http://first-stage.eu

Project Number	**688244**
Project Acronym	**first.stage**
Project Title	**Fast and Easy Previsualisation for Creative Industries**
Deliverable Number	D1.6
Deliverable Name	NUI Concept and Interaction Guideline for Previs Software
Type	Report
Status & Version	Version 2.0.3
Work Package Contributing to the Deliverable	WP1
Work Package / Task Responsible	Next Limit SL / University of Bremen
Other Contributors	-
Authors	Thomas Fröhlich, Thomas Muender, Dr. Dirk Wenig, Dr. Tanja Döring, Prof. Dr. Rainer Malaka
Keywords	Interaction concept, interaction guidelines, natural user interfaces (NUI), direct manipulation, previs

The research leading to these results has received funding from the European Union's H2020 Programme for research, technological development and demonstration under grant agreement n° 688244

Executive Summary

Aim of the Deliverable. *This deliverable contains the interaction concept for the prototypes of the first.stage project. This document will present a guideline for implementation of the interaction for all prototypes of the project.*

Brief Description of the Sections of the Document. *In the first section of this document, an introduction is given, followed by a detailed related work section. Reviewing the previous requirement analysis, we provide personas in the following section. The description of the natural user interface concept with interaction techniques, interaction principles, technologies and devices, guidelines and design methods represent the main part of this document. Concept implementations and scenarios conclude this document.*

Mayor Achievements. *An interaction concept utilizing natural interaction. A guideline for the implementation of the interaction is presented. Milestone MS3 ("Natural User Interface concept and guideline finalised") is accomplished with this document.*

Summary of the Conclusions Obtained. *In this document, we provide a natural user interface concept that is built around a set of interaction methods which are chosen specific to the nature of the diverse previs tasks. The concept is based on the requirements of the application areas and incorporates a detailed review of the literature.*

Content

1 Introduction 4

2 Related Work 7

2.1 Natural User Interfaces 7

2.2 Direct Manipulation 9

2.3 Reality-based Interaction 9

2.4 Natural User Interfaces for Performance Animation and Modelling 10

2.5 Spatially-aware Displays 12

2.6 Mobile Augmented Reality 12

2.7 Gesture Interfaces and Free-Hand Interaction 12

3 Personas 14

4 Natural User Interface Concept 20

4.1 Interaction Techniques 21

4.2 Interaction Principles 24

4.3 Technologies and Devices 27

4.4 Previs-Specific NUI Guidelines 30

4.5 Design Method 31

5 Concept Implementations 33

5.1 Main Prototype 33

5.2 Research Prototypes 34

6 Application Scenarios 39

6.1 Theatre Scenario 39

6.2 Film Production 40

6.3 Animation 41

6.4 Summary 42

7 References 43

8 Document History 47

1 Introduction

This document presents the general natural interaction concept for the first.stage previs toolset. The main goals of first.stage are intuitive and natural user interfaces (NUI) for practitioners that take into account their innate creative capabilities. First.stage empowers users in how they can express themselves in an optimal way depending on the task they want to perform in the previs cycle. This will lower the need and effort to translate user intents and actions through the interface but rather allows for more direct natural expressiveness.

In their current work, creatives are used to adapt their tools, workflows, and methods of expression depending on what they want to accomplish and feels the most familiar for them. For example, when creating a storyboard, drawing and sketching are preferred ways of expression in which practitioners can express themselves most efficiently, or when expressing in spatio-temporal media, video and sound editing are used to transport a vision of the project in progress.

For previs, this creative process often incorporates different tasks on various media such as 2D and 3D layout and animation. All of these media require expressive tools. The ideal previs user interface should support creative freedom and flexibility, so that users experience the previs tool as a natural extension of their own physical capabilities and can express their ideas in a seamless and intuitive way. However, no single interface can offer natural expression, precision, and task orientation at the same time. Further, due to the task diversity in previs, a natural user interface should consider a well-balanced workflow: moving between tasks quickly and without confusion. The first.stage user base also comprises of a broad range of users in different roles who have different workflows and stories: what is wanted for animation is not necessarily what is wanted for theatre.

We argue that users should be able to choose which way of interaction they want to use for different previs tasks, selecting the one that fits their creative needs the best. In this way, we make sure that artists can be creative and expressive and in turn feel natural in their work without forcing them to use an inflexible interface that doesn't account for what artists want to achieve.

Constructing a natural interaction concept, we centre all efforts around the needs of the users. For this, we are building upon the results of the deliverables D1.1 to D1.4 (first.stage D1.1 Authors, 2018; first.stage D1.2 Authors, 2018; first.stage D1.3 Authors, 2018; first.stage D1.4 Authors, 2018), where we identified user requirements for previs in the domains film, animation, theatre, and visual effects, and defined the core functionalities for previs in the deliverable D1.5 (first.stage D1.5 Authors, 2018). Based on these results, we create the natural interaction concept by mapping both users' needs and the previs task characteristics to meaningful and natural interaction techniques. Depending on what users want to accomplish, we integrate a range of interaction technologies into our concept, that allow for natural expression and interaction: direct manipulation via touch on 2D interfaces, spatially aware displays, tangible interaction, augmented reality, direct manipulation in 3D and via gestures, full body and embodied interaction, free-hand interaction, and speech. We further present which hardware is meeting the requirements for implementing our NUI concept, based on usability, availability, and costs.

As first.stage cannot cover all possible tasks for all previs scenarios, we will focus on the previs tasks identified in D1.5 that represent a core set of most important actions in many previs scenarios. Our research and development will focus on designing a core interaction concept for this set of core functionalities.

As described in D1.5, we identified 20 core functionalities as crucial for a previs toolset. However, not all of the core functionalities can be directly translated into previs tasks, because some of them relate to general functions like multi-user capability or seamless changes between devices. Therefore, we reduce the set of 20 core functionalities to crucial previs tasks that apply to all application areas, resulting in the following list of 9 core previs tasks:

Project Structure	Organisation of previs work in projects, including all resources like 3D models or textures, that comprise of scenes (scene graph) and shots (2D views on a scene).
Import/Export	For the integration of previs into production pipelines, different input and output options in the form of file format support are offered. For example, OBJ, FBX, or STL files are supported at both ends.
Shot Management	Creation and overview of shots on a 3D scene.
Sketching (Modelling)	Creation and modelling of 2D and 3D objects for outlining and dressing scenes.
Assets and Layout	Creation of 3D scenes as virtual worlds, sets, or places using assets, models, animations, effects, etc. Import, selection, and interaction with pre-made 3D objects and animations from a database.
Camera Control	Adding and interaction with virtual cameras in a scene for shot creation with different camera parameters, lenses, and camera path animations.
Visual Effects (VFX)	Create, edit, and arrange visual effects in a scene by selection from a database or by direct gesture control to the simulation.
Lighting	Adding and interaction with virtual light sources, creating different moods and scene styles.
Posing and Animation	Creation and application of animations onto characters and objects using prerecorded animations, physics animation, and motion capture. Posing on rigged characters and pose selection from library.

On the basis of these tasks we formulate and present our core interaction techniques that account for the individual task characteristics and allow creatives to express themselves. It should be noted, that these tasks show quite distinct features in the nature of their media types. While some of them are inherently one- or two-dimensional (import/export, project structure, shot management) structuring some flat data content, others are by nature three-dimensional such as 3D scene modelling, layout, camera, or lighting. This is also reflected in the non-digital counterparts on the respective activities like organizing data on a desk or set design in an atelier. Thus, first.stage needs to reflect the inherent nature of these tasks in affordable user interfaces and has to select appropriate technologies and interaction media for both 2D and 3D supporting the requirements of creative people.

The workflow-relevant tasks (Assets and Layout, Camera Control, Visual Effects, Lighting, Posing and Animation) will later also build the foundation for the pipeline integration concept in deliverable D5.1 (first.stage D5.1 Authors, 2018).

In the following sections, we first introduce and discuss related work and the state-of-the-art in relation to our work (Section 2). We will then present a set of descriptive personas that will guide our human-centred design process. The personas presented in Section 3 help to illustrate and personalize design decision and be used in explanatory scenarios. Our main concept is introduced in Section 4, where we develop the concept on basic interaction techniques for the described core previs tasks. Section 5 will present an overview of currently implemented previs tools that are currently evaluated. Finally, in Section 6 we conclude the deliverable with a set of three application scenarios that exemplify the envisioned use of our previs tools in practise.

2 Related Work

The term natural user interface (NUI) marks a dominating trend in current HCI research, and it is as well used in popular media. In general, it addresses the idea, that designed interaction styles integrate users' skills and experiences from real life to a much higher degree than command line interfaces or graphical user interfaces do[1]. Ideally, this generates user interfaces that can be easily understood and learned. The term started to become popular in 2007, when the first Microsoft Surface was announced, and the first Apple iPhone came to the market. A year before, in 2006, the Natural User Interface Group (NUI group), an open source initiative media community working on natural user interfaces had been founded[2]. In the beginning, the term NUI was mainly used for (multi-)touch gestural interaction. Further popular consumer products that have been mentioned a lot in the context of natural user interfaces are the Wii and the Kinect for Microsoft Xbox, both vision-based devices that indirectly or directly support full-body movements for system input (Francese, Passero, & Tortora, 2012). Overall, the term *natural user interface* is used in different ways and there does not exist one dominating definition yet. In the following, some of these views are presented.

2.1 Natural User Interfaces[3]

Until today, the term natural user interface lacks a common definition, because the understanding of the term is subject to the constant devolvement of the technology (Fu, Landay, Nebeling, Xu, & Zhao, 2018). However, in the literature, there are recurring aspects of NUIs that appear to establish a shared view on natural user interfaces. In this regard, NUIs have been defined as being only controlled by using the human body (García-Peñalvo & Moreno, 2017), as interfaces that make use of natural human actions and that do not require learning as in classical control (Zielke et al., 2017). Liu describes the characteristics of NUIs as being user-centred, multi-channel, inexact, high bandwidth, voice-based, image based, and behaviour based (W. Liu, 2010).

In their book "Brave NUI World" Wigdor and Wixon (2011) argue that not the interface itself is meant to be natural. Rather, they "see natural as referring to the way users interact with and feel about the product, or more precisely, what they do and how they feel while they are using it" (Wigdor & Wixon, 2011). Thus, the "natural" in natural user interface does not refer to the interface, but rather addresses the way users should behave and feel when using the UI (cf., Wigdor & Wixon, 2011, p. 11). Thus, NUIs should not mimic the real world but rather create experiences that, for expert users, can feel like an extension of their body (Wigdor & Wixon, 2011, p. 13). In order to illustrate this, Wigdor and Wixon give the example of the first Apple pad, the Newton Message Pad, which did not succeed. Among the problems it had was the handwriting recognition, that was aimed to recognizing real handwriting, which did not work. An opposite example is the Palm Pilot that came out 1997 and used a special handwriting language called "Graffiti". Graffiti is similar to single character handwriting but simplifies the symbols – this made the recognition easier but at the same time required some learning by the users. Nevertheless, for Wigdor and Wixon this is an example for a successful natural input technique that does works, because it does not mimic reality, but still feels natural to the users. In their book,

[1] Cf., "A natural user interface is a user interface designed to reuse existing skills for interacting appropriately with content." (Blake, 2012).
[2] http://nuigroup.com/log/about/
[3] This passage has been adapted from the PhD Thesis of Tanja Döring (Döring, 2016)

Wigdor and Wixon describe a number of further aspects they regard as crucial for natural user interfaces, among which are that NUIs should be enjoyable, leading to skilled practice and be appropriate to context (Wigdor & Wixon, 2011, p. 29).[4]

In most cases the term natural user interface is used for gestural interaction, both touch or mid-air interaction, pen interaction and sometimes also for speech input. Generally, it could encompass all possible interaction modalities, for input as well as for output (but so far, the focus was more on input) (Jain, Lund, & Wixon, 2011). As such, natural user interfaces are related to the research field of multimodal interfaces. Oviatt describes multimodal systems as follows:

> *"Multimodal systems process two or more combined user input modes – such as speech, pen, touch, manual gestures, gaze, and head and body movements – in a coordinated manner with multimedia system output. [...] This new class of interfaces aims to recognize naturally occurring forms of human language and behaviour, which incorporate at least one recognition-based technology (e.g. speech, pen, vision)." (Oviatt, 2012, p. 405)*

NUIs are often described to be "intuitive". This term also is strongly related to users' previous knowledge and the idea, that the focus of the user is on the content and not on the user interface. For example, the definition by Naumann et al.: "*A technical system is, in the context of a certain task, intuitively usable while the particular user is able to interact effectively, not-consciously using previous knowledge*." Thus, "*[...] intuitive use can only be attributed to the human-machine interaction in a certain context [...], but not to a technical system per se*". (Naumann et al., 2007).

In order to let an interaction be regarded as natural and to keep the learning overhead small, the design of appropriate mappings of commands to interactions, e.g. through gestures, is important. This is a central challenge in current user interface design and research. As Norman (D. A. Norman, 2010) argues, design principles as easy memorability, user feedback or consistent conceptual models are as important for natural user interfaces as for GUIs. The design of NUIs is challenging as "natural mappings" for commands in most cases do not already exist, they have to be learned by the user and established by the UI designers and technology producers. What feels natural, depends not only on the task, but also on the specific culture a user comes from, etc. Thus, Norman argues, that "natural user interfaces are not natural" and the term in this sense is misleading (D. A. Norman, 2010).

Nevertheless, novel technologies always need time to establish, and it is one of the current major challenges in user interface design to design interaction styles that feel as natural as possible – keeping in mind, that interaction techniques always have to be learned to some degree. The NUI approach dominates HCI research and can be found also in specific research fields such as 3D user interfaces (Bowman, McMahan, & Ragan, 2012). We are aware that the term NUI is a controversial one and that there currently is no common definition. Most recently, at the ACM CHI Conference on Human Factors in Computing Systems 2018, a special interest group (SIG) discussed the topic of "Redefining Natural User Interface" (Fu et al., 2018). The result of the SIG and the panel discussion was that there is currently no reliable definition of what a natural user interface is. Nevertheless, NUI has been established

[4] More recently, Preim and Dachselt published the following definition for natural user interface (in German): "Eine natürliche Benutzungsschnittstelle ist ein System zur Mensch-Computer-Interaktion, mit dem Benutzer mittels intuitiver und zumeist direkter Bedienhandlungen interagieren, die einen klaren Bezug zu natürlichem, realweltlichem menschlichen Alltagsverhalten aufweisen. Natürlich heißt dabei nicht angeboren, sondern bezieht sich auf dem Benutzer durch den Alltag vertraute und erlernte Handlungen bzw. auf solche Handlungen, die Benutzern im Moment der Interaktion als angemessen erscheinen." (Preim & Dachselt, 2015, p. 472).

as a pragmatic set of interface qualities and is widely used. In first.stage, we use this pragmatic approach towards NUI and understand NUI as interfaces that are very intuitive to use and learn, make use of familiar interaction patterns that integrate human behaviour and expressive capabilities.

2.2 Direct Manipulation

A well-established interaction concept that promotes intuitive control and easy to use and learn attributes is the direct manipulation (DM) concept (Shneiderman, 1982). It involves continuous representation of objects of interest and rapid, reversible, and incremental actions, as well as feedback. The intention of direct manipulation is to allow a user to manipulate objects presented to them, using actions that correspond loosely to manipulation of physical objects. No intermediate interface is needed to interact with an object of the domain space and actions have an immediate effect on the state of the object which is directly visible to the user.

DM systems implement continuous object representation, physical actions instead of complex syntax, rapid incremental reversible operations with immediate feedback on object impact, learnability, rapid task execution by experts, retainable operational concepts, no need for error messages, and immediate perception of goal achievement. Some limitations of the approach were identified by Frohlich (1993), which are also relevant for the application in previs software: manipulating small, distant, or attribute-rich objects under limited space, high density, or high precision; manipulating multiple objects simultaneously as a group (including group attributes) and manipulating intangible object properties.

DM interactions can be built for 3D and 2D interfaces. However, in many cases 3D content has to be presented on a 2D medium such as a tablet or interactive surface. In such scenarios, it is not straightforward to design interaction techniques for 3D content manipulation on 2D surfaces. Thus, it is necessary to investigate how 3D content manipulation can be interacted with on 2D interfaces. In first.stage, this is important as we provide access to 3D media on different devices such as tablet computers. For overall interaction with 3D content such as rotation, translation, and other tasks, several approaches and interaction techniques can be found in the literature (Herrlich, Walther-Franks, & Malaka, 2011; Herrlich, Walther-Franks, Schröder-Kroll, Holthusen, & Malaka, 2011; Reisman, Davidson, & Han, 2009).

2.3 Reality-based Interaction[5]

A concept that is closely related to natural user interfaces is Reality-based Interaction. With his framework, Jacob et al. (2008) addressed the lack of terms for the evolving class of “post-wimp” user interfaces by introducing the reality-based interaction framework that allows to discuss aspects of the multitude of current user interfaces “beyond the desktop” in a structured way by identifying and analysing aspects of reality and computational power that are useful for interaction. The authors’ assumption is that “*basing interaction on pre-existing real world knowledge and skills may reduce the mental effort required to operate a system because users already possess the skills needed*” (Jacob et al., 2008, p. 204). Jacob and colleagues argue that interaction designers should always start from the knowledge and skills humans have gathered in the real world. This potentially helps to reduce the gulf of execution (Hutchins, Hollan, & Norman, 1985). The authors suggest four themes of experiences and skills derived

[5] This passage has been adapted from the PhD Thesis of Tanja Döring (Döring, 2016)

from reality: (1.) *naive physics*, (2.) *body awareness and skills*, (3.) *environment awareness and skills*, and (4.) *social awareness and skills*.

Naive physics includes all the knowledge and experiences humans have collected about physical laws, like gravity, friction, mass, stiffness, weight etc. The theme body awareness and skills comprise the knowledge about the own body: how it can be moved to achieve a specific movement or how a certain movement feels. Environment awareness and skills contains human knowledge and skills regarding orienting and navigating in the environment, e.g., humans are able to use landmarks to orient themselves. The fourth topic, social awareness and skills, addresses the knowledge and skills humans have in relation to other humans, e.g., different ways of talking to others depending on the situation, different distances that are appropriate or not in specific situations, etc. These reality-based themes can play a role within interaction design. To give some examples, e.g., friction as a principle based on gravity has been metaphorically applied to graphically presented data on touch user interfaces. I.e., when the user slides in the address book on an android phone for example, the sliding movement slows down, to make the movement experience more realistic (naive physics). Knowledge and skills about how to move the own body, how to use kinaesthetic and proprioceptive feedback, is important for full-body or hand movement interaction, e.g., realised in Kinect Games (body awareness and skills). The human capabilities to orientate and navigate in the environment are very important design factors for map and navigation applications (environmental awareness and skills). And, last, social awareness and skills play a role in social networks or in the virtual world of Second Life for example, where the users' communication still is based on their learned social skills.

Although most of the novel user interfaces are based on humans' experiences from the real world to a stronger degree than it used to be, computer tools and applications would not be as powerful, if systems were based purely on reality-based themes. Computers have many advantages, called "computational power" by Jacob et al., among which are *Expressive Power, Efficiency, Versatility, Ergonomics, Accessibility and Practicality*. E.g., when searching for a document where the title is known, it is much more efficient to search in digital documents by typing in search terms than browsing through a pile of physical documents (which would be the more reality-based way to do it). Thus, one of the central challenges of current user interface design lies in the combinations of and trade-offs between computational power and reality. Jacob et al. "*propose that the goal is to give up reality only explicitly and only in return for other desired qualities*" (Jacob et al., 2008, p. 205).

2.4 Natural User Interfaces for Performance Animation and Modelling

Animating digital content in a believable and interesting way can be a complex and tedious undertaking. However, animation is crucial for previs and interfaces that provide a natural access to animation are highly needed. We introduce some relevant and fundamental works in the area of performance animation that successfully implement natural user interface controls and were mainly developed at the University of Bremen, building the foundation of the current research and development efforts in first.stage.

Implementations of full-body input as a natural interface to character animation are presented by Lee, Chai, Reitsma, Hodgins, & Pollard (2002) by extracting user silhouettes from camera images and Chai & Hodgins (2005) who implement a more complex setup based on optical marker tracking. An affordable approach to performance animation is presented by Walther-Franks et al. (2012) with the "Animation Loop Station", allowing users to create character animations layer by layer by capturing users' movement with Kinect sensors. For system interaction, a speech interface is included for system interaction so that users can fluently work on their animation without the need for a graphical interface. In another work by Walther-Franks et al. (2012). The Dragimation technique allows users to control timing in performance-based animation on 2D touch interfaces where they can directly interact on the characters instead on a timeline. In a comparative study, the authors found that Dragimation performs better with regard to learnability, ease of use, mental load, and overall preference compared to timeline scrubbing and a sketch-based approach. Interview results support their findings as professionals "*could well imagine benefits from using performance timing tools*" (Walther-Franks, Herrlich, et al., 2012) in their workflow. This system is inspired by the work of Moscovich et al. who introduced a rigid body deformation algorithm for multi-touch character animation (Moscovich, Igarashi, Rekimoto, Fukuchi, & Hughes, 2005).

An interesting approach to augment the own character animations with rich secondary animations is the combination of performance animation with physics simulation. However, the combination of both technologies is not trivial. An approach is presented in a natural user interface that combines motion capture using the Kinect sensor and physics simulation for character animation by Liu and Zordan (2011). They provide a framework for the combination of the two technologies in order to overcome the "*[…] potentially conflicting inputs from the user's movements and physics engine.*" (C. K. Liu & Zordan, 2011). Their approach and framework has been extended by Shum and Ho (2012) who present a more flexible solution to the problem of combining physical and motion capture information.

Another area of complex 3D interaction is the field of digital modelling and sculpting. There is a strong need for tools that allow for natural expression in digital content creation, especially for previs, as many productions start with zero assets. This means, that assets, objects, and props have to be created from scratch most of the times. Here, we take into account relevant literature and systems that investigated natural interaction aspect for these tasks. For example, Herrlich et al. investigated (Herrlich, Braun, & Malaka, 2012; Herrlich, Krause, Schwarten, Teichert, & Walther-Franks, 2008) interface metaphors for 3D modelling and virtual sculpting, further contributing an interactive table that supports users in creating virtual 3D models on 2D interfaces (Herrlich, 2013).

Regarding virtual sculpting, a first implementation is presented by Galyean and Hughes (1991). In their system, a custom force-feedback system with nine buttons is used for interaction, translating the absolute positions of the input device into a 3D mesh. This approach has been picked and extended by Chen and Sun (2002) and Galoppo et al. (2007) by implementing virtual sculpting using a stylus device and a polygonal mesh instead of a voxel approach. Wesson and Wilkinson implement a more natural approach by using a Kinect sensor for deformation of a virtual mesh, while also integrating speech commands for a more fluent user experience. Natural animation can also be approached by using VR technology. For example, Vogel et al. designed a VR system for animation where users work with a puppeteering metaphor for character animation. They evaluated their tools with animation experts and found that it improves the speed of the workflow and fast idea implementation (Vogel, Lubos, & Steinicke, 2018).

2.5 Spatially-aware Displays

Today's mobile devices come with a variety of sensors (e.g. accelerometers, gyroscopes and magnetometers) which can be used to determine the device's position and orientation. Mobile interfaces based on this information are called spatially-aware displays (Fitzmaurice, 1993) . Spatially-aware displays implement an eye-in-hand metaphor and act as windows onto the virtual information space. They respond to the user's movements and *"[...] serve as a bridge or porthole between computer-synthesized information spaces and physical objects*" (Fitzmaurice, 1993). Peephole interfaces (Yee, 2003) augment the space around the user with information. They "*[...] fall into the category of spatially aware displays, which differ [...] in that they create a positional mapping between the virtual space and the real world, enabling the use of spatial memory for navigation"* (Yee, 2003). More recently, practical implementations combine different use cases of spatially-aware displays. For example, "Huddlelamp" (Rädle, Jetter, Marquardt, Reiterer, & Rogers, 2014) is a system that tacks multiple devices and objects on a table, including tablets that display information based on their own location. Here, devices can be added and removed from the system at runtime in order to reconfigure the collaborative space. Spatially-aware displays represent a useful and natural interaction technique because users can navigate with devices in the virtual world by referencing the real world.

2.6 Mobile Augmented Reality

Mobile Augmented Reality (AR) systems integrate virtual information into a users' physical environment, thus providing people with information in the context of the real world (Höllerer & Feiner, 2004). Early systems explored the use of mobile AR for the visualisation of historical data (Kretschmer et al., 2001) and digital story telling (Malaka, Schneider, & Kretschmer, 2004). Possible applications integrate the approach for example in outdoor pedestrian navigation systems on smart watches (Wenig, Schöning, Olwal, Oben, & Malaka, 2017), indoor navigation (Mulloni, Seichter, & Schmalstieg, 2011), mobile games (Picklum et al., 2012), or environmental monitoring (Veas, Grasset, Ferencik, Grünewald, & Schmalstieg, 2013). Recently, Augmented Reality technology became widely available at a consumer level with the introduction of Apple ARKit in iOS and Google ARCore in Android Devices, making a widespread adoption of the technology possible. Mobile AR is an interesting interaction technology for first.stage because it can be used to bring together the virtual and physical world for multiple users if implemented on larger tablets and touchscreen devices.

2.7 Gesture Interfaces[6] and Free-Hand Interaction

Gestural Interaction is among the most popular trends within HCI and UI design for products nowadays. The term gestural interaction comprises different kinds of interaction forms that all realize gestures, but in varying ways and settings. For example, gestural interaction can be performed on interactive surfaces as done in single or multi-touch interaction. These gestural interactions on surfaces can either be performed with fingers but could also involve further objects as pens for example. Another approach to use gestures for interaction is to design mid-air gestures, this again could include arm and hand movement.

A well-known example for early gestural interaction is Bolt's "Put-that -there" interaction technique, developed in MITs Architecture Machine Group, where pointing gestures where combined with speech

[6] This passage has been adapted from the PhD Thesis of Tanja Döring (Döring, 2016)

interaction in order to interact with a wall display (Bolt, 1980). Further, multi-touch gestural interaction, although widespread only since Jeff Han's 2005 presentation (Han, 2005) as well as Apples first iPhone and the first Microsoft Surface table in 2007, had been applied in a large number of research prototypes for many years; for an overview see (Buxton & others, 2007).

Among the advantages of gestural interaction is that bimanual interaction can be realised within human-computer interaction; especially in the context of multi-touch interaction (Wagner, Huot, & Mackay, 2012).

Guiard (1987) studied bimanual interaction and found "an asymmetric division of labor": in many tasks, one hand sets the spatial referential frame for the other, e.g. we first fix the orientation of a piece of paper on a tabletop surface with the non-dominant hand and then write with the dominant hand. Guiard modelled hands as abstract motors that "correspond(s) to a temporal-spatial scale" (Guiard, 1987). Guiard's insights do not only mean that the two hands perform different tasks that relate to each other but also that the kind and frequencies of movements typically differ (see also discussion in Hummels, 2000). This should be considered when designing for bimanual interaction. Domains where Guiard's model has been applied to multitouch applications are, for example, 3D-modeling (Herrlich, 2013) and computer animation (Walther-Franks & Malaka, 2014).

Free-hand interaction or gestures-in-the-air interaction are interaction techniques that require no physical medium in order to be executed. For this, the hands and fingers are tracked by devices such as the Microsoft Kinect[7] or Leap Motion[8]. Free-hand interaction is especially interesting in combination with VR, where users can work without using the controllers. This can potentially increase the immersion and workflow of users. Free-hand interaction has been explored in virtual modelling (Kim, Albuquerque, Havemann, & Fellner, 2005), mobile augmented reality applications (Datcu & Lukosch, 2013), virtual handwriting (Vikram, Li, & Russell, 2013), or as user interfaces for stroke rehabilitation (Khademi et al., 2014).

In the next section, we will present a set of previs personas from the film, animation, and theatre domain. These will be used in later sections in order to ground application scenarios of our previs tools in the world of the users.

[7] https://developer.microsoft.com/de-de/windows/kinect
[8] https://www.leapmotion.com/

3 Personas

Personas cast light on specific actors of an otherwise diverse group of users to design for. The benefit of creating personas is "increasing the focus on users and their needs, being an effective communication tool, to having direct design influence, such as leading to better design decisions and defining the product's feature set" (L. Nielsen, 2013). Further, using personas we can accumulate knowledge regarding the users that we collected in the process of user requirement definition, already done in WP1 ("Requirements Analysis and Specification"). For each application area, we present a set of representative fictional personas that illustrate typical users. This tool helps to describe who is using the previs software and what their needs are.

3.1.1 Film Personas

Directors and directors of photography (DOP) are a main target audience for first.stage in the film industry. Although there are often whole teams working on previs, the aim is to empower directors and other creative personnel to directly perform previs instead of leading a team to do it. In this sense, we propose two personas for film, aiming at describing prototypical users. What stands out with both of the directors is that complex technology and time-consuming setup and operation of a previs tool has to be avoided. Both strive for innovation and aim at creating novel and interesting shots and sequences. For this, an easy to use solution with an easy to learn user interface would be ideal for them to experiment with camera and simple animations. Additionally, although both are technically experienced, a previs tool for both would have to offer ways of interaction that both can explore without having to go through a long learning phase or instruction. Another important aspect for film directors is to have a direct sense of being in the scene so that they can explore view angles and find interesting shots.

Barbara Schmidt

Age	55
Occupation	Director (Film)
Status	Married, 2 children
Tier	Professional
Archetype	Hero

Photo Credits:
Flickr user sdsc_ucsd under CC BY 2.0
Source: https://flic.kr/p/bo5DVA

Motivations

Goals

- Creativity through writing
- Communicate and ideate
- Compelling imagery

Frustrations

- Fiddling with technology

Bio

Attended film school, 20 years of experience. Directed in different locations and countries as a freelancer and contracted director, mainly working with small production staff and limited budget.

Personality

Hands-on type, no previs experience, not sure if previs is useful (expensive, difficult), likes to work the old way, is unsure about changing a functioning workflow

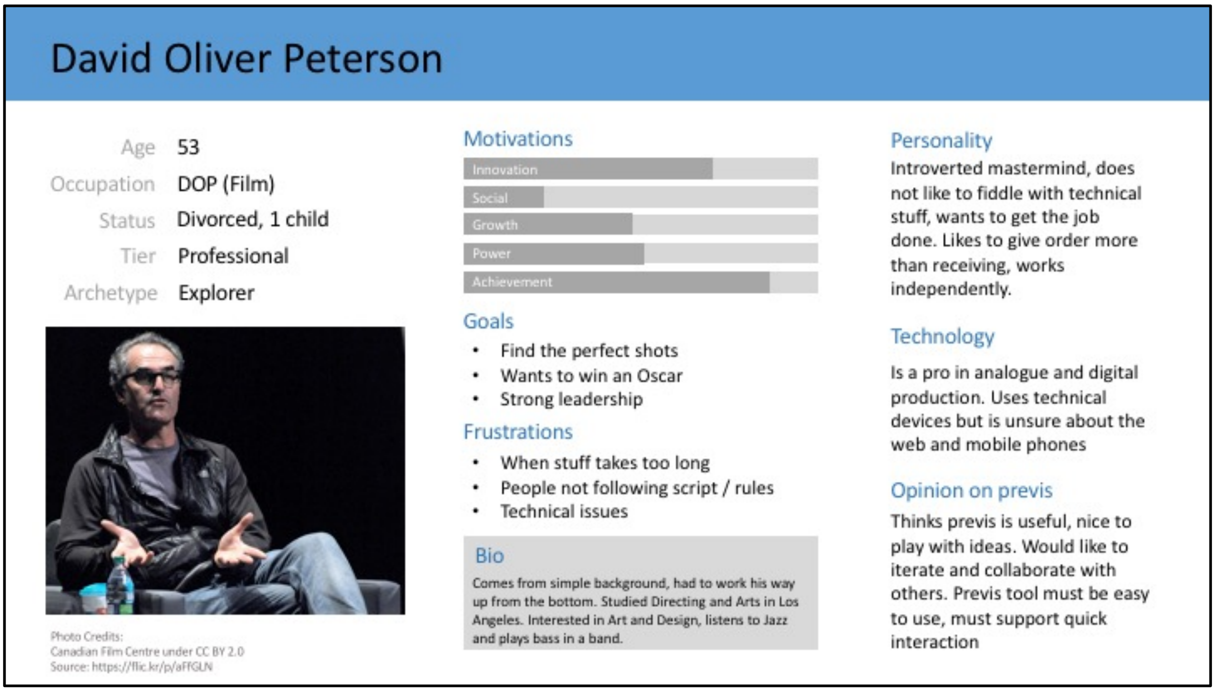
Technology

Open to technology, but must prove useful and effective, interaction with tech must be quick and easy and have clear benefit, does not like to play around

Opinion on previs

Previs is very time consuming and the benefit of it for small productions is not measuring up to the effort that has to be done

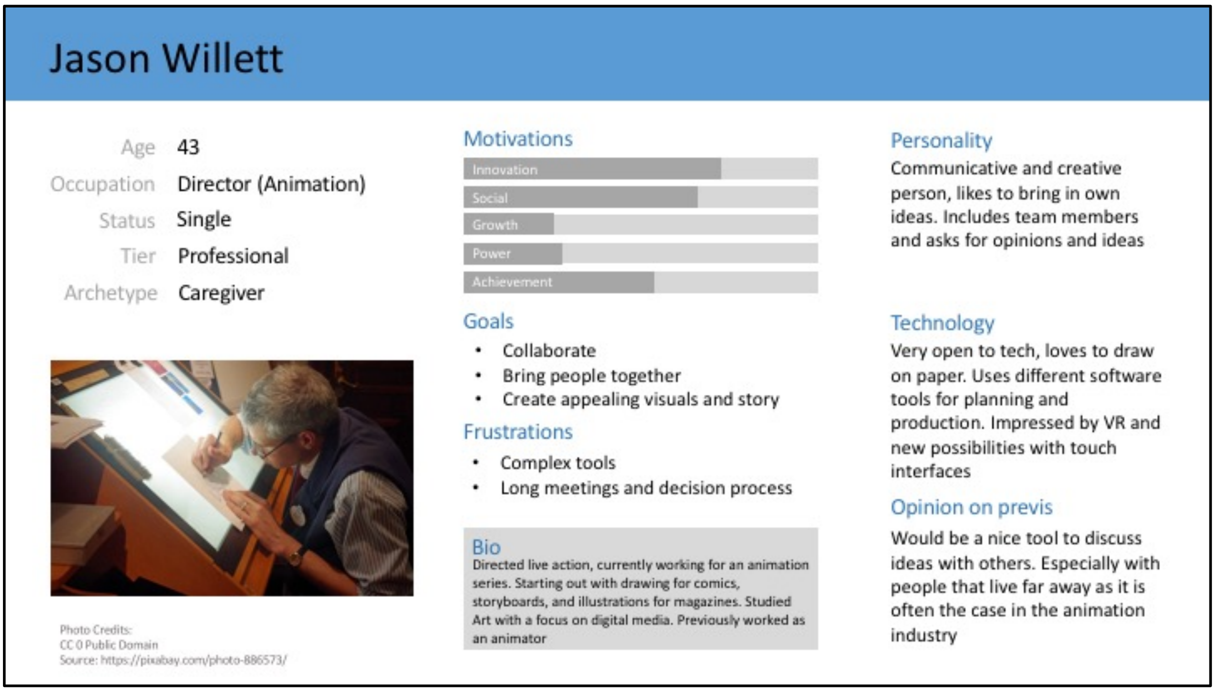
Barbara is a film director and has a strong focus on ideation. She likes to be inspired by writing and ideating with others. For this, she is generally open to previs, but is however sceptical about the technical challenges and possibilities that previs can offer. Tools that she would use have to be easy to use and have to support the task she is trying to accomplish in a way that using the tool improves her work.



David is a director of photography (DOP) in the film industry. He is working together with film directors, on one hand transferring the director's ideas and conceptions but also implementing his own vision of the film, giving it his own touch and feel. Mainly, he is working the crew, communicating his vision to actors, lighting personnel, or camera men.

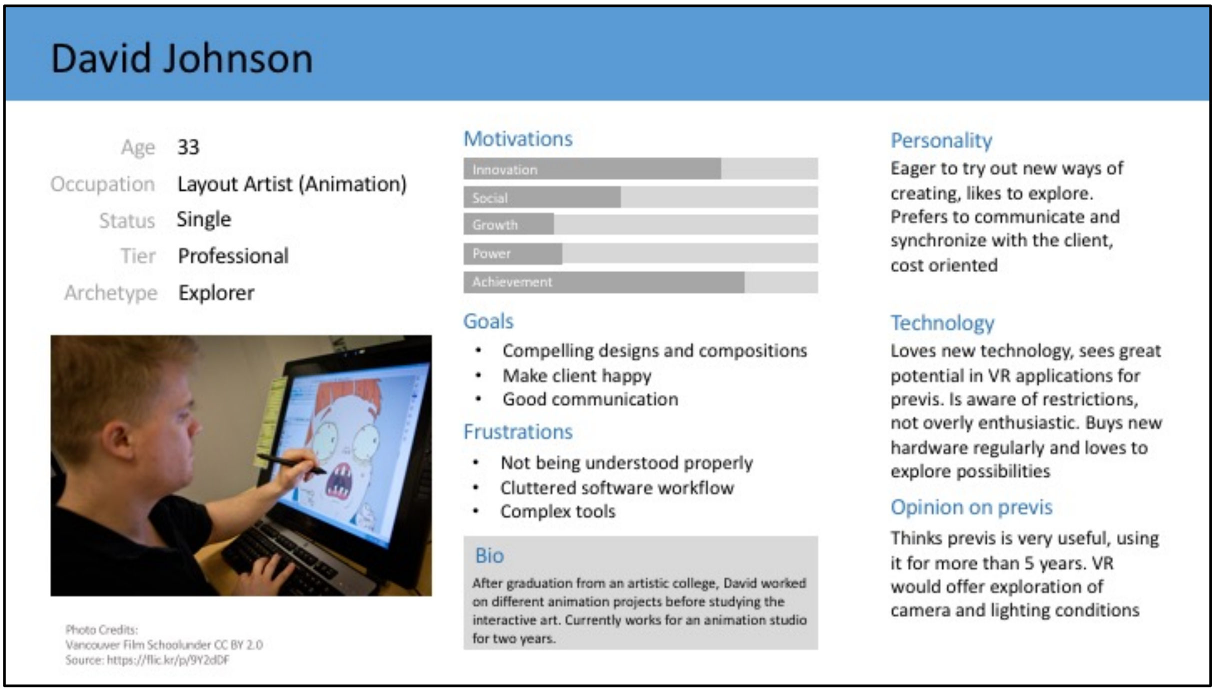
3.1.2 Animation Personas

In the animation production, layout artists, layout supervisors, directors and also clients are involved in the production process. In our example, animation directors and layout artists are two possible users of a previs tool. We consider 3D animation films and series as possible end-products. For these, previs is already being done in different steps of the production. For example, from the storyboard, artists create first shots in 3D to exemplify and extend the vision of a storyboard image. From there, several artists are working with tools like Maya directly in the 3D scene, creating and placing assets and characters. This blocking phase is actual previs already. However, for first.stage the goal would be to empower the directors, as in film, to go into the 3D scene and do the previs tasks themselves. The requirements are similar to film but differ in a sense that animation directors come with a more sophisticated background in computer graphics and 3D animation. The challenge for these users is to provide a workflow that fits into their currently used toolset, so that they can create assets in the common tools and integrate these into the previs software and also are able to continue to work with the results created in previs. In contrast to film, where the previs material is being replaced completely by real imagery, in animation, the resulting shots, compositions and set designs should be exportable so that the work done is not lost. Communication plays an important role as well. Animation teams are often distributed around the globe, so that regular telephone conferences and video conferences are commonly used tools in the production process. A previs tool has to offer a collaboration feature

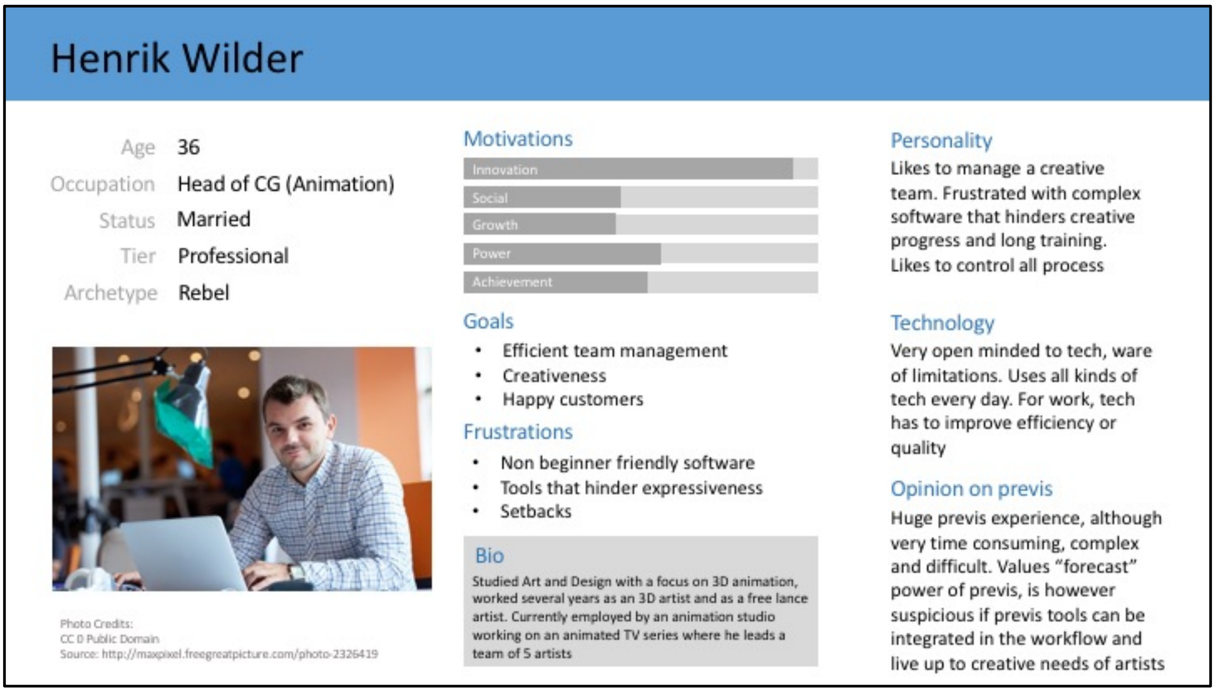


that allows customers, directors, artists, and other contributors like sound and voice actors to work together in a shared virtual environment.

Jason is an animation director. He has a strong artistic background, worked on his own animated short films before coming to be a director. He is impressed by different media that capture his ideas for storytelling and loves to experiment with new technology. The focus of improving previs for his productions is on collaboration and exchange of ideas. Jason uses a variety of different software solutions for his work, reaching from sketching, animation, organisation and collaboration. All these tools have to be integrated in the production process so that a smooth transition between the production steps is possible, also across different people who work on similar tasks like the animators.



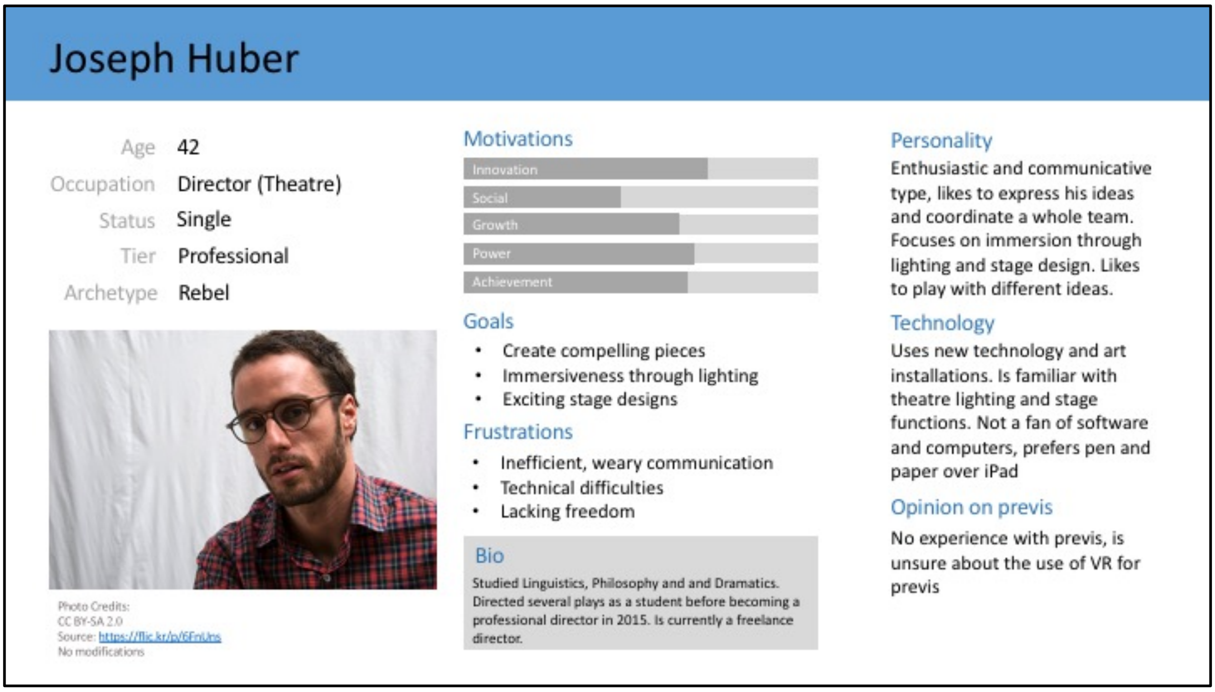
David is working as a layout artist. His job is to create digital scenes and sets, placing assets, characters and other props on the scene in order to create an appealing imagery which suits the storyboard and script. He is very technology oriented and sees great opportunities in Virtual Reality for his work, especially for collaboration and layout. However, due to his tight schedule, interaction has to be fast and efficient so that he can concentrate on his actual work. As in film the interaction has to be fluid and fast so that the workflow of going through dedicated previs does not limit or hinder the currently existing production pipelines.



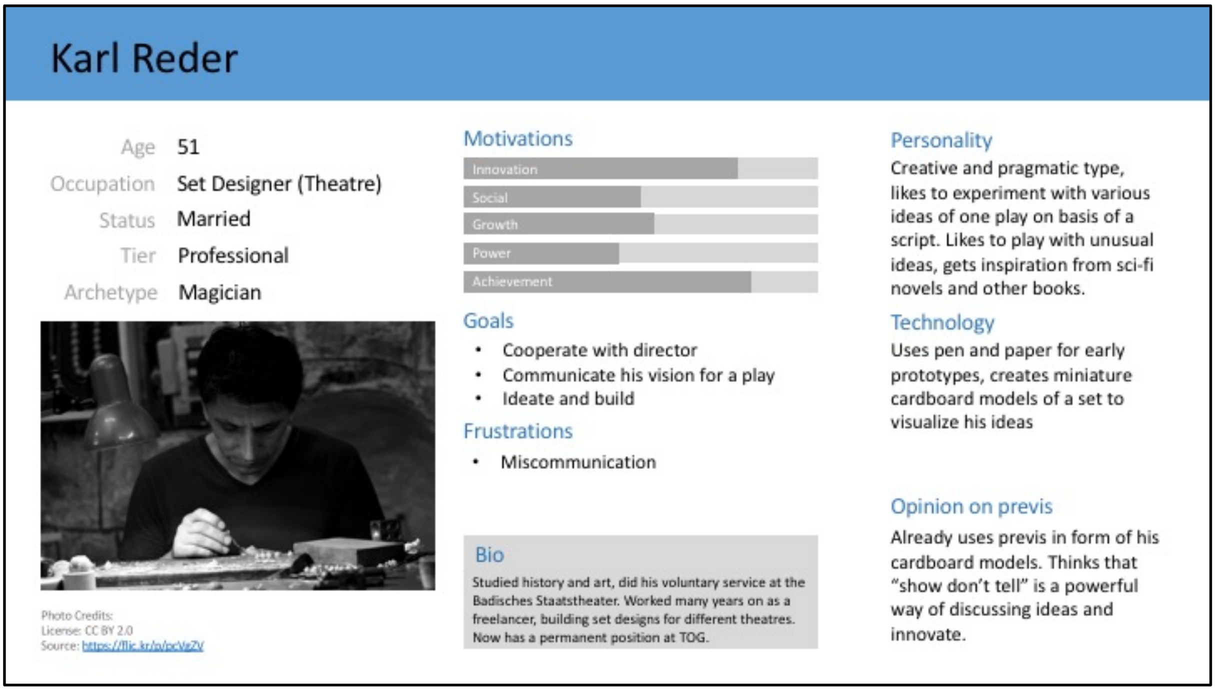
Henrik is a CG director and his main job is to manage a team of creative people. As a manager, he is very suspicious if another tool can be integrated in the overall production pipeline. He appreciates technology that improves the production process, makes it more transparent, especially for the client. One of his main goals is to deliver outstanding quality while also coming up with novel ideas that stand out from other companies, so that the clients want to expand their cooperation with his company. Henrik is sure that previs is a tool that can improve the production process in terms of ideation and creativity but is not sure if it can be integrated in the complex and ever-changing everyday work process.

3.1.3 Theatre Personas

Theatre productions and requirements of a previs tool differ the most from animation and film requirements. As with animation, previs is also already being done in the theatre field. For almost all productions, set designers build physical models of a stage to express and communicate their ideas based on a script, technical capabilities of the house and instructions of the producer and director. This is a very challenging process, because the creation of a set in the form of a physical model takes time and expertise. Physical models lack the ability to try different lighting conditions, which is important for the theatre context as much of the immersion in a play comes from the light in the theatre. Another aspect are the technical capabilities of the theatre itself. Every theatre is different on a technical level. Some houses can rotate their stage, can elevate and/or lower objects underneath the stage, others not. Further, fly towers have different setups, stages are limited in their capabilities to host different plays at the same time, height and depth or every stage is different and even the viewing angles from the audience room differ in every house. Thus, technical capabilities play a huge role in planning a play and even finding out what can be accomplished on a specific stage. Under these circumstances, previs for stage production is a great opportunity to create set designs in 3D, explore and max out technical capabilities of the houses and can improve the coordination process between many actors: the director, set designer, carpenters, lighting technicians, and managers.

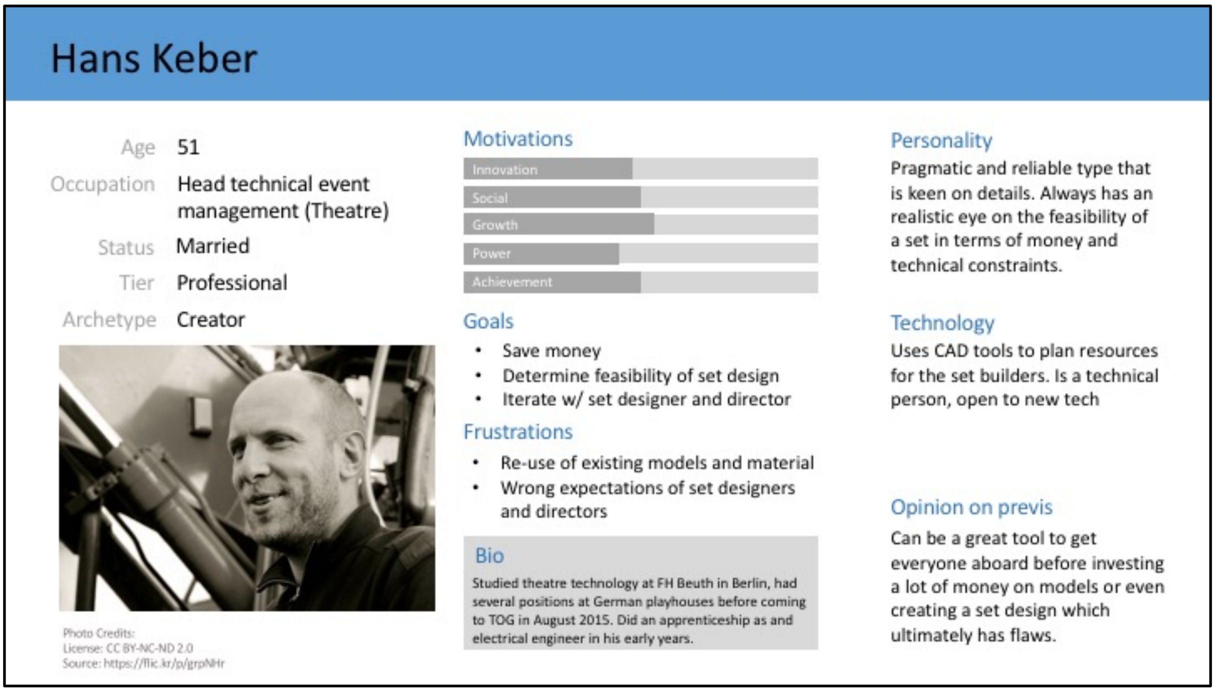


Joseph is a freelance director and creates plays for different houses. For him it is very important to get to know a house in order to create a play that not only has nice storytelling and acting, but also fascinates the audience in a visual sense. He is very open to try previs in his job because he feels that it can give him new ideas playing with the theatre in 3D or even in VR. This way, he is immersed and can



change the light conditions himself, creating a rough stage design in order to drive his creative potential to the limits. Having the opportunity to "go into the theatre" at any time gives him the freedom to work intensively on his projects whenever he wants, no rehearsals would interrupt his creative processes trying to create a new interpretation of a play based on his ideas.

As a set designer, Karl is very much used to work on stage designs, considering different houses, directors, and budgets. As the process of creating a physical setup of a set is very challenging, he finds it very annoying if his ideas are not well understood or received because of miscommunication with directors and producers or finding that his ideas are too expensive or technically not possible. Karl is a very visual person and is of the opinion that a picture is worth a thousand words. In this sense, he likes to convince his clients using real models. With a previs tool, Karl would be able to test exciting ideas before investing a lot of time for the production of a physical model. However, he would never substitute his models with sole previs as the visual impression of a touchable model allows for better perception of his ideas. In his work, cost estimation plays an important role and with previs, he could better adapt to his clients while being more expressive than using pen and paper.



As a technical event manager, Hans' job is to make sure that all the technical facilities in a theatre are considered by the creative personnel. His interest is also to calculate the costs of a play so that the theatre manager can meet the budget goals. He is further coordinating all technical aspect of a play and has to be involved in the production process at every step so that expectations do not get out of hand. With previs, Hans could give very early feedback on the planned play and direct what can be done on the stage and what not. Also, he could harmonize the production process with the workshop, director and producer much better if the vision for the play is clear.

3.1.4 Summary

The personas will guide the further development of the interaction concept so that the needs of the users are transparent based on their jobs and personal (technological) background and will serve to ground the following application scenarios in the world of the users.

4 Natural User Interface Concept

In this section, we present our NUI concept that builds the foundation for the implementation of previs tools in the first.stage project. Our ideas and concept development are driven by the variety of feedback on naturalness during the requirements elicitation that has been performed in the deliverables D1.1 to D1.5 (first.stage D1.1 Authors, 2018; first.stage D1.2 Authors, 2018; first.stage D1.3 Authors, 2018; first.stage D1.4 Authors, 2018; first.stage D1.5 Authors, 2018) and our first-hand experiences with users that tested our prototypes during the first two years of the project. The concept is also based on the related work and own previous work discussed in the Section 2. In addition, current technological trends also drive some of our concepts as today's consumer hardware is a powerful driver in user expectations and acceptance of interaction paradigms. In particular for interaction with 3D content, recent developments had a significant influence on the way experts and novices interact in 3D.

We start this section with presenting our core interaction techniques that we created on basis of the core previs tasks, each one considering the natural aspects of each task and context of use. We further present additional interaction methods that complement our concept generation. On the basis of our interaction techniques and methods, we contribute a resulting list of hardware and input technology that is suited for implementing our interaction techniques and methods. After this, we present specific NUI guidelines that frame the development and conclude the section with a description of the development method we employ in first.stage.

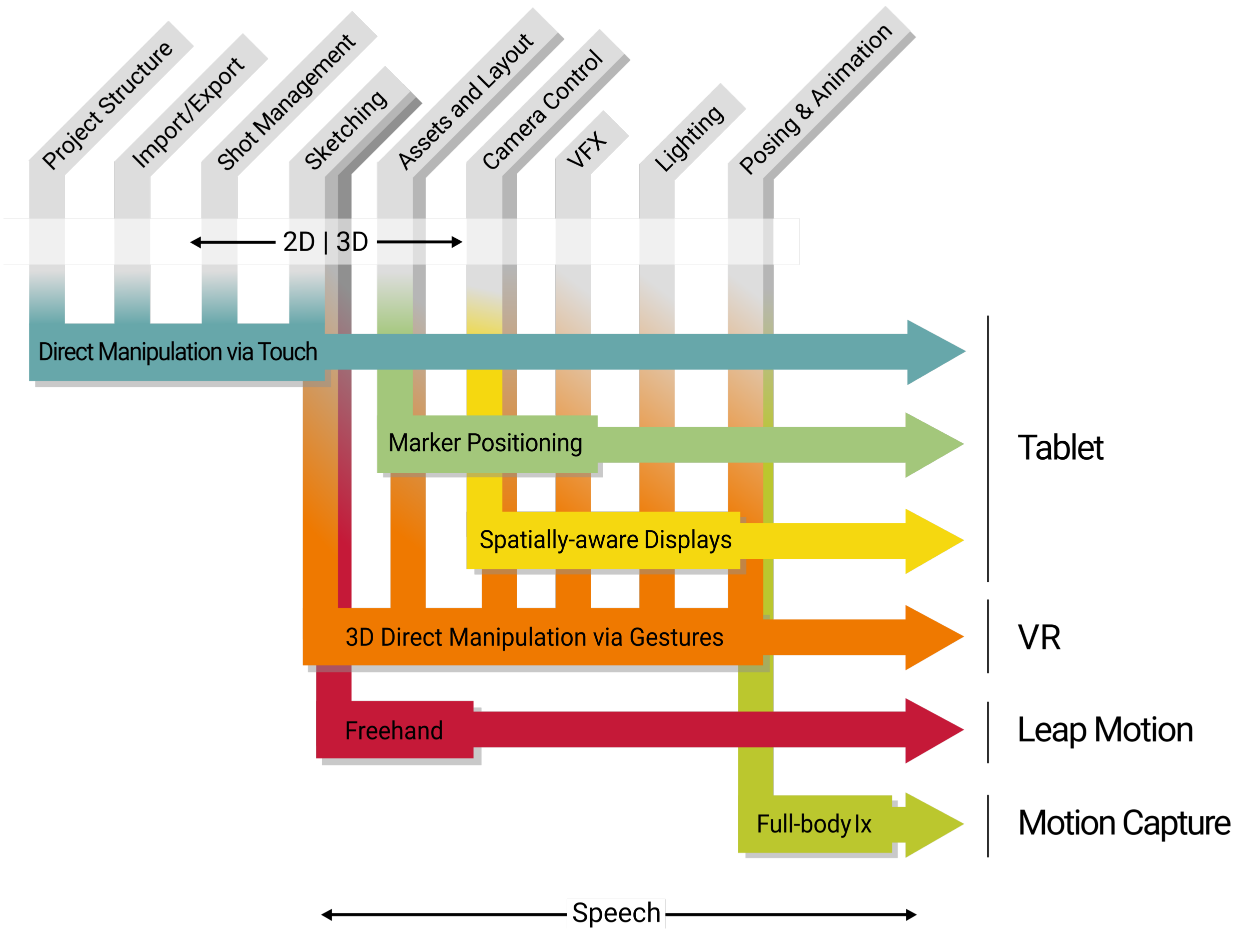


Figure 1: From previs tasks over interaction techniques to technologies

4.1 Interaction Techniques

In the following, we introduce the interaction techniques direct manipulation via touch on 2D interfaces, spatially-aware displays, tangible interaction, augmented reality (AR), direct manipulation in 3D and via gestures, full body and embodied interaction, and free-hand interaction in more detail, how they fit to the respective tasks, and which hardware can be used to implement these interactions.

In Figure 1 we provide a graphical overview of our overall interaction approach based on the previs tasks from the Introduction. We open up a 2D/3D space where we fit the task affordances to 2D and 3D interaction techniques and select the hardware correspondingly.

Direct Manipulation via Touch: 2D interfaces that offer direct manipulation are intuitive, easy to learn and offer a high degree of precision and overview. As with the high penetration of iOS and Android touch devices, these interfaces prove to be very intuitive for a large user base and naturally capture user intent on 2D screens via direct manipulation metaphors and 2D gestures. For the tasks project structure, import/export, and shot management, 2D touch interaction on different devices like tablet computers or mobile phones is very natural because we make use of already present mental models that users employ for 2D content manipulation.

3D Direct Manipulation: Direct manipulation is well suited for the interaction with 3D content as it is presented in a model-world interface which reflects to the real world. Having real-world metaphors for objects and actions can make it easier for a user to learn and use an interface, and rapid, incremental feedback allows a user to make fewer errors and complete tasks in less time, because they can see the results of an action before completing the action, thus evaluating the output and compensating for mistakes. The user of a well-designed model-world interface can wilfully suspend belief that the objects depicted are artefacts of some program and can thereby directly engage the world of the objects. This is the essence of the "first-person feeling” of direct engagement. This interaction technique is well suited and natural for all tasks that require object manipulation in 3D space using stereoscopic view: modelling, assets and layout, camera control and motion, animation and posing, and lighting.

In practise, there can be situations when it is not possible for users to interact in a direct manner. Objects might be too far away, can be occluded by other objects, also interacting with multiple objects at the same time can be hard to achieve. In order to overcome these issues, we implement an addition to the direct manipulation concept: surrogate objects. Surrogate objects represent one or more objects that the user wants to interact with. Other than the object within the scene, a surrogate object can always be presented in the field of view of the user and within the interaction space, always reachable by touch or VR controllers. This is especially relevant for groups of objects which can be spread across the scene, objects which are far away from the user and too small for direct interaction and objects that are completely or partly occluded by other objects. The idea of surrogate objects is presented in “Direct Manipulation Through Surrogate Objects” (Javed, Elmqvist, Yi, & others, 2011) and shows that by extending the direct manipulation concept by such objects, the limitations of direct manipulation can be overcome. This concept is also represented in the object-oriented interface described below by presenting a three-dimensional interface for intangible properties of the objects. A similar approach has been introduced by Schröder-Kroll at al. (Schröder-Kroll, Walter-Franks, Herrlich, & Malaka, 2012) with proxy-based 3D selection.

Two aspects play a key role in designing a user interface: the visual appearance and the interaction itself. Both aspects can get very complicated when the software provides a huge amount of functionality, that have to be presented and controlled by the user. Further, as our results from D1.1 to D1.4 show, most users have little technical knowledge and do not want to learn new complicated tools. In all application areas, previs tools are used infrequently as it is only needed in the first stages of a larger production project. Previs tools must be quick and easy to use, with no demand for extra staff, production time and additional costs. However, state-of-the-art 3D tools often come with overloaded Windows, Icons, Menus und Pointer (WIMP) interfaces, see Figure 2.

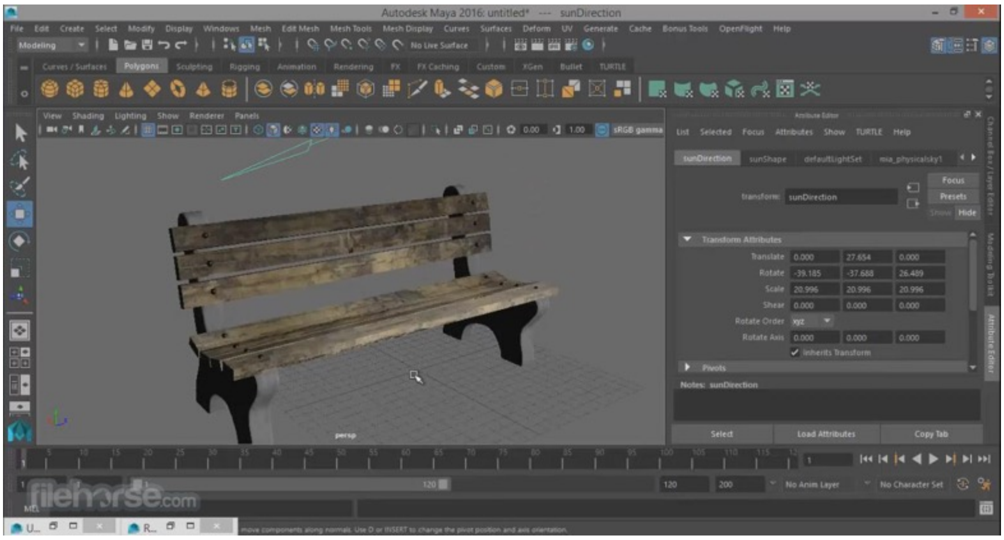


Figure 2: Complex and cluttered Maya user interface

Such interfaces are not suitable for typical users of previs software as they were described in the personas. Previs tasks require specialised training and are often time consuming. This explains the need for additional technical staff, mentioned in the interviews. In addition, the first.stage tools will support different devices according to the requirements from D1.5. As different input and output devices often require a totally different interface, it would be required for the user to learn different interface variants. These arguments contradict the requirement that the software is suitable for non-technical persons and can be learned quickly.

A common practice in standard WIMP interfaces is to use buttons for all possible actions. If an object is selected, the actions that are not supported for this object are deactivated (in grey colour) but still presented to the user. This leads to cluttered interfaces, which can easily overwhelm users. In addition, buttons are commonly spread across the interface leading to a loss of association with the object the user wants to manipulate. Buttons are rather grouped by function which is often only useful for experienced and advanced users. Novice users have no benefit at all because if they do not know the function of near buttons they cannot infer the function of a specific button.

Based on these observations, we implement object-oriented user interfaces (OOUI). These are types of user interfaces in which the user interacts explicitly with objects that represent entities in the domain of the application. It can be seen as the counter approach for function-oriented interfaces normally used in 3D applications. We propose a system in which objects have a dedicated interface only displaying the actions available for this specific object. The interface is positioned relative to the object it is corresponding to, rather than fixed button positions in a static WIMP interface.

Object oriented user interfaces are proven more user friendly compared to other interface paradigms and provide several advantages in terms of usability (Raskin, 2000). OOUI adopts a 'noun-verb', rather than a 'verb-noun' style of interaction. In addition, the relation to the object which will be affected by an action is better understandable. Users can also classify objects based on how they are presented and behave. In the context of what users are trying to do, all the user interface objects fit together into a coherent overall representation. OOUI can reduce the learning curve for new users as only relevant actions and options are displayed.

It is possible to display an object-oriented interface with a similar visual representation on different output devices like tablet or VR. It can be rendered as part of the application and does not need a WIMP GUI. Furthermore, the interaction with such an interface can be designed to be similar with different input devices. The aspects described here make the OOUI concept very suitable for a previs application as it solves many problems and user demands stated above.

Spatially-aware Displays: These are displays that have information about their position and orientation in the room either by relative differences (gyroscope and compass data) or access to absolute position and rotation data using tracking devices. With these displays, which are mainly tablets and tablet computer devices, it is possible to create a direct access to a virtual scene by putting the control over the virtual camera directly in the hands of the users where they can move the device across the room in order to change position and orientation of the virtual camera. This approach can be complemented with basic 2D gestures in order to extend the interaction space. Previs tasks that make the most out of this interaction technique are camera work and camera motion. Using spatially-aware displays, users can position cameras in a virtual scene and navigate in 3D while at the same time having the resulting shot visible through the display at all times for high efficiency. Thus, it is very natural for frame a shot in a virtual scene using real-world references by walking in a real room and orienting the device as needed.

Marker Positioning (Mobile AR): With Augmented Reality, real images and projections of real scenes on screens like mobile and tablet devices can be augmented with digital information. This bridges the real world with the virtual world, allowing for an immersive experience. We use marker-based AR in our first.stage tools for layout tasks that require multiple users to work together in a virtual scene. The markers connect real objects with digital media and enable interaction with the data via interaction with the objects. These interfaces are very intuitive and easy to learn because users can learn them directly by using them, observing actions on the digital data that have a direct mapping to their actions.

Full-body Interaction and Embodiment: Embodied interaction relies on the integration of interaction between humans and computers into the material and social environment. In our specific case, the recording of motion capture data via the Rokoko SmartSuit allows for full body interaction and embodied performance animation though inertial sensors that can be worn in the form of a suit. With this, it is very easy and natural to record character animations because users can use their own body as input, directly transferring their motion data to virtual characters. In many previs cases and especially for character animation, this makes complex and expensive keyframe animation techniques superfluous. This interaction technique is very natural because it doesn't rely on a translation between user intent and action. There are almost no graphical interfaces needed and the approach has a low learning curve regarding the animation.

Free-hand Interaction: With free-hand interaction users can perform manipulations to digital objects using their own hands and without digital tools by tracking the hand and finger motion and applying the data for the manipulation of the virtual object. We employ free-hand interaction for rapid prototyping and modelling of 3D content and objects in the modelling and layout tasks. Here we focus on the workflow of modelling and layout where we support users in "getting their ideas quickly out of their head" by removing a graphical user interface (noUI) where they can concentrate on implementing their vision in either a 3D model or scene layout.

Speech: In our previs tools, speech is used as a supportive layer that can be integrated as a secondary interaction to complement a primary interaction with speech commands. For example, in layout, users can filter and select the asset library using speech commands. Generic commands include deletion, menu interaction, and basic manipulation tasks that are expressive by speech but hard to express through manual interaction, for example flipping an object, switching to a specific camera view, or switching between different views and zoom factors.

To sum up, the first.stage natural interaction techniques centre at providing natural expression by implementing interaction methods using state-of-the art technology that is optimal for natural expression and fits best the specific task context and needs of creatives. This way, users can adapt the interface to their needs, putting the work in the focus, working directly on the content and not a complex interface, and in turn express themselves naturally. As depicted on the right-hand side in Figure 1, the resulting technologies and devices are listed. These state-of-the-art technologies and devices offer the required technological foundation for the realisation of our interaction techniques. Before we are introducing these technologies and devices in Section 4.3, in the following we present interaction principles that are relevant for first.stage.

4.2 Interaction Principles

Having introduced our natural interaction techniques, we now present interaction principles that enhance the interaction by implementing the notion of "how" the interaction in our tools should be. The principles are grounded in the requirements defined by the application partners in the deliverables D1.1 to D1.5 and research done in the field of interaction design, usability, and user experience. Standard usability guidelines (Malaka, 2008; J. Nielsen, 1994; D. Norman, 2013) should be followed as well.

Task Context: Depending on what users are doing, it is important to understand what natural interaction means in that specific scenario. For example, there is a difference between managing a previs project on a high level by sorting and naming scenes or adding users to a project and creating 3D content using the own hands. The distinction that we draw here is that naturalness is coupled to the context of use. Organizing a digital project like organizing files in the Windows Explorer or Finder is certainly more natural on a 2D interface because this kind of interaction stems from the digital domain of creating graphical user interfaces that most computer users are familiar with, because they provide a known context, look and feel, and interaction. It would be very different when trying to organize, copying, duplicating or moving files in VR using other than 2D metaphors. An example for this is the “minority report vision” that has often been used to exemplify the use of gesture and 3D interfaces for desktop tasks. Studies have shown that these kinds of tasks are slower to perform and are cognitively more complex to achieve than with 2D GUIs (Bérard et al., 2009). We pick up on this notion and motivate a natural use depending on the context of use. For example, arranging 3D objects in space is most natural done in VR, as well as exploring spaces, getting sense of scale, picking shots for cameras, etc. On the other hand, project organisation is best done on 2D interfaces like tablets or computers. Transforming and working with digital content for organisation is, as previously stated, most natural using GUIs. Another example is motion capture. Animating humanoid characters can be done in different and more or less natural ways. It can be done by keyframing a 3D character or by drawing frame by frame. Both has advantages and disadvantages, but looking for a natural way of interaction, using the own body is the only way of having a direct 1-to-1 relationship between user intend and desired outcome. Using the SmartSuit Pro, users can work with their own body without having to understand complex keyframing or drawing techniques, making it more natural to create animation content and providing a low learning curve.

Multi-User: Handing project to others is complicated and requires a lot of management. But working together in the same physical context is very natural like in sailing where a common understanding is shared in the physical space through observation and understanding of intention and tasks. We support this natural interaction by providing input to the system that is not bound to one device. For example, in a multi-user scenario, different devices can be used in a shared context in order to achieve a common goal. When designing a 3D scene, one user can be set in VR so that he can arrange objects in 3D very easily in his own space while using the teleport functions to reach other destinations. In this scenario, another user can join the scene with another device, for example a tablet computer. Here, the tablet user can provide overview and another kind of assistance by manipulating larger structures or roughing out a broader scene design while the VR user keeps on working in s smaller space. The tablet device can also be tracked and internal sensors like the gyroscope can be used to track the movements of the tablet user so that he can have an own representation in the virtual world that can be seen by the VR user in order to increase collaborative work and co-presence.

Playful and Fun: Playing is intrinsically motivated and autotelic (Deci & Ryan, 1985). When we play, show a playful attitude towards a task, or even are provided with a user interface that supports playful expression, humans can explore, be creative, try different solutions, find joy and amusement, even in productive contexts, increasing long-term motivation and lower the frustration barrier (Deterding, Dixon, Khaled, & Nacke, 2011). This is achieved by more pleasurable interaction rather than optimizing for speed and being goal oriented. Playful interaction also invites the user to discover features rather than frustrating him with an overloaded interface. Therefore, it is suitable to support novice users and users with little technical knowledge.

Rapidness: Interaction should be rapid and avoid time-based interactions and large motions. It should be possible to quickly produce rough results. Time-based interaction should be avoided as it disrupts the workflow, especially for experienced users. For example, “look and hold” or “point and hold” interactions should not be used, see Figure 3.

Accuracy and Precision: In contrast to the general schema of rapid interaction, it should be possible to make things precise with additional effort. This means it is possible to scale up and work on detailed structures or position things according to measurements using additional tools that first have to be activated.

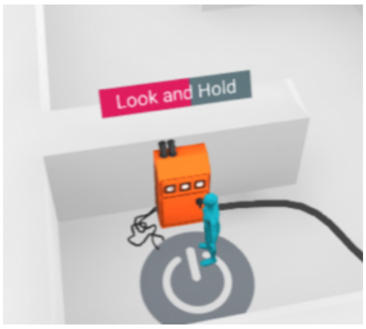


Figure 3: Look and hold should not be used

Consistency: Interaction should be consistent within the application. This is important so that the user does not get confused by different interaction schemes for the same tasks in a different context and only have to learn a minimal set of interactions. For example, positioning an object by "grab and place" should work the same way in layout mode and in animation mode.

The interaction should also be as consistent as possible across different input devices. This will help the user to seamlessly switch between devices and do not have to learn or remember special interactions for this input method. In some cases, this will not be possible as the different input devices provide different input modalities and degrees of freedom. But a primary interaction (left click, tap or trigger button) should perform the same action on all devices.

Furthermore, the interaction should be consistent with other applications from the field, so that the user who has learned to interact with another tool does not get confused by completely different interaction schemes. This means the application should not break with interaction standards from the field, e.g. support drag and drop.

Easy to Use: In order to make the first.stage tools accessible to users with little technical knowledge, special aspects should be considered to make the software easy to use. Interactions should not be designed to be as fast as possible but plausible and intuitive to the user. Actions should always display the outcome as text or preview when hovering over an interactive object. In general feedback to all actions should be provided. This helps the user to follow his intent. This is especially important when an action will involve multiple sequential interactions. In order to not overwhelm the user with options, only a minimal amount of possible options should be shown. This can be achieved by using interaction scaffolding and nested interactions.

Physical Manipulation: Objects behave according to physical laws, e.g. mass, acceleration, friction, gravity (pushing an object against another moves both objects but a little slower). Physical behaviour is restricted to assist the intended action of the user. Falling objects stop when in contact with the ground or other objects (no bouncing), objects do not have correct mass, e.g. pushing a building is as easy as a chair, goal-based physics, break from physical behaviour to match user intent, interact with object like in the real world or better than real, e.g. scaling of objects.

Reality-based: The interaction in 3D should orient towards the interaction with objects in the real world. As the interaction in 3D space is novel and users are not always familiar with it, it should utilize the everyday knowledge of the users (Jacob et al., 2008). This applies to positioning and rotating objects: Small objects like a bottle can be grabbed, rotated and placed with one hand. Bigger objects like a table or houses on the other have to be pushed and rotated using two hands. This behaviour comes natural to the user and should be supported. Another aspect that should be reality based is the interaction with buttons, knobs and slider elements from the object-oriented interface. A clear 3D representation of the intractable object should be provided comparable to real life light switches, music volume knobs and radio controls. The visual representation should present feedback of the current state of the control. Audio and haptic feedback should be provided when changing the state.

Feedback: The system should provide feedback to all actions of the user. Visual feedback is the most important one, which should be provided on all hardware platforms. If the object the user is interacting with is currently not in the field of view, and visual feedback is not visible, assisting indicators at the edges should indicate where the interaction is happening. Haptic feedback should be utilised when interacting in virtual reality using the vibration functions of the controllers. Audio feedback can be helpful in certain situations especially when something is happening outside the field of view.

Nested Interaction: Rather than putting all actions onto different buttons of controllers, mouse and keyboard overloading the interface, the system should utilize nested interactions which chains the selection of an action, parameter finding and performing the action into a series of small lightweight interactions.

Sequenced actions should be designed so that experienced users can perform them very quickly and do not perceive them as impairing. The position of buttons in a sequenced action for example should be positioned in close distance to each other, following the direction of motion.

Sitting Interaction: The interaction should be able to be done sitting in a chair. This is important for daily and long-term use. Physical exertion should be reduced to a minimum, so that it is acceptable to use for longer times at a workplace. Interface positions and scales should adapt to the users' posture.

Creativity Support: In order to support the creative process of the user creating a virtual set, animation or camera shot, the interaction with the software can be designed accordingly. Interactions should invite users to interact with objects in a natural manner instead of telling them what to do. This can be achieved using affordances and signifiers (D. Norman, 2013). Designing subtle affordances invites the user to discover through exploration. Presenting the user with different viewpoints on a scene or with the sequence of his/her actions the user can be supported in his/her iterative and evaluative process. When designing the interaction, this aspect should not be top priority as it only provides supporting features but rather be considered as an additional aspect that can help the user.

4.3 Technologies and Devices

Based on the interaction techniques and principles that we defined in Sections 4.1 and 4.1, we now present interaction hardware and devices that fit and support the interaction.

4.3.1 Virtual Reality

Based on interaction techniques we identified in Section 4.1, Virtual Reality (VR) hardware emerges as our core technology because it provides high-precision direct interaction and manipulation for 3D content using tracking controllers and natural depth perception through stereoscopic view with head-mounted displays. VR is grounded in how humans interact with objects in the real world, making use of the human's capabilities of interaction and perception that everybody is familiar with and an expert in. This is especially prominent when looking at the HTC Vive room-scale tracking where users are able to use their body in a larger context, being able to physically walk to another location in VR instead of having to completely rely on teleport functions. Users are also able to express themselves in a very natural way using and working with their own arms, hands, and head movement. These capabilities make the use of VR very intuitive as the majority of tasks (orientation, locomotion, relocation) can be used and performed without the need to use any interface. More importantly, the visualisation capabilities that recent head-mounted displays (HMD) offer, are a strong fit for natural 3D interaction. Having the possibility to experience immersive 3D worlds is one of the core assets that VR offers. No other system can provide such a realistic and believable first-person experience. Because of these capabilities, in the scientific literature, VR is recently explored as a tool for Virtual Reality Exposure Therapy (VRET) (Anderson et al., 2013), demonstrating immersive visualisation capabilities.

Another important aspect of interaction in VR are the controllers. While it can be argued that these offer only rudimentary control and interaction, we argue that in fact those devices offer a rich and natural experience. This is manifested in the large variety of interaction tasks that those devices can be used for. For instance, users can directly point, move, grasp, touch, and perform gestures using the controllers without having to switch modes or tools for these very common operations.

In a practical sense, there are further supporting arguments for the use of VR. First, VR is especially useful for previs as many tasks centre on 3D content interaction and manipulation. For example, for animating characters, which is one of the main activities in previs, VR offers a first-person experience of the scene and editing can be done directly in place with the animation control being directly on the characters and not obfuscated through a complex user interface (cf. D1.1). Second, it can also be useful for rehearsals and digital production where today actors rehearse in front of a green screen with non-existent characters (cf. D1.2). Using VR in the previs process at this point, actors get a sense of the scene by experiencing it themselves and in turn have a much better understanding of the context. Directors can switch into any character in the VR scene, thus discovering new perspectives and foreseeing issues and chances in the production. Third, what is most important in the end is the final result that is presented to the audience (cf. D1.3). With VR, every director can instantly take this perspective, being it in the theatre or an animated or non-animated film. This kind of perspective change is not that easily possible in any other medium than VR. Moreover, VR and 360° technology are the future and many more productions will focus on this technology being it in production or as a medium for audience experience (cf. D1.1). Fourth, when planning and previsualising exhibitions, customers and stakeholders can get a direct and immersive experiences of scale, visual and technical appearance of the final design and even interact on content, depending on the detail of the previs.

The application of VR for previs is further supported by our findings in D1.1 to D1.4 where practitioners stated that they envision VR as a useful future interaction method for previs. However, we also collected some less optimistic feedback where users stated that they fear ergonomic disadvantages for their work. Specifically, users are uncertain is the head-mounted display (HMD) can be worn over a longer period of time and if the HMD might cause cyber sickness or other uncomfortable bodily reactions (cf. D1.2). While we take these concerns seriously by implementing our VR-based tools in the latest fashion of ergonomic VR guidelines, we further claim that these problems will be solved in future iterations of the hardware, making the technology even more easy and pleasant to use.

Virtual Reality technology has recently become widely available, providing excellent tacking and visual quality at reasonable prices. VR itself is not a new invention and many usage patterns and interaction metaphors have been established in the first VR wave in the 90s. However, the novelty now is that we can usefully build for VR with the expectation that small production teams can afford the necessary equipment. Not only that, with precise tracking of the headset and controller position as well as orientation we get room-scale VR which really is new anywhere outside an expensive lab or studio. Even with simple controls, the HTC Vive, Oculus Rift, and other solutions like the Windows Mixed Reality Devices[9] are remarkably expressive precisely because what you see in VR is exactly what your hands are doing: you do not have a displacement between hand and screen. It could be questioned if VR is in itself natural or not, however, we argue that this is not a relevant question. As with other input technology, user interaction and user experience can be crafted in more or less suitable ways. For example, touch interfaces are very natural for 2D manipulation as it is now the gold standard implemented in tablets, cell phones, watches, public displays, etc. There are always ways to provide less natural user experiences such as command line inter-faces, however, VR is especially suited for natural interaction through its inherent capabilities of immersion, stereoscopic 3D, direct manipulation, and gesture capabilities.

However, although VR has many advantages for previs, we also believe that the technology has some limitations that are crucial in the productive context. For example, when looking at how multiple people can work together in a room on one scene, sharing ideas and brainstorming, VR is limited in this way because it excludes the VR user from the others because of the head-mounted display (HMD). Although we offer multi-user access via VR, allowing distributed groups to work together in VR, in this scenario, the most natural way of interaction is a shared medium that all users can view and work on at the same time. This problem can be solved by using tablets and mobile devices that implement AR technology.

4.3.2 Tablet computers

Touch interaction on tablets and tablet computers is proven to be the most natural way of interaction with 2D content, as manufactures such as Apple and Google proved with iOS and Android respectively. By using natural touch interaction that mobile devices offer, users are able to employ multi-touch usage patterns that they are already familiar with, making the transition of the technology even easier. Further, devices such as the iPad Pro or the Microsoft Surface offer high computational power, flexibility, and can be shared among multiple users. As these devices are also equipped with a series of sensors, these can be used to implement spatially aware capabilities that allow for 3D tracking of position and orientation. These devices are affordable and widely available, making it an ideal technology for the implementation of 2D direct manipulation interfaces, and as input devices for 3D interaction, as for example in the camera task. Further, as with VR, AR technology became recently widely available through the integration of ARKit and ARCore in iOS and Android tablets and mobile phones. Working with AR in previs is a very natural experience because users share the spatial context both in the virtual and real world, embedding advanced previs concepts directly in the real world. Directors, artists and producers can stand on a stage, in an outdoor scenery or film set and plan the production using the direct mapping that AR provides.

---

[9] https://www.microsoft.com/de-de/store/b/virtualreality

4.3.3 Rokoko SmartSuit

When creating an animated scene, users can rough out the character animations in VR by drawing a walking path on the floor in order to direct the characters to walk in a certain direction. If users miss animation features like crouching or jumping, they can put on the Rokoko SmartSuit in order to create the missing piece. Here, we use the technology of our project partner Rokoko that provides an easy approach to performance animation. As already stated, users can record their own character animations using their own body. It is also possible to include more advanced systems like the OptiTrack systems that offers even more precision but is also more expensive.

4.3.4 Leap Motion, Kinect, and OptiTrack

Virtual modelling can be done in VR using crafting tools, but if more precision is needed, users can switch to a digital sculpting mode where exact hand tracking is used for more precise modelling. Here, we use the Leap Motion sensor as an affordable and practical device for rapid scene prototyping and sculpting tool, making asset creation a natural task. If more precision is needed, these tools can also be used in combination with an OptiTrack system, again allowing for more precision and performance at higher costs. We further integrate Kinect Sensors as input devices for surface tracking in the modelling task. Specifically, three sensors are fused in order to create a more precise 3D model of a tracked surface such as a sandbox, allowing for natural terrain modelling.

To sum up, we use VR as our core technology because of the natural interaction paradigms the technology provides for previs and the fact that it is well suited for all basic 3D previs tasks. If needed, users can switch to touch and gesture interaction as adjacent interaction methods that complement the natural interaction capabilities of our first.stage toolset by providing collaborative features in VR, AR, and shared spaces, more expressive tools for modelling (free hand and gesture interaction) and animation (motion capture). With cross device and multi user interaction, the first.stage tools further enhance natural collaboration when working remotely and with multiple users on different devices. In the next section, we present a set of previs-specific NUI guidelines that we distilled from our experience with users in Work Package 1 as well as in the requirements elicitation in the same context.

4.4 Previs-Specific NUI Guidelines

We present the following first.stage notion of natural interaction guidelines for previs in the creative industries:

a) Design interaction methods that are in line with accustomed natural human experiences and behaviour
    a. Employ familiar motor patterns for expression
    b. Make use of innate spatial orientation and awareness
    c. Use 3D perception capabilities (2D screens are not optimal for 3D content)
    d. Employ familiar patterns of expression and object manipulation that creatives and artists are used to
    e. Provide a direct translation of user intent to action instead of outsourcing interaction to visual and graphical user interfaces and GUI elements.

b) Make hardware selection task and context dependent. Every task has its own characteristics and preconceptions and different hardware options are more suited than others for different tasks. Design interfaces that allow a natural user experience by providing the right modality for the task so that users can express themselves in a natural way.

c) Make use of multiple ways of expression that combine different modalities for fluent and natural user behaviour. For example, we can give speech commands while modelling with our hands to augment the interaction and fluency of the task.

d) Direct manipulation and reality-based interaction as main interaction metaphor. For VR this is natural, for different modalities, we make use of tracking, motion sensing hard-ware in order to facilitate one-to-one mappings in the interaction.

e) Content orientation. In previs and creative disciplines, users mainly manipulate content and should not be confronted with user interfaces where not necessary.

4.5 Design Method

In this section, we outline our design methodology for our previs tools. We follow an iterative approach in the development where we create a series of project demonstrators as required by our description of work (DOW, first.stage consortium, 2016) that are named D3.1, D2.2, for example. In order to achieve short implementation iterations, we will implement incremental releases of the demonstrators that are named $R_1 \ldots R_n$, where R stands for release. These releases are combined and form the project demonstrators (e.g. D3.1, D2.2, …) that include all functions that are being developed throughout the project and that also include the several demonstrators that are being delivered in the work packages 2 and 3. In our iterative approach, we implement the *double diverge-converge pattern* that was first introduced by the Design Council in 2005 (Design Council, 2005) and was later adapted by others, see Figure 4. This design method highlights the importance of generating ideas in a discovery and research phase, where man alternatives are explored. In the definition phase, the ideas converge to possible solutions in order to implement different solutions in a divergent manner. Finally, in the delivery phase, working solutions are selected for implementation in the context of use.

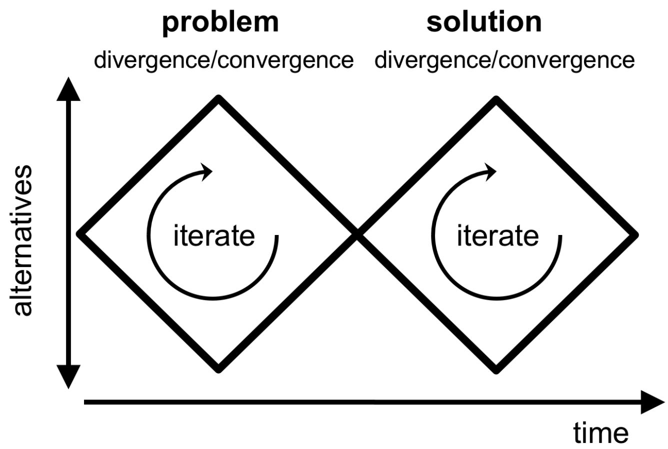


Figure 4: Double Diamond Framework, adapted from Design Council, 2005

4.5.1 Main Prototype Development

The main demonstrators and R_n prototypes are developed and integrated by MovieStorm Limited in cooperation with the University of Bremen. Here, the main track of functions is integrated based on the requirements collected in D1.1 to D1.5. These releases are continuously created during the project runtime, resulting in at least three iterations per year. Additional to the implementation of the requirements, we perform technical workshops with the project partners where we assess and evaluate the technical features and user experience and plan the upcoming releases concerning which features have to be implemented and what changes we have to integrate. The latter is important as we have to consider the evaluation plan that is being developed in WP6 (“Evaluation”). These discussions are organised in the form of focus groups and brainstorming sessions.

4.5.2 Research Prototype Development

As first.stage is a research and development effort, different research prototypes are being developed throughout the project by University of Bremen in order to test and evaluate new natural user interface systems and paradigms. Evaluations are performed in addition to the regular evaluations of the releases in lab studies including qualitative and quantitative methods. Upon positive evaluation, the research prototypes are integrated in the R_n builds in order to create a direct impact of the research results into the main development and can be subsequently integrated in the project demonstrators.

4.5.3 Iterative Testing and Evaluation

While different prototypes are created in the project, we employ the following user experience methods for continuous testing and evaluation of our implementation efforts: heuristic evaluation, focus groups and user Interviews, quantitative surveys, and field studies. These evaluations are performed in Work Package 6 (“Evaluation”) on a regular basis, starting in year 2 of the project. The prototypes will be distributed to the application partners for a heuristic evaluation in the context of the application area. These evaluations will be performed following a standardised procedure that will be developed in Work Package 6 including an evaluation of the release against standard use cases. Qualitative and quantitative data will be collected in the form of questionnaires, user observations, think aloud, and semi-structured interviews. The results of these evaluations will be used to improve the subsequent versions of the releases.

5 Concept Implementations

In this section, we present different previs tools that we created in the first two years of the first.stage project. All these tools follow our NUI concept as we will demonstrate. This section subdivides in the presentation of our main prototype and research prototypes, as introduced in Section 4.5.

5.1 Main Prototype

We already highlighted the advantages of VR for previs as introduced in Section 4.3.1 and thus created VR-based tools that cover the core previs tasks modelling, layout and assets, camera, lighting, and animation. A detailed description of the different functions and interactions can be found in the deliverables of Work Packages 2 and 3. Here, we are presenting a condensed overview of the main aspects of the systems. We will first list all demonstrators that implement the NUI concept and show how the R_n prototypes relate to those demonstrators. According to our DOW (first.stage consortium, 2016), the following demonstrators have to be delivered that integrate this NUI concept:

D2.1 – Demonstrator for Asset Creation (M14) implemented in R_{1-5}
D2.2 – Final Demonstrator for Asset Creation (M24) implemented in R_{5-9}
D3.1 – Demonstrator (Mock-Up, M14) implemented in R_{1-5}
D3.2 – 2^{nd} Demonstrator (M24) implemented in R_{5-9}
D3.3 – Final Demonstrator (M35) future implementation in R_{10+}

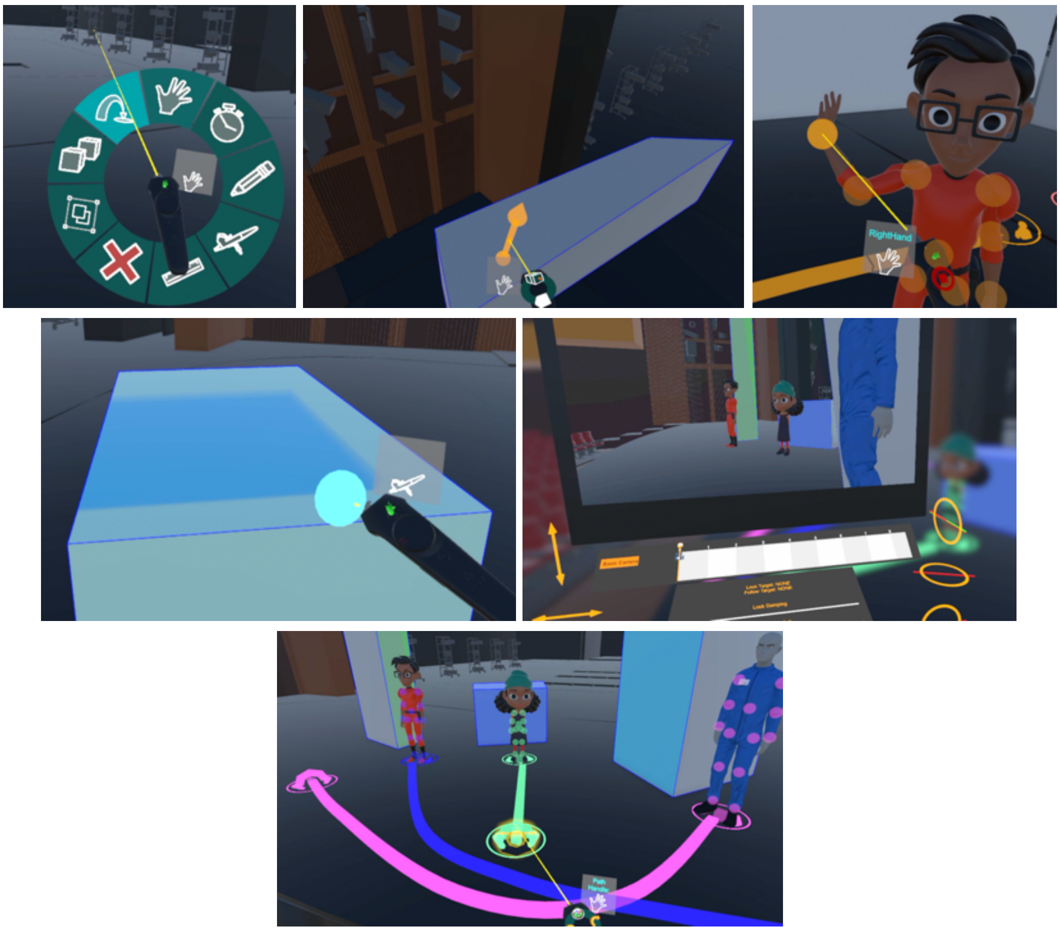


Figure 5: Selected functions of the VR tools, top left to bottom: VR menu, crafting custom shapes, posing a character, painting an object, first-person-camera-view, animating different characters

5.2 Research Prototypes

5.2.1 Virtual Reality Motion Capture

VR user interface for capturing embodied animations. This is natural because embodied interactions resemble bodily experiences that every human is familiar with. In contrast to traditional interfaces for motion capture, this system enables users to record animations from the perspective of their own body, to slip in any other body (human or not) and perform animations from their perspective. With this, one single user can record a whole crowd of people just by himself in a matter of minutes, see Figure 6.

Figure 6: VR Motion Capture, left to right: preparation for capture, recording high-five of first character, recording high-five of second character, final result

5.2.2 Anticipation Cues for Orchestrated VR Animation

Anticipation support for capturing embodied animations in VR. With this prototype, we solve the problem of syncing actions of embodied motion capture avatars in real time, circumventing post-editing. Users can orchestrate different animations in order to record synchronised actions between different animated avatars, for example as in a dance performance. This is natural because as in VR motion capture, people can use their own body as input, having a one-to-one mapping of action and result. Circumventing post-editing is further natural as it makes working in other applications obsolete.

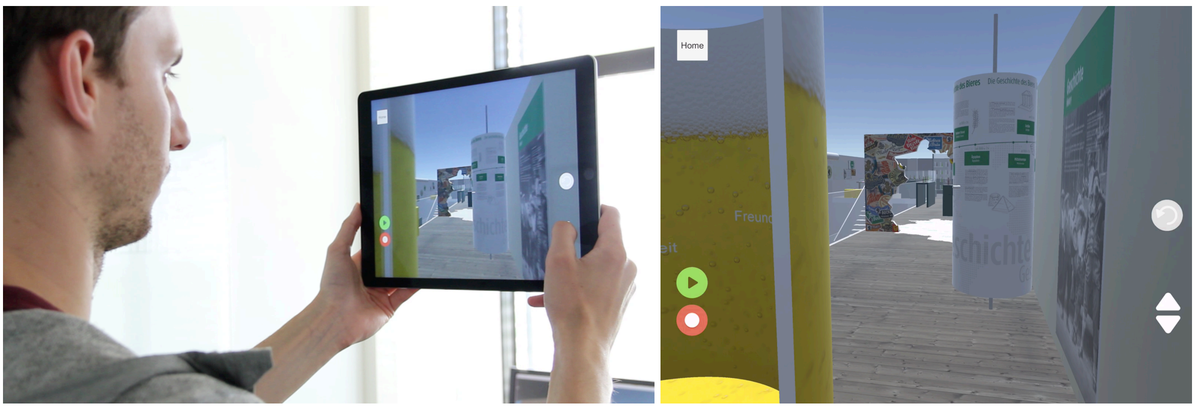


Figure 7: The tablet camera prototype (left) with app screenshot (right)

5.2.3 Tablet Camera Prototype

Camera work is a task that requires the operation of a camera in a real world setting by switching perspectives, places, views, positions, and many other things. On set, camera work is important, and it is costly to start to try out different camera positions with all actors on set. Thus, virtual cameras can be used to plan camera positions and translations before the actual day on the set. We use tablets as "virtual windows" to the scene by integrating orientation and position of the tablet into the experience. With this, users can move the tablet through the room in order to change perspectives and use on-screen controls to further move in the scene. This is natural because by moving the tablet with the own hands through the room, users can directly control the camera orientation in 3D and also easily move through the 3D world with simple additional gestures. The combination of moving the tablet and using simple gestures is easy to comprehend as only a very small set of gestures has to be memorised. For moving, we use the innate capabilities of humans for spatial orientation, Figure 7.

5.2.4 Multi-User Scene Design

This tool focuses on natural distribution of 3D tasks on different devices to support collaboration. This is natural because for complex 3D tasks that can be shared among multiple users (scene design), different input modalities have certain strengths and weaknesses. Here we combine the strengths of VR (3D interaction, immersion, first-person view, 3D placement) and tablet interaction (2D interaction, overview, distance, sharing) in order to improve collaboration and fluid interaction a shared 3D task, see Figure 8.

Figure 8: Resulting scene design (top) of two users created collaboratively using tablet and VR (bottom)

5.2.5 Augmented Reality Scene Creation

Planning and dressing a 3D scene together with multiple people can be a difficult task because only one person can work on a computer, while the others have to instruct that person to place object in the scene using verbal instructions. We present a system that uses augmented reality that allows multiple users to create a 3D scene collaboratively in a table-top manner using tablets and mobile phones. This is natural because locations, scale, and scene layout can be changed in a tangible manner, with every user having the option to alter the scene and add objects to it. It is not needed to operate a complex 2D interface and work in a linear fashion.

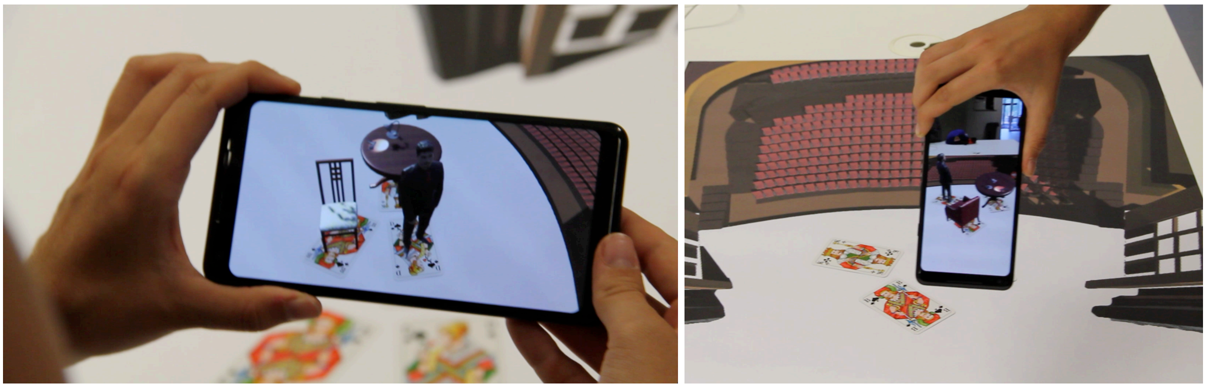


Figure 9: The AR scene creation prototypes allows to layout scenes (left) using cards as markers (right)

5.2.6 Playful 3D Scene Design Shooter

Playful approach to 3D scene design that allows users solve the blank paper problem by shooting a 3D scene that is easy, fun, and playful. It further supports serendipitous exploration through the game mechanics (see Figure 10). This is natural because users can quickly and effortlessly move in 3D (using intuitive game controls), get a good sense of the scene (scale, depth, immersion) and use different tools (weapons) that each have unique features in placing objects. This combination leads to improved creative behaviour, easy and fun access to 3D content creation and serendipitous creations that spark ideas and allow for creative re-framing and context switching. This research prototype is an exception to our NUI concept because it is developed as a desktop application. The reason is that we explored first-person-shooter game mechanics for scene design, which are mainly operated using mouse and keyboard.

Figure 10: Screenshots of the scene design shooter; spawning multiple trees using a grenade (left), final scene design (right)

5.2.7 Tactile VR-Cameras

With virtual cameras, users can explore different shot options and viewing angles in VR for previs. However, camera men miss the tactile expression of a real camera rig in order to create immersive experiences when working with virtual cameras. In this project, we augment the camera experience with tactile sensation using approximate camera models and weights in order to improve immersion and presence for camera men in VR. This is natural because camera men are used to different weights and tactile sensations when working with cameras, giving the film making process a natural feeling. In VR, this is still missing, and we overcome this problem by creating a real camera sensation in VR where we provide natural experience and let camera men perceive the situation to be filmed as more immersive in order to create realistic shots.

5.2.8 Sandbox

Physically interacting with natural materials like sand is a natural way of expression because of the rich haptics, immersion and physical affordances. Crafting with sand triggers emotional associations and have a deeper meaning that are associated with playfulness, creativity, exploration and early childhood memories. We created an augmented sandbox that is complemented by VR for greater immersion and can be used for in-situ exploration of the virtual worlds, see Figure 11.

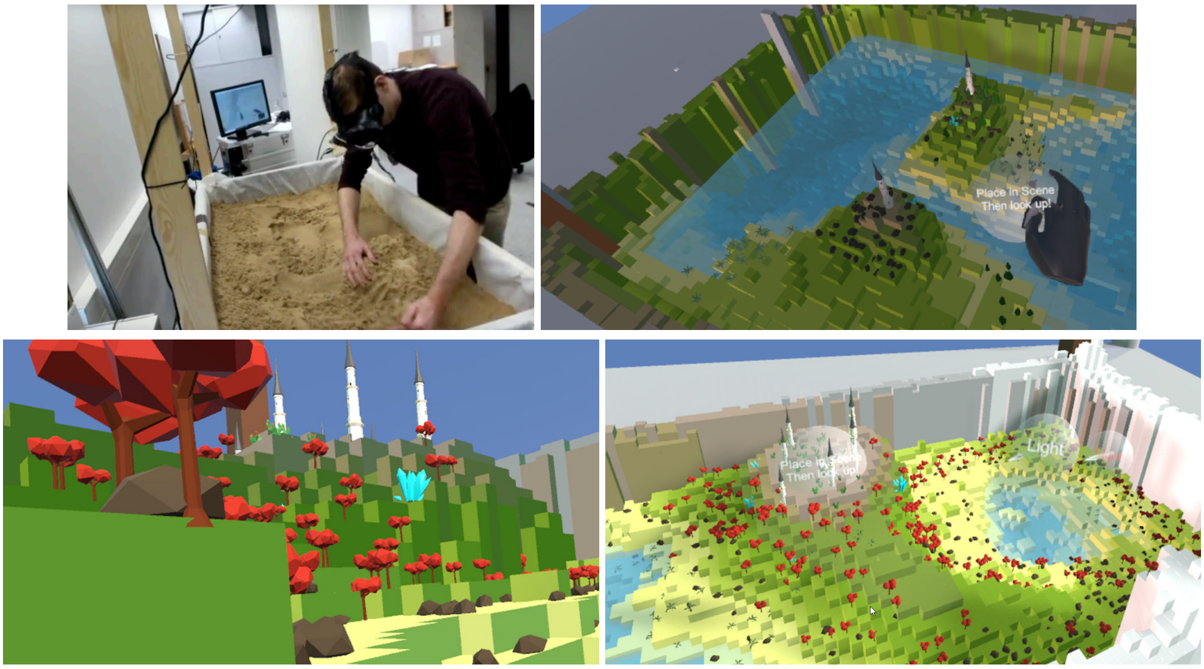


Figure 11: VR Sandbox, top left to bottom right: user crafting a landscape using hands, ego-perspective, first-person view within the landscape, table-top overview

5.2.9 Rapid Prototyping using Free-hand Interaction

Intuitive and fast interaction for rapid creation of 3D objects, see Figure 12. This is natural because users are not required to access graphical interfaces for rapid 3D prototyping and in turn are able to interact without distractions and context/menu switching in order to improve workflow and creative expression, focusing on the work and idea implementation rather on UI interaction that pulls the focus on the work away from the user. We employ hand tracking with the Leap Motion sensor and a VR environment.

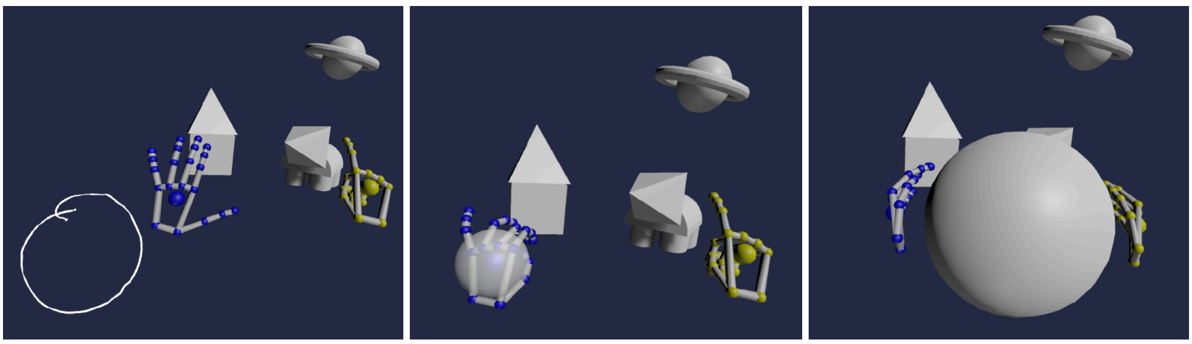


Figure 12: Free-hand interaction for rapid prototyping. Gestures can be used to spawn different objects. In this case, a circle is drawn (left), spawning a circle in the user's hand (middle), which can be scaled using a pinch gesture (right).

5.2.10 Asset Selection through Sketching

Finding suitable assets in a large database for 3D scene creation is a problem because of the sheer number of objects available. Especially in VR, this is problematic because filtering and sorting are difficult because of the missing text input and GUI capabilities that are not suited for a natural VR experience. In this project, we overcome these issues by providing a sketch-based interface where users draw a rough version of the required asset on a 2D plane and are presented with a selection of matching 3D objects. This is natural because users can scribble a rough version without having to be very precise and correct about the visual appearance. The users can make use of their natural way of expression by scribbling, are not overwhelmed by a large database, can select similar objects, and can explore different new objects through scribbling.

5.2.11 Digital Free-Hand Sculpting Bench

Modelling in 3D is complex and hard for inexperienced users because of the 2D perspective, view switching, tool selection, and many other things. Our approach to virtual sculpting centres around VR and free-hand interaction for natural sculpting. We further include haptic feedback for improved precision, as it is sometimes hard to recognize when a tool is touching the model surface. This is natural because users do not have to translate their intent through an interface, instead they can "touch" the model, feel the surface, effortlessly rotate the model with gestures, and can see their creation directly in 3D, making depth perception easy.

In the following section, we present application scenarios that exemplify the use of the different interaction concepts and prototypes in a practical manner. We employ the personas from Section 3 for a more meaningful picture.

6 Application Scenarios

In order to conclude the NUI concept, we present four user scenarios that exemplify how we envision our natural previs tools to be used in practise.

6.1 Theatre Scenario

Imagine a virtual previs studio where users can start from simple sketches and work the scene up to the point that it is ready for production:

To begin with, the theatre director Joseph goes in VR and picks a suitable scene from the library, in his case a theatre setting as he is planning a small piece for his local theatre. He starts to sketch and paint in space using the HTC Vive controllers to rough out his initial vision where he sketches houses, a small forest and some characters directly on stage. Thanks to the 3D perception and interaction, he is able to effortlessly estimate scale, distances, and can directly put his ideas in the 3D space around him. In the next step, Joseph would like to add some more realistic props on stage, giving the whole scene more context by browsing the asset library for curtains, walls, and chairs using his voice, while continuously switching perspectives from actor to audience. As he filters the library using voice commands, the props appear on the stage by a simple button press on the controllers. He continues to add some basic animations for the first scene by letting the characters he placed walk around the stage. He teleports himself to the theatre ceiling, adjusting some basic lighting to the scene in order to create the general mood and feeling. For now, Joseph is happy and exits VR, saving his project. As he is dependent on what the Intendant would say to his ideas, the walks up to the office of Frank, who is sitting at his desk. Both agree to look at the scene by starting the first.stage tools on Franks tablet computer. Together, they inspect the scene by using the tablet as a virtual window to the scene, moving it in the room to naturally explore the scene. With touch gestures, they navigate to different locations in the scene, adding text comments as they go. Frank noted that the animations are not expressive enough to transport the vision both have on the project. After the meeting, Joseph wants to enhance the character animations and puts on his motion suit and VR glasses and finds himself in the body of one of the characters on stage. He uses his body to animate each character after another. Being happy with the animations, but not the orchestration of the different actors, he uses the Vive controller to reposition and retime the animations while standing directly in the scene. After a short inspection with Frank via multi user scene sharing (Frank again on his tablet and Joseph can see the tablet position and orientation directly in his VR scene), both agree on the sketch and add two of the potential actors to the call. Sven and Hagen are participating on their Laptops as they have no other hardware available. Both can enter a spectator mode and chat naturally with Frank and Joseph, discussing the upcoming rehearsal session.

This example represents the continuous and natural switching of input and output modalities that allow for each task (modelling, animation, inspection, acting…) the optimal and most natural way of interaction. Our vision is a system that complements the capabilities of the users by providing dynamic entry points in the system, allowing for flexible use that enables creative expression and tools that artists actually need in order to transport their vision. We further complement the natural usage in the first.stage ecosystem by providing interfaces to traditional devices in order maximise the compatibility.

6.2 Film Production

David is a director of photography working in a medium size film company that is specialized in image films and commercials. His company is tasked with creating a commercial for a new exhibition of an event centre. The problem is: The exhibition is not completely build yet and still needs some time to be finished before they can start shooting. Nevertheless, the director and David already want to plan the commercial in order to give the customer a first impression what they envision and how it will look, so that the customer will approve the budget for the commercial.

David and his team got the building plans and a 3D-model of the exhibition beforehand in order to be able to make a first draft for the commercial. They used the first.stage toolset to create a first previsualization by importing the 3D-model into the software as a new scene. After writing a first script of some persons walking through the exhibition and doing various activities the exhibition has to offer, they place some characters in the virtual scene and animate them. The animation is done by drawing the walking path of the characters and detailed animations are created by recording the motion of a real person with a motion suite. The previs now gives a good impression of what will happen in the commercial and the team wants to show it to the customer.

David and the director have an appointment with the customer at the building site of the exhibition. They could have brought just a video of the previs they have created to convince the customer, but they feel that an interactive presentation in the real exhibition space will be more convincing. David brings his tablet with the first.stage application to the meeting in order to show the previs. He opens the augmented reality application in which only the content of the scene that they created for the previs is shown, but not the scene itself. This is projected on top of the camera image of the tablet moves according to the motion of the tablet. David uses touch gestures to align the virtual model with the real world and then gives it to the customer. He can now watch the virtual previs scene projected onto the real exhibition space through the tablet. Two virtual characters are pointing at a virtual exhibit where in reality is still a blank space. The customer is impressed but there is some minor aspect that bothers him. The plan David's team used to create the previs is outdated and the exhibit will be a few meters further away now. David quickly takes the tablet and selects the exhibit and the people pointing at it by tapping them on the screen. Then, with a finger gesture he moves the selected virtual objects in the virtual space to the position the customer pointed to. Now everything fits perfectly, but the customer has no idea how the final commercial film will look. To show it to the customer, David switches to the camera mode in the software and touches the record button. The animations of the scene start from the beginning and David operates the tablet as if it was a real camera filming the action. He follows the main characters of the commercial through the exhibition and films them from different angles. In the end, he stops the recording and hands the customer the tablet to watch what he just created. The customer is convinced of the result and the presentation and immediately grants the full budget to David's company.

This example shows that new technologies like augmented reality applications with easy and intuitive interfaces that allow for on-site customizations can make a big difference when transporting a vision and convincing others.

6.3 Animation

David Johnson is a Layout Artist at a medium size animation production company, working on their new animated TV-series. The director and some of his colleges have already started creating a virtual scenes with the first.stage tools. David now wants to use the tools to find some interesting shots in the scene and save them to share them with the director.

David first wants to get a realistic impression of the set, in order to be able to better plan his shots. For this he starts the first.stage tools, loads the project and puts on the VR headset. As the set is completely virtual, without VR he would have no chance to get a realistic impression of the scene. He finds himself standing in a futuristic street with flying cars in the air and cartoon-style-persons crowding the street. David feels completely immersed into the scene and immediately picks up the mood. He switches to the time manipulation tool by twisting his hand slightly to the right. The buttons of the controller now represent standard video player controls. He presses play and watches the action of the scene happening around him. The cars start flying in straight lines in the air and the virtual people around him start walking by. The main character of the series pushes trough a crowd of people in front of David. For him it feels like he is really in the virtual world and immediately has some ideas for different shots coming to his mind. He puts up the asset menu by pressing the menu button on the controller and chooses the camera tag. A few different cameras models appear in mid-air in front of him. He chooses his preferred camera model which he also uses in the real world. A new camera appears in front of David in the scene. He grabs it with the controller and moves the virtual camera like it would be attached to the controller. He moves it behind the main character to make a shot closely following the main character. As the character is pushed to left and right while making his way through the crowd, David wants to create the same left and right motion with the camera as if it would be another person following the main character. For this he points at the camera with the left-hand controller and with a press of a button, he is "inside" the camera. In this inside view, he has a big viewfinder with the camera image in front of him and the rest of the scene is slightly blurred. He can focus on the camera image but still has reference to the scene around him. Around the viewfinder there are motion handles for moving and rotation the camera. He grabs the two motion handles with the left and right controller and starts navigating the camera by the motion of his arms as if he would steer a real heavy camera. He now presses play on the controller, starting the action of the scene again. David navigates the camera as he intended it in his vision following the main character through the crowd. After finishing the shot, he goes back to the start time and replays the recorded motion. He watches his shot play trough and presses pause at one point. He is not completely satisfied at this particular point as his motions were a bit too fast when recording. As the motion is recorded by keyframes, he looks at the timeline displayed under the viewfinder. It displays the recorded keyframes at the current time. David reaches to the representation of the keyframe at the current time and pushes it slightly further to the right, meaning it will take a bit more time to reach this keyframe and the camera motion will be slowed down. David is now happy with this shot and continues to make some more shots from different locations.

After David has made a series of shots in the scene, he leaves the VR application, picks up his bag with his tablet and leaves the office. He wants to meet with the director of the series in a cafe nearby. David arrives there a bit early and still has some time to work on the shots he made in VR. He picks up his tablet and opens the first.stage app. After selecting and opening the movie project he is presented with a rough timeline and a list of all the shots with a little preview. He recognizes his shots at the bottom of the list. He drags the shots on the timeline and arranges them in the order David envisioned the scene. He cuts some of the shots and adds some transitions in between using touch gestures. After a few minutes, he has a build a short sequence that can be viewed in the play mode. The director arrives at the cafe and David shows him the sequence he has created. The director is very pleased with the first results but also has some comments. Tapping the "comment" button on the shots which the director commented, David types a few remarks he wants to change later when he is back in the office. When using the VR application again to edit his shots these comments will appear in the "inside the camera" mode and David will remember to work on the remarks of the director.

In this example, we show that virtual reality can help certain tasks to get a realistic and immersed view into a scene which is necessary to judge the scene correctly and have the right foundation to create creative ideas. In the animation domain VR can help this process a lot as the sets are completely virtual and can never be explored in real life. Further, we show that 2D tasks like the arrangement of shots on a timeline are preferably performed on a 2D device like a tablet and not in VR. It also allows for cooperative work between multiple persons.

6.4 Summary

In this section, we presented use case scenarios that are based on the natural interaction concept we introduced in this paper. The use case scenarios should exemplify the concept and aim to give an overview of what possibilities for design emerge when working with the concept and design guidelines. However, it is important to point out that this is only a small selection of possible use cases. In reality, the use cases differ to a large extend based on the production type, personnel involved, and other factors.

7 References

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual Reality Exposure Therapy for Social Anxiety Disorder: A Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*, *81*(5), 751.

Bérard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J. R., & Cooperstock, J. R. (2009). Did Minority Report Get It Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task. In *IFIP Conference on Human-Computer Interaction* (pp. 400–414). Springer.

Blake, J. (2012). *Natural User Interfaces in .Net*. Manning Publications Company. Retrieved from https://books.google.de/books?id=mmMCkgEACAAJ

Bolt, R. A. (1980). *Put-that-there: Voice and Gesture at the Graphics Interface* (Vol. 14). ACM.

Bowman, D. A., McMahan, R. P., & Ragan, E. D. (2012). Questioning Naturalism in 3D User Interfaces. *Communications of the ACM*, *55*(9), 78–88.

Buxton, B., & others. (2007). Multi-touch Systems That I Have Known and Loved. *Microsoft Research*, *56*, 1–11.

Chai, J., & Hodgins, J. K. (2005). Performance Animation from Low-dimensional Control Signals. In *ACM Transactions on Graphics (ToG)* (Vol. 24, pp. 686–696). ACM.

Chen, H., & Sun, H. (2002). Real-time Haptic Sculpting in Virtual Volume Space. In *Proceedings of the ACM symposium on Virtual reality software and technology* (pp. 81–88). ACM.

Datcu, D., & Lukosch, S. (2013). Free-hands Interaction in Augmented Reality. In *Proceedings of the 1st symposium on Spatial user interaction* (pp. 33–40). ACM.

Deci, E., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-determination in Human Behavior*. Springer Science & Business Media.

Design Council. (2005). *Eleven Lessons: Managing Design in Eleven Global Brands: A Study of the Design Process*.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). ACM.

Döring, T. (2016). *A Materials Perspective on Human-Computer Interaction*. University of Bremen, Bremen.

first.stage consortium. (2016). first.stage Grant Agreement - Annex 1 (Part A).

first.stage D1.1 Authors. (2018). Deliverable 1.1: Requirements Definition for Animation Productions.

first.stage D1.2 Authors. (2018). Deliverable 1.2: Requirements Definition for Film Productions.

first.stage D1.3 Authors. (2018). Deliverable 1.3: Requirements Definition for Stage Productions.

first.stage D1.4 Authors. (2018). Deliverable 1.4: Requirements for Visual Effects based on Computer Simulation.

first.stage D1.5 Authors. (2018). Deliverable 1.5: Core Functionality.

first.stage D5.1 Authors. (2018). Deliverable 5.1: Pipeline Integration Concept.

Fitzmaurice, G. W. (1993). Situated Information Spaces and Spatially-aware Palmtop Computers. *Communications of the ACM*, *36*(7), 39–49.

Francese, R., Passero, I., & Tortora, G. (2012). Wiimote and Kinect: Gestural User Interfaces Add a Natural Third Dimension to HCI. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 116–123). ACM.

Frohlich, D. M. (1993). The History and Future of Direct Manipulation. *Behaviour & Information Technology*, *12*(6), 315–329.

Fu, L. P., Landay, J., Nebeling, M., Xu, Y., & Zhao, C. (2018). Redefining Natural User Interface. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (p. SIG19:1–SIG19:3). New York, NY, USA: ACM. https://doi.org/10.1145/3170427.3190649

Galoppo, N., Tekin, S., Otaduy, M. A., Gross, M., & Lin, M. C. (2007). Interactive Haptic Rendering of High-resolution Deformable Objects. In *International Conference on Virtual Reality* (pp. 215–223). Springer.

Galyean, T. A., & Hughes, J. F. (1991). Sculpting: An Interactive Volumetric Modeling Technique. In *ACM SIGGRAPH Computer Graphics* (Vol. 25, pp. 267–274). ACM.

García-Peñalvo, F. J., & Moreno, L. (2017). Special Issue on Exploring new Natural User Experiences. *Universal Access in the Information Society*. https://doi.org/10.1007/s10209-017-0578-0

Guiard, Y. (1987). Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. *Journal of Motor Behavior*, *19*(4), 486–517.

Han, J. Y. (2005). Low-cost Multi-touch Sensing through Frustrated Total Internal Reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (pp. 115–118). ACM.

Herrlich, M. (2013). *3D-Modellierung mit interaktiven Oberflächen* (PhD Thesis). Staats-und Universitätsbibliothek Bremen.

Herrlich, M., Braun, A., & Malaka, R. (2012). Towards Bimanual Control for Virtual Sculpting. *Mensch & Computer 2012: Interaktiv Informiert–allgegenwärtig Und Allumfassend!?*.

Herrlich, M., Krause, M., Schwarten, L., Teichert, J., & Walther-Franks, B. (2008). Multitouch Interface Metaphors for 3D Modeling. *Proceedings of IEEE Tabletops and Interactive Surfaces*, 23–24.

Herrlich, M., Walther-Franks, B., & Malaka, R. (2011). Integrated Rotation and Translation for 3D Manipulation on Multi-touch Interactive Surfaces. In *International Symposium on Smart Graphics* (pp. 146–154). Springer.

Herrlich, M., Walther-Franks, B., Schröder-Kroll, R., Holthusen, J., & Malaka, R. (2011). Proxy-Based Selection for Occluded and Dynamic Objects (pp. 142–145). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22571-0_15

Höllerer, T., & Feiner, S. (2004). Mobile Augmented Reality. *Telegeoinformatics: Location-Based Computing and Services. Taylor and Francis Books Ltd., London, UK*, *21*.

Hummels, C. C. M. (2000). Gestural Design Tools: Prototypes, Experiments and Scenarios.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct Manipulation Interfaces. *Human-Computer Interaction*, *1*(4), 311–338.

Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., & Zigelbaum, J. (2008). Reality-based Interaction: a Framework for Post-WIMP Interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 201–210). ACM.

Jain, J., Lund, A., & Wixon, D. (2011). The Future of Natural User Interfaces. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 211–214). ACM.

Javed, W., Elmqvist, N., Yi, J. S., & others. (2011). Direct Manipulation through Surrogate Objects. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 627–636). ACM.

Khademi, M., Mousavi Hondori, H., McKenzie, A., Dodakian, L., Lopes, C. V., & Cramer, S. C. (2014). Free-hand Interaction with Leap Motion Controller for Stroke Rehabilitation. In *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing*

Systems (pp. 1663–1668). New York, NY, USA: ACM. https://doi.org/10.1145/2559206.2581203

Kim, H., Albuquerque, G., Havemann, S., & Fellner, D. W. (2005). Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling. In *IPT/EGVE* (pp. 191–199).

Kretschmer, U., Coors, V., Spierling, U., Grasbon, D., Schneider, K., Rojas, I., & Malaka, R. (2001). Meeting the Spirit of History. In *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage* (pp. 141–152). ACM.

Lee, J., Chai, J., Reitsma, P. S., Hodgins, J. K., & Pollard, N. S. (2002). Interactive Control of Avatars Animated with Human Motion Data. In *ACM Transactions on Graphics (ToG)* (Vol. 21, pp. 491–500). ACM.

Liu, C. K., & Zordan, V. B. (2011). Natural User Interface for Physics-based Character Animation. In *International Conference on Motion in Games* (pp. 1–14). Springer.

Liu, W. (2010). Natural User Interface- Next Mainstream Product User Interface. In *2010 IEEE 11th International Conference on Computer-Aided Industrial Design Conceptual Design* (Vol. 1, pp. 203–205). https://doi.org/10.1109/CAIDCD.2010.5681374

Malaka, R. (2008). Intelligent User Interfaces for Ubiquitous Computing. In *Handbook of Research on Ubiquitous Computing Technology for Real Time Enterprises* (pp. 470–486). IGI Global.

Malaka, R., Schneider, K., & Kretschmer, U. (2004). Stage-based Augmented Edutainment. In *International Symposium on Smart Graphics* (pp. 54–65). Springer.

Moscovich, T., Igarashi, T., Rekimoto, J., Fukuchi, K., & Hughes, J. F. (2005). A Multi-finger Interface for Performance Animation of Deformable Drawings. *Proc. of User Interface Software and Technology (UIST'05), Seattle, WA, ACM Press*.

Mulloni, A., Seichter, H., & Schmalstieg, D. (2011). Handheld Augmented Reality Indoor Navigation with Activity-based Instructions. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* (pp. 211–220). ACM.

Naumann, A., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A., & Hu\s slein, S. (2007). Intuitive Use of User Interfaces: Defining a Vague Concept. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 128–136). Springer.

Nielsen, J. (1994). *Usability Engineering*. Elsevier.

Nielsen, L. (2013). Personas in The Encyclopedia of Human-Computer Interaction, 2nd Ed. In M. Soegaard & R. F. (eds.). Dam (Eds.). Retrieved from http://www.interaction-design.org/encyclopedia/personas.html

Norman, D. (2013). *The Design of Everyday things: Revised and Expanded Edition*. Basic Books (AZ).

Norman, D. A. (2010). Natural User Interfaces Are Not Natural. *Interactions*, *17*(3), 6–10. https://doi.org/10.1145/1744161.1744163

Oviatt, S. (2012). Multimodal Interfaces. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, 405–429.

Picklum, M., Modzelewski, G., Knoop, S., Lichtenberg, T., Dittmann, P., Böhme, T., … others. (2012). Player Control in a Real-time Mobile Augmented Reality Game. In *International Conference on Entertainment Computing* (pp. 393–396). Springer.

Preim, B., & Dachselt, R. (2015). *Interaktive Systeme: Band 2: User Interface Engineering, 3D-Interaktion, Natural User Interfaces*. Springer-Verlag.

Rädle, R., Jetter, H.-C., Marquardt, N., Reiterer, H., & Rogers, Y. (2014). Huddlelamp: Spatially-aware Mobile Displays for Ad-hoc Around-the-table Collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (pp. 45–54). ACM.

Raskin, J. (2000). *The Humane Interface: New Directions for Designing Interactive Systems*. Addison-Wesley Professional.

Reisman, J. L., Davidson, P. L., & Han, J. Y. (2009). A Screen-space Formulation for 2D and 3D Direct Manipulation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (pp. 69–78). ACM.

Schröder-Kroll, R., Walter-Franks, B., Herrlich, M., & Malaka, R. (2012). Proxy-based 3D Selection. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on* (pp. 161–162). IEEE.

Shneiderman, B. (1982). The Future of Interactive Systems and the Emergence of Direct Manipulation. *Behaviour & Information Technology*, *1*(3), 237–256.

Shum, H., & Ho, E. S. (2012). Real-time Physical Modelling of Character Movements with Microsoft Kinect. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology* (pp. 17–24). ACM.

Veas, E., Grasset, R., Ferencik, I., Grünewald, T., & Schmalstieg, D. (2013). Mobile Augmented Reality for Environmental Monitoring. *Personal and Ubiquitous Computing*, *17*(7), 1515–1531.

Vikram, S., Li, L., & Russell, S. (2013). Writing and Sketching in the Air, Recognizing and Controlling on the Fly. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 1179–1184). ACM.

Vogel, D., Lubos, P., & Steinicke, F. (2018). AnimationVR - Interactive Controller-based Animating in Virtual Reality. In *Proceedings of the 1st Workshop on Animation in Virtual and Augmented Environments*. IEEE. Retrieved from http://basilic.informatik.uni-hamburg.de/Publications/2018/VLS18b

Wagner, J., Huot, S., & Mackay, W. (2012). BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2317–2326). ACM.

Walther-Franks, B., Biermann, F., Steenbergen, N., & Malaka, R. (2012). The Animation Loop Station: Near Real-time Animation Production. In *International Conference on Entertainment Computing* (pp. 469–472). Springer.

Walther-Franks, B., Herrlich, M., Karrer, T., Wittenhagen, M., Schröder-Kroll, R., Malaka, R., & Borchers, J. (2012). Dragimation: Direct Manipulation Keyframe Timing for Performance-based Animation. In *Proceedings of Graphics Interface 2012* (pp. 101–108). Canadian Information Processing Society.

Walther-Franks, B., & Malaka, R. (2014). An interaction approach to computer animation. *Entertainment Computing*, *5*(4), 271–283. https://doi.org/10.1016/j.entcom.2014.08.007

Wenig, D., Schöning, J., Olwal, A., Oben, M., & Malaka, R. (2017). WatchThru: Expanding Smartwatch Displays with Mid-air Visuals and Wrist-worn Augmented Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 716–721). ACM.

Wigdor, D., & Wixon, D. (2011). *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Elsevier.

Yee, K.-P. (2003). Peephole Displays: Pen Interaction on Spatially-aware Handheld Computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1–8). ACM.

Zielke, M. A., Zakhidov, D., Hardee, G., Evans, L., Lenox, S., Orr, N., … Mathialagan, G. (2017). Developing Virtual Patients with VR/AR for a Natural User Interface in Medical Teaching. In *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on* (pp. 1–8). IEEE.

8 Document History

Ver.	Date	Changes	Author
V0.1	25.04.2017	Creation of the document, first aspects of use, hard-ware and initial introduction	Thomas Münder
V0.2	27.04.2017	First concept description, user section	Thomas Münder, Thomas Fröhlich
V0.3	02.05.2017	Extended aspects of use	Thomas Münder
V0.4	09.05.2017	Integration of VR survey results	Thomas Fröhlich
V0.5	10.05.2017	Integration of various figures, object-oriented user interfaces section	Thomas Münder
V0.6	11.05.2017	Added Personas and descriptions	Thomas Fröhlich
V0.7	12.05.2017	Extension of object-oriented user interface section and added direct manipulation section	Thomas Münder
V0.8	17.05.2017	Extended user interface concept, hardware input modalities added	Thomas Münder
V0.9	19.05.2017	Application examples, high level task description	Thomas Münder, Thomas Fröhlich
V1.0	26.05.2017	Introduction added, personas edited, survey edited, integration in project template, references, sent for first review	Thomas Fröhlich
V1.1	29.05.2017	Introduction	Thomas Fröhlich
V1.2	29.05.2017	Final formatting and proofreading	Brigitte Fass, Peter Knackfuß
V1.3	31.05.2017	Final editing and conclusion added	Thomas Fröhlich
V2.0	31.05.2018	NUI concept 2.0	Thomas Fröhlich, Thomas Muender, Dirk Wenig, Tanja Döring, Rainer Malaka
V2.0.1	01.06.2018	Minor corrections	Dirk Wenig
V2.0.2	20.06.2018	Minor editing	Thomas Fröhlich, Dirk Wenig
V2.0.3	18.07.2018	Fixed reference	Dirk Wenig