Embodied Interactions for Novel Immersive Presentational Experiences

Fabrice Matulic

Lars Engeln

Christoph Träger

Raimund Dachselt

Interactive Media Lab Dresden
Technische Universität Dresden
{firstname}.{lastname}@tu-dresden.de

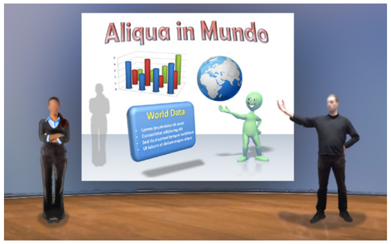


Figure 1: Two presenters represented by an avatar (active presenter, right) and a silhouette (passive presenter, left) in the digital background content.

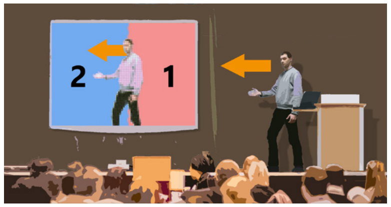


Figure 2: A virtual representation of the speaker is shown within the projected presentation content. Here, the presenter walks from right to left to initiate a transition between two scenes, where the separation line corresponds to the presenter's body.

Abstract

In this work, we introduce and propose a preliminary realisation of a concept to enhance live multimedia presentations, where presenters are directly integrated in their presentation content as interactive avatars. Using multimodal input, especially body gestures, presenters control those embedded avatars through which they can interact with the virtual presentation environment in a fine-grained fashion, i.e. they are able to manipulate individual presentation elements and data as virtual props. The goal of our endeavour is to create novel immersive presentational experiences for live stage performances (talks, lectures etc.) as well as for remote conferencing in more confined areas such as offices and meeting rooms.

Author Keywords

Live gesture-controlled presentations; embodied avatars; embodiment; interactive presentational experiences; gestural interaction.

ACM Classification Keywords

H.5.2. User Interfaces

Introduction

When giving multimedia presentations, lectures or video-conferences, presenters typically control the sequencing of slides or views through classic (indirect)

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892501

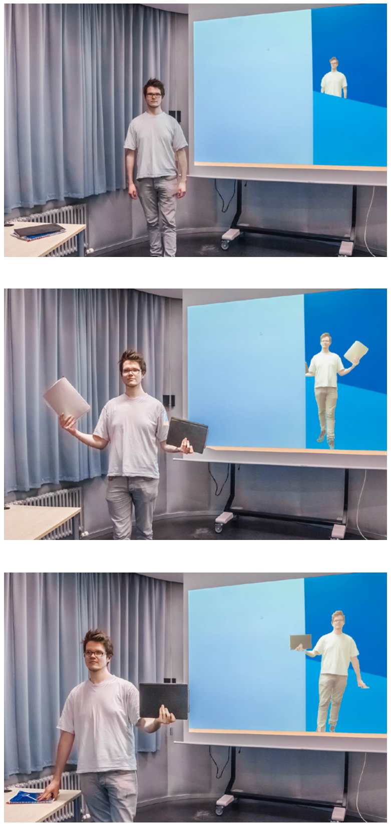


Figure 3: Presenter integrated as video avatar in the presentation 3D space. This allows, among other things, depth-based pixel thresholding to create appearance effects.

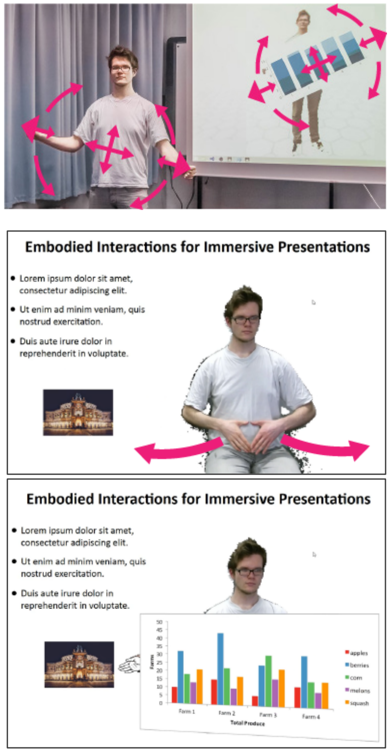
input methods such as keyboard, mouse or remote controls, this, in a more or less scripted manner. The presentation content and the presenter live in entirely different realms and thus the latter is not able to directly and freely interact with the former and its data elements. This separation is also a problem for audiences, as it is difficult for them to focus both on the presenter and the presentation content displayed on a screen, especially when the two are far apart, or worse, when only one can be displayed at a time (as in videoconferencing).

The aim of this work is to significantly enhance the expressiveness and level of integration in such presentations by allowing presenters to virtually immerse themselves in the digital supporting environment and to actively engage with their presentation content via a wide range of multimodal interactions, especially body gestures, but also speech, gaze and tangible-based input. One of the main application scenarios is to empower public speakers, lecturers, teachers, demonstrators etc., who traditionally rely on slideware and the like for live presentations, to become an active part of their digital content through avatar-embedding. Through those persona proxies, presenters would be able to smoothly and naturally interact with the data and individual graphic elements of their presentations. Another target setting is telepresence, where, in addition to manipulating presentation elements in various ways, remote participants would be able to perform elaborate interactions with each other to support their collaborative work and enrich their communication.

Related Work

Probably the easiest method to embed a presenter in their presentation content, which has now become a standard feature of web-conferencing platforms, is to include a live video feed of the speaker inside a frame located in a corner of the presentation screen [2, 4]. However, because that frame is separate from the window containing the presentation data, it is not possible for presenters to directly point at items and perform deictic gestures in relation to that content.

A modern way to integrate the presenter in the content for live presentations is to use a 3D holographic system such as the Musion Eyeliner [6]. Such appliances allow video feeds of people or content to be projected on a transparent film or screen placed on the stage to create the illusion that, from the audience's perspective, they are integrated within it [24]. In more confined meeting spaces, systems like Personify [7] and ImmerseBoard [18] overlay an image of the presenter on top of the presentation content so that the latter appears in the background, similar to computer-generated maps shown behind news presenters in weather forecast broadcasts. In those solutions, the presenters can point at presentation items on the hologram or the overlay and perform well-timed body movements to follow scripted animations; however, they are not able to dynamically interact with individual elements. Presenters are not an integral part of the presentation and the only gestures available to them are to control the general flow (e.g. wave a hand to move to the next/previous slide etc.), or, as in [17], to insert digital content inside a physical rectangular frame held by the user. Those types of control gestures are also implemented in a number of Kinect-based presentation tools, which do not support person embedding [13, 21, 25]. As for dynamic presenter integration within the presentation content but without any virtual representation, a rudimentary proof-of-concept prototype of a text-based



Embodied Interactions for Immersive Presentations

- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Ut enim ad minim veniam, quis nostrud exercitation.
- Duis aute irure dolor in reprehenderit in voluptate.

Embodied Interactions for Immersive Presentations

- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Ut enim ad minim veniam, quis nostrud exercitation.
- Duis aute irure dolor in reprehenderit in voluptate.

Figure 4: The presenter can manipulate presentation elements such as images and charts with hand gestures. In the two examples above (one in a standing and another in a sitting condition), the user grabs a virtual chart with his two hands and rotates/scales it to present it to the viewers. The chart can also be summoned using an associated trigger gesture at predefined time intervals or locations in the virtual space. Once he is finished with the chart, the presenter can make it disappear by clapping his hands.

presentation system exists, where paragraphs wrap around the presenter physically standing in front of the screen and lines of text can be "pushed" off slides [22]. Since only the direct influence of the physical body as an occluding entity is considered, more complex interactions involving the interplay of a virtually embedded presenter-avatar and presentation elements are however not possible.

Another popular use of virtual presenters is in agent-based systems, i.e. where the presenter image is not the real-time embodiment of a human speaker, but an independent virtual persona obeying pre-recorded or programmed behaviours [5, 8, 9]. Elementary kinds of lifelike agents have often been used at exhibitions, trade shows, airports etc. to guide the public [1, 3]. In the research literature, there are examples of virtual presentation agents (namely 3D articulated figures) that can read out text and perform a variety of gestures following scripted commands [23, 28]. Some more elaborate systems make it possible for their agents to react and adapt according to spectators' attention [16] and input [20]. A hybrid approach, including a virtual agent and a real presenter, where the former serves as a training assistant to help the latter, has also been proposed [27]. As with the live streaming solutions, however, those agents are hardly capable of fine-grained interactions with the presentation elements and their animation patterns are limited by their underlying programs or scripts.

The ability of multi-sensor input devices to record multimodal data of human activity (video, speech and skeletal joints) has also been exploited for the evaluation of presenter performance. Specifically, body language, visual cues and speech patterns obtained from sensor data can be automatically analysed to estimate people's presentation skills [12, 15]. Some prototypes even include feedback loops with virtual audiences [11] or information displayed on head-mounted displays [26, 29] to assist the presenter on-the-fly.

Concept and Research Challenges

As mentioned above, in live multimedia presentations, the existence of two centres of attention, i.e. the presenter and the presented content on the screen, makes it difficult for the audience to focus on both at the same time. The introduction of a live embodied representation of the speaker within the digital presentation content (as with Personify [7]) is a first step towards providing a solution to that problem. That representation can be a direct video feed of the presenter as captured by the camera, but —depending on the situation, the presentation content and the presenter's style (among other factors)— it could also take other shapes, e.g. avatars, silhouettes, symbolic figures etc. Figure 1 shows an example of a two-person presentation, in which the active presenter is represented by a solid 3D avatar and the passive presenter by a semi-transparent silhouette.

Within this vision of immersive embodied presentations, the main contribution we would like to make is an exploration of rich interaction possibilities with the presentation data, that is, the ways the immersed presenter can meaningfully and efficiently connect with the different digital elements contained in the virtual presentation world, and trigger special effects through appropriate input modalities. Those interactions and modalities likely vary depending on the data elements to manipulate and their associated effects as well as the presentation context (live stage presentation where

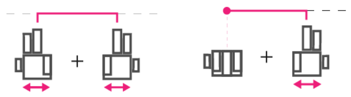


Figure 5: Example of hand postures defining symmetric (left) and asymmetric (right) two-handed range-controlling interactions.

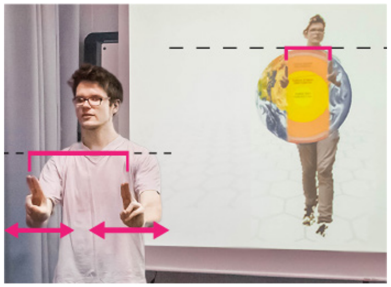


Figure 6: Using a symmetric bimanual expand gesture with lasso hand poses, a user "opens" a section of an earth model in order to reveal its stratigraphic layers. The exposed section corresponds to the space between the two hands.

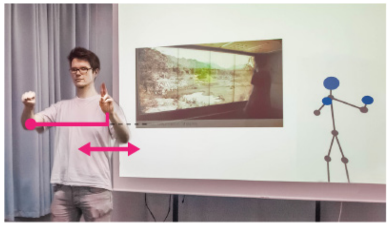


Figure 7: Example of an asymmetric bimanual gesture to control the playback position of a video.

all limbs are available for gesturing or sitting position in which only the upper body can perform limited movements). The greatest challenge there is to design robust interactions that can easily blend in the natural presentation flow, with little risk of being accidentally activated by casual gestures, while at the same time not appear incongruous to the viewers.

Tightly associated with the aforementioned issues is the matter of feedback provided to presenters. Since they interact with digital presentation elements only indirectly (via avatars), we anticipate that adequate feedback and feedforward mechanisms will be necessary to respectively show the system's response to detected input and indicate what actions can be performed to trigger available effects. Those indications could appear within the presenter view on the presenter's notebook screen or manifest themselves through other means when the presenter is moving on the stage, e.g. audio/haptic feedback on phone or wearable device, visual aids on smart glasses.

Finally, we are interested in evaluating how such a system is perceived by people, specifically, whether the presenter-audience dialogue is indeed enriched and to what extent that is the case. In particular, we wish to assess how both parties are able to relate to the presentation content beyond the likely initial "wow" effect and possible distractions caused by the animations. In other words, how do presenters and listeners perceive this kind of immersive experience after they become used to it and how does it enhance communication? Another important usability aspect concerns the learnability of the system, i.e. how easy it is for presenters to master the new skills required to efficiently interact with the virtual presentation environment and its elements. What are those learning costs compared to classic presentations or, simply put, what is the learning overhead?

Realised Interaction Concepts

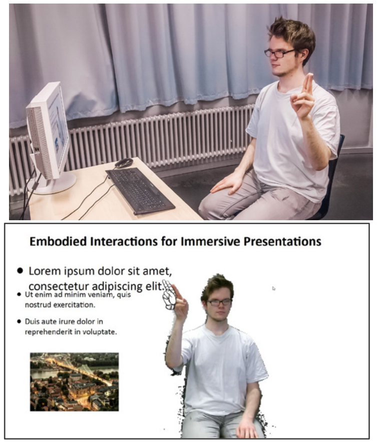
Towards realising the above vision, we started developing a prototype, where for this iteration we focused on body-based interactions captured by a Kinect V2. The camera is placed in front of the presenter on a pulpit or table (for example next to the laptop used for the presentation) or directly on the stage. In the future, we envision arrays of unobtrusive sensors integrated in mobile devices and/or in the environment that are able to reliably track multiple presenters moving on large stages.

Appearance of Embedded Presenters

In our current implementation, the user is materialised by RGB pixels mapped to the 3D pixel data obtained by the depth camera. Thus, we can represent the presenter with a direct video projection in 3D space (Figure 3), similar to Personify [7] and ImmerseBoard [18], or plain pixels for silhouettes (whose intensity we can also modify depending on the distance to create subtle depth effects). We are in the process of integrating 3D models rigged to the Kinect skeleton to support character-based avatars. We are also investigating other visual metaphors, in which presenters, for example, can directly embody data elements themselves.

Example Interactions

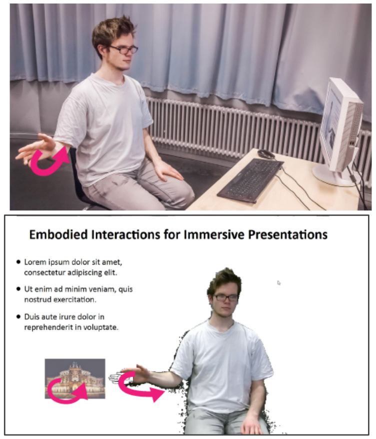
Central to the immersion of presenters in their digital environment is their ability to interact with the elements contained in it. Where prior work involving real-time body-controlled avatars in virtual environments tended to focus on applications such as games and ar-



Embodied Interactions for Immersive Presentations

- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Ut enim ad minim veniam, quis nostrud exercitation.
- Duis aute irure dolor in reprehenderit in voluptate.

Figure 8: Presenters can point at particular items and cause them to pop out or become highlighted to indicate the current topic.



Embodied Interactions for Immersive Presentations

- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Ut enim ad minim veniam, quis nostrud exercitation.
- Duis aute irure dolor in reprehenderit in voluptate.

Figure 9: Repeated hand-closing gestures above an image stack can be used to flip through the different pictures.

tistic performances, our exploration of the avatar-content dialogue takes place in the context of presentations and their multimedia components. Thus, our approach concentrates on typical presentation data elements such as text blocks, images, videos, charts, shapes etc., for which we seek suitable manipulative gestures and effects. We implemented a number of such gestures to demonstrate those presenter-data interaction possibilities.

Arm/Hand Interactions

The most straightforward manipulation of presentation elements in a virtual environment is probably the ability to grasp and move them in the 3D space. This is naturally done with arm/hand interactions, which, it should be noted, can be performed both in standing and sitting positions and therefore are theoretically suitable for stage presentations as well as video-conferences in an office. In the latter case, however, one must make sure that very little physical body movement is required to be able to reach desired objects.

To take hold of or engage interaction with virtual objects, users simply perform a hand-closing gesture on them (grasp). Acquisition of floating objects can be facilitated by allowing hand-closing gestures above or in the vicinity of their projection on the 2D screen. The target objects then fly to their fist as if attracted by a Star-Wars-like force. Those grabbed objects can then be moved around in the 3D world using natural arm movements. Furthermore, while one hand is "holding" the object, the other hand can be used to point at specific content on the figure for explanatory purposes, similar to holding a real cardboard chart to an audience. Using a two-handed grab-expand motion, users are also able to rotate and stretch elements between their hands (Figure 4). Within a presentation performance, we imagine that presenter avatars could be seen grabbing minimised presentation objects lying or floating around them, bringing them to the foreground and enlarging them to show them to the audience. When finished with an element, the presenter could dismiss it with a throwing gesture, whereupon the element would either automatically return to its original place or simply disappear.

To allow the user to explicitly engage manipulation modes (and thus distinguish those interactions from casual arm movements that naturally occur during presentations), we make use of specific hand postures. The Kinect SDK provides support for the recognition of three different hand states: closed, open and lasso (a pointer hand with both the index and middle finger up). We support symmetric and asymmetric hand gestures using respectively same and different hand states for the two hands (Figure 5). For instance, we use the closed posture for both hands to initiate bimanual grabbing and stretching actions (Figure 4). Another example of symmetric two-handed interactions is a double lasso gesture to create temporary visualisation windows into particular sections of a multi-layer object. Specifically, the user joins their two hands in the lasso state and moves them apart to reveal a section of a hidden layer between them (Figure 6).

Asymmetric hand gestures, for their part, allow differentiation of hand roles and can be used, for instance, to define start and end points of a sequence or an action. Figure 7 shows an example of such an asymmetric bimanual range-defining gesture, in which a closed hand pose signifies a starting or anchor point and the second hand in a lasso state moves to set the playback posi-

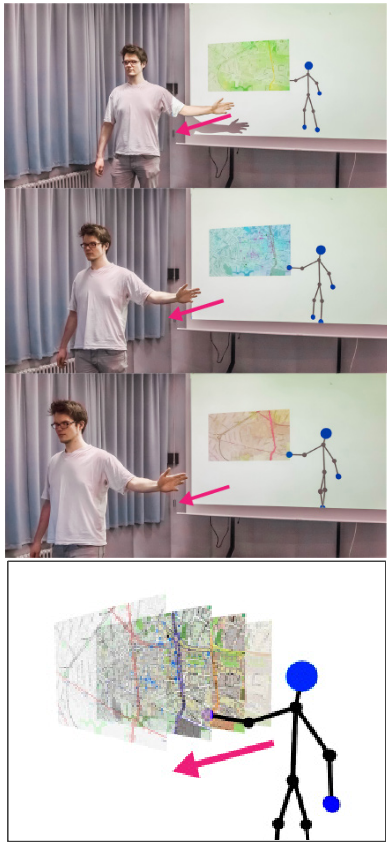


Figure 10: Through the avatar, the presenter can slice through a stack of images, where the position of the body or a stretched arm selects the picture to be shown to the audience. The presenter view displayed on the presenter's personal computer (bottom) shows a different perspective of the scene so that he can see the images available for selection in the stack.

tion of a video. This technique can be easily extended to 2D gesturing by using the orthogonal direction in order to support variable scrubbing speeds [10].

Hand pose-based unimanual gestures can naturally also be used to interact with the content. Most obviously, pointing gestures can be taken advantage of to create emphasis effects. For instance, pointing with a lasso hand at an element such as an image or a text block highlights it to indicate the topic that the presenter is currently talking about (Figure 8). We also developed a feature based on repeated hand-closing/opening gestures allowing users to flip through sequences of images (Figure 9).

Other hand-based actions that we have not yet implemented but are considering include waving/brushing gestures to scroll through a list of items, pushing and throwing actions to dismiss elements.

Body-based Interactions

Those interactions involve the whole body of the presenter and thus can only be performed on a stage. Other than mapping physical movements of the presenter to the avatar in the 3D world, we wish to create effects triggered or controlled by body motions. Figure 2 shows an example (not yet implemented) of a body-based technique to transition between two scenes. After activating transition mode through a particular trigger action, the presenter walks across the stage, thereby gradually uncovering the next scene, as if pulling out a new panel from the side with their body.

In our current prototype we support only one type of full body-driven effect, namely a slicing interaction, where the user can "walk through" a stack of objects (typically images) in the virtual world and select an item with their body and stretched arm (Figure 8). This technique is similar to distance-based interactions for wall-displays to set data visualisation types [14, 19], but likely allows better control since the user is directly integrated in the dataset through the avatar and the feedback view shows all items contained in the stack.

Conclusion

We presented an immersive presentation concept and a first implementation of a prototype, where presenters are tightly integrated in their presentation data and are able to intuitively interact with those elements through a range of body gestures (and later other types of multimodal input). We believe that if carefully designed and integrated in a coherent experience, data-centric body interactions can enhance presenters' expression and establish new communication dimensions with their audiences. Furthermore, we think that affording presenters with the possibility to freely manipulate individual data components will enable new styles of non-linear presentations, where the course and articulation of the performance are determined by such body postures, motion, speech patterns etc.

In a nutshell, we ambition to leverage the range of kinesic expression and multimodal input paradigms to create new immersive synergies between presenters/actors/performers and their digital content and measure their impact on viewers. We imagine that a wide range of presentational experiences can thus be created depending on the level of immersion and interaction infused in those digital realms.

References

1. 3M Virtual Presenter. (March 8, 2013). Retrieved February 15, 2016 from http://news.3m.com/press-release/company/3m-debuts-interactive-virtual-presenter-south-southwest-interactive
2. Adobe Connect. Retrieved February 15, 2016 from http://www.adobe.com/products/adobeconnect.html
3. Casio Virtual Presenter. Retrieved February 15, 2016 from http://www.casioprojector.com/features/applications/virtual_presenter
4. Cisco WebEx. Retrieved February 15, 2016 from http://www.webex.com/
5. Living Actor Presenter. Retrieved February 15, 2016 from http://www.livingactor.com/Presenter
6. Musion Eyeliner. Retrieved February 15, 2016 from http://www.eyeliner3d.com
7. Personify Inc. Retrieved February 15, 2016 from http://www.personify.com/
8. TVnima. Retrieved February 15, 2016 from http://www.tvnima.com
9. Vydeo Presenters. Retrieved February 15, 2016 from http://vydeopresenters.com
10. Caroline Appert and Jean-Daniel Fekete. 2006. OrthoZoom scroller: 1D multi-scale navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, 21-30.
11. Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency and Stefan Scherer. 2013. Cicero-towards a multimodal virtual audience platform for public speaking training (IVA 2013). In *13th International Conference on Intelligent Virtual Agents*, 116-128.
12. Lei Chen, Gary Feng, Jilliam Joe, Chee Wee Leong, Christopher Kitchen and Chong Min Lee. 2014. Towards Automated Assessment of Public Speaking Skills Using Multimodal Cues. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI 2014)*, 200-203.
13. Stefania Cuccurullo, Rita Francese, Sharefa Murad, Ignazio Passero and Maurizio Tucci. 2012. A gestural approach to presentation exploiting motion capture metaphors. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2012)*, 148-155.
14. Jakub Dostal, Uta Hinrichs, Per Ola Kristensson and Aaron Quigley. 2014. SpiderEyes: designing attention- and proximity-aware collaborative interfaces for wall-sized displays (IUI 2014). In *Proceedings of the 19th international conference on Intelligent User Interfaces*, 143-152.
15. Vanessa Echeverría, Allan Avendaño, Katherine Chiluiza, Aníbal Vásquez and Xavier Ochoa. 2014. Presentation Skills Estimation Based on Video and Kinect Data Analysis. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge (MLA 2014)*, 53-60.
16. Tobias Eichner, Helmut Prendinger, Elisabeth André and Mitsuru Ishizuka. 2007. Attentive Presentation Agents. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA 2007)*, 283-295.
17. Dan Gelb, Anbumani Subramanian and Kar-Han Tan. 2011. Augmented reality for immersive

remote collaboration. In *2011 IEEE Workshop on Person-Oriented Vision (POV 2011)*, 1-6.

18. Keita Higuchi, Yinpeng Chen, Philip A. Chou, Zhengyou Zhang and Zicheng Liu. 2015. ImmerseBoard: Immersive Telepresence Experience using a Digital Whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 2015)*, 2383-2392.
19. Ulrike Kister, Patrick Reipschläger, Fabrice Matulic and Raimund Dachselt. 2015. BodyLenses - Embodied Magic Lenses and Personal Territories for Wall Displays. In *Proceedings of the 2015 ACM international conference on Interactive tabletops and surfaces (ITS 2015)*.
20. Stefan Kopp, Lars Gesellensetter, NicoleC Krämer and Ipke Wachsmuth. 2005. A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In *5th International Working Conference on Intelligent Virtual Agents (IVA 2005)*, 329-343.
21. N. H. Lehment, K. Erhardt and G. Rigoll. Interface design for an inexpensive hands-free collaborative videoconferencing system. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)* (2012), 295-296.
22. Haruki Maeda, Yuya Kurosawa, Kazutaka Kurihara and Homei Miyashita. 2011. MAEDE: A Presentation Style that the Presenter Stands in Front of the Screen. In *19th Workshop on Interactive Systems and Software (WISS 2011)*, 164-166.
23. Tsukasa Noma, Norman I Badler and Liwei Zhao. 2000. Design of a virtual human presenter. *IEEE Computer Graphics and Applications*, *20*, 4, 79-85.
24. Don't Panic - The Truth About Population*.* Wingspan Productions. BBC. November 7, 2013. https://vimeo.com/79878808
25. Worapot Sommool, Batbaatar Battulga, TimothyK Shih and Wu-Yuin Hwang. 2013. Using Kinect for Holodeck Classroom: A Framework for Presentation and Assessment. In *12th International Conference on Advances in Web-Based Learning (ICWL 2013)*, 40-49.
26. M Iftekhar Tanveer, Emy Lin and Mohammed Ehsan Hoque. 2015. Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th ACM Conference on Intelligent User Interfaces (IUI 2015)*, 286-295.
27. Ha Trinh, Lazlo Ring and Timothy Bickmore. 2015. DynamicDuo: Co-presenting with Virtual Agents. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 2015)*, 1739-1748.
28. Herwin van Welbergen, Anton Nijholt, Dennis Reidsma and Job Zwiers. 2005. Presenting in Virtual Worlds: Towards an Architecture for a 3D Presenter Explaining 2D-Presented Information. In *1st International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2005)*, 203-212.
29. Telmo Zarraonandia, Ignacio Aedo, Paloma Díaz and Alvaro Montero Montes. 2014. Augmented Presentations: Supporting the Communication in Presentations by Means of Augmented Reality. *International Journal of Human-Computer Interaction*, *30*, 10 (2014/10/03), 829-838.