Demonstrating *PilotAR*: A Tool to Assist Wizard-of-Oz Pilot Studies with OHMD

NUWAN JANAKA, Smart Systems Institute, National University of Singapore, Singapore

RUNZE CAI, School of Computing, National University of Singapore, Singapore

SHENGDONG ZHAO*, School of Creative Media & Department of Computer Science, City University of Hong Kong, China

DAVID HSU, School of Computing, National University of Singapore Smart Systems Institute, National University of Singapore, Singapore

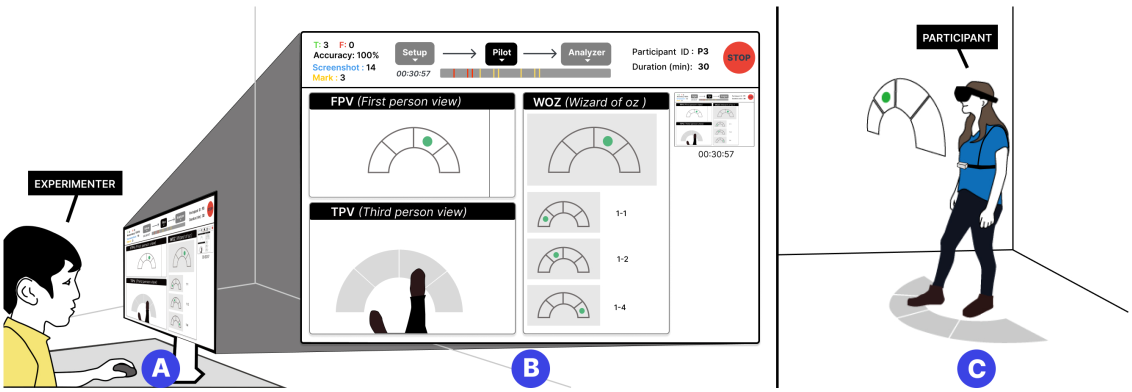

Fig. 1. (A) The experimenter employs *PilotAR*, a desktop-based experimenter tool, for OHMD-based pilot studies. (B) *PilotAR* facilitates real-time monitoring of participants' experiences from both first-person and third-person perspectives, enabling experimenters to track ongoing studies dynamically. In addition, the tool's annotation features allow for the precise marking and capture of significant moments in a photo or video format. Quickly logging quantitative metrics, such as event time, can be done using shortcut keys. Furthermore, a real-time summary of the observed moments and recorded data, available for post-study interviews, promotes in-depth discussions, insights, and support for collaborative review and interpretation. (C) In a separate room, the participant interacts with the simulated AR system, maintaining communication with the experimenter.

While pilot studies help to identify potential interesting research directions, the additional requirements in AR/MR make it challenging to conduct quick and dirty pilot studies efficiently with Optical See-Through Head-Mounted Displays (OST HMDs, OHMDs). To overcome these challenges, including the inability to observe and record in-context user interactions, increased task load, and difficulties with in-context data analysis and discussion, we introduce *PilotAR* (https://github.com/Synteraction-Lab/PilotAR), a tool designed iteratively to enhance AR/MR pilot studies, allowing live first-person and third-person views, multi-modal annotations, flexible wizarding interfaces, and multi-experimenter support.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **User interface toolkits**; **Mixed / augmented reality**.

*Corresponding Author.

Authors' addresses: Nuwan Janaka, nuwanj@u.nus.edu, Synteraction Lab, Smart Systems Institute, National University of Singapore, Singapore; Runze Cai, runze.cai@u.nus.edu, Synteraction Lab, School of Computing, National University of Singapore, Singapore; Shengdong Zhao, shengdong.zhao@cityu.edu.hk, Synteraction Lab, School of Creative Media & Department of Computer Science, City University of Hong Kong, Hong Kong, China; David Hsu, dyhsu@comp.nus.edu.sg, School of Computing, National University of Singapore and Smart Systems Institute, National University of Singapore, Singapore.

Additional Key Words and Phrases: toolkit, tool, pilot study, heads-up computing, augmented reality, OST-HMD, smart glasses, evaluation, interaction

1 INTRODUCTION AND RELATED WORK

Quick and dirty pilot studies validate research concepts, identify usability issues, and guide design decisions without extensive resource commitments [25, 26]. However, conducting pilot studies in Augmented Reality (AR) and Mixed Reality (MR) using optical see-through head-mounted displays (OST-HMD, OHMD, or AR smart glasses) poses significant challenges [22, 23] due to their unique characteristics [15] such as personal, near-eye displays.

Compared to traditional studies on 2D UIs in desktop/mobile, which mainly observe users from third-person perspective, AR/MR requires both observations from a first-person perspective to understand users' interactions with digital content and a third-person perspective to understand user interactions with the physical world [2, 22, 24]. Besides observing a multifaceted environment, the task load for experimenters involved in AR/MR pilot studies can also be increased by the need to optionally perform wizard-of-Oz tasks [2, 7, 9, 10], thus necessitating methods to reduce their multitasking burden [2, 4, 11]. Furthermore, there is insufficient support for in-context data analysis [3, 8, 21] during the pilot studies, especially for quantitative data, which are typically collected in an informal and raw way. This hinders real-time analysis and deeper discussions in post-study interviews.

Given the absence of an integrated solution for AR/MR pilot studies, despite the development of many specialized tools for individual steps in experiments (e.g., content authoring [12, 20], rapid prototyping [11], gesture interaction [17, 27, 29], experiment setup [8, 21], video analysis [16, 30], 3D and MR visualization [5, 13], immersive experiment environments [2, 22, 23]), we created *PilotAR* (See Appendix A-Table 1 for comparison). It offers experimenters the flexibility to use familiar prototyping or wizarding interfaces rather than requiring the construction of an immersive system with specific skill sets (e.g., [2, 21, 22] requires Unity3D background), during the early stage of research. Similar to Momento [6], *PilotAR* supports the entire study conduction life cycle: setting up, experimentation, analysis and summarizing, and repeating. However, *PilotAR* caters to unique challenges of OHMDs (e.g., context, interface [15, 28]), including multiple observation viewpoints real-time synchronization, which is not supported by Momento [6] as it focuses on applications on mobile phones and desktops.

PilotAR (Fig 1) is an open-source desktop-based tool for experimenters to conduct AR/MR iterative pilot studies with OHMDs. It streamlines the pilot process from situated observations to results sharing. It incorporates first-person and third-person video observations to help experimenters understand users' in-situ relationship with visual content and environment in real-time and automatically record them for post-analysis. It enables annotations, allowing manual or automatic tagging of significant events during the experiment to prevent tedious post-study analysis and missing labeling. *PilotAR* also allows for task distribution among multiple experimenters, reducing multitasking load and making remote monitoring possible. Finally, *PilotAR* enables real-time data summaries, encouraging a deeper discussion during post-pilot interviews and facilitating results sharing with collaborators by exporting data report. For detailed evaluation, please refer to our original paper, *PilotAR* [14].

2 *PILOTAR* TOOL

In this section, we outline the functions and a typical usage scenario of *PilotAR* (Figure 2). *PilotAR* integrates features that streamline processes and support replication and innovation in AR/MR pilot studies using the wizard-of-oz [1, 2, 10, 11, 17] approach. See Appendix B for implementation details.

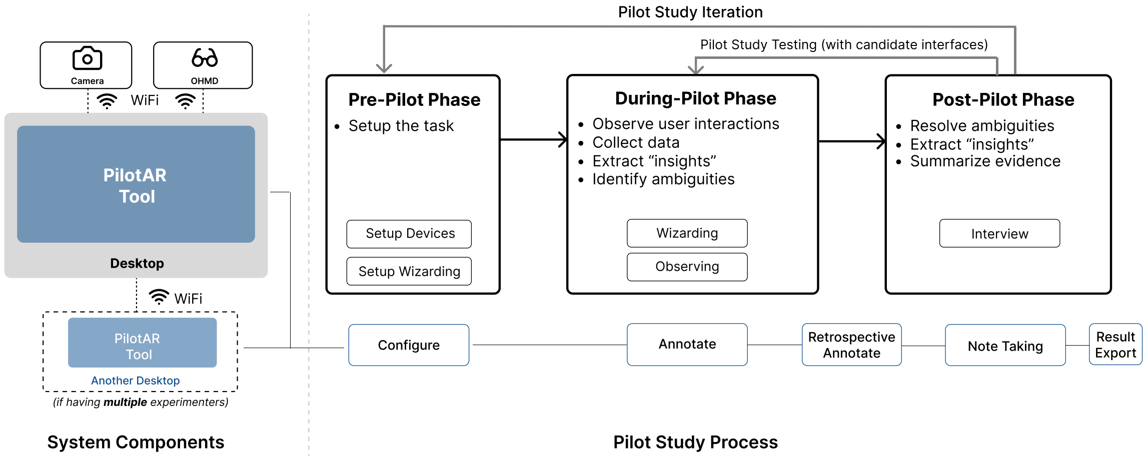

Fig. 2. Overview of the system components and workflow with *PilotAR*.

2.1 Major Functions

FPV and TPV Live Streaming (Support Observations in Situated Contexts): Although relatively straightforward in design, we enabled experimenters to observe participants wearing OHMD in situated contexts through the live first-person view with grids (FPV) and third-person view (TPV), as depicted in Figure 1. Simultaneous video recorded for subsequent analysis is enabled. Specifically, FPV streams the overlay of digital content and the realistic environment rendered by the OHMD. TPV streams video from a user-attached camera or one positioned by experimenters.

Annotations with Function Shortcuts (Reduce Task Load of Experimenters): To facilitate important information documentation during pilot study observations, we enable a variety of annotations. These encompass *Screenshot* (to capture the screen, optionally with a colored block highlighting a specific Region of Interest (ROI)), *Focus* capturing only a selected screen region), *Correct* and *Incorrect* (for accuracy calculations), and *Counter* (for tracking interaction attempts). The communication between experimenters and participants is recorded and transcribed to *Voice Annotation* in text format. During pilot studies, experimenters can use customized keyboard shortcuts to activate *Annotation* functions. These shortcuts can be mapped to UI, user, or experimenter actions for automatic annotations. Additionally, each *Annotation*'s color can be customized for easy identification, and all annotations are time-stamped for later review.

Multi-experimenter Support (Reduce Task Load of Experimenters): To reduce task load during pilot studies, we support multi-experimenter scenarios alongside traditional single-experimenter setups. In a single-experimenter scenario, the experimenter concurrently manipulates the wizarding interface, conducts observations, and makes annotations. In the multi-experimenter configuration, one experimenter can act as the wizard, adjusting the interface based on users' actions observed via FPV and TPV, and another experimenter can focus solely on observation and annotation. After the pilot, annotations from both experimenters can be synchronized.

Analyzer (Expedite Data Recording, Analysis, and Generation of Creative Insights): To allow experimenters to get a real-time summary of the collected data, we implemented the *Analyzer* view. By reviewing the annotation index on the recording's timeline, experimenters can identify key moments and use video playback to assist participants in recalling their experiences. Experimenters can adjust annotations recorded during the pilot session (e.g., change timestamp, modify manipulation correctness, modify notes), add new notes, and take screenshots. The analyzer also briefly summarizes accuracy and the time duration between two indices of *Annotation* and corresponding events.

Summary Review (Expedite Data Recording, Analysis, and Generation of Creative Insights): A comprehensive review of the pilot results can be exported from the analyzer to facilitate information sharing among collaborators, including overall descriptive statistics, selected annotation timestamps, notes, and screenshot images. Raw data (e.g., video) can be shared for subsequent analyses.

2.2 *PilotAR* Usage Scenario

Experimenters might adopt various strategies with *PilotAR*. Here, we outline a basic approach for conducting a pilot study using *PilotAR*, with the replication of 'Mind the Tap' [19] as an example to highlight its usage.

*Mary, an AR researcher, conceives a novel idea employing foot-tapping as an input interaction for OHMDs (**Figure 1**). She identifies two potential interactions:* direct *(i.e., the menu appears on the floor within leg's reach) and* indirect *(i.e., the menu displays in front of the eyes, requiring users to use proprioception to associate it with their foot, Figure 1C). She aims to discern the strengths and limitations of each foot-tap interaction. Choosing a within-subject design for an initial comparison, Mary opts to employ the wizard-of-oz technique to minimize developmental efforts in a tangible system (e.g., Unity development with optical tracking) and to persuade colleagues to explore this concept further.*

2.3 Interface and Workflow

The main workflow using *PilotAR* is divided into three phases: *pre-pilot*, *during-pilot*, and *post-pilot*. This section demonstrates how Mary can utilize *PilotAR*'s interfaces throughout these phases.

2.3.1 Pre-pilot Phase. See Appendix C for details of setup UI.

Mary quickly crafts a wizarding interface using Google Slides with a 2x4 menu, where the target location randomizes on subsequent slides. She mirrors these slides to the HoloLens 2 (HL2) via Google Meet on a browser. She uses a phone camera as the TPV by linking it to Google Meet. For direct *interactions, the mirrored WOz interface is fixed on the floor. Conversely, for* indirect *interactions, it's positioned in front of the users' eyes.*

Mary initiates the PilotAR, selects 'Single User' (Figure 5A), and sets up the devices (Figure 5B) with the HL2 IP address for FPV, a Google Meet link for TPV, and Google Slides for the Wizarding Interface (Figure 5C1). She then adds a "Check foot visibility" checklist item (Figure 5C2) to verify the FPV setup is accurate before each pilot session. To ascertain accuracy and usability, she enables (Figure 5C3) Correct, Incorrect, Counter, and Screenshot annotations.

2.3.2 During-pilot Phase. After setting up and confirming the checklist, experimenters can enter the anticipated duration and participant and session ID, and initiate the "Pilot" phase by clicking the "Start" button (Figure 3A4).

Top Bar (Figure 3A). The top bar displays session-related metadata, including live statistics of measures (e.g., count of Annotations, Figure 3A1), session progress (e.g., duration and timeline, Figure 3A2), and session information (e.g., participant info, anticipated duration, Figure 3A3). Experimenters will receive a notification when the anticipated time has elapsed and can stop the session by clicking the "Stop" button located at the right corner of the top bar.

Main Working Panel (Figure 3B). The working panel displays FPV (Figure 3B1), TPV (Figure 3B2), and *Wizarding Interfaces* (Figure 3B3), with a layout that can be customized according to the experimenter's preferences. In the right corner of the working panel, the captured *Screenshot* and *Focus* annotations using keyboard shortcut keys (e.g., "3" key

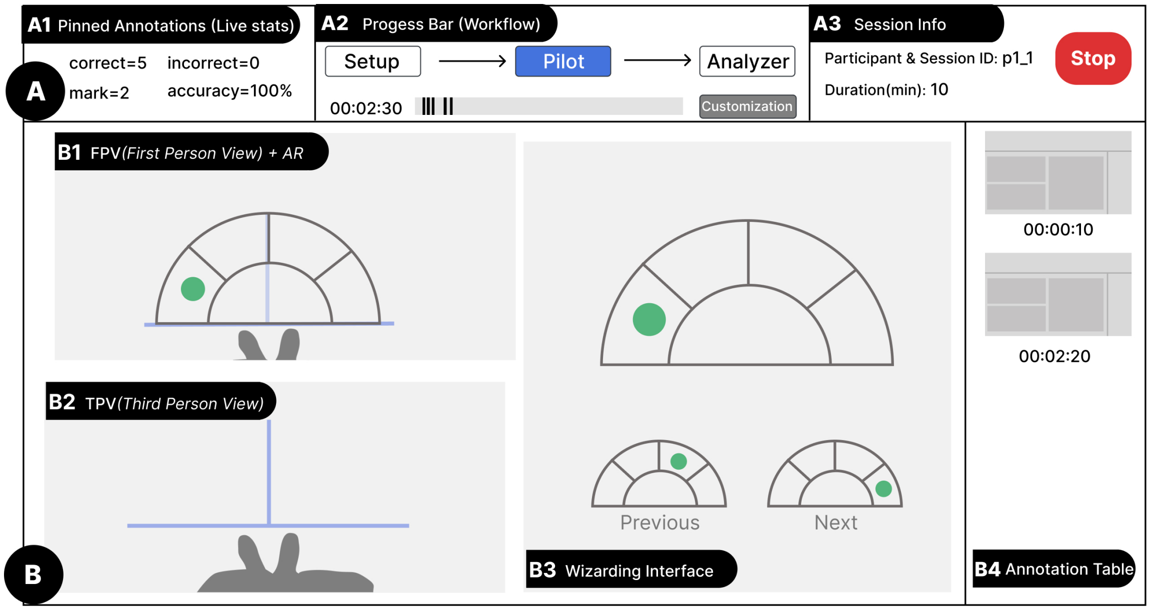

Fig. 3. Pilot interface, which includes two major areas. Area (A) is the Top Bar showing (pinned) *Annotations*' live statistics (A1), the session progress (A2), and session information (A3). Area (B) presents the main working panel housing the FPV (B1, which shows the digital interface and user's feet from their FPV), TPV (B2), *Wizarding Interface* (B3), and a sidebar for the annotation table (B4).

key) are shown as images with timestamps in the Annotation Table (see Figure 3B4). Clicking on these images opens a pop-up window, allowing the experimenter to add notes to the annotations.

[Piloting with the First Interface] Mary starts the pilot with direct *interface (Figure 1C). Adjusting the target location on the Wizarding Interface (Figure 3B3), she annotates accuracy across ten trials, taking screenshots of any interesting behavior (Figure 3B4). Mary also monitors the trial count and accuracy via the live statistics dashboard (Figure 3A1).*

2.3.3 Post-pilot Phase. Upon completion of the pilot session, the *Analyzer* window appears (Figure 4), displaying the video panel on the left (Figure 4A) and *Annotations* panel on the right (Figure 4B).

Video Panel(Figure 4A). It can play the recorded video (Figure 4A1) and navigate to any timestamp by clicking the timeline (Figure 4A2) or using three buttons to rewind, pause, and fast-forward. Experimenters can create new *Annotations* with notes in the "New Note" area below the video timeline (Figure 4A2).

Annotation Panel (Figure 4B). It features an annotation preview (Figure 4B1), annotation filtering options (Figure 4B2), an annotation table (Figure 4B3), and an exporting button. The annotation preview (Figure 4B1) provides an overview of the pilot, including its duration, manipulation accuracy, and collected screenshots. Experimenters can click on these screenshots to pinpoint annotated moments in the recorded video.

Within the Annotation table (Figure 4B3), experimenters have the capability to view and adjust annotation details by double-clicking on a cell. Additionally, specific Annotations can be highlighted by clicking the corresponding icon in the first column or applying the filters available (Figure 4B2). The tool also facilitates the export of summaries and selected *Annotations* in both PDF and CSV formats (Figure 4C).

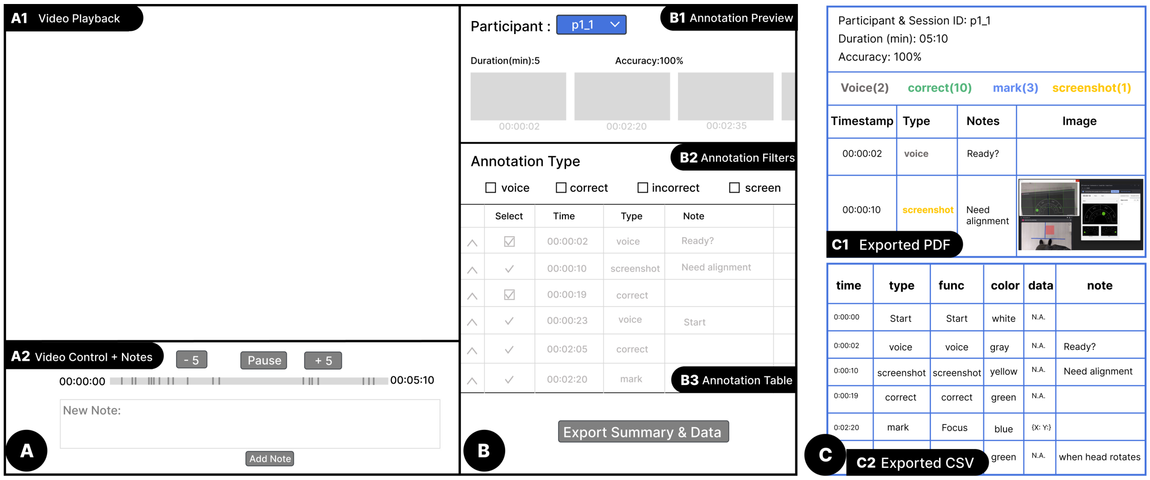

Fig. 4. The Analyzer interface comprises two main panels: the video panel (A) and the annotation panel (B). The video panel includes video playbacks of the pilot (A1), video controls, and a new note panel (A2). The annotation panel features an annotation preview (B1), annotation filtering options (B2), an annotation table (B3), and an exporting button. The Analyzer supports exporting the annotations (C) in PDF format (C1) and CSV format (C2).

[Analysis] Upon finishing the session, the Analyzer activates, presenting screenshots, accuracy data, and annotations (Figure 4). Before the interview, Mary reviews these annotations and accuracy (Figure 4B1-B3), devising questions for further inquiry. For clarity on specific screenshots, she replays footage from 5 seconds prior (Figure 4A1-A2). She then conducts the interview, discussing the participant's experiences and challenges, and incorporates their feedback into the annotation notes (Figure 4B3).

Experimenters can return to the "Pilot" session for subsequent pilot studies and initiate new recordings. All interactions in the *Analyzer* are stored, enabling experimenters to switch between different pilot recordings using the drop-down menu in Figure 4B1.

[Piloting with the Alternative Interface] After assessing the direct *interface, Mary tests the* indirect *interface in the same approach.*

[Overall Analysis] After piloting both interfaces, Mary invites the participant for an overall interview, utilizing the Analyzer to toggle between pilot recording sessions or view them simultaneously (Figure 4B1). This comparison offers insights into "rough" accuracy and usability variations, which are noted in Analyzer (e.g., direct one is slightly more accurate while causing neck pain for long usage, (Figure 4B3).

[Repeating] Mary replicates this process with three more participants, counterbalancing the interface. Mary exports participant data summaries in PDF (Figure 4C1) and shares them with colleagues to convince the differences between direct *and* indirect *interfaces. She cites participant feedback and replays specific recordings for context when queried for details.*

[Further Exploration: Multi-experimenter] Seeing the team's interest, Mary broadens their exploration to assess how interaction accuracy and speed vary between two interfaces as menu size changes. She trains a colleague to act as the wizard, thus reducing the wizarding workload and focusing more on observations. After creating additional slides for varied menu sizes (e.g., 1x2, 2x4, 3x6), they conduct pilot tests with four participants using a between-subjects design. To calculate the speed of interactions, they combine Correct/Incorrect annotations with custom annotations that automatically mark target changes (linked to slides' changes). After each pilot session, data is exported to CSV (Figure 4C2) for graph generation in Excel, which facilitates comparing relationships among speed, accuracy, and menu size. Convinced that their pilot study has uncovered a notable trend, the team decides to transition to a formal study.

[Summary] Employing the wizard-of-oz methodology with PilotAR, the team expedites (e.g., less than one week as opposed to a full-fledged motion tracking application, which can take several weeks to months) the identification of viable research directions. Using PilotAR, experimenters can overcome challenges in rapidly evaluating diverse concepts, gathering preliminary quantitative measures for comparison, and convincing colleagues, significantly shortening the knowledge discovery phase.

3 CONCLUSION

As AR/MR technology is poised to shape the future immersive world, including the metaverse, facilitating interactions between digital and physical entities becomes paramount. This underscores the importance of tools tailored for refining these interactions through pilot studies. As an initial step, we introduce *PilotAR*, an open-source tool (https://github.com/Synteraction-Lab/PilotAR) designed to support such studies. It enables real-time and retrospective multi-viewpoint observations, notes, and filters of crucial observations, thus facilitating comprehensive discussions with participants and researchers to discover insights effectively. Additionally, it can share the pilot study process, data, and insights with the larger research community. Its all-in-one capability can be applied as a standalone observation tool or a video analyzer tool to border studies beyond pilot studies or OHMD-based studies.

ACKNOWLEDGMENTS

We would also like to thank Tan Si Yan and Siddanth Ratan Umralkar for developing specific system components. Additionally, we wish to thank the anonymous reviewers for their valuable time and insightful comments. This research is supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-016). The CityU Start-up Grant 9610677 also provides partial support. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not reflect the views of the National Research Foundation, Singapore.

REFERENCES

[1] Günter Alce, Mattias Wallergård, and Klas Hermodsson. 2015. WozARd: a wizard of oz method for wearable augmented reality interaction—a pilot study. *Advances in Human-Computer Interaction* 2015 (June 2015). https://doi.org/10.1155/2015/271231

[2] Andrea Bellucci, Telmo Zarraonandia, Paloma Díaz, and Ignacio Aedo. 2021. Welicit: A Wizard of Oz Tool for VR Elicitation Studies. In *Human-Computer Interaction – INTERACT 2021 (Lecture Notes in Computer Science)*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 82–91. https://doi.org/10.1007/978-3-030-85607-6_6

[3] Steve Benford, Andy Crabtree, Martin Flintham, Adam Drozd, Rob Anastasi, Mark Paxton, Nick Tandavanitj, Matt Adams, and Ju Row-Farr. 2006. Can you see me now? *ACM Transactions on Computer-Human Interaction* 13, 1 (March 2006), 100–133. https://doi.org/10.1145/1143518.1143522

[4] Jack Brookes, Matthew Warburton, Mshari Alghadier, Mark Mon-Williams, and Faisal Mushtaq. 2020. Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods* 52, 2 (April 2020), 455–463. https://doi.org/10.3758/s13428-019-01242-0

[5] Wolfgang Büschel, Anke Lehmann, and Raimund Dachselt. 2021. MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445651

[6] Scott Carter, Jennifer Mankoff, and Jeffrey Heer. 2007. Momento: support for situated ubicomp experimentation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 125–134. https://doi.org/10.1145/1240624.1240644

[7] Marco de Sá and Elizabeth Churchill. 2012. Mobile augmented reality: exploring design and prototyping techniques. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 221–230. https://doi.org/10.1145/2371574.2371608

[8] Arindam Dey, Mark Billinghurst, Robert W. Lindeman, and J. Edward Swan. 2018. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI* 5 (2018). https://doi.org/10.3389/frobt.2018.00037

[9] Steven Dow, Jaemin Lee, Christopher Oezbek, Blair MacIntyre, Jay David Bolter, and Maribeth Gandy. 2005. Wizard of Oz interfaces for mixed reality applications. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1339–1342. https://doi.org/10.1145/1056808.1056911

[10] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J.D. Bolter, and M. Gandy. 2005. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing* 4, 4 (Oct. 2005), 18–26. https://doi.org/10.1109/MPRV.2005.93 Conference Name: IEEE Pervasive Computing.

[11] Gabriel Freitas, Marcio Sarroglia Pinho, Milene Selbach Silveira, and Frank Maurer. 2020. A Systematic Review of Rapid Prototyping Tools for Augmented Reality. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*. 199–209. https://doi.org/10.1109/SVR51698.2020.00041

[12] Maribeth Gandy and Blair MacIntyre. 2014. Designer's augmented reality toolkit, ten years later: implications for new media authoring tools. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14)*. Association for Computing Machinery, New York, NY, USA, 627–636. https://doi.org/10.1145/2642918.2647369

[13] Sebastian Hubenschmid, Jonathan Wieland, Daniel Immanuel Fink, Andrea Batch, Johannes Zagermann, Niklas Elmqvist, and Harald Reiterer. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3491102.3517550

[14] Nuwan Janaka, Runze Cai, Ashwin Ram, Lin Zhu, Shengdong Zhao, and Yong Kai Qi. 2024. PilotAR: Streamlining Pilot Studies with OHMDs from Concept to Insight. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Sept. 2024). https://doi.org/10.1145/3678576

[15] Nuwan Janaka, Jie Gao, Lin Zhu, Shengdong Zhao, Lan Lyu, Peisen Xu, Maximilian Nabokow, Silang Wang, and Yanch Ong. 2023. GlassMessaging: Towards Ubiquitous Messaging Using OHMDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023). https://doi.org/10.1145/3610931

[16] Michael Kipp. 2014. ANVIL: The Video Annotation Research Tool. In *The Oxford Handbook of Corpus Phonology*, Jacques Durand, Ulrike Gut, and Gjert Kristoffersen (Eds.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199571932.013.024

[17] Minkyung Lee and Mark Billinghurst. 2008. A Wizard of Oz study for an AR multimodal interface. In *Proceedings of the 10th international conference on Multimodal interfaces (ICMI '08)*. Association for Computing Machinery, New York, NY, USA, 249–256. https://doi.org/10.1145/1452392.1452444

[18] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. 2004. 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 (UIST '04)*. Association for Computing Machinery, New York, NY, USA, 197–206. https://doi.org/10.1145/1029632.1029669

[19] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300707

[20] Michael Nebeling and Maximilian Speicher. 2018. The Trouble with Augmented Reality/Virtual Reality Authoring Tools. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 333–337. https://doi.org/10.1109/ISMAR-Adjunct.2018.00098

[21] Michael Nebeling, Maximilian Speicher, Xizi Wang, Shwetha Rajaram, Brian D. Hall, Zijian Xie, Alexander R. E. Raistrick, Michelle Aebersold, Edward G. Happ, Jiayin Wang, Yanan Sun, Lotus Zhang, Leah E. Ramsier, and Rhea Kulkarni. 2020. MRAT: The Mixed Reality Analytics Toolkit. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376330

[22] Alejandro Rey, Andrea Bellucci, Paloma Diaz, and Ignacio Aedo. 2021. A Tool for Monitoring and Controlling Standalone Immersive HCI Experiments. In *The Adjunct Publication of the 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 20–22. https://doi.org/10.1145/3474349.3480217

[23] Alejandro Rey Lopez, Andrea Bellucci, Paloma Diaz Perez, and Ignacio Aedo Cuevas. 2022. IXCI:The Immersive eXperimenter Control Interface. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3531073.3534489

[24] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300767

[25] Lehana Thabane, Jinhui Ma, Rong Chu, Ji Cheng, Afisi Ismaila, Lorena P. Rios, Reid Robson, Marroon Thabane, Lora Giangregorio, and Charles H. Goldsmith. 2010. A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology* 10, 1 (Jan. 2010), 1. https://doi.org/10.1186/1471-2288-10-1

[26] Edwin R. van Teijlingen and Vanora Hundley. 2001. The importance of pilot studies. (2001). https://aura.abdn.ac.uk/handle/2164/157

[27] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Yuanzhi Cao, and Karthik Ramani. 2021. GesturAR: An Authoring System for Creating Freehand Interactive Augmented Reality Applications. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 552–567. https://doi.org/10.1145/3472749.3474769

[28] Xuhai Xu, Anna Yu, Tanya R. Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, Te-Yen Wu, Junxiao Shen, Ting Zhang, Narine Kokhlikyan, Fulton Wang, Paul Sorenson, Sophie Kim, and Hrvoje Benko. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–30. https://doi.org/10.1145/3544548.3581500

[29] Hui Ye and Hongbo Fu. 2022. ProGesAR: Mobile AR Prototyping for Proxemic and Gestural Interactions with Real-world IoT Enhanced Spaces. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3517689

[30] Patrick H. Zimmerman, J. Elizabeth Bolhuis, Albert Willemsen, Erik S. Meyer, and Lucas P. J. J. Noldus. 2009. The Observer XT: A tool for the integration and synchronization of multimodal signals. *Behavior Research Methods* 41, 3 (Aug. 2009), 731–735. https://doi.org/10.3758/BRM.41.3.731

A COMPARISON

Table 1 highlights the differences and similarities between *PilotAR* and prior tools.

Table 1. Summary of the feature comparisons between the tools for conducting AR-related studies. Here, FPV = first-person view, TPV = third-person view, AR = virtual content view. Note: This list is not exhaustive. Although DART [12, 18] is meant for authoring AR/MR content, we have added it here for comparison as it supports various functions that could also be used for conducting AR/MR experiments.

Tool/Toolkit	Lee et al. [17]	Rey et al. [22, 23] (IXCI)	MacIntyre et al. [12, 18] (DART)	Nebeling et al. [21] (MRAT)	Proposed tool (***PilotAR***)
Purpose	Identify multimodal inputs for AR manipulation tasks and how AR display conditions affect them	Support research by streamlining immersive user studies	An authoring tool enabling rapid prototyping of AR applications by designers/non-technologists	An experimenter support tool for analyzing MR experiences	An experimenter support tool for conducting AR/MR **pilots**, data collection, and analysis
Target studies	WOz studies	Unity3D-based studies	AR studies	Unity3D-based studies	Pilot studies in AR/MR, including WOz
Prototype fidelity	High	High	Low-High	High	Low-High
Multiple experiment support	Single	Multiple	Multiple	Multiple	Multiple
Observation support	FPV, TPV	AR	FPV, AR, TPV	Interaction datapoints	FPV with AR, TPV
Recording support	✓	✗	✓	Processed spatial-temporal interaction data points	✓
Note taking	✗	✗	✗	✗	✓
Post-analysis	✗	✗	Not applicable	✓	✓
Summarizing and exporting	✗	✗	Not applicable	✓	✓

B IMPLEMENTATION

We used Python (3.9) as our primary programming language due to its cross-platform compatibility (e.g., Windows, MacOS). To achieve the tool's functionalities, we incorporated several third-party packages. The user interface (UI) was developed using Tkinter[1] and related theme packages, such as CustomTkinter[2]. The *PilotAR* utilizes Pynput[3] to monitor user inputs and FFmpeg[4] to handle screen recording. For video playback, we used Python-VLC[5] and audio transcription we used Whisper[6]. FFmpeg and websocket were incorporated to enable video and data streaming between the wizard and the observer in multi-experimenter settings. Detailed information about the **open-source** implementation can be found in https://github.com/Synteraction-Lab/PilotAR.

C *PILOTAR* SETUP

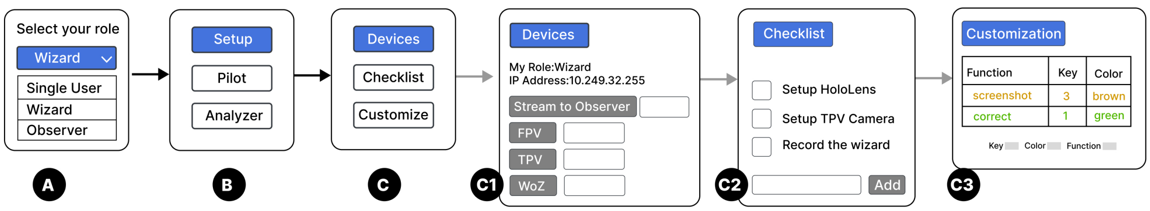

Fig. 5. Workflow of Setup UI. Upon starting the tool, the experimenter is prompted to select the role (A), including single- and multi-experimenter (wizard/observer). Then, menu (B) indicates the three major steps of conducting a pilot study: Setup, Pilot, and Analyzer. In Setup (C), there are three sub-steps, including device configurations (C1), checklist configuration (C2), and annotation customization (C3).

Role Selection (Figure 5A). Upon launching the tool, the experimenter is prompted to select their role: *single-user* for single-experimenter pilots or *wizard/observer* for multi-experimenter pilots.

Device Configuration (Figure 5C1). This task allows the experimenter to input essential information such as FPV and TPV connections (e.g., IP address, credentials), *Wizarding Interface* (e.g., Google Slides URL link or python file path), and screen recording inputs (e.g., video and audio source), making them all displayed on the monitor.

Checklist Creation (Figure 5C2). The checklist aids in remembering crucial steps during the pilot study, such as confirming OHMD, TPV camera, and recording. Customizable items can be added by typing in the provided space at the bottom.

Shortcut Key Customization (Figure 5C3). Experimenters can manage which *Annotations* are displayed during the pilot session (known as Pinned *Annotation*) and customize aspects like color, name, and shortcut key.

[1]https://docs.python.org/3/library/tkinter.html
[2]https://github.com/TomSchimansky/CustomTkinter
[3]https://pypi.org/project/pynput
[4]https://ffmpeg.org
[5]https://pypi.org/project/python-vlc/
[6]https://openai.com/blog/whisper/