

Gestures From the Point of View of an Audience: Toward Anticipatable Interaction of Presenters With 3D Content.

Andrey Krekhov

High Performance Computing
Group
University of Duisburg-Essen
andrey.krekhov@uni-due.de

Katharina Emmerich

Entertainment Computing
Group
University of Duisburg-Essen
katharina.emmerich@uni-
due.de

Maxim Babinski

High Performance Computing
Group
University of Duisburg-Essen
maxim.babinski@stud.uni-
due.de

Jens Krüger

High Performance Computing
Group
University of Duisburg-Essen
jens.krueger@uni-due.de

ABSTRACT

Presenting content to an audience is important in several fields, including education, marketing, and entertainment. Therefore, the main goal of the presenter is to transport messages to the audience.

The paper aims to improve the process of message transportation by providing audience-friendly and anticipatable gestures for the presenter to be used for 3D interaction with the content. For this purpose, we first gathered input from a potential audience through a Wizard of Oz experiment and implemented three coherent gesture sets using the Kinect. We conducted an online survey to evaluate the hypotheses regarding the anticipation rate and perceived user experience. In particular, two of our three gesture sets show tendencies to be intuitively predictable by an untrained, uninformed audience. As the three sets differ significantly in the anticipation level, we conclude that future improvements of such gestures would enhance the audience's ability to predict the intended actions even further.

Author Keywords

Gestures; Presentation; Audience; Kinect

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g., HCI): User Interfaces

INTRODUCTION

Being able to transport certain key messages to the audience is what distinguishes excellent presenters. Hereby, the behavior of the presenters is as important as their messages. Improving presentation skills has been the subject of recent research, not to mention the plethora of workshops and schools that provide

courses designed to improve these skills. The latter teach us how to breathe, when to speak up, how to move, when to stop, how to engage, how not to lose the audience, and many more. Other hot topics target the presented material: how to build pitch decks, which font size to use, impacts of images, and how to present live content.

Gestures are an important part of the presenters behavior. Those gestures which can be easily interpreted by the audience greatly enhance the understanding of the content as shown in various studies [15, 39, 4]. On the other hand, less intuitive gestures increase the cognitive load of the audience as the attendees need time to process and understand these movements, even if they might be more ergonomic for the presenter. For that reason, having gestures that are easily linked to what currently happens in the application (what we refer to as *anticipation*) provides a significant improvement for the audience.

Thus, good presenters are willing to learn new techniques, if it rewards them with reduced cognitive load for their audiences. The audience, on the other hand, is not prepared for misleading and counter-intuitive gestures. An ergonomic interaction for one person, the presenter, might therefore still result in misunderstandings on part of many attendees. All these reasons motivate our proposed change of the perspective, i.e., to look at the audience and not only the user (presenter) while designing the gestures.

We focus on the case when the presenter interacts with an application—in 2D or 3D—and the audience both listens and spectates. One common scenario is a medical school where the presenter, or teacher in that case, performs a virtual autopsy on a dataset and explains different specs to the students while interacting with the data. Another scenario, mostly entertaining in its nature, is a digital planetarium. The presenter interactively navigates through the universe while giving insights into astrophysics. In addition, the audience might ask live questions, which then results in, for example, navigation tasks being carried out by the presenter.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2017, May 6-11, 2017, Denver, CO, USA.

Copyright © 2017 ACM ISBN 978-1-4503-4655-9/17/05 ...\$15.00.
<http://dx.doi.org/10.1145/3025453.3025641>

Based on input from planetarium content presenters, we decided to investigate the gestures from the point of view of the audience, instead of focusing our attention to the person who performs the interaction. Our goal is to find out whether gestures can help the audience anticipate the interaction of the presenter with 3D content. That goal poses a first and crucial step to answering the final question: whether it is a benefit to include gesture-based interaction in our daily presentations when it comes to software demonstrations rather than PowerPoint slides. For that purpose, we first ask a potential, yet untrained, audience in a Wizard of Oz study to gather ideas of how the target group would perform these 3D interactions. We then implement and group the gestures in three coherent sets. Finally, we perform an online survey regarding the anticipation rate and perceived user experience to validate the respective hypotheses.

RELATED WORK

Our contribution mainly benefits from related work in two areas: enhancing presentations and improving gesture-based interaction. The former includes techniques and systems that aid a presenter during the presentation. The latter deals with the question of crafting intuitive, comprehensive interaction methods as well as technically improving gesture recognition.

Enhancing Presentations

Giving an engaging presentation is a challenging task and the presenters behavior is as important as her message. This is illustrated in various prominent works on presentations such as Haider et al. [14] or Keith et al. [7]. That literature points out that the behavior of the presenter (such as gestures) has a huge impact on how the presentation is perceived by the audience.

Cuccurullo et al. [6] proposed enhancing public presentations by allowing a natural user interface for the presenter. Their approach, Kinect Presenter (KiP), adds gesture based interaction metaphors for presentation software such as PowerPoint. Several other publications also aim to enhance and improve the default PowerPoint presentation scenario by, e.g., using the Wiimote as an input device [40]. However, to our knowledge, there is no similar work with regard to presenting 3D applications. To improve presentations in large auditoriums, Tan et al. [34] demonstrate an approach relying on a live video view that combines the presenter and the presented material. In addition, the system leverages the possible amount of interaction between the presenter and the local or remote audience.

Roth et al. [28] propose seven techniques that enhance the slide-based transmission of information and, thus, have a positive impact on the quality of a presentation. Another system is *Fly*[18], a prototype presentation system that enhances presentations by adding a visual structure to the underlying content. In particular, a spatial organization based on Mind Maps is applied. In contrast, Trinh et al. [37] focus on the rehearsal stage of the presentation. In particular, their system *PitchPerfect* has been shown to have a positive impact on the overall presentation quality.

Another interesting approach to generate comprehensible gestures is to rely on popular science fiction movies. Filmmakers

have to create futuristic interactions that the audience should be able to understand intuitively. In this area, the work of Figueiredo et al. [9] aims to establish a compilation of hand interactions in Sci-Fi movies and, thus, to generate useful input for further research in that area. Shedroff et al. [31] also point out that Sci-Fi interfaces often mirror the actual expectations of potential users.

Crafting Gestures

Nielsen et al. [26] describe suitable approaches to craft a gesture-based interaction. They also emphasize the lack of a universal gesture vocabulary, which reinforces us conducting a Wizard of Oz experiment for our particular scenario. Similar to our Wizard of Oz procedure, a crowd-sourcing approach was applied by Grijincu et al. [13] to annotate video-based gesture data sets.

Rovelo Ruiz et al. [29] focused their work on gesture-based interaction with omni-directional video. Similar to our approach, the authors asked participants to execute mid-air gestures that they consider to be appropriate. Based on such input, a user-defined gesture set for the given scenario was established. Fikkert et al. [10] explored how uninstructed users gesture when asked to perform basic tasks in the context of large public displays. After rating these outcomes in a second study, a gesture set for explicit command-giving to large displays was established.

Intuitiveness and effectiveness of gestures for large displays in the public space was researched by Hespanhol et al. [16]. The authors focus on simple gestures for the execution of the basic actions of selecting and rearranging items in a large-scale dashboard. Grandhi et al. [12] also aim for naturalness and intuitiveness in gesture production by providing guidelines for the characteristics of gestures and according user mental models in their work.

The work of Soh et al. [32] focuses on the establishment of a set of 3D object manipulation gestures that can be realized with the Microsoft Kinect. The authors note that the focus of gesture crafting is on either reproducing real-world interaction or on the technical possibilities, i.e., gestures that are well recognized by the Kinect. Intuitively, the former approach seems to be more suitable with the audience in mind.

Song et al. [33] propose a handle bar metaphor to resolve the mode switching issue with regard to object manipulation. Since the Kinect is not able to track fine-grained hand rotations robustly, the authors propose to skewer an object with a bimanual handle bar and rely on relative 3D motion instead.

From a more technical point of view, several studies have been conducted to improve the gesture recognition process. Alexander et al. [2] created the *Gestur* framework to improve the optical recognition of hand gestures. A neural network (semi-supervised Fuzzy ARTMAP) allows for incremental online learning and recognition of static gestures, and hidden Markov models serve the same purpose for dynamic gestures. Ren et al. [27] use the Kinect to create a hand gesture recognition system that is especially suited for uncontrolled environments and ignores hand variations and distortions by considering both the color and depth information. Oikonomidis et al. [20]



Figure 1. The original planetarium scenario used as inspiration for our research. The presenter interactively maneuvers through the universe, visiting a number of points of interest. The audience is often asked about the next destination. Without gesture support, an assistant is required to execute the respective commands and stay in sync with the presenter.

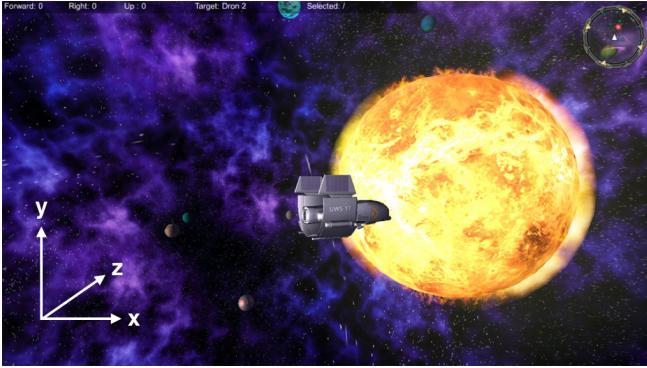


Figure 2. A simplified planetarium scene used for our surveys. The setting includes a number of planets, the sun, and a space ship in the foreground. The space ship is used to accomplish object manipulation tasks. The displayed left-handed coordinate system with z pointing into the scene is used throughout the paper to describe the interaction axes.

aim to recover the full articulation of a human hand captured with the Kinect. Therefore, they handle that task as an optimization problem and solve it with a variant of Particle Swarm Optimization.

Further Application Areas

Aside from the presentation context, other situations that are potential candidates to be enhanced with natural user interfaces have been studied. Particularly in a medical context, the touch-free nature of gesture-based interactions provides significant benefits for aseptic environments and can be applied in operating rooms. The work of Gallo et al. [11] presents such a Kinect-based system for the interactive exploration and manipulation of medical data sets. The system automatically determines the dominant hand and provides interactions such as browsing through data sets, zooming, translating, rotating, and windowing. Instead of using the Kinect, Tani et al. [36] work with a glove-driven interface to meet the requirements of 3D visualization applications for radiological workstations. Combining trajectory recognition with hand posture provides a broader set of domain-specific functionalities.

Mixing Kinect-based posture recognition with virtual reality is also applicable for evaluations of balance training performance [19]. Another way to enhance the Kinect experience is

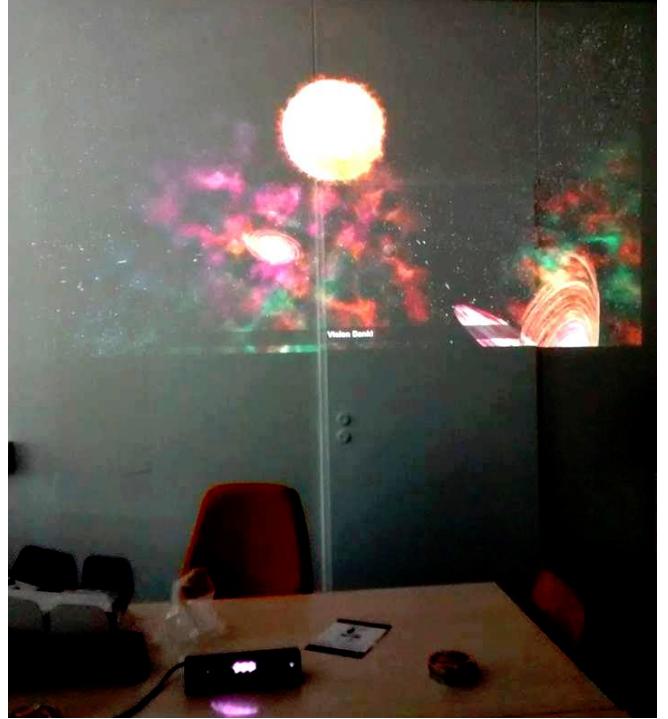


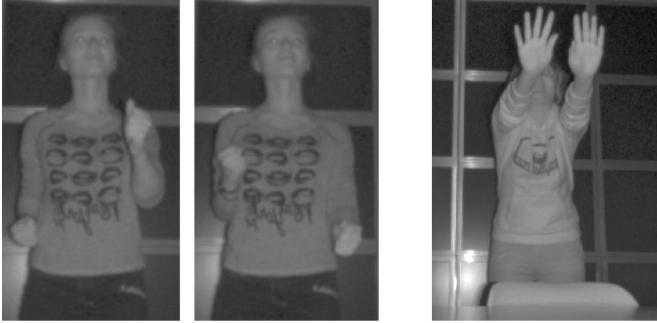
Figure 3. The setup of the Wizard of Oz experiment. The participants faced the Kinect and had to invent gestures to complete given camera and object manipulation tasks for the application projected in front of them. They were told that the Kinect automatically detects and understands their interaction and, thus, allows them to control the application by their preferred gestures. In truth, a wizard, located in another room, monitored the subjects with the Kinect camera and triggered interactions in a timely manner within the target application.

presented in the work of Nebelig et al. [25]. The authors evaluate user-defined Kinect gestures and speech commands with regard to controlling a wall-projected web browser. Carter et al. [5] evaluated combining gestures with gaze. Using the Kinect, the authors describe a novel lounge-style remote-interaction and conduct a user experience and preference study in order to provide a number of design recommendations.

Another area that increasingly relies on natural user interfaces is the car industry. Zobl et al. [42, 41] study the reasonability of using hand gestures to control a number of car systems such as the speaker volume. Other areas often involve multi-user setups. Tang et al. [35] use three Kinect sensors to create a collaborative virtual ball playing environment. Such scenarios might benefit from comprehensive gestures to enhance the player experience.

BUILDING THE GESTURE SETS

We consider the planetarium scenario similar to that shown in Figure 1 to investigate the anticipation level of various gestures. Our underlying scene in Figure 2 represents a similar, simplified space setting including the sun, planets in the background, and a space ship. The tasks of the presenter are limited to navigation in 6 degrees of freedom (6-DoF) and 6-DoF object manipulation. In our case, manipulated object is given by the space ship. Furthermore, a selection option is needed to enter the manipulation mode for the specific object.



moving the camera forward



tilting the camera



rotating the camera



rotating the object

Figure 4. An extract of the proposed gestures. All interactions were captured on video and used to craft the three gesture sets. The superman-like pose (center of the top row) was one of the most prominent input methods to navigate forwards. Certain gestures, such as the shown object rotation, were not included in our final sets due to their poor recognition rate by the Kinect.

As already mentioned by Nielsen et al. [26], there is no universal gesture vocabulary. Apart from relying on the discussed related work, we also consider the input of a potential audience for our precise scenario. For this purpose, a Wizard of Oz study is applied.

Wizard of Oz Study

A first step toward building appropriate, anticipatable gesture sets was to gather ideas about suitable gestures from a typical audience, i.e., subjects who are unexperienced both with natural user interfaces and the target application. For this purpose, we conducted a Wizard of Oz experiment [3, 8, 22, 30]. The overall setup is depicted in Figure 3. The application was displayed on a wall. The Kinect was located on the table in the middle of the room and faced the subject.

Upon entering the room, we introduced the subjects to the application. The story was based on the real planetarium example, i.e., we told the subjects that the software interactively renders the universe and that they would be able to explore an excerpt of it. In the next step, we informed them that their task was to use body movements to complete a set of tasks within the application. Each task would be displayed on the screen and we would notify the subjects upon task completion in the same way. We informed the subjects that their gestures would be automatically recognized by the Kinect and transformed into appropriate actions within the target application. In addition, we notified the subjects that all interactions would be recorded and used for an offline evaluation.

The tasks to be completed consisted of 6-DoF navigation, object selection, and 6-DoF object manipulation. Hence, the subjects received on-screen task commands before each task, such as “*move forwards*” or “*tilt to the left*”. The wizard was located in another room and monitored the subjects via the Kinect camera. The task of the wizard was to trigger the corresponding behavior within the target application in a timely manner. Most subjects experienced a trial period as they were trying out different approaches. Hence, the wizard waited for consistency, i.e., until a consistent gesture was applied for at least 10 seconds. At this point, the wizard initiated an on-screen notification that signaled the task completion to the subject.

Overall, 18 subjects (10 females) between 24 and 27 years old participated in our study. We invited subjects with no experience with the Kinect and gesture-based interfaces in general because an audience is anticipated to contain attendees with no gesture-based interaction experience, and we wanted the presentation to be comprehensive for all subjects.

All interactions were recorded and used as input for the development of our gesture sets. Sample outcomes of the experiment are depicted in Figure 4. Surprisingly, a number of subjects did not limit the interaction to arm gestures and used the whole body instead. On the other hand, various gestures known from related work were also represented. Furthermore, continuous movements dominated over discrete gestures in the experiment.

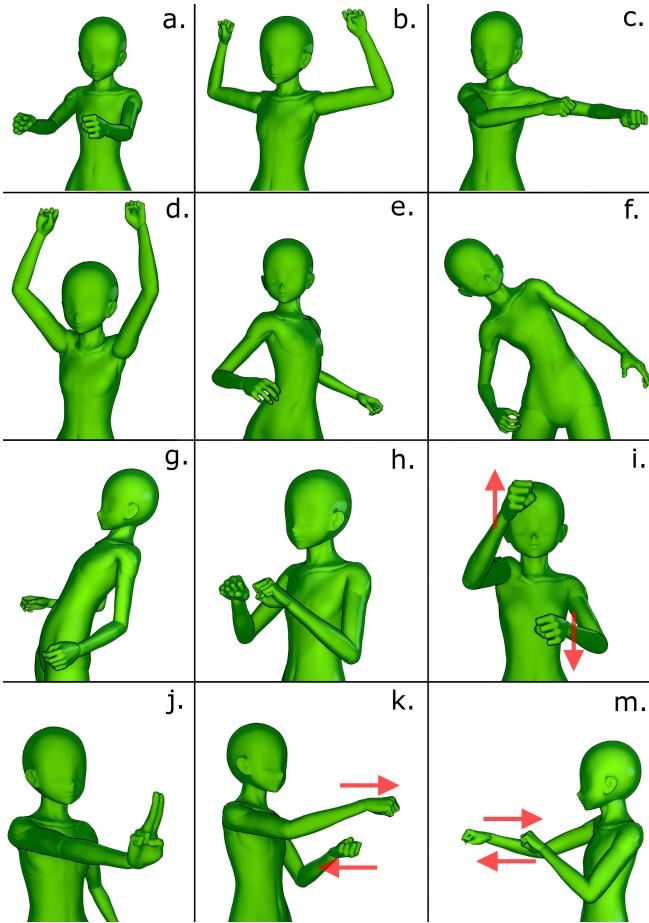


Figure 5. The *green set* is mostly based on bimanual gestures. Gestures in a) - d) are used to move the camera. e) - g) present camera rotation by moving the upper body in the same direction. After selecting an object with j), the presenter moves her clenched fists as shown in h) to move the object. Changing the distance between the fists as in i), k), and m) results in object rotation. Object deselection is performed with the same gesture as the selection, i.e., with j).

Establishing Three Gesture Sets

Based on the results of the described experiment, three different gesture sets were elaborated. For simplicity reasons, we color coded the sets and referred to them as the *green*, *red*, and *blue sets*. To illustrate the gestures, the sketch doll *DesignDoll* [1] was used. Each set covers the four subtasks for camera rotation, camera translation, object rotation, and object translation. Additionally, the *green set* contains an explicit gesture for object selection.

Two criteria were dominant for the creation process. Firstly, the gestures should be consistent within the four subtasks, i.e., a gesture should work for all three axes. Otherwise, it would not be possible to, e.g., modify the object position on multiple axes simultaneously, which is clearly a notable drawback. For instance, these cases often occur in the depicted scenario to achieve certain kinematical effects. Secondly, driven by that same real-world motivation, we favored gestures that permit execution of multiple subtasks simultaneously, e.g., rotating and moving the camera.

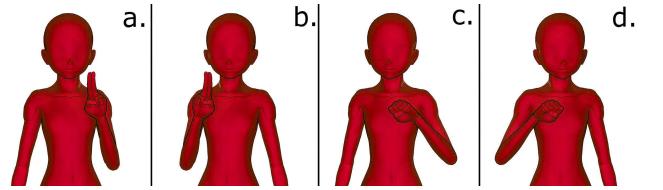


Figure 6. The set consists of minimalist, one-armed gestures. The finger gestures determine whether the camera or the object is manipulated. The functionalities for both arms differ, i.e., the left arm is responsible for translation and the right arm for rotation. a) translates the camera, c) translates the object, b) rotates the camera, and d) rotates the object.

We applied a clustering approach to determine the gesture sets. In the first stage, we filtered out all gestures that did not span a continuous 3D space and, thus, did not allow interaction on all axes. Based on the remaining gestures, three similarity groups were identified: bimanual gestures, one-armed gestures, and full-body movements. Finally, we determined the cluster representatives by accounting for majorities, while keeping the mutual interference as minimal as possible.

Green Set

The *green set* mostly consists of bimanual gestures as depicted in Figure 5. To move the camera, the user has to clench both fists and move them in the corresponding direction. Due to anatomical reasons, the magnitude of drawing back one's arms, i.e. moving the camera backwards, is more limited compared with other directions. Camera rotation is coupled to the movements of the upper body. In particular, the presenter needs to tilt to the sides, bend forward/backward, and rotate sideward.

To select an object, the presenter has to move the right hand toward the object while holding the forefinger and the middle finger stretched out (*lasso gesture*). After the successful selection, the presenters use their clenched fists to move the object in space, such as they would hold it. To rotate the object, the steering wheel metaphor is applied, i.e., the presenters change the distance between their fists in the desired rotation plane.

Red Set

The *red set* contains minimalist, one-armed gestures as shown in Figure 6. The movement of the left arm is dedicated to translation, and the right arm movement controls rotation. Clenched fist and lasso with the appropriate hand trigger between object and camera manipulation, respectively. Hence, left-armed lasso is used to translate the camera, whereas the right clenched fist allows the presenter to rotate the object. No extra gesture is needed for selection since the interactions do not interfere and thus, are distinguishable by the Kinect.

Blue Set

The *blue set* combines the shifting of weight with bimanual gestures and is outlined in Figure 7. Camera translation is performed by shifting the weight in the corresponding directions. Leaning forward/backward results in moving the camera in the corresponding direction. Sidelong camera movement is performed by a weight shift to the appropriate leg. To move the camera upward and downward, the presenter has to perform tiptoeing and a slight squat, respectively. The presenter uses the movement of her open hands to rotate the camera

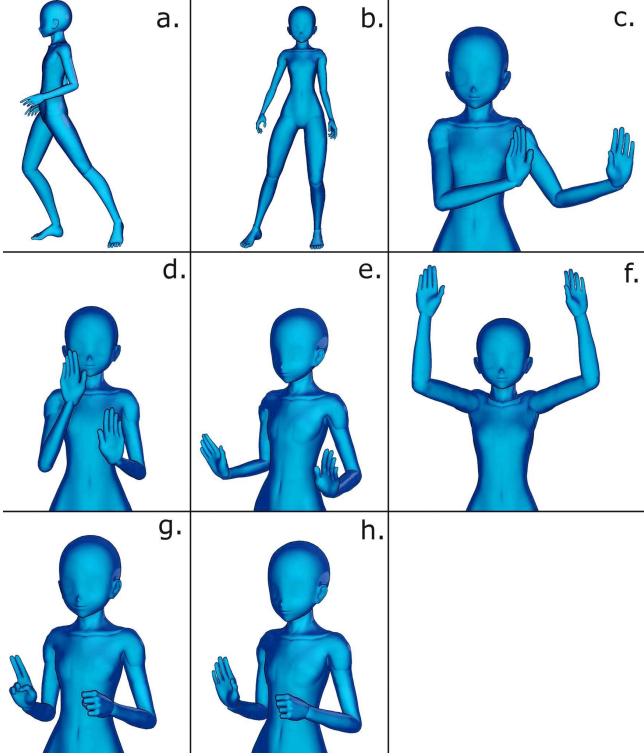


Figure 7. The set relies on weight shift and bimanual gestures. a) - b) are examples for camera movement, c) - f) are gestures executed with two open hands that control the camera rotation. Object manipulation is performed by grabbing the object with the left hand. The lasso gesture in g) is then responsible for object translation, whereas the open right hand performs object rotation.

up, down, left, and right. In order to tilt the camera (rotation around the axis pointing into the screen), one hand has to be moved up, while the other is being moved down.

Object manipulation mode is triggered by clenching the left fist and stands for grabbing the object. The lasso gesture of the right hand can then be used to translate the object in space. To rotate the object, an open hand gesture is used to remain consistent with the camera rotation gestures.

Technical Realization

Finally, to validate the practical feasibility of the established sets, all gestures were implemented and connected to our target application written in Unity3D [38]. Having our real-life planetarium example in mind, we relied on consumer devices to track and recognize the gestures of the presenter. In particular, our implementation uses the Kinect 2 [23] as underlying hardware in combination with the Kinect plugin [24] for Unity.

EVALUATION

We conducted a survey to evaluate the established gesture sets. Our main goal was to study the effects of the gestures on an untrained audience. Therefore, our focus was on the potential anticipation rate, i.e., whether the audience would be able to predict the impact of the gestures on the target application solely by observing their execution by the presenter. To examine that question, a between-subjects design was applied. For

each of the three gesture sets described in the previous section, a dedicated group of subjects was established.

Hypothesis

Our main hypothesis is that the audience, to a certain degree, is able to anticipate effects of the presenters gesture-based interactions with the 3D content. We base the hypothesis on the fact that our gestures were crafted based on user input in our Wizard of Oz experiment and, thus, should be intuitively understandable even by an untrained and unexperienced audience.

Furthermore, we want to evaluate whether the anticipation level and the perceived user experience significantly differ among the three established sets. Such differences could point out potential weaknesses and strengths of the corresponding sets and pave the way for future improvements.

Procedure and Applied Measures

We executed our survey online by sharing the participation link over available social media channels such as Facebook. Each subject was randomly assigned to one of the three groups, i.e., to one of three gesture sets.

After informing the subjects that the study would take about 20 minutes and focus on different gesture-based interactions from the point of view of an audience, the demographical data was gathered. We asked subjects to indicate their gender, age, their highest level of education, their current employee status, and their experience with the Microsoft Kinect. The latter question consisted of the following options: no experience, having heard of it or seen it, having played games with it, and having developed software for it.

In the next step, we instructed the subjects to watch a video of the target application. The video lasted one minute and implicitly contained all possible interactions, i.e., 6-DoF camera manipulation and 6-DoF object manipulation.

Subsequent to the introduction video, the subjects entered the main part of the survey. Since each set consists of 6-DoF interactions for both camera and object manipulation, 12 interactions were presented for each group. The selection gesture of the *green set* has been implicitly built into the corresponding object manipulations to maintain consistency between groups. Furthermore, since each gesture was symmetrical along each axis, only one randomly picked direction was demonstrated. Example interactions include *camera movement to the right* or *object rotation around the y-axis*. A detailed list is given in Figure 9, Figure 10, and Figure 11.

We applied the following process for each demonstrated interaction. Firstly, we showed the subject a screenshot of the initial setup of the target application. For consistency reasons, we used the same initial setup (see Figure 2) for all interactions for all groups. Secondly, we showed a video of the presenter executing the corresponding gesture in front of a blank canvas as depicted in Figure 8 on the left. We asked The subjects to write a free form text about what impact the gesture would have on the initial setup. Thirdly, we demonstrated the same interaction in front of the target application as shown in Figure 8 on the right such that the subjects could observe the real

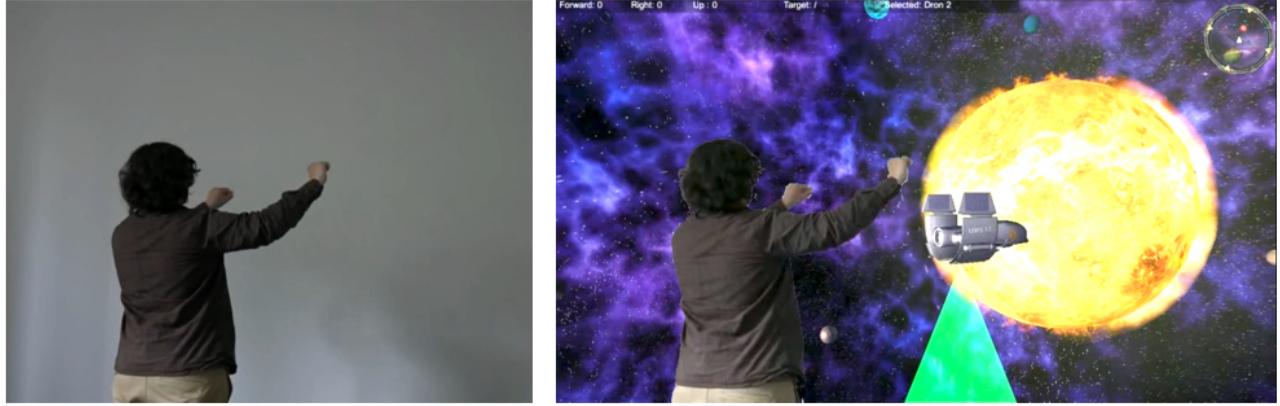


Figure 8. After displaying the initial situation (see Figure 2, the subjects were shown a video of the presenter executing a certain gesture (left image). Subsequent to writing a free form text about their assumption on the outcome, the real outcome was presented in a second video (right image). The green ray occurs only for object manipulation and helps the subject to perceive the outcome. After the second video, the subjects assessed their anticipation rate on a seven-point Likert scale. That setup was repeated for all 12 gestures of the underlying gesture set.

effects of the gesture. At this point, we asked the subjects once again to write a free form text and describe what actually happened in the target application. Finally, we posed the most important question: the subjects were asked to assess how much their predicted outcome matched the real effects. Therefore, we used a seven-point Likert scale, with 1 being *total coincidence* and 7 indicating *no coincidence at all*. For simplicity, we will refer to that scale as our *anticipation scale*.

Subsequent to demonstrating these 12 interactions, the subjects were asked to fill in the User Experience Questionnaire (UEQ) [21]. For our survey, we used the UEQ from [17], which contains the dimensions *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, and *novelty*. The UEQ relies on a seven-point Likert scale ranging from -3 to +3 with larger numbers indicating a more positive outcome for the user experience. As the main goal of this study is to investigate whether the gestures are predictable and understandable, the dimensions *perspicuity* (*is it easy to get familiar with the gestures?*) and *dependability* (*does the system react/work as intended?*) are of special interest in the following analysis.

Results

The survey was completed by 70 participants, 41 female and 29 male. The average age of the subjects was 24.3 years ($SD = 5.48$). Most participants were students or employees. About half of the participants had no prior experience with the Microsoft Kinect ($n = 36$), 15 participants reported to have seen it before, and 19 persons had already used the Kinect for playing or developing before. As participants were randomly assigned to one of the three experimental groups, the number of participants in each group differs slightly with $n = 23$ in the *green set*, $n = 22$ in the *red set*, and $n = 25$ in the *blue set*. However, the groups did not differ significantly regarding age or gender distribution or prior Kinect experiences and, thus, are comparable.

To test our main hypothesis, we focus on the evaluation of the anticipation scale. Keep in mind that lower values indicate higher anticipation, i.e., 1 stands for *total coincidence*. The

average outcomes for the *green set* ($M = 2.99$, $SD = 1.09$) and the *blue set* ($M = 3.28$, $SD = .71$) are within the positive scope.

The mean values of the dimensions *perspicuity* and *dependability* of the UEQ confirm that outcome as they are above zero (on a scale ranging from -3 to +3) for the *green set* (perspicuity: $M = 1.18$, $SD = 1.02$; dependability: $M = .40$, $SD = .66$) and *blue set* (perspicuity: $M = .56$, $SD = 1.17$; dependability: $M = .30$, $SD = .63$). In contrast, our *red set* is within the negative scope regarding its anticipation score ($M = 4.00$, $SD = .86$) with the UEQ dimensions confirming that tendency (perspicuity: $M = .05$, $SD = 1.54$; dependability: $M = -.41$, $SD = .70$).

Comparing the anticipation scores among the three experimental groups using a one-way ANOVA leads to the conclusion that the sets significantly differ with regard to the provided anticipation level ($F(2, 67) = 7.53$, $p = .001$). A detailed comparison using post hoc Bonferroni tests shows that this difference is due to a significant difference of the anticipation score of the *red set* compared to the *green set* ($p = .001$) and the *blue set* ($p = .023$), whereas the *green* and the *blue sets* do not differ significantly ($p = .811$). Hence, the *red set* shows a significantly worse anticipation rate than the other two sets. An ANOVA comparing the scores of the perspicuity and the dependability dimensions shows a similar outcome: there is a significant difference between the three sets regarding perspicuity ($F(2, 67) = 4.65$, $p = .013$) and dependability ($F(2, 67) = 10.02$, $p < .001$). Post hoc tests show that perspicuity differs significantly between the *green set* and the *red set* ($p = .010$), but not between the other sets. Dependability, on the other hand, differs significantly between the *red set* and the *green set* ($p < .001$) as well as between the *red set* and the *blue set* ($p = .001$).

Figure 9, Figure 10, and Figure 11 show detailed outcomes for each gesture for each set. All three sets contain positive and negative peaks, i.e., certain gestures that led either to a high level of anticipation, or to no or rather faulty anticipation.

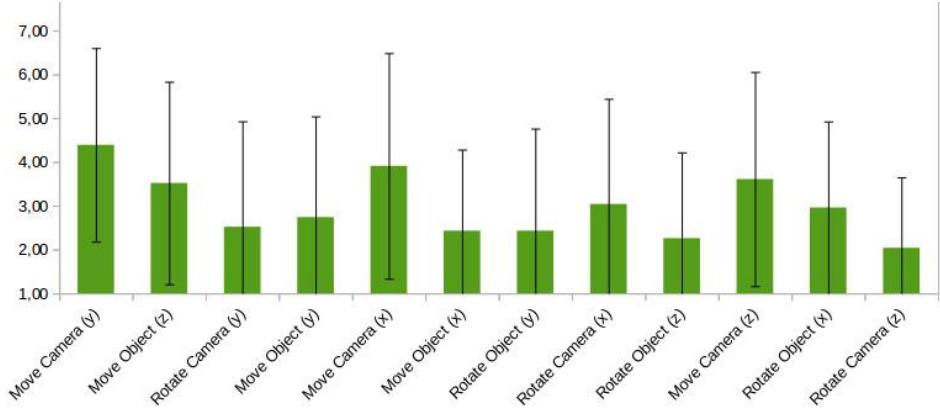


Figure 9. Mean values and standard deviation for the *green set* with regard to our anticipation scale. The order of interactions is encoded on the horizontal axis from left to right and was randomly chosen for each set, yet remained consistent for all subjects within that group. Refer to Figure 2 for the orientation of the axes.

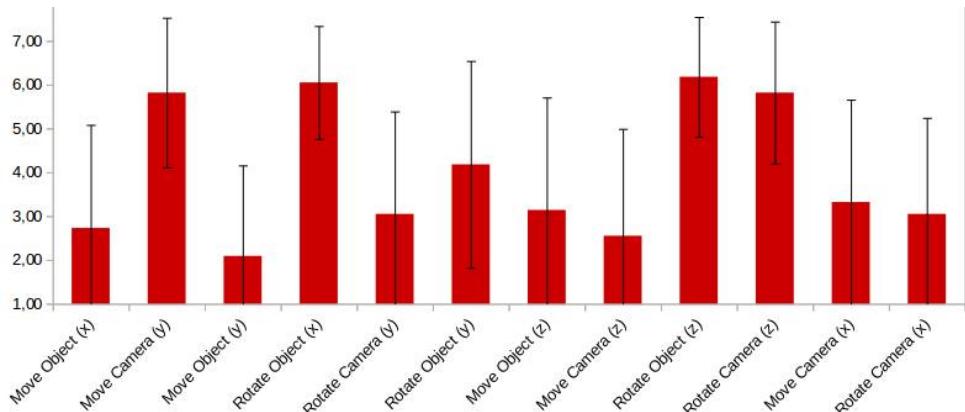


Figure 10. Mean values and standard deviation for the *red set* with regard to our anticipation scale. The order of interactions is encoded on the horizontal axis from left to right and was randomly chosen for each set, yet remained consistent for all subjects within that group. Refer to Figure 2 for the orientation of the axes.

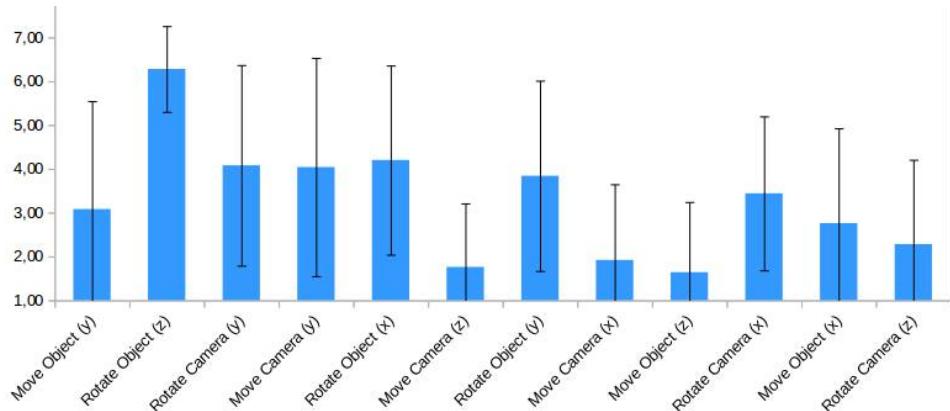


Figure 11. Mean values and standard deviation for the *blue set* with regard to our anticipation scale. The order of interactions is encoded on the horizontal axis from left to right and was randomly chosen for each set, yet remained consistent for all subjects within that group. Refer to Figure 2 for the orientation of the axes.

DISCUSSION

Our anticipation hypothesis holds for two of the three gesture sets. Hence, relying on gestures for 3D interaction tasks might be a valid option to enhance the presentation. In particular, certain gestures render the upcoming interaction predictable, which, in turn, could improve the process of transporting messages to the audience.

The *blue set* offered three interactions with an average anticipation rate below 2 (remember that 1 stands for *total coincidence*): moving the camera forward (*weight shift toward the application*), moving the camera sideward (*weight shift to the appropriate leg*), and pulling the object to oneself (*grabbing with the left hand, pulling with the lasso*). In particular, the weight shift seems to be robust in terms of possible misinterpretations. Hence, our future gesture sets might rely even more on intuitive full body movements instead of hand gestures. However, shifting the weight comes with the disadvantage of being more exhausting, which, in combination with constantly speaking, might make the job as a presenter even more stressful.

The minimalist, *red set* often leads to completely wrong assumptions of the triggered effect. The fact that different arms lead to different interactions despite executing the same gesture seems to be hard to grasp. This result suggests that further gestures should be as distinct as possible and provide analog behavior for both arms.

Surprisingly, for all three sets, we did not observe any learning effects over time. However, the main reason might also be the small number of iterations, since each gesture was shown only once and similarities across different axes did not pay off within the survey. As an example, the object rotation of the *red set* can be considered. After observing object rotations on the x- and the z-axis, the subjects completely failed to anticipate the third rotation (z-axis): the anticipation rate was worse even compared to the first rotation. Other examples where the third interaction of the same kind was anticipated to be at least slightly worse than its predecessors include object rotation (*green and red sets*), object movement (*green and blue sets*), and camera movement (*red set*). Only the anticipation of camera rotation improved in all three sets in its third occurrence.

Hence, we conclude that taking the *green and blue sets* and fixing the negative peaks might further improve the anticipation rate. Furthermore, our survey required a certain commitment with regard to the audience. Firstly, each subject was exposed to the same angle of vision. This is usually not the case in a presentation and might have a heavy impact on how much of the actual gesture is being perceived. Second, the presenter was seen from behind in order to naturally interact with the content on the projection and have the same orientation toward the content as the audience. A number of presenters, however, would prefer to not turn their back to the audience. In that case, certain issues with the directions will occur, even if the presenter does not have to look over same shoulder and uses a dedicated monitor in front of same. In particular, the axis pointing into the scene (moving backward and forward) will

always have a different spatial orientation for the audience and the presenter.

CONCLUSION AND FUTURE WORK

The goal of our paper is to improve and shape the future of 3D content presentations. We investigated gesture-based interaction from the point of view of an audience based on the planetarium example. Certainly, our research question is also transferable to many other 3D content presentation contexts. We established the hypothesis that such an input method for the presenter is able to enhance the overall anticipation rate of the audience. To prove the hypothesis, we first conducted a Wizard of Oz experiment to gather input from a potential audience in order to craft three distinct gesture sets. Finally, an online survey was executed. The results show that two of the three gesture sets have a rather positive outcome regarding the anticipation scale. The poorly comprehended set indicates that the audience benefits from easily distinguishable gestures and pays less attention to whether a gesture is executed by the left or right arm.

The next step toward helping the audience anticipate the interaction would be to carefully analyze each positive and negative peak of the three sets and craft an optimal set for our target application. Once completed, several additional variables have to be introduced. The same survey could be executed offline rather than online, simulating a real audience in order to converge toward the real-world example. Hence, factors such as the point of view, the auditorium layout, partial occlusions, and distance to the presenter might have a significant impact on the usefulness of the gestures. Such adaptations of our research will significantly enhance presentations for large audiences as they will better understand the transported message.

Another direction for further research is to explicitly study the learning curve of the audience. In our setup, the subjects were exposed to each gesture three times, each time on a different axis. A real presentation would, however, contain each gesture multiple times and, thus, the aspect of how easily a gesture can be learned might play an even more important role than the initial anticipation.

REFERENCES

1. DesignDoll. <http://terawell.net/>. Accessed: 2016-21-19.
2. Alexander, T. C., Ahmed, H. S., and Anagnostopoulos, G. C. *An Open Source Framework for Real-Time, Incremental, Static and Dynamic Hand Gesture Learning and Recognition*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, 123–130.
3. Althoff, F., McGlaun, G., Lang, M., and Rigoll, G. *Evaluating Multimodal Interaction Patterns in Various Application Scenarios*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, 421–435.
4. Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training. In *International Conference on Intelligent Virtual Humans*,

- Lecture Notes on Computer Science (Edinburgh, UK, Aug. 2013), 116–128.
5. Carter, M., Newn, J., Velloso, E., and Vetere, F. Remote gaze and gesture tracking on the microsoft kinect: Investigating the role of feedback. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, ACM (New York, NY, USA, 2015), 167–176.
 6. Cuccurullo, S., Francese, R., Murad, S., Passero, I., and Tucci, M. A gestural approach to presentation exploiting motion capture metaphors. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, ACM (New York, NY, USA, 2012), 148–155.
 7. Curtis, K., Jones, G. J. F., and Campbell, N. Speaker impact on audience comprehension for academic presentations. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI 2016, ACM (New York, NY, USA, 2016), 129–136.
 8. Dahlbäck, N., Jönsson, A., and Ahrenberg, L. Wizard of oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*, IUI '93, ACM (New York, NY, USA, 1993), 193–200.
 9. Figueiredo, L. S., Gonçalves Maciel Pinheiro, M. G., Vilar Neto, E. X., and Teichrieb, V. An open catalog of hand gestures from sci-fi movies. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, ACM (New York, NY, USA, 2015), 1319–1324.
 10. Fikkert, W., van der Vet, P., van der Veer, G., and Nijholt, A. *Gestures for Large Display Control*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, 245–256.
 11. Gallo, L., Placitelli, A. P., and Ciampi, M. Controller-free exploration of medical image data: Experiencing the kinect. In *Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on* (June 2011), 1–6.
 12. Grandhi, S. A., Joue, G., and Mittelberg, I. Understanding naturalness and intuitiveness in gesture production: Insights for touchless gestural interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 821–824.
 13. Grijincu, D., Nacenta, M. A., and Kristensson, P. O. User-defined interface gestures: Dataset and analysis. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, ACM (New York, NY, USA, 2014), 25–34.
 14. Haider, F., Cerrato, L., Campbell, N., and Luz, S. Presentation quality assessment using acoustic information and hand movements. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (March 2016), 2812–2816.
 15. Hartley, R., Elsayed, A., and Pesheva, M. Communication and design in multimodal presentations. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (July 2007), 923–924.
 16. Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., and Kay, J. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, PerDis '12, ACM (New York, NY, USA, 2012), 6:1–6:6.
 17. Hinderks, A., Schrepp, M., Rauschenberger, M., and Thomaschewski, J. Konstruktion eines Fragebogens für jugendliche Personen zur Messung der User Experience. Usability Professionals 2012, September 2012.
 18. Holman, D., Stojadinović, P., Karrer, T., and Borchers, J. Fly: An organic presentation tool. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, ACM (New York, NY, USA, 2006), 863–868.
 19. Hsieh, W., Chen, C., Wang, S., Chen, Y., Hwang, Y., and Lai, J. Combination of the kinect with virtual reality in balance training for the elderly. *Engineering*, 5 (2013), 171–175.
 20. Iason Oikonomidis, N. K., and Argyros, A. Efficient model-based 3d tracking of hand articulations using kinect. In *Proc. BMVC* (2011), 101.1–101.11. <http://dx.doi.org/10.5244/C.25.101>.
 21. Laugwitz, B., Held, T., and Schrepp, M. *Construction and Evaluation of a User Experience Questionnaire*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, 63–76.
 22. Maulsby, D., Greenberg, S., and Mander, R. Prototyping an intelligent agent through wizard of oz. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, ACM (New York, NY, USA, 1993), 277–284.
 23. Microsoft. Kinect hardware. <https://dev.windows.com/en-us/kinect/hardware>, 2016.
 24. Microsoft. Kinect v2 plugin for unity. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&sqi=2&ved=0ahUKEwiDjcft6szNAhX CtRQKHZobB1cQFggcMAA&url=http%3A%2F%2Fgo.microsoft.com%2Ffwlink%2F3FLinkID%3D513177&usg=AFOjCNGZ4Z6J1e9X99V18nod1smaYQTQOO&sig2=J6RXEWs-J0bfY9bM-t5tSw&bvm=bv.125801520,d.d24&cad=rja>, 22.10.2014. Accessed: 2016-21-19.
 25. Nebeling, M., Huber, A., Ott, D., and Norrie, M. C. Web on the wall reloaded: Implementation, replication and refinement of user-defined interaction sets. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, ACM (New York, NY, USA, 2014), 15–24.
 26. Nielsen, M., String, M., Moeslund, T. B., and Granum, E. A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction (2003). 409–420.

27. Ren, Z., Meng, J., Yuan, J., and Zhang, Z. Robust hand gesture recognition with kinect sensor. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, ACM (New York, NY, USA, 2011), 759–760.
28. Roth, S., and Schulz, C. A pattern-based approach to presentations using slide facilities. In *Proceedings of the 18th European Conference on Pattern Languages of Program*, EuroPLoP '13, ACM (New York, NY, USA, 2015), 15:1–15:12.
29. Rovelo Ruiz, G. A., Vanacken, D., Luyten, K., Abad, F., and Camahort, E. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 4077–4086.
30. Salber, D., and Coutaz, J. *Human-Computer Interaction: Third International Conference, EWHCI '93 Moscow, Russia, August 3–7, 1993 Selected Papers*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993, ch. Applying the Wizard of Oz technique to the study of multimodal systems, 219–230.
31. Shedroff, N., and Noessel, C. Make it so: Learning from sci-fi interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, ACM (New York, NY, USA, 2012), 7–8.
32. Soh, J., Choi, Y., Park, Y., and Yang, H. S. User-friendly 3d object manipulation gesture using kinect. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '13, ACM (New York, NY, USA, 2013), 231–234.
33. Song, P., Goh, W. B., Hutama, W., Fu, C.-W., and Liu, X. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 1297–1306.
34. Tan, K.-H., Gelb, D., Samadani, R., Robinson, I., Culbertson, B., and Apostolopoulos, J. Gaze awareness and interaction support in presentations. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, ACM (New York, NY, USA, 2010), 643–646.
35. Tang, T. Y., Winoto, P., and Wang, R. Y. Having fun over a distance: Supporting multiplayer online ball passing using multiple sets of kinect. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, ACM (New York, NY, USA, 2015), 1187–1192.
36. Tani, B. S., Maia, R. S., and Wangenheim, A. v. A gesture interface for radiological workstations. In *Proceedings of the Twentieth IEEE International Symposium on Computer-Based Medical Systems*, CBMS '07, IEEE Computer Society (Washington, DC, USA, 2007), 27–32.
37. Trinh, H., Yatani, K., and Edge, D. Pitchperfect: Integrated rehearsal environment for structured presentation preparation. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 1571–1580.
38. Unity. The leading global game industry software. <https://unity3d.com/public-relations>, 08.06.2016. Accessed: 2016-21-19.
39. Xiang Cao, Eyal Ofek, D. V. Evaluation of alternative presentation control techniques. In *ACM SIG CHI 2005*, ACM (January 2005).
40. Yang, Y., and Li, L. Turn a nintendo wiimote into a handheld computer mouse. *IEEE Potentials Volume: 30, Issue: 1* (2011), 12–16.
41. Zobl, M., Geiger, M., Bengler, K., and Lang, M. A usability study on hand gesture controlled operation of in-car devices. In *Poster Proceedings HCII 2001 (New* (2001).
42. Zobl, M., Nieschulz, R., Geiger, M., Lang, M., and Rigoll, G. *Gesture Components for Natural Interaction with In-Car Devices*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, 448–459.