



HAL
open science

Welicit: A Wizard of Oz Tool for VR Elicitation Studies

Andrea Bellucci, Telmo Zarraonandia, Paloma Díaz, Ignacio Aedo

► To cite this version:

Andrea Bellucci, Telmo Zarraonandia, Paloma Díaz, Ignacio Aedo. Welicit: A Wizard of Oz Tool for VR Elicitation Studies. 18th IFIP Conference on Human-Computer Interaction (INTERACT), Aug 2021, Bari, Italy. pp.82-91, 10.1007/978-3-030-85607-6_6 . hal-04291275

HAL Id: hal-04291275

<https://inria.hal.science/hal-04291275>

Submitted on 17 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Welicit: A Wizard of Oz Tool for VR Elicitation Studies

Andrea Bellucci¹[0000-0003-4035-5271], Telmo
Zarraonandia¹[0000-0003-3574-0984], Paloma Díaz¹[0000-0002-9493-7739], and
Ignacio Aedo¹[0000-0001-5819-0511]

¹Universidad Carlos III de Madrid (UC3M), Avenida de la Universidad, 30, 28911,
Leganés, Madrid, Spain
{firstname.secondname}@uc3m.es

Abstract. Recent advances and availability of consumer hardware has enabled the proliferation of voice, head-gaze, haptic and gesture-based interactions in VR applications. However, interactions are often the result of idiosyncratic designs and little has been investigated about user preferences and behaviour. Elicitation and Wizard of Oz (WOz) studies showed potential to design user-informed intuitive and discoverable interactions in a wide range of domains. In this paper we introduce Welicit, a WOz tool to support researchers in running VR elicitation studies. The system provides tools to adapt the elicitation methodology to immersive environments by allowing users to experience the result of proposed interactions. We discuss a use case in which the tool has been used to uncover multimodal speech+gesture interactions in VR.

Keywords: Elicitation studies · Wizard of Oz · Design tools.

1 Introduction

The development of interactive systems outside the desktop pushed the narrative of hand gestures, body movement and voice as natural forms of interacting with a computer. However, the desired “naturalness” does not solely rely on the properties of an interaction modality, technique or device: it is delivered by a skilfully crafted user interface that makes the interaction feeling seamless and meaningful in the context it is performed [11]. Early research on gestures, voice and body movement was heavily influenced by technological limitations (e.g., Charade [1]) and designers prioritized interactions that were easy to recognize by the system, but that did not reflect user preferences and behaviour. Wizard of Oz (WOz) [9, 19] and elicitation methodologies [20] were proposed to counter this issue and gather feedback from the user to investigate an interaction space we don’t know much about. Modern consumer VR hardware provides voice, head-gaze, haptic and gesture-based interactions that are manufacturer- or application-specific. As research has not yet developed a common framework and guidelines for designing interactions with 3D virtual objects, elicitation and WOz studies could help to understand users’ preferences, expectations and behaviours

in different VR scenarios. Our work contributes to tools for interaction design in the form of enabling technology [8]. We present a visual tool that aims at helping researchers to setup (pre-experiment), conduct (during the experiment), and analyse (post-experiment) WOz studies for exploring user interaction for VR experiences. We discuss a walkthrough use case in which the tool has been used to uncover user preferences for speech+gesture interaction in immersive VR editing.

2 Background

The design of the tool has been informed by previous research on interaction elicitation and Wizard of Oz studies as well as software tools that support such methodologies.

2.1 Elicitation studies and tools

Elicitation studies refer to a methodology, originally designed for surface computing [20], in which the experimenter shows a referent —the effect of an action— and ask a participant to perform the interaction that would produce that effect. The methodology, not only aims at producing unique interaction vocabularies that are easier for the users to discover, use and remember [20]. More broadly, it has been used in think-aloud experiments to establish a dialog with the participants and learn about novel or underexplored interaction spaces [14, 16, 17]. While various elicitation studies on AR interactions can be found in the literature [11, 17, 18], very few studies have explored user-defined immersive interactions for VR [15]. Running elicitation studies for novel interactive environments, such as the immersive VR medium, can be a non-trivial task: a substantial technical effort and time is required to setup and run the experiment as well as to analyse the results [13].

2.2 Wizard of Oz studies and Tools

In WOz studies, the experimenter, through a purposefully crafted interface, simulates systems’ functionality that is not already implemented, thus giving participants the illusion they are interacting with a functioning system. Researchers have used this approach to conduct user studies in a wide array of contexts [3, 4, 9, 12, 19]. Relevant for this work is the combination of WOz simulation with the interaction elicitation methodology to inform the design of interactions for AR/VR/MR [7, 18]. WOz interfaces allow the experimenter to manipulate 3D objects in the virtual environment [11], which can be used to present the referent and, more importantly, to show participants the result while they are performing the proposed interaction [18]. The literature of WOz elicitation studies is, however, still scarce. Current studies primarily discuss the resulting interaction vocabulary without providing insights on the technical and methodological considerations of WOz tools [7]. Common issues of WOz systems can be amplified

when the methodology is applied to immersive VR environments and the WOz system has to be specifically designed to reduce the cognitive and physical load of the wizard [6].

3 The Welicit system

Based on the review of the literature and our own experience in conducting WOz studies, we identified the following design requirements (R) for a WOz tool to support elicitation studies for VR.

R1. Reduce technical knowledge. Researchers might not have the technical knowledge needed to install and run complex scripts or plugins, tweak configuration files, or to setup the environment to cast the VR experience from standalone headsets as well as send commands. The WOz system should be easy to install and configure (possibly no installation) and it should work with known technology (e.g., web-based applications in a browser).

R2. Support collocated as well as remote studies. VR environments enable different time/space configurations, e.g., single user [4], collocated multi-users [5], remote multi-users [10]. The system should support the researcher in the definition of a wide range of experiment configurations, including the case in which multiple users are interacting in different cross-reality interactions between physical and virtual worlds [2].

R3. Contextual richness. Novel interactions in VR are not limited to gestures, but they encompass a wide range of modalities that can be combined in multimodal interfaces. Very little is known about how best combine different modalities or to what extent the semantic of the objects would affect the user interaction. A WOz tool, therefore, should support exploration of different modalities, e.g., hand-free, controller-based, voice, or head-gaze, as well as interaction with tangible objects. To this end, the tool should also implement strategies to reduce the effects of legacy bias, such as encouraging participants to propose more than one interaction for each referents, or priming participants towards to the use of a new device or sensing modality.

R4. Reduce the experimenter/wizard cognitive and motor load. The major drawback of the WOz technique is the high cognitive and physical demands it introduces [4]. WOz interfaces can be difficult for the experimenter to manipulate, and, in order to provide meaningful results, the wizard has to provide responses as quick and accurate as possible. The system should support the experimenter in the definition of the wizard interface, possibly providing tools to easily modify the interface and support more than one wizard. In elicitation studies, the experimenter has also to show the referents to the users, take note of the different interactions proposed for the referents as well as the user ratings for the interactions (questionnaires). The system, therefore, should support the experimenter in these tasks.

R5. Facilitate data analysis. The literature stresses that analysing results from an elicitation study is a complex task [13]. The situation is even more complex for multimodal interaction in VR: the experimenter has to analyse data

from different sources such as audio/video recording as well as the resulting interaction in the virtual world. The system should provide an interface that facilitates this process and help the experimenter to organize data for analysis.

R6. Generate interaction datasets. Contributions to the field also comes with the generation of datasets of interactions in different application contexts that other researchers can use and compare. Generated data can be used to train AI recognition algorithm, and also to speed the design of the user experience in case developers do not have access to an actual device.

3.1 Architecture and implementation

The Welicit tool is built on top of the A-Frame Inspector¹, an opensource visual tool that provides user interface widgets to build 3D scenes and tweak entities' components and properties. We wanted to integrate a flexible tool to support WOz interfaces into a design environment for prototyping VR experiences, with the goal to reduce the technical setup (R1) and provide a unique interface to build a 3D scene, and define the wizard interface as well as the connections between widgets' behaviour and 3D objects (R3). Figure 1 shows the architecture of the

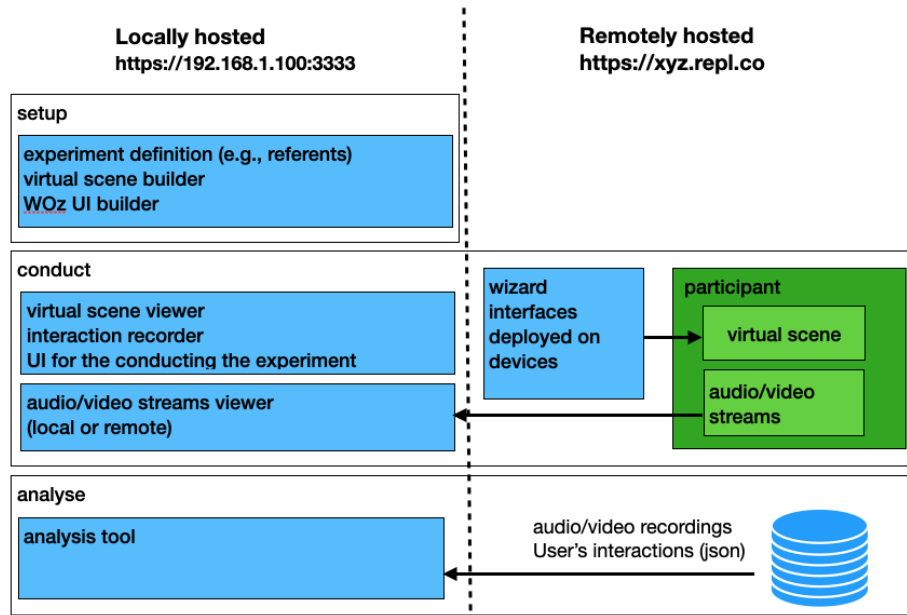


Fig. 1. Architecture of the Welicit system.

system. On a local server runs the application that the experimenter uses to

¹ <https://github.com/aframevr/aframe-inspector>

design the virtual environment, define the details of the experiment, and build the wizard interface. The system exposes the local development server for other computers to access, as well as to stream the audio and video (R2). The virtual scene, the wizard interfaces as well as the interface to run the experiment are exposed as web applications through a Node.js server, which can be locally or remotely hosted (e.g., on services such as Replit.com). Having a remote server with a fixed URL eases the setup of collocated studies (R1). At the same time, it offers the possibility to have more than one human wizard (R4), and it allows to run WOz experiment with multiple participants accessing from remote locations (R2) as well as to run remote users studies with one participant (R2), without requiring extra configurations (R1). The tool has been implemented using Web technologies such as React for building the user interface, Node.js for the server, Websocket and WebRTC API for the communication, WebXR API for hand joints visualization and recording on supported hardware, and Sensor API to retrieve data from the absolute orientation sensor of mobile devices.

3.2 System design

The Welicit system provides tools to setup and conduct an experiment as well as to analyse data.

Setup. The system provides a step-by-step interface to define details of the study (R1), such as the study id, the name, the starting and closing date and the list of participants. In the next screen, the experimenter enters the list of referents, the maximum number of alternative interactions, the post-interaction questions, and the number of scenarios to test. We decided to include scenarios because we envision that for VR interactions it would be meaningful to test the same set of referents under different conditions, as also reported in [17]. In case of multiple conditions, the system will automatically counterbalance the conditions in the experiment to avoid learning bias. Figure 2 shows the main interface of Welicit for building WOz interfaces. The system supports cross-device distributed interfaces, that is, the experimenter can select different WOz components and widgets to compose an interface across different devices. This has the goal to enable multiple human wizards at the same time, provide extra flexibility to the interface setup and avoid to clutter the interface of a single device with too many widgets (R4). Welicit implements a WYSIWYG visual editor for the composition of the wizard interface (R1): widgets are selected from a dropdown list and they can be (re)arranged through drag-and-drop interaction. Once a 3D object in the scene has been selected (mouse click), a panel appears with the properties of the object that can be controlled through the widgets of the wizard interface. The widgets that are currently implemented are: button, toggle, sliders (linear and circular), digital thumbpad, canvas for \$1 gesture recognizer (R3, R4). Additionally, quaternion data from the absolute orientation sensor of mobile devices (using the Sensor Web API) are supported. This has the goal to enable embodied controls for the experimenters (R4). Each widget has its own attributes that can be defined by the experimenter. For example, for a slider, the following attributes can be defined from the interface: the default value, the type

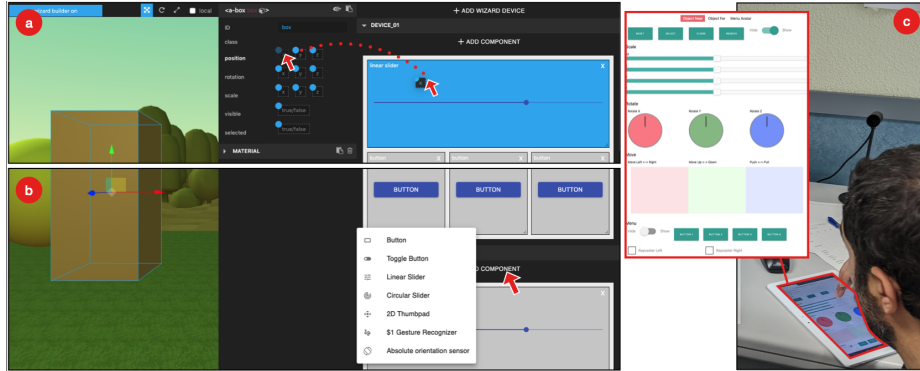


Fig. 2. The Welicit user interface to build the Wizard interface. (a) Drag-and drop properties of a 3D object to a component of the Wizard interface that will be used to control the property. (b) Add a new component to the Wizard interface. (c) An example of a deployed Wizard interface.

(continuous or discrete), the step, the min and max values and the orientation.

Conduct. Once the experimenter has defined the virtual scene and the wizard interface, all the code is automatically generated and made available on the remote server. A URL is generated to open the virtual scene on a VR device. Other URLs are generated and codified into QR Codes to access the wizard interface from the browser of the target devices (Figure 2.c). When running the experiment, the locally hosted server receives data from the virtual scene as well as audio and video streams from remote or local clients, which the experimenter will use to observe the interaction of the user (R1, R2). During this phase, Welicit provides an interface to view the interaction of the participant in the virtual scene, local as well as remote video streams and buttons to go to the next referent and to record the interaction. The system implements a functionality to record the interaction as it takes place in the virtual world, exploiting the motion capture tools of the A-frame Inspector, which allows to save *json* files describing *xyz* position and rotation of camera and tracked controllers with timestamps. The motion capture tool was extended to record hands joint data on the Oculus Quest 2 (WebXR API) and Oculus Rift (Leap Motion), and it was tweaked to support two modes of recording: automatic and manual. In the automatic mode, interactions in the virtual world are recorded as long as hands or controllers movements are in the field of view of the participants. Auto recording could lower the cognitive load of the experimenter (R4) and it facilitates unsupervised recording of interactions [16]. However, it does not capture camera-only movements, which are important to study head-gaze interaction. To give the experimenter more control over recordings, the manual mode allows to start and stop recording by pressing the <space> key.

Analyse. In this last stage, the Welicit system offers an interface (Figure 3) that shows video and audio streams matched with the interaction performed

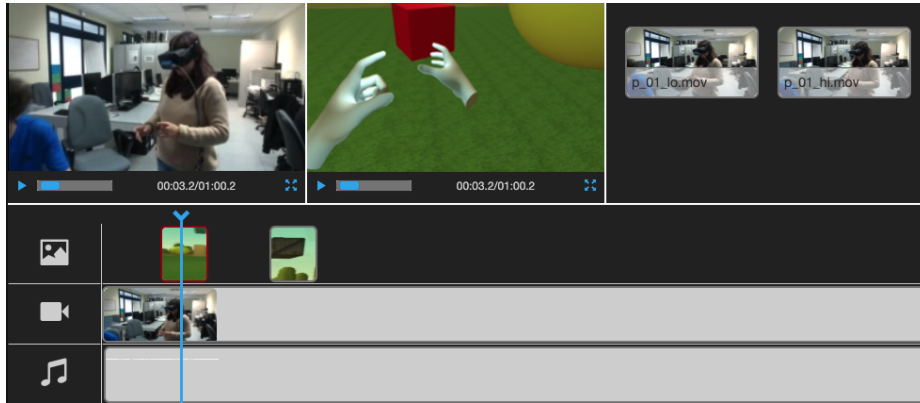


Fig. 3. The Welicit interface to analyse video, audio and VR interaction recordings.

in the virtual world. This functionality aims at supporting analysis of complex interactions (R5), for instance multimodal interactions with speech, head-gaze and gestures, for which the experimenter needs to observe, at the same time, the participant interaction in the real world and the resulting interaction in the virtual world. The analysis tool automatically organizes recordings according to the referents and produces one set of recordings for each scenario.

4 Walkthrough use case and lessons learned

We used the Welicit tool to run a WOz study to understand user preferences for multimodal speech+gesture interactions for VR editing environments. Our initial goal was to understand how properties of the objects, such as their size and position with respect to the user (within or out of reach) would affect the interaction. We discuss the user study as a walkthrough scenario and to describe the potential of the tool to support WOz studies as well as lessons learned.

Setup. In the setup stage, we designed a 3D scene of a small flat apartment with a cup on table and a bed, using existing 3D models. Two researchers defined the referents to use in the study, namely: select, clone, delete, change color, move, scale, rotate and undo. Two scenarios were also defined: one for the small object close to the participant (the cup) and another one for the big object out of reach (the bed). We then built the user interface for the WOz to run on tablet device, with two tabs, one for each object, and widgets that allowed to perform manipulations of the objects: linear sliders for changing the xyz scale, circular sliders for xyz rotation, 2D digital thumbpads for moving the object on xy -, xz -, and yz -axis and buttons to hide/show, duplicate or delete the objects. We also designed a fake menu: an object with buttons that would appear close to the object in case the participants felt the need to use virtual UIs. Welicit proved useful in this phase, since it allowed the two experimenters to quickly run pilot studies to decide the final design of the virtual environment and the wizard in-

terface and, at the same time, gain experience in the use of the interface itself. **Conduct.** We conducted the experiment with 16 participants, using the Oculus Rift VR headset and Leap Motion for hand-tracking. Automatically counterbalancing the scenarios as well as presenting the next referent on button press was helpful: it reduced the cognitive load of the experimenter, who did not have to worry about remembering or reading on a piece of paper what was the next referent. We also discovered that, in cases when experiments are not conducted by the same researcher, the interaction to show to the user should be pre-recorded. This would assure consistency among studies, since different researchers could show the interaction in a slightly different way, which could bias the study. Recording the interaction to present for a referent would also allow to show only *before* and *after* states of an object, and not an animated transition [17]. As we expected, allowing participants to experience their interactions encouraged them to think out loud and to propose more than one interaction for each referent.

Analyse. The manual recording feature was helpful to extract the audio corresponding to an interaction in VR and the visualization tool allowed to easily identify and classify it. We discovered that it would be useful to provide the experimenter with an interface to add time-based annotation during the study. In the case of voice interaction only, in fact, it was difficult to find when they were performed and such annotations would solve this problem.

5 Limitations and future development

The Welicit tool has technological limitation that will be addressed in future development. The system supports only single-user immersive experiences on Oculus Quest 2 and Oculus Rift. There is no mechanism to easily extend the tool to support additional widgets and sensors; new components could be added but this would require programming work and we do not describe in this paper how it could be accomplished. Lastly, user studies with researchers outside our lab are needed to evaluate the effectiveness of the tool.

6 Conclusions

We introduced a tool to setup, conduct and analyse elicitation studies with the Wizard of Oz paradigm. Easing the technical and methodological issues of these studies is an important step to generate user and interaction requirements for novel VR interactions in different contexts. To this end, we plan on making the tool available under open-source license and create a space where researchers and designers can contribute by sharing their data and experience.

7 Acknowledgments

This work was supported by the Spanish State Research Agency (AEI) under grants Sense2MakeSense (PID2019-109388GB-I00) and CrossColab (PGC2018-101884-B-I00).

References

1. Baudel, T., Beaudouin-Lafon, M.: Charade: remote control of objects using free-hand gestures. *Communications of the ACM* **36**(7), 28–35 (1993)
2. Bellucci, A., Zarraonandia, T., Díaz, P., Aedo, I.: End-user prototyping of cross-reality environments. In: *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. pp. 173–182 (2017)
3. Connell, S., Kuo, P.Y., Liu, L., Piper, A.M.: A wizard-of-oz elicitation study examining child-defined gestures with a whole-body interface. In: *Proceedings of the 12th International Conference on Interaction Design and Children*. pp. 277–280 (2013)
4. Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J.D., Gandy, M.: Wizard of oz interfaces for mixed reality applications. In: *CHI’05 extended abstracts on human factors in computing systems*. pp. 1339–1342 (2005)
5. Grandi, J.G., Debarba, H.G., Nedel, L., Maciel, A.: Design and evaluation of a handheld-based 3d user interface for collaborative object manipulation. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. pp. 5881–5891 (2017)
6. Grill, T., Polacek, O., Tscheligi, M.: Conwiz: a tool supporting contextual wizard of oz simulation. In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*. pp. 1–8 (2012)
7. Höysniemi, J., Hämäläinen, P., Turkki, L.: Wizard of oz prototyping of computer vision based action games for children. In: *Proceedings of the 2004 conference on Interaction design and children: building a community*. pp. 27–34 (2004)
8. Hudson, S.E., Mankoff, J.: Concepts, values, and methods for technical human–computer interaction research. In: *Ways of Knowing in HCI*, pp. 69–93. Springer (2014)
9. Klemmer, S.R., Sinha, A.K., Chen, J., Landay, J.A., Aboobaker, N., Wang, A.: Suede: a wizard of oz prototyping tool for speech user interfaces. In: *Proceedings of the 13th annual ACM symposium on User interface software and technology*. pp. 1–10 (2000)
10. Lang, T., MacIntyre, B., Zugaza, I.J.: Massively multiplayer online worlds as a platform for augmented reality experiences. In: *2008 IEEE Virtual Reality Conference*. pp. 67–70. IEEE (2008)
11. Lee, M., Billinghurst, M.: A wizard of oz study for an ar multimodal interface. In: *Proceedings of the 10th international conference on Multimodal interfaces*. pp. 249–256 (2008)
12. MacIntyre, B., Gandy, M., Dow, S., Bolter, J.D.: Dart: a toolkit for rapid design exploration of augmented reality experiences. In: *Proceedings of the 17th annual ACM symposium on User interface software and technology*. pp. 197–206 (2004)
13. Magrofuoco, N., Vanderdonckt, J.: Gelicit: A cloud platform for distributed gesture elicitation studies. *Proceedings of the ACM on Human-Computer Interaction* **3**(EICS), 1–41 (2019)
14. Morris, M.R.: Web on the wall: insights from a multimodal interaction elicitation study. In: *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. pp. 95–104 (2012)
15. Nanjappan, V., Liang, H.N., Lu, F., Papangelis, K., Yue, Y., Man, K.L.: User-elicited dual-hand interactions for manipulating 3d objects in virtual reality environments. *Human-centric Computing and Information Sciences* **8**(1), 1–16 (2018)

16. Nebeling, M., Ott, D., Norrie, M.C.: Kinect analysis: a system for recording, analysing and sharing multimodal interaction elicitation studies. In: Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. pp. 142–151 (2015)
17. Pham, T., Vermeulen, J., Tang, A., MacDonald Vermeulen, L.: Scale impacts elicited gestures for manipulating holograms: Implications for ar gesture design. In: Proceedings of the 2018 Designing Interactive Systems Conference. pp. 227–240 (2018)
18. Piumsomboon, T., Clark, A., Billingham, M., Cockburn, A.: User-defined gestures for augmented reality. In: IFIP Conference on Human-Computer Interaction. pp. 282–299. Springer (2013)
19. Salber, D., Coutaz, J.: Applying the wizard of oz technique to the study of multimodal systems. In: International Conference on Human-Computer Interaction. pp. 219–230. Springer (1993)
20. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1083–1092 (2009)