) get interpolated to zero to avoid discontinuities when passing through a virtual speaker object.

The Ambisonics rotator is able to rotate the whole sound field in the Ambisonics domain. It enables head rotations of the virtual listener.
At the decoding stage Ambisonics signals get decoded to a regular distribution of virtual speakers. The number of virtual speakers depends on the Ambisonics order N: For periphonic Ambisonics a t-design [22] of degree $t = 2N+1$ and for planar Ambisonics a circular distribution of $2N + 2$ speakers are used. For HRTF individualization arbitrary SOFA[8] (Spatially Oriented Format for Acoustics) HRTFs are supported.

D. *User interface*

Figure 5 shows the user interface of the application. Before scene playback can be started, an auditory scene and an HRTF set need to be chosen by using the blue dropdown menus. Optionally, the Ambisonics type, Ambisonics order N and image source order can be adjusted to fit the scene-specific needs and computational possibilities. The Ambisonics type can be switched between 2D, 3D and 2D in combination with first order 3D components. The restriction of a maximum of 32 channels per audio node by the WAA and a highest supported *t-design* of degree $t = 21$ by JSAmbisonics yields maximum Ambisonics orders of $N = 4$ for 3D, $N = 15$ for 2D and $N = 10$ for 2D with first order 3D components. The navigation of the listener (depicted by a head, cf. figure 5) is accomplished by mouse dragging or using the up and down arrow keys. The left and right arrow keys as well as the azimuth slider are used to turn the head of the listener in the horizontal plane. The elevation slider is used to perform up and down head movements which are not graphically depicted as the scene is represented from a 2D perspective. Alternatively, head movements can be controlled via a low-cost open-source MIDI headtracker[9] [23] which is integrated using the Web MIDI API[10] and WebMidi.js[11]. The usage of a headtracker is currently only possible either using Chrome or Opera browsers, supporting the Web MIDI API. A volume slider and a start/pause toggle allow controlling the playback. A grey canvas below the settings section represents the room. Inside the canvas the listener and the virtual speaker objects are depicted. A virtual speaker object of type *mono* is depicted by a single speaker symbol. Virtual speaker objects of type *fourChannelArray* are represented by four speaker symbols arranged in a circle.

VI. Conclusion and Outlook

After informal listening tests the presented application creates a promising impression of a virtual concert scene. Localization of single sound sources works well, especially if *mono* sources (corresponding to spot microphones at recording stage) are used. The use of *fourChannelArray* sources (corresponding to microphone arrays sampling a part of the sound field) enhances the immersion. Therefore, a combination of *fourChannelArray* and *mono* sources leads to the best results. Crosstalk between spot microphones should be avoided as much as possible as it may split the perceived direction of a sound source. The perceived immersion due to a valid room impression can be further improved by using image sources. Unfortunately, for big auditory scenes it is often not possible to activate image sources as the number of simultaneously processed audio channels rises by a multiple for every image source order. For big auditory scenes containing a high number of audio channels the computational power of an average personal computer may then be insufficient. The efficiency of the program might be improved by using an underlying C++ implementation integrated through a JavaScript wrapper like in WAA audio nodes. New drafts of the WAA also contain *AudioWorkerNode* classes which might be able to enhance the performance. Further challenges occur when embedding the application into a website: The limited download speed might prohibit the playback of big auditory scenes due to the big amount of audio data which needs to be downloaded.

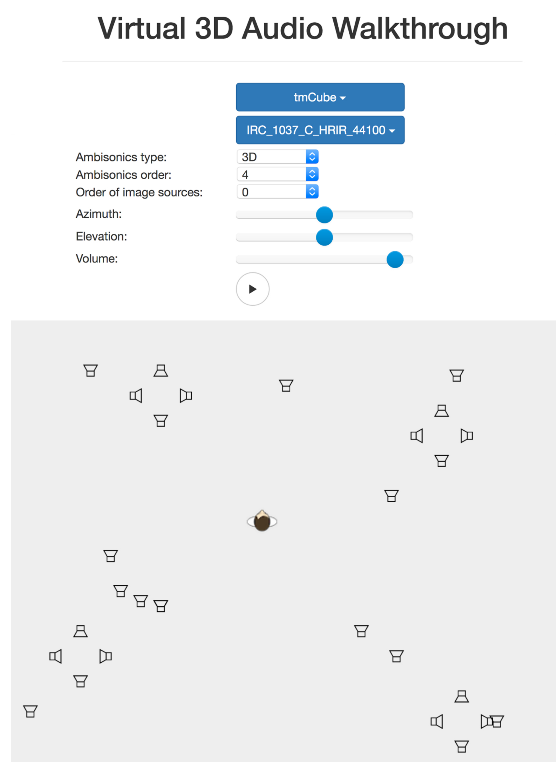


Fig. 5. User interface of the application.

[8] https://www.sofaconventions.org/
[9] https://git.iem.at/DIY/MrHeadTracker
[10] https://www.w3.org/TR/webmidi/
[11] https://github.com/cotejp/webmidi

References

[1] "Method and apparatus acoustic scene playback," European patent application: PCT/EP2016/075595, applicant institution: HUAWEI Technologies CO. LTD. (China), inventors: Schörkhuber Christian, Zotter Franz, Frank Matthias, Höldrich Robert, and Grosche Peter.
[2] J. Blauert, *Spatial Hearing: The psychophysics of human sound localization*. MIT Press, 1997.
[3] E. Macpherson and J. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *Journal of the Acoustical Society of America*, vol. 111, no. 5, 2002.

[4] H. Moller, M. F. Sorensen, D. Hammershi, and C. B. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc*, vol. 43, no. 5, pp. 300–321, 1995.
[5] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *AES 16th International Conference on Spatial Sound Reproduction*, 1999.
[6] P. Zahorik, D. Brungart, and A. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, pp. 409–420, 05 2005.
[7] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources. II. Localization of a broadband source," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1956–1968, 1999.
[8] T. Carpentier, "Binaural synthesis with the Web Audio API," in *1st Web Audio Conference (WAC), Paris, France*, 2015.
[9] M. Kronlachner, "Ambisonics plug-in suite for production and performance usage," in *Linux Audio Conference*, 2013.
[10] T. Pihlajamäki and V. Pulkki, "Synthesis of complex sound scenes with transformation of recorded spatial sound in virtual reality," *J. Audio Eng. Soc*, vol. 63, no. 7/8, pp. 542–551, 2015.
[11] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc*, vol. 55, no. 6, pp. 503–516, 2007.
[12] A. Allen and W. B. Kleijn, "Ambisonic soundfield navigation using directional decomposition and path distance estimation," in *4th International Conference on Spatial Audio, Graz, Austria*, 2017.
[13] F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays," Dissertation, University of Music and Performing Arts, Graz, A, 2009.
[14] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *J. Audio Eng. Soc., Vol. 60, No. 10*, 2012.
[15] M. Gerzon, "General metatheory of auditory localisation," in *Preprint 3306, 92nd Conv. Audio Eng. Soc.*, 1992.
[16] C. Travis, "A new mixed-order scheme for ambisonic signals," in *Ambisonics Symposium*, Graz, A, 2009.
[17] J. Ivanic and K. Ruedenberg, "Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion," *The Journal of Physical Chemistry*, vol. 100, no. 15, pp. 6342–6347, 1996.
[18] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich, "3d binaural sound reproduction using a virtual ambisonic approach," in *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 2003.
[19] A. Politis and D. Poirier-Quinot, "JSAmbisonics: A Web Audio library for interactive spatial sound processing on the web," *Interactive Audio Systems Symposium, York, UK*, 2016.
[20] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *116th Conv. Audio Eng. Soc.*, 2004.
[21] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
[22] F. Zotter, M. Frank, and A. Sontacchi, "The virtual t-design ambisonics-rig using vbap," in *1st EAA Euroregio Ljubljana*, 2010.
[23] M. Romanov, P. Berghold, D. Rudrich, M. Zaunschirm, M. Frank, and F. Zotter, "Implementation and evaluation of a low-cost head-tracker for binaural synthesis," in *Paper 9689, 142nd Conv. Audio Eng. Soc.*, 2017.

All Around Audio Symposium

Ultrasonic Communication: Risks and Chances of a Novel Technology

Matthias Zeppelzauer, St. Pölten UAS, AT

The ultrasonic frequency band represents a novel and so far hardly used channel for the communication of different devices, such as mobile phones, computers, TVs, and personal assistants like Google Chromecast. Ultrasonic communication is a promising technology since it requires only a standard loudspeaker and a microphone (as built into our phones) for communication. While offering a number of opportunities for innovative services (e.g. in the domain of Internet of Things), the technology, however, also bears a number risks. Companies like Silverpush employ ultrasonic data exchange to track users across devices and to collect information about their behavior without their knowledge. In my talk I will present the novel technology of ultrasonic communication, show how it works and which risks and chances are linked to it. Additionally, I will present the project SoniControl which aims at the development of an ultrasonic firewall to protect the privacy of users as well as the project SoniTalk which aims at developing a safe and privacy-oriented protocol for ultrasonic communication.

Modular Synthesizer Ensemble

gammon, Vienna, AT

The Modular Synthesizer Ensemble performs with fixed instruments and variable orchestration. 12 Modular Synthesizers provide the starting point for this participatory music project, with the aim to present electronic music live as an ensemble. With the analog Modular Synthesizer the participants are able to shape the process of electronic sound formation by themselves, even with no previous knowledge. Proceeding with the originated sound material we will invent, prove, execute, improvise and compose. A simultaneous prizes of composing and executing electronic music is evolving. The aim of the Project is, to perform the musical result live as an ensemble. The Installation of the modular synthesizers at the hall will be supervised by Gammon and Jessica and Thomas from `http://schneidersladen.de`.
`http://www.gammon.at`

On Models and Pragmatic Features in Digital Musical Instruments

Cornelius Pöpel, Ansbach UAS, DE

The digitalization of objects, methods and working procedures is a big topic in our times. In the field of audio, digitalization has been done in many areas already. A core issue in digitalization is the development of models and their transformation by formalization. According to the model theorist Stachowiak (1973, p. 131) the term "model" can be understood to include three features: a) the feature of mapping, b) the feature of reduction, c) the pragmatic feature. In order to create a model the essentials of an object (which the model maps to) have to be figured out. The model only includes those essentials. Things that are not essential are left out. The question which properties of the object are essential is coupled to what Stachowiak calls "pragmatic feature". The model was created from a specific group of persons, in a specific time and for a specific reason. Given the precondition of the pragmatic-feature-settings in which the model was created and for which the model was of use and so to say valid, the question comes up what may happen to the model in case the pragmatic-feature-setting has changed. What does it mean for a model if persons, time and reason differ in comparison to when the model was created? One of the reasons musicians give for why they do unplugged music is that they want to get beck to the essentials when making music. Since the models implemented in synthesizers were always thought to cover the essentials of tones, sounds and playing an instrument it is questionable in how far the models used in the digitalization of audio really do cover the essentials those musicians are talking about. The digitalization in audio has brought a huge mass of new opportunities in working with sound. According to a seemingly loss of essentials in music it may be seen as a need to do research on qualities in sound that have been forgotten, unseen or lost. One question may be what the essentials are that have not been covered yet. A second question might be what factors play a role when it comes to models that do not cover the essentials needed by musicians. Another question can be in how far this loss plays a role for the younger generation of digital natives who may be more interested in the new opportunities of digital musical instruments than in a loss which does not play a bigger role for them. The paper will cover selected parts of the author's findings when doing research on the development and usage of models for musical purposes with a specific focus on the pragmatic feature. It will include as well results of a study on how digital natives did couple musical ideas with the difficulty of creating a digital musical instrument.

References

Stachowiak, H. (1973). Allgemeine Modelltheorie [A general theory of models] Vienna, Austria: Springer.

3D Audio: Sculpting with Sound - Report on an Artistic Research Project

Sabine Breitsameter, Darmstadt UAS/Soundscape- & Environmental Media Lab (SEM-Lab), DE

Although 3D audio is considered a novel way of producing, the aesthetic desire, and the capabilities for a three-dimensional positioning of sound in the 360 degree sphere can be traced back to the antiquity and even to the time before. Numerous 20th century composers tried to implement their 3D audio 'visions', but the full technological possibility to accomplish sonic plasticity has come up quite recently only by the availability of innovative 3D sound systems. However, applying these systems in a technically correct way does not automatically lead to convincing artistic results. So, what needs to be clarified and explored in order to create plausible artistic 3D audio productions? In this presentation we would like to give an overview on selected topics of our 3D audio artistic research at Darmstadt's SEM-Lab. It assumes that 3D audio needs distinct aesthetic concepts and criteria, in order to prove its necessity, beyond just providing hyped-up versions of already familiar artistic phenomena. Based on the rich cultural history of 3D sound creation, this presentation will point out major categories and main criteria which reflect the specifics of 3D audio. It will point out why the approach given by the concept of soundscape can be crucial. Trendy terms like immersion, tangibility, illusion, and virtuality are questioned and investigated in reference to overused aesthetics, naive realism and the lack of providing the position of critical distancing. We will suggest and point out that a huge artistic potential for specific 3D audio productions can lie in dramaturgical approaches like fragmentization, deconstruction, as well as in the careful conceptualization of auditory materials and their representational potential.

Acoustic holograms: Artistic approach to 3D-Audio

Natascha Rehberg, Darmstadt UAS/Soundscape- & Environmental Media Lab (SEM-Lab), DE

Emerging 3D-Audio technologies locate and treat sounds as three-dimensional, virtual sound sources with a certain position, dimension and shape - acoustic holograms, that provide an increasingly tangible experience (in particular referring to my experiences in working with the SpatialSound Wave System (SSW) by Fraunhofer IDMT Ilmenau, during an ongoing research project at Darmstadt UAS). The emancipation of sound from a speaker is altering the role of the listener, as well as the role of the listening: the frontal stage disappears and the auditory perception becomes an omnidirectional experience, in which the listener interacts with the acoustic environment. With the objective of expanding artistic means of expression by the use of such an Apparatus, this writing situates 3D-Audio within the conceptual framework of soundscape and hints at aspects of conceptualization and practical implementation.

Soundscape: concept of 3D-Thinking

Understanding 3D-Audio compositions as soundscapes has many implications for conceptualizing and composing. The term and concept of soundscape refers to the appearance of all sounds in a room, place or landscape within a 360 degree sphere - an acoustic envelop, shaped by all properties of the environment [1][2][4]. Based on the premise, that hearing is an environmental form of perception, it indicates a non-selective, omnidirectional method of listening, which is a prerequisite for comprehensive 3D-Audio composition [3]. Moreover, the associated terminology contextualizes sound in its interdependent relations, identifies functional categories for the elements of a soundscape, such as keynote sound, sound mark, signal sound and provides design-related criteria, that are helpful to (re-)evaluate proportions [3].

The artist as sound architect and choreographer The arrangement of virtual sound sources creates figures, structures and forms – an architecture of sound, in which artistic intentions are expressed through construction, deconstruction and transformation of spatial relations. Thus, perspective and proportions are crucial criteria and fundamental design issues. From a conceptual viewpoint, the perspective significantly determines, whether the listener literally is immersed or in a more distant position. The implementation assumes a material concept, which takes in account the object-based production principle: sound is not assigned to a certain speaker, but to a to an object, that is positioned via graphic interface or other, even interactive devices. Objects consist of audio signal and meta-data (room-coordinates and other data) [5]. Meta-data potentially can be delivered by any device or software (sensors or game engines). The container-format technically allows to define various parameter as meta-data, such as volume or sound effects. The audio material constitutes the microstructure (inner structure) of a virtual sound source, which may contain a single note or a whole soundscape. A collapsing tree thus can be implemented as one virtual sound source (e.g. a distant event) or as a complex figure: a spatial construct of several virtual sound sources (e.g. an immersive situation). As overlapping sounds in an audio file cannot be spatially separated, a distinguished spatial polyphony requires thoughtfully prepared audio material. The collapsing tree also hints at the expanded possibilities of artistic expression through motion of sounds. Beyond illustrative or narrative aspects, sonic motion performances can generate fascinating, unheard structures and forms - advanced features such as programmable motion patterns prospectively enable to animate sound like a choreographer and therefore intensify a media-specific aesthetic.

Conclusion

Through holographic spatialization of sound, polyphonic textures and figures acquire a sophisticated manifestation. Moreover, the object-based production is best suited for interactive settings. Consequently, 3D-Audio offers the potential to create new forms of sonic or multimedia art with more holistic notions of auditory experience [4]. The soundscape approach is a valuable tool to develop dramaturgical expressiveness of spatialization, that goes far beyond reproduction or naturalistic, simulative and illusionary representation. It's a whole field of artistic exploration to create appropriate, object-oriented implementation methods. It's up to us, the forward-listeners, to take 3D-Audio as a gift to create an artistic and social value.

References

[1] Sabine Breitsameter, Hörgestalt und Denkfigur – Zur Geschichte und Perspektive von R. Murray Schafers Die Ordnung der Klänge. An introductory essay in: R. Murray Schafer, Die Ordnung der Klänge. Eine Kulturgeschichte des Hörens (published and translated by Sabine Breitsameter), Mainz – Berlin: Schott International, 2010.

[2] Schafer, R. Murray, Voices of tyranny: temples of silence, Ontario, Canada: Arcana Editions, 1993.

[3] Sabine Breitsameter, Sculpting with Sound: 3D

Audio and Its Aesthetic Specificity - Selected Problems from an Artistic Research Project, in: Proceedings of Sweep - Symposium on Sound Research at the University of Kassel, May 18-19, 2016 Kassel, Germany, 2016.
[4] R. Murray Schafer, Die Ordnung der Klänge. Eine Kulturgeschichte des Hörens (published and translated by Sabine Breitsameter), Mainz – Berlin: Schott International, 2010.
[5] Jürgen Herre, Johannes Hilpert, Achim Kuntz, Jan Plogsties in: MPEG- H Audio-The New Standard for Universal Spatial / 3D Audio Coding, AES 137th Convention, October 2014.
[6] Milena Droumeva, Understanding immersive Audio: A historical and socio-cultural exploration of auditory displays, in: Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005.

Steps Toward an A/R/Tography of Sound

Hans Ulrich Werner / HUW, Offenburg UAS, DE

AllKlang

Qualitative scholarship, artistic research, and research-based learning bring together insights from practice and experience. In the autoethnography of own auditory workshops, and of the cultures of other studios, I evaluate the new interdiscipline of sound (studies) and extend it with ideas for practice and theory, from the (still unknown) a/r/tography to a future a/r/tophony: artistic research in music and through sound composition, radio art and visual music.

EinKlang

The waveform symbolizes artistic research as a complex resonance in music, sound composition, and radio art. The picture was taken by Dan Curticapean at the Technorama in Winterthur. Curticapean is a physicist with a passion for art, who does research in photonics and creates through photography. His image highlights the interweaving of the methods of perception: through sound itself, in the highly developed discipline of sound art, and as the core of our work in all media, including those as yet unknown to us.

VielKlang

In the medium, sound unfolds as both a material and a workflowmadeupofmatter, in timeandspace, fromsoundscapetosound design, from perception to form and effect. Early soundscape models and "acoustic communication" (Truax 2001) encounter the now global phenomenon of sound studies: from natural to technological sound, cultural to societal audio image, always characterized by mediality, mediation, mediology, mediamorphosis (Smudits 2002). As a system, such transformation is a whole, but also fractal, in the practice of many sound artists, researchers, and educators. The Canadian discipline of a/r/tography is an especially intensive explora- tion of the trio of a/rtist, r/esearcher, and t/eacher, with the transitions between them deliberately included (Werner 2015). It proceeds from the sound-generating person to his or her aural environment, from sonic moments to sonic spaces: from research to creation, analysis, synthesis, and experience (cf. Dewey 1934). The composer Murray Schafer – the "great ear of Canada," as Klaus Schöning called him – titled his 1977 artistic instruction manual The Tuning of the World; in Sabine Breitsameter's 2010 translation, this became Die Ordnung der Klänge (The ordering of sounds). The two complement each other, becoming a third thing. The triad of practice, forward-thinking inquiry (Krippendorff 2011), and education keeps us close to the protagonists and lends itself to autoethnography as well. There are also connections here to ideas on the activity of music (Stroh 1984), "reflection-in-action" (Schön 1983), and learning through research (Huber 1970). Similarly, the thematic emphasis on "Künstlerisches Forschen in der Musik" (Artistic research in music) at the University of Münster Conservatory) does not come across as a variation on the global theme of artistic research as new system (currently on the rise in Germany as elsewhere); the verb form (forschen) foregrounds doing. The focus is on aural activity; what matters is audio art and the auditory in a broader concept of music and sound, rather than a general account of a future system of art and research. We are exhorted to "follow the actors" (Latour 2005, 12) – a call embodied by the sociologist of music Howard S. Becker in his decades of practice as teacher, researcher, and jazz musician. This gives rise to transitions in which "the actions of the scientist begin to approach artistic action" (Hildebrand 1994, 13). With Germany's longest-running series of media-research publications, the University of Siegen's MuK (Medien und Kommunikation), I have been experimenting with methods of analysis and "microtheories" on ways of presenting, describing, and dealing with sound. As an active participant, I query practitioners who work as scholar/artist/ educators (few of whom identify themselves as such). Basic research, too, is increasingly coming into contact with the aural: see for example Max Ackermann's 2003 dissertation on the "culture of hearing," or the symposium on "audio media cultures" held at Siegen in 2010 (Volmar and Schröter 2013). This reflection of sound as theorem has the potential to affect the creative act and the creators themselves. In the triad of content and aesthetic, communication and organization, and technology and actor, it can be put into practice in any studio and any sound. In dialogues and diagonals, the materiality of sound meets its mediality: temporality plus mediation (Debray 1996), sonic space plus culture and mediamorphosis.

Situating Performance in the Performing of Situation: The Effect of Situational Context on Performer Expressivity

Hans-Peter Gasselseder, Maria Kallionpää, Aalborg University, DK

How to articulate what is believed to be the fundamental artistic idea, and more arguably, the representative character of a state of mind or situative quality that is ascribed to a music composition? Apart from actually applying the operating instructions of a score to an instrument, several aspects of acoustic scene, ergonomics, attention focus and mood need to be taken into account when adapting to the situative affordances of a particular piece of music. But what if a performer lacked intuition and expertise to adapt to these contextual variables? Or in other words, what if one lacked the ability to adapt the handling of an instrument in different contexts or under varying acoustic conditions? Interpreting the current situation and selecting an appropriate action in a real-time performance setting often proves to be a challenging task. This is even more the case when thoughts and actions require an extra step of mediation [the instrument]. In order to bypass this step towards non- mediated representations of control, extensive practising allows the building of mental models detailing interactional patterns that are implicitly activated by environmental cues. The detection of these cues may vary depending on a performers' awareness of situational context; a cognitive representation of how we relate to our surroundings and give purpose to actions. Thus, we expect situational context to affect mental models of performer-instrument interactions and expressivity. In order to test this hypothesis, we examined to what extent specific parameters of acoustic scenery alter a performers' rendition of contemporary piano works. Utilising a combination of binaural DSP microphone/earphone setup, we were able to present subjects with life-like, immersive acoustic sceneries decoupled from their visual appearance. Data gathered from audio- and MIDI recordings as well as focus interviews with seven professional pianists illustrate how alterations of spectral- dynamic features and room acoustics affect the performing under varying situational demands.

When More is More: How to Supersize Musical Expression

Maria Kallionpaa, Hans-Peter Gasselseder, Aalborg University, DK

"Super" or "hyper" instruments are sometimes mentioned within the discussions among musicians but both terms are used relatively flexibly. Whereas some composers and performers refer to them with regards to certain software (for example, the hyper score software by Machover), our research regards the "super instrument" as a piece-specific concept or phenomenon. Rather than referring to any particular instrumentation or technological solution, the super instrument comes to be defined as a bundle of more than one instrumental lines that achieve a coherent overall identity when generated in real time. On the basis of our own personal experience of performing the works discussed at this lecture concert, super instruments vary a great deal but each has a transformative effect on the identity and performance practice of the pianist. An increasing number of composers, performers, and computer programmers have thus become interested in different ways of "supersizing" acoustic instruments in order to open up previously-unheard instrumental sounds. This leads us to the question of what constitutes a super instrument and what challenges does it pose aesthetically and technically? We argue that the essence of the super instrument lies in the enhancement of the technical and expressive capabilities of the performer and composer, as well as in the better interaction between the performer, instrument, and live-electronic systems in a concert situation. Our presentation explores the effects that super instruments have on the identity of a given solo instrument, on the identity of a composition and on the experience of performing this kind of repertoire. The purpose of this lecture concert is to showcase the essence and role of piano or toy piano in a super instrument constellation, as well as the performer's role as a "super instrumentalist". We consider these issues in relation to case studies drawn from our own compositional work and a selection of works by other contemporary composers.

Breaking The (Imaginary) Wall between Performers and their Audience in Live Music

Oliver Hödl, University of Vienna, AT

Breaking The Wall is a research project at the intersection of art and technology in live music. Its goal is to explore how to use technology to involve the audience in live music performances, or metaphorically speaking, how to break the imaginary wall between performers and their audience in live concerts. The project is a collaboration between the Vienna University of Technology, the University of Applied Arts Vienna and the University of Music and Performing Arts. Throughout the project, the research team and the involved artists developed four performances. These four performances were showcased at the music-event Breaking The Wall in Vienna in June 2017 and two of them additionally at the Ars Electronica Festival 2017 in Linz. During their concerts, the three musicians Electric Indigo, null.head and Johannes Kretz call for participation in the interplay of artist, audience and technology. The artists played electronic, electro-acoustic and industrial music and the audience participated through robots, smartphones and laser tracking. The fourth performance was not music-based and provoked the audience to make them aware of surveillance aspects in technology-mediated audience participation. This talk presents the development process of the performances and the actual technologies used for the concerts. Furthermore, you learn about the results of the scientific evaluation and how to use the new knowledge in future projects around technology-mediated audience participation.

Line & Hemisphere – A Hybrid Studio Setup for Immersive Experiments in Spatial Audio and Music

Paul Modler, Hochschule für Gestaltung Karlsruhe, DE

The development of new audio reproduction systems are based on multichannel speaker setups to apply recent distribution techniques such as Higher Order Ambisonics (HOA), Vector Based Amplitude Panning (VBAP) or Wave Field Synthesis (WFS). The presented studio setup aims to combine audio projection approaches to provide a test bed for experiments in order to investigate new possibilities of increased immersive perception of spatial audio and music. For this a hemispheric speaker setup is extended with a horizontal linear speaker arrangement. The hemispheric setup is based on standard high quality active loudspeakers, whereas the WFS is based on multi speaker boxes combined with 16 channel audio amps developed as a low budget feasibility study. According to the number of channels the system can operate from one CPU or of two remote CPUs controlled through network sockets. The system is implemented in a standard class room with no or very basic acoustic treatment, to showcase achievability with restricted resources found in normal environments.

AudioAllAround: Immersive Audio – Evolution of Techniques and Tools

Martin Mayer, Diana Mayer, Mister Master, Klosterneuburg, AT

This short talk about our work in the last 20 years, wants to show the evolution of our techniques and tools by presenting individual projects. The spectrum ranges from early experiments with analog 4-channel technology, through large scale outdoor opera productions, to recordings and concerts in full 3D Audio. Today's technologies provide a level of realism being impossible until recently. This opens up new areas beyond obvious applications in music, theater, cinema, TV, museums and exhibitions, which were our main-fields of interest in the past and present. But immersive audio is now also increasingly gathering interest from areas such as recreation and health-care, with promising new approaches in therapies against dementia, tinnitus and different phobias in artificial but completely realistic 3D audio wave-fields. Our new ATMIX 3D Audio Lab has opened it's doors in early 2017 as a new space to experience, experiment and evaluate in a full-dome speaker setup using WFS and other immersive technologies and formats.

MED-EL Hearing Implants and the Science Center AUDIOVERSUM in Innsbruck

Eckhard Schulz, Ewald Thurner, MED-EL GmbH, Innsbruck, AT

The sense of hearing is besides the important aspect of human communication and interaction also a major channel to express our emotions. The sophisticated anatomy of the ear (outer, middle and inner ear structures) plays an important role in processing sound information. Hearing loss is caused by damage to one or multiple parts of the ear. The sense of hearing is the only human sense, which can be replaced and/or reproduced by means of technology. Hearing implants may enable relief for those affected by hearing loss. The Austrian company MED-EL with its headquarter in Innsbruck, dedicated the past 27 years of focused research to overcome the barrier of hearing loss by developing an innovative and wide-ranging product portfolio. The commitment of its founders, Ingeborg and Erwin Hochmair, in fostering a company culture of excellence, advanced MED-EL to the industry's technology leader in implantable hearing solutions for a variety of indications. The ScienceCenter AUDIOVERSUM, which opened in 2013, was initiated by MED-EL and aims to raise awareness for hearing loss among the society by giving a combination of medical, technical, educational and art exhibitions with regard to the sense of hearing. The AUDIOVERSUM is unique in Europe and fascinates its visitors with interesting facts about hearing and the accompanied senses. The learning objectives of the given presentation comprise the anatomy of the ear and the physiology of hearing, different types and degrees of hearing loss and how they can be treated with MED-EL hearing implants, as well as the interactive ScienceCenter AUDIOVERSUM in Innsbruck with its various exhibitions. The presentation will be held by Dr. Eckhard Schulz, former Managing Director of MED-EL Germany and founder of the AUDIOVERSUM and by DI Ewald Thurner, Area Manager of MED-EL Vienna.

Heart Sound – how sound and radio can help to improve the relationship between people with dementia and their carers

Christine Schön, Berlin, DE

Imagine the sound of happily screaming children, splashing water, a light breeze and chirring crickets. Can you feel the summer? Now imagine the sound of a stiff breeze blowing the icy branches and crunching steps in the deep snow. Can you feel the cold? Sound translates directly into an emotion. The ear is a very sensitive organ: hearing is the first sense we develop in the womb and from this point onwards our ears can never be closed again. Sounds are deeply rooted in us – when we listen to a familiar sound, it triggers an emotional memory. That's why sound is so suitable for people with dementia whose reactions to emotional stimuli are much stronger than to cognitive ones. Collective listening is a very familiar thing for today's elderly: in their youth, radio was the most common medium – families and friends got together to listen to entertainment programmes, sportscasts and concerts. Dementia – The current situation: 46.8 million people worldwide live with dementia; due to estimations it will be 131.5 million in 2050 (statistic presented at the Alzheimer Europe Conference 2017). In Germany, there are 1.5 million people living with dementia. The German Alzheimer Foundation estimates that this number will have reduplicated itself by 2050. Every year, 300.000 people are diagnosed with dementia. This effects their friends, family and carers, too. People with dementia have the right to live a fulfilling life with their impairments and to play an active role in society. To ensure this, there need to be offers specially tailored to their needs and capabilities. It can be very difficult for carers and relatives to get into emotional contact with people with dementia who often live in their own world. Motivated by this challenging situation we developed Hörzeit – Radio wie früher ("Listening Time – radio like in the old days") and Herzton ("Heart sound"). The major concern of our sound projects is to strengthen the relationship between people with dementia and their carers.

How do "Heart Sound" and "Listening Time" work?

Hörzeit – Radio wie früher is a worldwide unique radio programme especially designed for people with dementia. It is produced in the style of the 1950s radio entertainment shows. Each programme focuses on a different subject, such as children, travel and professions – timeless topics to delve into a conversation with people with dementia. Christine Schön and Frank Kaspar lead the listeners through the programme; they speak about their personal experiences – about their children, their most impressive journeys or their dream jobs. They present sound collages, feature reports, famous pieces of music, proverbs and rhymes. The presentation converts the communication techniques of validation: it applies a genuine and deep appreciation of people with dementia, takes them seriously with their feelings and emotional states and doesn't give too much information. Helga Rohra, a person with dementia, has a regular column in every issue. The programme for people with dementia is around 50 minutes long; a following programme for relatives and caregivers is around 20 minutes long. In this second programme Schön and Kaspar review books, films and games, present institutions and interview experts (`http://www.hoerzeit-radio-wie-frueher.de`). The non-profit sound-based web portal Herzton ("heart sound") helps to activate people with dementia individually using all acoustic means: e. g. self-sung songs, dialects, interviews with contemporary witnesses and easy accessible soundscapes. They are recorded and produced by sensitive journalists, sound artists and musicians. Relatives and caregivers can select pieces individually for the people with dementia entrusted to their care: for example people who grew up in the countryside may enjoy the sounds of a farm; for someone from Bavaria, it might be a pleasure to listen to a story told in the Bavarian dialect. Herzton will be launched in late December 2017 (`http://www.herzton.org`).

Philology of electronic music - New methods, strategies, falsifications and historic cleansing: Stockhausen, Xenakis, KRAFTWERK

Reinhold Friedl, Goldsmith University, London, GB

It is astonishing the classical philological methods have not been adapted for electronic music so far. This lecture will discuss how this can be made and that applying this new methods astonishing results can be achieved. This will be shown on three prominent musical examples: Karlheinz Stockhausens "Konkrete Etude" that is actually not his piece, Iannis Xenakis hiding of the real sound sources in his multitrack compositions and KRAFTWERKs historical cleansing of their body of work. This research is part of a PHD project at Goldsmiths University London.

<<cresc >>Worte werden Raum

Eva Paulitsch, Coburg University of Applied Sciences And Arts, DE

"Fast könnte man sagen, dass vom Tempo, der Geduld und Ausdauer des Verweilens beim Einzelnen, Wahrheit selber abhängt" (Theodor W. Adorno)

"One could almost say that truth itself is dependent on an individual's tempo, patience and endurance in lingering." (Theodor W. Adorno)

From 2006 to 2016, the artist duo Eva Paulitsch und Uta Weyrich collected mobile phone videos shot in public space by teenagers and young adults; with this material, they created a video archive that is unique in the world – the Mobile Video Archive. The skewed and unexpected fragments of the world from young adults' perspective – which simultaneously open up space for associations and belie a fascination for moving images – were Paulitsch and Weyrich's motives for speaking to young people in public places about their videos of daily life. The artists asked for the videos as a gift and began creating an archive with them. Their interest was in the mobile phone videos that were "resting" on the smartphones' memory cards, and not the consciously staged videos made for YouTube, for example. Their collection campaign saved the videos from being deleted and declared them to be basic artistic material. The artistic transformations usually took place in cooperation with experts from the fields of music, computers or the theater. In walk-through video installations, they created spaces that visitors could approach from many different perspectives. In contrast to the video installations, in # CRESCENDO, it is not the moving images that are in the focus, but rather the respective audio tracks of specific videos. In this work, the artists only explore the mobile phone videos' audio tracks. They had the original sound of all the videos in their collection transcribed and thus expanded their "no-story video" archive to include a "no-story audio" collection. In the transcription process, it is possible to represent the spoken language as well as the context of the speaking situation beyond the content of what is said. Abbreviations, punctuation marks and special characters frame individual words. Meaning is only constructed when reading the text: these are acts of speech, dialogues – teenage slang. The translation from sound to a manuscript has its own power, which already exists in the texts' unusual codification. In the materialization in script, language itself becomes an image – the ephemeral, often incomprehensible but perceivable sounds are paused; new spaces and new meanings develop. By decoupling the soundtrack from the film level, the fragmentary dialogues become singular and achieve an autonomous reality.

DaVinci Head project: The best price/performance binaural head

Vytenis Gadliauskas, LT

There are a lot of binaural microphones for consumers in the market. Some of them are dedicated to professionals, others look like they were created as a toy. Though all them use slightly different approach to record Interaural Level Differences (ILD's) and Interaural Time Differences (ITD's), the goal is the same – immersive spatial audio experience to the end user. Binaural head is one of the most accurate but also pricey approach. DaVinci Head project started as home build binaural head with no intention to go worldwide. It was the final prototype, test results and creator's motivation, that later set the goal of the project – DaVinci Head have to be the best price/performance binaural head in the market.