JollyGesture: Exploring Dual-Purpose Gestures in VR Presentations

Gun Woo (Warren) Park
warren@dgp.toronto.edu
Department of Computer Science, University of Toronto
Toronto, Ontario, Canada

Anthony Tang
tonyt@smu.edu.sg
School of Computing and Information Systems, Singapore Management University
Singapore, Singapore

Fanny Chevalier
fanny@cs.toronto.edu
Department of Computer Science, University of Toronto
Toronto, Canada

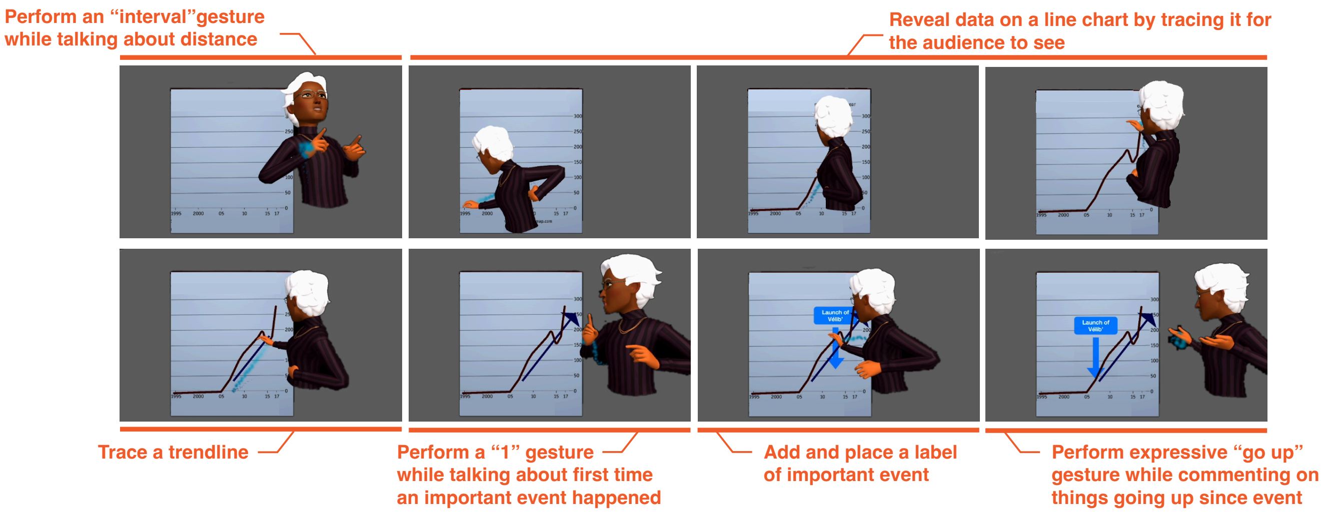

Figure 1: A speaker performing a presentation in VR using JollyGesture. The speaker can use expressive, meaningful gestures that are visible to their audience to enhance their narrative, e.g. expressing a distance with their hands apart. Such gestures have dual-purpose: they can also control the VR presentation system to alter content in the presentation as the speaker performs the gesture. For instance, the speaker can trace data and trendlines in charts, or move and place elements on the slide for emphasis.

ABSTRACT

Virtual reality (VR) offers new opportunities for presenters to use expressive body language to engage their audience. Yet, most VR presentation systems have adopted control mechanisms that mimic those found in face-to-face presentation systems. We explore the use of gestures that have dual-purpose: first, for the audience, a communicative purpose; second, for the presenter, a control purpose to alter content in slides. To support presenters, we provide guidance on what gestures are available and their effects. We realize our design approach in JollyGesture, a VR technology probe that recognizes dual-purpose gestures in a presentation scenario. We evaluate our approach through a design study with 12 participants, where in addition to using JollyGesture to deliver a mock presentation, we asked them to imagine gestures with the same communicative and control purpose, before and after being exposed to our probe. The study revealed several new design avenues valuable for VR presentation system design: expressive and coarse-grained communicative gestures, as well as subtle and hidden gestures intended for system control. Our work suggests that VR presentation systems of the future that embrace expressive body language will face design tensions relating to task loading and authenticity.

KEYWORDS

Gestural input, Virtual Reality, Presentation.

ACM Reference Format:
Gun Woo (Warren) Park, Anthony Tang, and Fanny Chevalier. 2018. JollyGesture: Exploring Dual-Purpose Gestures in VR Presentations. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 14 pages. https://doi.org/XXXXXXX.XXXXXXX

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Virtual and Mixed Reality (VR/MR) technologies offer new opportunities for presenters to engage audiences with presentation content.

Hans Rosling [6] offers a vivid example of this, where digital, information content is overlaid atop the video as if being prompted by Rosling's hand movements and gestures. While these effects may have been added in post-production, it has been inspiring for the HCI community. Several researchers have been motivated to explore MR approaches for augmenting delivery of presentations. For example, Saquib et al. [30] consider how custom body gestures can be used to directly manipulate digital content. Liao et al. [22] and Hall et al. [14] build on this approach to trigger animations and other elements based on voice and hand tracking, enabling presenters to engage in rich, dynamic storytelling in their presentations.

While presentations delivered with such flourish are undoubtedly engaging, they require that the presenter is aware of "what should happen next" in the presentation. For instance, a presenter must recall the specific gesture or phrase to enact in order to trigger the interaction with the digital content. For many types of presentations, this is extra load and can be understandably demanding on a presenter who has "*a lot to figure out on the fly*" [14].

In this work, we are similarly inspired by gesture-driven animation by presenters, particularly in the mode of Hans Rosling's presentation [6]. Thus, our goal is to design presentation systems that offer similar types of flourish in terms of how they are seen by an audience (as illustrated in Figure 1). Yet, rather than focusing on the output as perceived by the audience, our focus is on giving the presenter flexibility to present visual ideas and information in the order they would like. Here, our goal is to allow presenters to flexibly drive the presentation content through gestures (i.e. what the audience will see), all while providing an understanding of how to make these flourishes happen (i.e. gesture guidance).

Our design approach was to explore dual-purpose gestures, by leveraging an established vocabulary of gestures that could be perceived by the audience as being communicative (i.e. with an intention in relation to the content), as something that could also be used as input to the system. Presenters can use this set of simple-yet-expressive gestures to communicate effectively with the audience, as well as to control the presentation system. We also explore how a guidance/preview mechanism (illustrated in Figure 4) can be integrated in such system, so presenters can make *ad hoc*, in-the-moment decisions about what content to share and when.

We realize our approach in a technology probe called JollyGesture, a working prototype that allows users to explore these design concepts in a VR presentation system. To scope our work, we consider slide-based presentations rather than the more freeform storytelling contexts of other prior work (e.g. [30]). We evaluated our design approach through a three part study with 12 participants, who were asked first to brainstorm possible communicative gestures that also control a presentation system, then to use JollyGesture to deliver a presentation, and then finally to reflect on these gestures again in light of their experiences.

The main focus of this research is to gain insights to answer our research questions through the use of a concrete technology probe, enabling participants to share their perspectives on a technology that they might not have significant background experience with. Through the evaluation, we aim to provide formative answers to three specific research questions:

- **RQ1:** How can gestures that possess communicative purposes simultaneously serve as a control mechanism for the VR presentation systems (dual-purpose gestures)?
- **RQ2:** What are the needs and challenges of presenters when using dual-purpose gestures during VR presentations to create engaging presentation experiences?
- **RQ3:** How should the VR presentation systems designed to enable presenters to control visual content dynamically, in a flexible way, through the use of gestures?

On the basis of our study, we found evidence that presentation systems aiming to provide engaging experiences for audiences need to consider presenter needs carefully. If systems are to be driven by explicit, visible gesture-based interactions, then these need to resonate with how presenters think about the content itself—much as we have considered in JollyGesture. At the same time, designers need to be cautious about how much effort presenters are already putting into delivering presentations, and not add burden.

This work makes two contributions: first, we contribute JollyGesture, a technology probe that explores how communicative gestures visible for the audience can also be used to control a presentation; second, we contribute new empirical findings and insights into presenters' needs for using presentation systems that are focused on dynamic visual experiences for audiences, with particular foci on using dual-purpose gestures, and guiding users to use dual-purpose gestures in VR presentation systems.

2 RELATED WORK

Our work builds upon prior work in two areas: gesture control in extended reality and gestural expression in traditional presentations.

2.1 Gestural Presentation Control

In the context of delivering presentations, visionary communicators such as Hans Rosling [6] have explored reifying presentation content as floating objects in the 3D space, with post-production effects giving the illusion that this content is directly manipulated through gestural interaction. Recent work has realised this approach in interactive systems where the presenter interacts with visual content using full body gestures [14, 16, 22, 27, 30], or voice and/or hand gesture [22] to place, activate, or manipulate visual elements through real-time video compositing. Particularly, mid-air control interactions, like full body/hand gesture interactions, recently became feasible because of the wider availability of better camera/sensor technologies [8, 18]. RealityTalk allows the presenter to dynamically make a label or image appear at one's fingertips upon recognition of hand gesture and speech [22]. Hall et al.'s approach allows the presenter to interact with dynamic visualizations through hand tracking [14]. Saquib et al. [30] support visually-enriched storytelling though custom body landmark-action mapping, e.g. touching a virtual object with a hand triggers a pre-defined animation or allows to grab and move the object around.

This prior work gives us an exciting glimpse into possibilities offered by VR/MR in a variety of contexts, including storytelling [22, 30], visualization walkthrough [14], and slide-based presentations [5]. Without question, the audience experience of these systems is compelling. Our work builds on this by even more thoroughly exploring the experience of presenters. Building on the insights that rich gesture control may come at the expense of adding extra burden on the presenters to learn the gestures [14], or re-define mappings in the moment [9], we explore how to support presenters with a gesture vocabulary that serve a dual-purpose.

While gestures can be presented primarily to support accurate recognition [5], our work explores whether it is possible to reconcile gestures for system control with communication goals. Hall et al. [14] do this by using a *What you see is what I see* (WYSIWIS) paradigm [34]. Yet, this approach only works well if the presentation is scripted with no *ad hoc* deviations from the planned script. Our work contributes to this line of research, through the exploration of dual-purpose gestures and gesture guidance. We explore how presenters can use an established set of communicative gestures as control gestures to modify the presentation content, along with presenter-specific content to support flexibility and memory aids, in the form of gesture guides.

While we employ a VR setup in our study, we use a 1:1 mapping between movement in the real world to the virtual world, making our findings with regard to gestural interaction relevant for other extended reality systems.

2.2 Communicative Gestures in Presentations

Gestures are an important part of the communication involving speech [28, 35]. How people use gestures to enhance communication is well explored through observational studies of weather forecasting [12, 31], pedagogy [37], or television newscasts [41]. While a complete review is beyond the scope of this paper, of particular relevance to our context are gestures used for presentations involving forward facing audience. Researchers have embraced the importance of gesture support in communication [7, 24, 38], and evidenced that gestures have a significant role in communication in real life [10, 22, 40], as well as in animated agents [25]. In particular, these works have highlighted the importance of coordinating the gestures based on speech or eye gaze. Stronger results have been found that the high level of embodiment of gestures with respect to visual elements can enhance student achievement after watching a pedagogical animation with an avatar [21].

Prior research shows that gestures are an important communicative delivery tool for presenters, and that seasoned presenters are deliberate about their use of body language to not only communicate their message, but also to ensure that it does not take away from this message. While we know that there is diversity in how people use gestures, previous systems supporting presenters still tend to propose novel gesture vocabularies which can be recognized by the system (e.g. [14, 30]), but may at times appear unnatural for both presenters and their audience. Taking motivations from prior work on communicative gestures and existing gesture sets [1, 12, 13, 21, 25, 31, 37, 41], we explore how we could optimize gestural input, with respect to a dual-purpose: we focus on how well gestures allow presenters to conveying meaningful information to the audience, while meaningfully controlling visual content on the slides, and vice versa.

3 JOLLYGESTURE: PROBING THE DESIGN OF A VR PRESENTATION SYSTEM

We realized our design explorations for VR presentations in an iteratively designed technology probe called JollyGesture. Our goal was to develop a working prototype that would allow one to experience delivering mock presentations within a VR environment. Through the use of this prototype, we provide concrete, formative answers to research questions that we suggest. We designed two types of artefacts (the presentation interface tool, and a set of presentations) situated within a VR presentation room. To scope our exploration, we focused on canonical slide-based presentations, where the presenter's avatar stands before a display surface that shows slides. We first describe the set of ideas we explored with the JollyGesture probe before describing the version of the probe that participants used in our study.

We explored three design concepts (**DC**) in our probe:

Design Concept 1: Dual-Purpose Gestures JollyGesture explores the concept of dual-purpose gestures for *ad hoc* control and communication. Our focus was to identify gestures to control the slideshow (e.g. advancing slides or animations) that are consistent with a communicative gesture that a presenter would make for the audience's perception. Our goal with this design concept is to merge the idea of performing gestures for communicative purposes (as illustrated by Hans Rosling), with gestures for control (as illustrated by Hall et al. [14]), thereby allowing presenters to deliver presentations with *ad hoc* (non-linear) presentation flows.

Design Concept 2: Gesture Guidance for Presenters JollyGesture explores the idea of gestural guidance in VR. Our early explorations with JollyGesture showed us that for a presenter, recalling every one of the spatially mapped gestures would be difficult. Our probe uses a proactive and responsive feedforward mechanism to provide cues to the presenter about possible actions [4, 20, 26]. Our goal here is to address the challenge of presenters needing to remember exactly what gesture would enact which animation in the moment.

Design Concept 3: Heads-Up Displays (HUDs) for Presenters Finally, JollyGesture also explores the provision of a teleprompter, a monitor which allows presenters to know how they appear, and next/previous slide views. In principle, showing specific controls only to the presenter (and not to the audience) allows for many new types of slide controls and hints that we have not yet seen.

While an exhaustive description of our design process is beyond the scope of this document, we emphasize that the purpose of the probe was formative and not prescriptive. Given that such systems do not exist (or exist only in primitive forms), our goal was to consider how these design concepts, based on lessons learned from prior work, could be embodied within the context of a VR presentation system. By realizing them as a probe to elicit thoughts from user participants, we were also able to use the tool and try out the features, which allowed us to iterate on design ideas and features before conducting our user study.

3.1 Version of JollyGesture used by Participants

We describe our probe through a presenter's experience, for example, a professor who uses JollyGesture to deliver a presentation to a remote undergrad class. The presenter wears a head-mounted display.

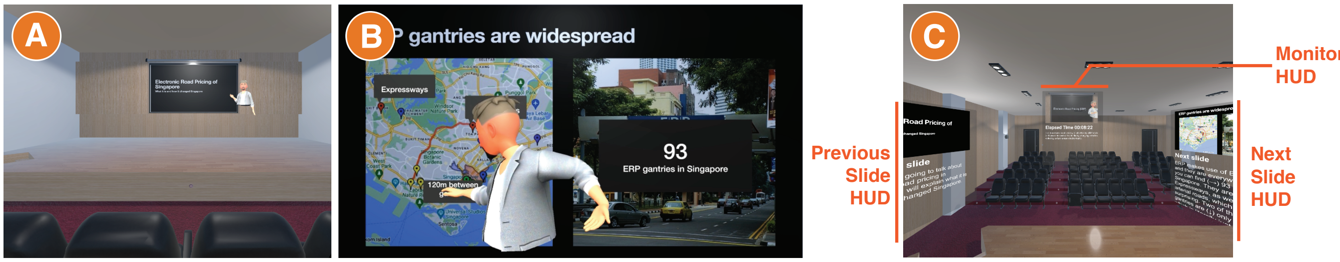

Figure 2: (A) View of a presenter's avatar from one of the audience seats. (B) As they circle around an area of importance, the audience can see a visual emphasis and an important message related to it, triggered by the presenter's gesture. (C) At any time, the presenter can see what the audience would see, elapsed time, script (front), as well as previous (left) and next (right) slide through heads-up displays (HUD).

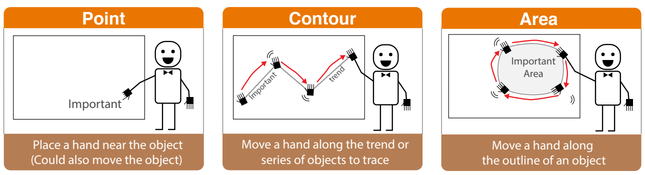

Figure 3: JollyGesture gesture types.

Virtual Presentation Room and Presentation Tools. In the virtual world, the presenter faces an auditorium (Figure 2A), and stands on a stage before a projection screen showing presentation content. The presenter's avatar is controlled by their body movements: when they physically walk, their avatar moves; when they reach their arm out, their avatar correspondingly reaches out to touch the screen (Figure 2B). When the presenter looks towards the seats, they see heads-up displays (HUD) (Figure 2C), that show the previous and next slides, along with the current timing, and slide notes. The presenter also sees a simulated monitor showing how their avatar look live to the audience.

Dual-Purpose Gestures in JollyGesture. As the presenter delivers the presentation, they use dual-purpose gestures to communicate their ideas to the audience as well as to control the presentation. For instance, to emphasize an area on a map or image, they perform an area-style gesture around that area on the slide, which creates a highlighting effect. Similarly, to add emphasis to a term in their slides (i.e. to underline it), they trace a line under the term, which triggers the underline animation on the slide. These control actions (and the corresponding gesture) were chosen to be highly visible and meaningful to the audience (e.g. point at an element, emphasize an area). Thus, as the presenter moves their hands as part of their presentation delivery, the system recognizes these as gestures, and triggers animations or transitions as appropriate (Figure 2B).

JollyGesture relies on a relatively simple set of established gestures as in Figure 3. These are applied to a set of communication intentions that we developed based on study of prior literature studying gesture use in presentations [1, 12, 21, 25, 31, 37, 41], and triggered by gestures as described in Table 1.

Design Rationale and Approach. We first constructed the set of intentions outlined in Table 1 based on a review of prior literature supporting slide-based interaction (e.g., [30]) as well as current, existing slide-based tools (e.g. PowerPoint, Google Slides, Keynote). We also considered the many domain-specific "during presentation" actions presenters try to illustrate and show as they guide learners through concepts on presentation slides. To map these to gestures within JollyGesture, we iteratively designed these based on inspirations from literature [1, 12, 17, 21, 25, 31, 32, 37, 41, 42].

Gestural Interaction Zones. We define several zones of interaction based on the presenter's hand proximity to the presentation screen. When the hand is within 0.13m (5"), gestures here can trigger effects specific to the slide (if there are any to be triggered). When the hand is between 0.13m ~ 0.26m from the presentation screen, gestures navigate between the slides. When the presenter's hand is beyond 0.26m of the presentation screen, the presenter can gesture freely without any impact on the presentation slides.

Design Rationale. We base our design from prior work in proxemics [23], drawing particular inspiration from Vogel and Balakrishnan [39] who also employ zones of interaction that depend on the user's proximity from the display.

Gesture Guidance. The presenter may not recall what gestures are available and what the result of them may be. JollyGesture provides guidance using visual previews and gesture paths. Figure 4A illustrates a moment in the presenter's presentation where they can show either a purple, yellow, green or blue element on the screen. These floating elements are not seen by the audience, but the presenter can see that if they follow the purple snake-path (Figure 4B), the corresponding element will be selected and appear to the audience as they follow the path to place the element (Figure 4C); if they were to follow the blue path, the blue element would appear. As in OctoPocus [11], the guidance mechanism automatically prunes branches that become irrelevant as the gesture is being performed to reduce visual clutter. Unlike OctoPocus, options are presented in a marking menu configuration [3, 19]. By showing multiple interaction options with spatial mapping, the presenter sees available gestures provided (and associated slide changes), and can thus plan which interactions to perform when delivering the presentation.

Design Rationale. JollyGesture's gesture guides focus on positional accuracy. We based our visual approach primarily on Fennedy et al.'s [11], and drew inspiration from extensive prior work (e.g. [2, 4, 15, 33, 36]). JollyGesture shows both a snake-like gesture guidance path, and a target point where to bring one's hand to.

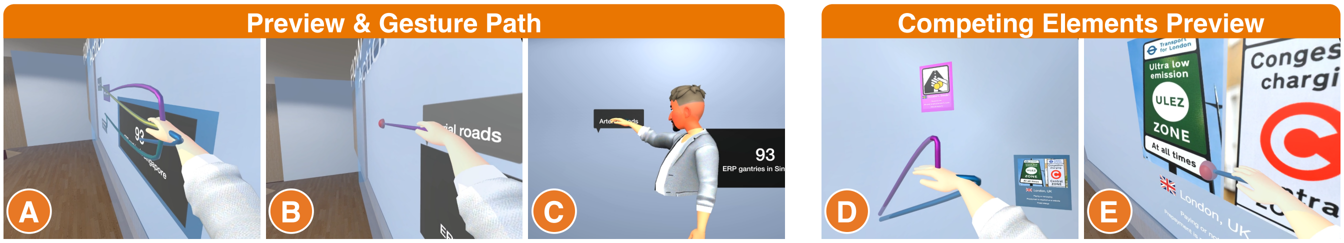

Figure 4: The preview mechanism in JollyGesture. (A) A presenter can see the previews of elements, and gesture guidance path to follow to place an element. (B) If the presenter follows the purple path, (C) the corresponding element appears at the presenter's fingertip and is placed where intended. Sometimes, several elements could be placed on the same part of the screen. In those competing cases, previews appear along the gesture guidance path (D); following the blue path here selects the associated element (E).

Intention	Transition / Animation	JollyGesture Gesture
Move a slide forward/backward	Transition	Contour (right to left / left to right)
Emphasize a word/term	Animation	Contour (line under the term)
Trace a data/trend line in a chart	Animation	Contour (along the data/trend line)
Add description to a term	Animation	Point
Add label on an image/map	Animation	Point
Add a step in sequence	Animation	Point
Choose an item between options	Animation	Point
Emphasize an area on an image/map	Animation	Area

Table 1: We identified several presenters intentions. These involve communicating an idea or controlling the slide deck in some way. Intentions are linked to a transition or animation, and we mapped these to a particular gesture in JollyGesture.

3.2 Mock Presentation Designed for Probe

As part of the probe, we built three slide decks that allow people to experience JollyGesture without having to author a presentation. Each deck contains seven slides and includes an explanation of terminology, a map, a three-step process explanation, a side-by-side comparison, and charts containing trends to be compared. We chose topics that we expected participants would have no direct knowledge of, but that they would find interesting: Electronic Road Pricing (ERP), a congestion charge system in Singapore; Vélib' Métropole, a bike-sharing scheme in France; and Reinsurance, an insurance for insurance companies. Slide decks had equivalent versions for JollyGesture and Keynote, with similar animation effects. Slides can be found in the Appendix A. Below, we describe how presenters could add communicative flourish for three sample slides, illustrated in Figure 5. These are triggered *ad hoc* via gestures but do not need to be (i.e. the presenter can trigger animations out of order or skip over them altogether if they so desire).

Example slide interaction: Map/Image. In the bike sharing slide deck, a presenter could highlight the high density of the bicycle hiring stations around Paris (Figure 5A). The slide first displays an image and map, to which the presenter can progressively add information, i.e. a quantitative value on top of the image, and important facts at the top of the map (Figure 5A1). The presenter does so by following the corresponding snake paths, as in Figure 4A-B. To emphasize an area of interest in the map, the presenter can draw the audience's attention to through an area gesture (Figure 5A2), which triggers visual emphasis (Figure 5A3).

Example slide interaction: Comparison. A comparison slide allows the presenter to summarize the concept that has been discussed, and compare it with another concept. In the bike share slide deck, two options for comparison against the Parisian system can be chosen from, i.e. similar sharing systems in London and Toronto (Figure 5B). Based on the audience reactions, or other factors as the presenter sees fit, they can choose either option by performing the corresponding gesture as in Figure 4D-E.

Example slide interaction: Trendlines. The presenter can emphasize interesting trends about bicycle users (Figure 5C). To achieve this, they can trace the chart as in Figure 5C1-2.

Unlike conventional presentation tools which only allows one pre-authored contents to be presented linearly, JollyGesture emphasizes the *ad hoc* nature of presentations. Here, presenters can choose in the moment what content and how to show during the presentation, allowing them to "author" their presentation live, in the moment, in response to the reaction of the audience.

4 STUDY DESIGN

We conducted a study to understand presenters' reactions to the design ideas embedded in the JollyGesture technology probe, and to explore how to improve the gesture support within the system. To this end, we recruited participants to deliver a mock presentation with JollyGesture. For our presenters, this would be the first time they would use the dual-purpose gestures, and so we wanted to understand whether they understood the concept easily and could make use of the gestures easily. Similarly, we wanted to understand whether they could make use of the gesture guidance system to enable *ad hoc* delivery of the presentations. Next, we wanted to use the study to understand the variety of gestures and bodily actions that they would use as part of their presentations to inform future design iterations of JollyGesture.

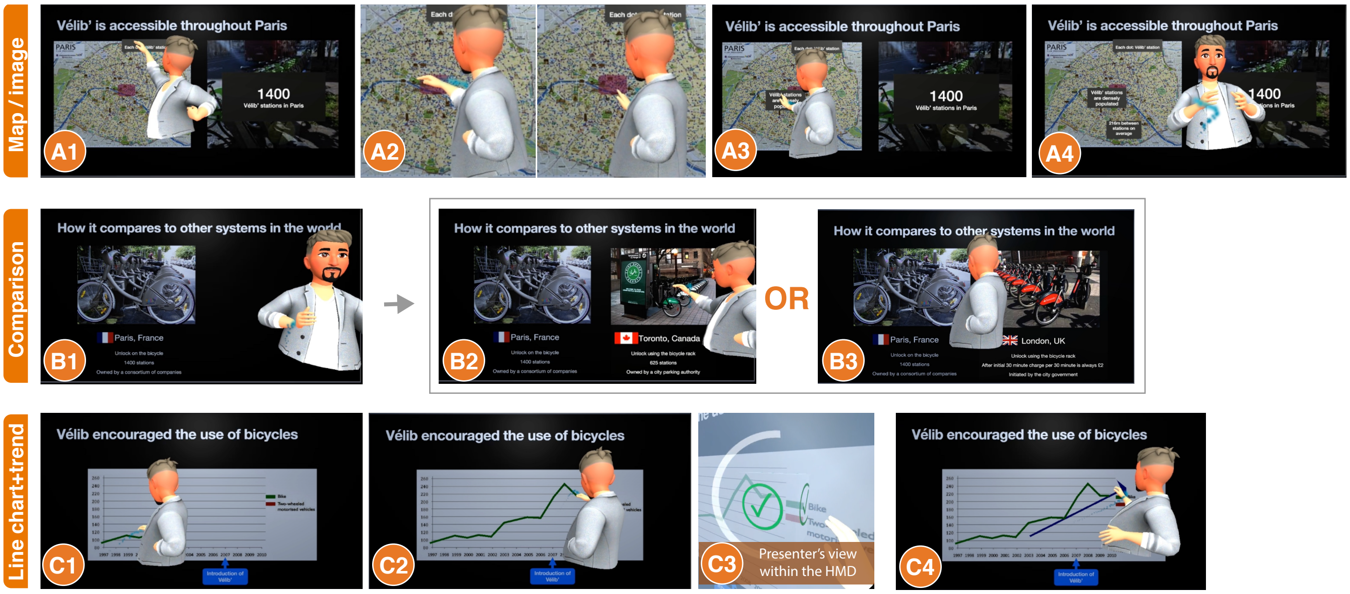

Figure 5: Examples of interaction flow in our probe. A: Slide with a map where presenter can add information through pointing gestures (A1, A3), emphasize areas through area gestures (A2), and free-form gesture to support communication (A4). B: Slide with two alternate stories. Following one or the other snake-path, presenter can choose to bring up a different image for their narrative. C: Slide with a chart, where presenter can progressively sketch the data line (C1-3), and a trend line (C4) through contour gestures.

4.1 Apparatus

Participants wore a Meta Quest 2 VR headset running JollyGesture on a PC to deliver their mock presentation[1]. JollyGesture was a custom-built technology probe built with Unity, Meta's Meta Avatars SDK, and the Oculus Integration SDK (for hand tracking).

To explore gesture design opportunities, participants used a large digital SMARTBoard[2].

4.2 Procedure

Each study session (~1h30) had three phases: (1) naïve brainstorm; (2) JollyGesture mock presentation, and (3) informed brainstorm.

In Phase 1 (naïve brainstorm), we asked participants to imagine they were delivering live face-to-face presentations in front of the large digital SMARTBoard, and to help us create gestures that would enact the intentions described in Table 1. We asked participants to imagine and perform these gestures. The experimenter then re-enacted each gesture to verify the gesture and to understand the participants' thought process.

In Phase 2 (JollyGesture mock presentations), participants delivered a mock presentation (randomly chosen amongst our set) with the JollyGesture prototype. Participants were given time to familiarize themselves with JollyGesture beforehand, watching a live demonstration before donning the HMD, and engaging in short practice tasks. Following this phase, we conducted a short semi-structured interview to understand participants' experiences with the technology probe, particularly focusing on challenges they experienced using the probe.

In Phase 3 (informed brainstorm), we used a similar process as in Phase 1, with the goal of gathering further insights on gestures that participants could think of. We expected here that their ideas would be informed by their experiences with JollyGesture—either to incorporate the ideas from their experience, or to produce gestures that contrasted with those in JollyGesture.

Study design rationale. We chose to include an initial brainstorm session (Phase 1) to help cement participants' personal opinion on the problem and solution space, grounded in their own practice. This phase allows us to (i) provide a well-formed initial anchoring reference in participants which they can compare JollyGesture against, and (ii) reduce any confounds introduced from prior exposure to the system. We include a second brainstorming session (Phase 3) to re-evaluate participants' initial ideas and additional thoughts which working with the probe would have inspired and/or informed.

Pilot Studies. We ran three pilot sessions before the procedure was formalized. This allowed us to introduce re-enactment captures, and optimize room configuration/calibrations.

4.3 Participants

We recruited 12 computer science graduate students (6 male, 6 female) through convenience sampling (age: 27.0±2.9). Nine reported giving presentations regularly; 11 had experiences with VR, though the majority of them (5/11) interacted just a few times; 11 had

[1]OS: Microsoft Windows 10, CPU: Intel Xeon E5-2667 v4, RAM: 32 GB, and GPU: NVIDIA RTX 2080 Super
[2]SMART SPNL-6065 65″

experience trying to make presentations interactive by using animation effects on PowerPoint/Keynote/Prezi, Myo Gesture Control Armband[3] (P9), or a laser pointer (P4,P7).

4.4 Analysis

We conducted a thematic analysis on the data collected in Phase 2, primarily focusing on participants' reactions and responses during the interview regarding our design concepts (see section 5).

We coded the gestures collected in Phases 1 and 3, by intention and type (Table 1), and further analyzed those which did not match any code, as we were interested in how people would conceptualize and think about gesture use in presentations (see section 6).

5 FINDINGS: RESPONSE TO JOLLYGESTURE

We conducted user studies to answer how dual-purpose gestures could be utilized for VR presentations, and what are the associated needs and challenges of presenters. During the user study, participants were able to deliver the mock presentations using JollyGesture. The rich functionality provided by our prototype (compared to a traditional slide-based tool) was considered useful, albeit at the expense of added complexity in its use. We organize participants' reactions to JollyGesture based on the three design concepts.

5.1 Dual-Purpose Gestures (DC1)

Making Established Communication Gesture Dual-Purpose is Meaningful and Useful. Participants appreciated the coupling of gestures for control of the presentation and for communicative effect. For example, they found it useful to be able to draw trend lines, or place labels to emphasize certain concepts with gestures—conceptually, such gestures are useful as a pedagogical approach (P3, P4, P5, P6, P7, P11, P12). For many, it was consistent with their existing approach of gesturing at the slide content itself. P11 commented: *"The gestures themselves are actually gestures that I would perform if I was doing a presentation. [...] So in terms of naturalness or how intuitive it is, I think it's very good."* Similarly P6 explains that they frequently circle and point at things on slides, and that the added visual emphasis that the system provides is effective: *"I really like the graph lines because I already do that so often for images, for example, I circle this, circle that. So I already enjoy the interaction of pointing stuff out and using my hands to emphasize trend lines."* In principle, the ability to add this emphasis on the fly would enable a high level of creativity in how a presenter delivers a presentation, along with where and when visual emphasis is added, as argued by P3: *"I think they can get pretty creative—even placing a bubble or drawing a circle. I can see a lot of potentials with how that interaction can be built into a lot of different types of transitions."*

Physically Needing to Move. JollyGesture relies on a 1:1 mapping between movement in the real world to the virtual world. While this allows presenters to directly use communicative gestures as they would typically do in a face-to-face presentation, making it a requirement to trigger flourishes raised some concerns for participants. Several of them found the need to physically move around in space, as well as to lift their arms to control content, to be potentially problematic in the long run. For instance, in one scenario, participants needed to move from one side of the screen to the other to reveal a line chart. Six participants made mention of this extra locomotion being problematic. P5 describes this extra movement as disruptive to their flow: *"I didn't really like how much walking there was involved. It feels like when I want to talk about something, I have to wait for a long time because I need [to execute on the gesture]."* Fatigue was also a concern, *"My hand will get tired if the presentation is an hour and a half, and [requires me] to activate all this stuff. I'm sure my hand and my arm will be super tired."* (P12)

Gestural Zones Can Support Gestural Flexibility and Control. The use of gestural zones gave participants a level of flexibility in how to think about their gestures being recognized and for what purpose. P12 explains that the zones releases them from needing to be concerned that a gesture would be inadvertantly recognized: *"I like the idea of having to play [with] two modes. One mode is like very high level, which is like navigating slide. The next level is more like content-based level, and you interact with the actual content. So I really like that separation."*

Reliably Triggering Commands. While participants generally valued the ability to freely gesture as a way of communicating to their audience, our approach of using fixed proxemic zones for triggering commands was also somewhat challenging to use in practice. Participants found that JollyGesture both had false positives (triggering when not desired), as well as false negatives (failing to trigger at the right time). This caused some visible frustration, and the absence of a quick, easy-to-use "Undo" function was also problematic. P12 describes the challenge as having gestures unintentionally being recognized as commands: *"I think that the most unsatisfying experience was the false positive for activations."* Participants might, for instance, think they were outside the activation zone and inadvertently trigger an effect or vice versa. The proxemic zones were not intuitively obvious from the perspective of a presenter *during* a presentation. P1 contrasted the experience with explicit sensory feedback when using traditional input devices: *"One of the problems that I had here was that I wasn't sure if things would work or not. If I had a clicker, the feedback from this making the click sound and then releasing it is everything that I need to know."*

Seamlessly Choosing Alternative Content. We were able to find support for the fact that our intention to support multiple options were correctly communicated. P2 clarifies that this functionality helps presenters to confer content in a non-linear way, as opposed to the pre-defined presentation flow that we usually see, and mentioned that it is nice to have the option of choosing between alternative contents for the comparison slide.

Summary Dual-purpose gestures are perceived by participants as intuitive and useful, facilitating natural interaction and creativity. However, the need for physical movement to execute these gestures can lead to fatigue and disrupt flow, necessitating design refinements. Gesture zones can enhance flexibility in control while utilizing dual-purpose gestures, with reliability in command triggering, more intuitive feedback, and the implementation of "Undo" functions being essential.

[3] An electomyogrphy sensor-based gesture recognition arm band

5.2 Gesture Guidance for Recognition (DC2)

Gesture Guidance is a Valuable Aid. Each slide in the mock presentations had multiple possible gestures, each with different effects (e.g. next slide, add label, add contour, etc.). Based on questionnaire responses, participants understood the gesture guidance paths, and generally felt that the assistance helped them to execute the gestures (8/12 participants). Quotes from participants, such as *"I think it's very straightforward, easy to use. I don't have any issue of learning how to use it."* (P12), *"I can follow that line to do some gestures that I want to highlight."* (P4), and *"I liked the idea of being able to follow lines and stuff like that. That's super cool. I think that works pretty well."* (P8) further supports this result.

Increased Control Complexity. The gesture guidance, however, introduced more visual clutter into the scene. This meant that participants would need to be thoughtful and careful about how to interpret what they saw while in the midst of their presentation. For instance, P5 describes having to follow different paths to determine the result of each action: *"It took me quite a bit of time to figure out sometimes which line is which, which line refers to which part. And I feel like especially if you have more components on the slide, this might be difficult."* This adds to the complexity of using the system during a presentation, which P9 noted is problematic as it adds cognitive load right when their effort is being placed on delivering a great presentation: *"The control logic has such a huge mental load there, which actually makes those materials useless for me because I don't have the bandwidth to solving those issues."* JollyGesture's heavy reliance on visual feedback made it difficult to turn away from it while interacting with the system. In fact, it forced participants to focus their attention on the presentation screen rather than look at their audience. P3 explains that on reflection, working with the system did not allow them to visually engage with their audience: *"And I didn't really look at the audience, because I was focused on making the slides progress."*

Perception of Restricted Flow. One consequence of the presence of the guidance paths was that participants felt that they *needed* to follow these, rather than being free to deliver the presentations as they saw fit. To wit, rather than feeling like the possibilities enabled ad hoc decision making about presentation flow, P1, P2, P9, and P10 described the guides as constraining on the flow of their presentation delivery. P1 described that *"[JollyGesture is] controlling me in terms of what I should do and what not"*, which P9 explains makes them feel like an automaton without agency: *"[JollyGesture] gives me the feeling I am actually like a robot following something. It's not something I'm making in presentation."*.

Summary Gesture guidance can aid presenters in performing dual-purpose gestural interactions. However, gesture guidance mechanisms themselves can cause visual clutters, resulting in increased control complexity. Guidance mechanisms can also cause a false impression that the flow of presentation is restricted.

5.3 Heads-Up Displays for Presenters (DC3)

See-Through Presenter Screens Helps Attend to the Audience. We designed JollyGesture also to answer how should the VR presentation systems be designed to control the visual contents. JollyGesture makes use of heads-up displays (HUDs) that are visible only to presenters and not audience members. Participants were unanimously in favor of placing these HUDs as hovering see-through panes that allowed the presenter to see the audience. For instance, P1 and P12 describe this as a way of attending to the audience even while devoting attention to presenters' aids: *"You're both facing towards the audience, and looking at the slide notes."* (P1), *"You have a see through screen script. And you're actually reading a script that it looks like you're looking at the audience."* (P12).

Display Location Matters. In the version participants used, the HUDs followed the presenter's head. While this had the convenience of being within reach at all times, it had the effect of making the view presenters had cluttered at all times—even when the content or information being presented was not useful in the moment. To resolve this problem, some participants suggested allowing the HUDs to remain fixed to a certain location (much like a laptop on a podium). P12 explains that this approach would allow the HUDs to function like a secondary display for a desktop computer: *"It can always be here when I really need to look at it. I can look at it and I can just read [briefly]."*

Summary Additional tools like heads-up displays are concrete examples that demonstrate how VR presentation systems should be designed to enable presenters to control visual contents.

6 FINDINGS: GESTURE BRAINSTORM

To answer how dual-purpose gestures could be utilized in VR presentation environments, and what are the needs and challenges associated, our focus was to understand which gestures participants do use already, or would imagine using for the intentions in Table 1. We used gesture brainstorming to obtain user insights first, as opposed to running a direct user study in evaluating the JollyGesture technology probe. This decision was informed by our literature review, which revealed inconclusive findings about communicative and control gestures in VR environments[4]. Specifically, there was a lack of significant insights into how people use gestures in VR to communicate. By using brainstorming sessions, we aimed to gain concrete insights into the design and functionality of dual-purpose gestures, ensuring that they are practical and intuitive for VR presentations. We collected a total of 252 gestures in Phase 1 for these intentions, similar to prior work on user-defined gestures [29]. We also collected gestures in Phase 3, but found that these were mainly refinements of gestures already covered in Phase 1, which were mainly matching with our expectations; for clarity, these are excluded from the quantitative analysis. We coded each of the participants' described gestures based on their type and intention through closed coding of speech and movement data.

We report on three main insights that emerged from our analysis: first, that the gestures participants chose *before* being exposed to the probe were largely consistent with those built into JollyGesture; second, that participants generated many dual-purpose gestures on their own—that is, the concept was aligned with their mental model of how gestures could be used and interpreted; finally, that

[4]There was some literature which provided a basis for the design decisions made for our JollyGesture technology probe e.g., [14, 22, 30].

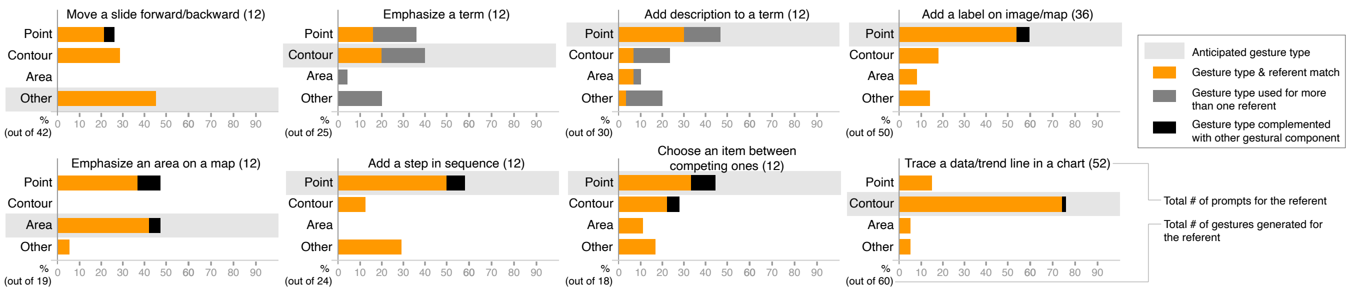

Figure 6: Distribution of gesture types imagined by participants, compared against the gesture used in JollyGesture.

participants still desired both communication-focused and control-focused gestures in certain circumstances.

6.1 A Consistent Set of Gestures

Figure 6 shows distributions of gestures participants came up with for each intentions, before seeing JollyGesture. The three gesture types (point, contour, area) captured 76% (192/252) of the corpus. For most intentions (7/8), the dominant gesture type was consistent with what our probe uses (showed with light gray background). Across all intentions, we observed an average 79±15% of the gestures imagined by participants are of a type that matches the JollyGesture's gesture type for that intention.

For a small number of intentions, participants gravitated around two different gesture types (e.g. "emphasize an area" drew an equal share of area and point gestures). We also noted a small portion of gestures (6%, 14/252) where participants extended a base gesture with a complementary component (e.g. point at an element while shaking the head)—showed in black in Figure 6; and instances (6%, 16/252) where participants described one gesture to achieve two intentions (e.g. emphasize a term and add a description to the term in a single pointing gesture)—showed in dark gray. These suggest that there are rich nuances in how people would like to use, combine, or extend gestures, as well as variety in terms of the scope they imagine these gestures have. Both aspects have implications regarding robustly recognizing gestures (i.e. gestures can be complex) and attributing the proper intention(s) (i.e. people may mean different things with a similar gesture).

Participants' gestures made it clear that they understood their bodies to be part of what the audience would perceive as part of the presentation. We observed this awareness in how deliberate participants were in choosing gestures for different purposes: some gestures were dual-purpose—for the purpose of system control and audience communication; others were mainly communicative—that is, mainly for the audience; while others were subtle—mainly for controlling the system without distracting the audience.

6.2 Dual-Purpose Gestures

Participants described several dual-purpose gestures beyond the ones embedded in JollyGesture. These gestures are those that have clear communicative value/purpose in their appearance (sometimes spatially in the context of the slide, other times not), while also clearly having an implication for system control.

Exaggerated or Large Movements. Some participants performed exaggerated movements to draw an audience's attention to specific ideas. As illustrated in Figure 7A, P12 traced a graph with their hand, but exaggerated height of the beginning, middle and end points of the chart to emphasize the trends. P8 used the speed with which they traced the chart to emphasize the slope–they moved their hand faster when the slope was steeper, but slower when the slope was shallower. Similarly, P7 described using large footsteps toward the audience could be used to draw the audience's attention to the items in a sequential list. Here the idea was that the system would respond by showing (or hiding) each item based on each step (step forwards: show; step backwards: hide).

Highly Expressive Gestures for Control. In many cases, participants produced gestures that had a clear spatial relationship with the content, and are derived from conversational communicative gestures. For instance P1, P2, and P8 all counted with the digits of their fingers as they revealed items in a sequence (Figure 7B, C). In P8's case (Figure 7-C), they also incorporated a two-handed gesture, where one hand represented the "running" count in the sequence, while the other hand would touch the next digit to indicate the next step. Finally, P12 performed a variation of this gesture when explaining terminology, where they pointed a closed fist towards the audience, and unfolded fingers to enhance engagement (Figure 7D): *"When you are giving a talk, if your palm is facing the audience, they will feel more engaged, and [they will feel that] we are on the same side."* Such conversational hand gestures can be used also for system control. P12 suggested when explaining trends on three lines on the line chart, they could unfold three fingers on their dominant hand and move their index finger on their non-dominant hand from one of these three "reference" fingers to explain the trend regarding that line (Figure 7E).

Such gestures were also produced to bring communicative emphasis to concepts being illustrated in the slide content. For instance, P1, P6, P7, P10, P11, and P12 performed two-handed gestures to emphasize certain distances in a figure. P1 and P10 placed their hands on either side of an important term that needed to be underlined, or where a label should appear. P10 described an idea of placing one concept in each of their hands to compare the two elements. P6 and P7 used smaller pinching-style gestures to emphasize the small differences in a chart, while P1, P11, P12 used large two armed gestures to emphasize large differences in the chart or quantities (e.g. Figure 7F). P11 used a variation of this, where one hand stayed still (to mark the maximum value), while the other moved toward the other with waving to emphasize growth (illustrated in Figure 7G). Finally, P12 used a similar style of gesture, where the two hands would represent a timeline. They then performed

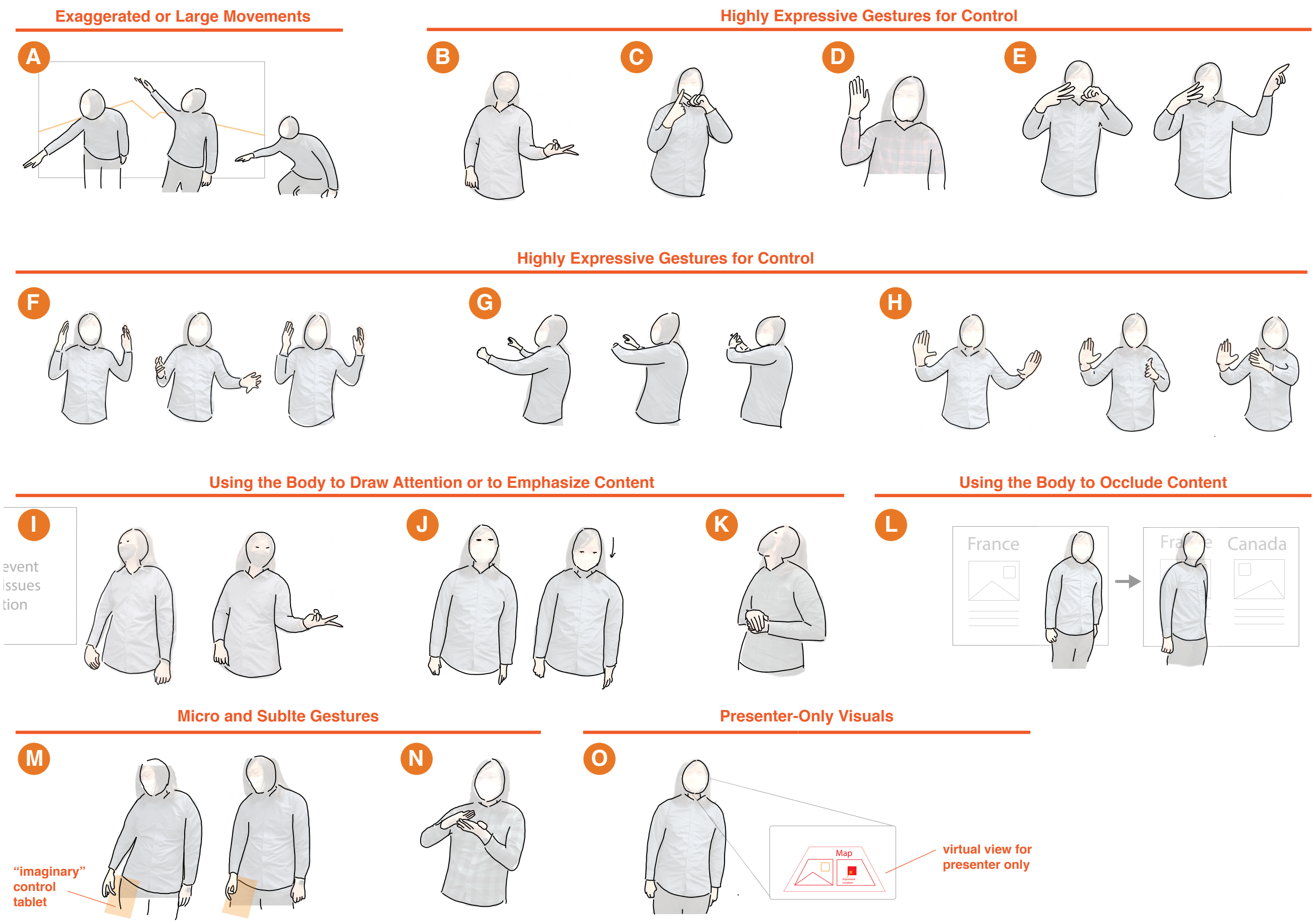

Figure 7: Examples of rich gestures observed during the gesture brainstorm phase of our study. (A-H): Dual-purpose gestures were proposed as communication tools to support the narrative, and as a mechanism to trigger changes in slide content. (I-L): Participants also generated gestures mostly intended to the audience (without any effects on the presentation slides), and (M-O), on the other side of the spectrum, they also proposed control gestures and features aimed to be invisible to audience.

a "karate chop" style movement in the middle to emphasize an incident that happens in the middle of the timeline (Figure 7H).

Semantically-Resonant, "Custom" Gestures. Participants also suggested the possibility of designing specialized "custom" gestures. For instance, P11 suggested the use of an imaginary globe as a metaphorical cue to change the geographic context. In one example, when discussing the comparison between bike share in Paris and Toronto, P11 suggested being able to rotate a virtual globe. Such a virtual proxy would only be relevant to this particular slide, but it would add additional context and richness to the presentation.

6.3 Communication-focused Gestures

We also saw many gestures that were mainly intended for audience communication, and decidedly not for system control.

Using the Body to Draw Attention or to Emphasize Content. P1, P6, P7, P11, and P12 used their head orientation to emphasize content by changing their gaze direction (6 gestures in total). As illustrated in Figure 7I, P1 looked at the screen as part of a hand enumeration gesture (where they are counting items 1, 2, 3, and so on on their fingers). The deliberate head re-orientation in combination with the enumeration gesture was intended to draw the audience's attention to the text of each enumerated item on the slide. P7 also used head nod as in Figure 7J, or conspicuously looking in the direction of an item on the slide, as in Figure 7K. In each of these cases, the head orientation is used to bring audience's attention to either the slide, or elements/items on the slide itself.

Using the Body to Occlude Content. Participants also performed gestures to occlude content deliberately with the intention to draw audience's attention towards the elements that the presenter was explaining in the moment. For instance, P1 and P3, used this type of approach when comparing two different concepts. As illustrated in Figure 7L, P1 and P3 moved their location from right half to the left half and vice versa to explain each concept to focus the audience's attention in the moment.

6.4 Control-focused Gestures

Some participants produced gestures that were deliberately not intended for audience communication, and strictly for system control. In producing these gestures, it is clear that participants intended for them to be minimally visible from the audience's perspective.

Micro and Subtle Gestures. In general, we observed that these were small or subtle in nature. For instance, P8 described tapping on their body as a way of controlling the slide presentation. As illustrated in Figure 7M, they imagined a control tablet on their thigh to trigger actions through finger taps, swipes, and place labels with taps. P10 similarly imagined using their non-dominant hand as a touch pad to control slide content (Figure 7N). P8 and P10 produced gestures using foot tapping gestures for slide control—again, largely invisible to the audience.

Clutching. Finally, two participants saw value in "turning off" gestures for a period of time to allow the audience to read the data in the slides. This is similar to the concept of clutching. P6 wanted to have such a feature when they show line charts to the audience. P1 suggested providing this kind of a sustained pause when showing both hands palm with fingers unfolded.

Summary. The two brainstorm phases of the study provided us with several pieces of insight. First, participants generally produced dual-purpose gestures that were consistent with the expectations we had in designing JollyGesture—that is, gestures with both communicative intent (for the audience) and control intent (for the system). All these ideas were collectively exhausted *before* being exposed to the probe, i.e. Phase 3 did not result in additional gestures. Second, the participants were keenly aware that their bodies were part of the presentation as seen by audience members. To this extent, they also generated a broad range of gestures: new dual-purpose gestures that we had not considered (and some of which might be considered "custom" for the content), gestures that should not be interpreted by the system (communicative intent-only), and also gestures that were subtle/hidden—strictly for system control, and not for the audience to see. Together, these findings suggest that even more exploration needs to be done—perhaps with a wider range of slide content, as these may elicit new ideas (i.e. custom dual-purpose gestures) that we had not considered at all.

7 DISCUSSION

Our experiences designing, experiencing and studying JollyGesture as a probe revealed several important aspects of supporting VR presentations that suggest paths for further exploration.

7.1 Gesture Design Continuum: Control, Communication, and Everything in Between

Our findings from the first "naïve" brainstorm session showed that presenters are cognisant of the fact that their bodies are viewable by the audience, and very much a part of the presentation. Thus, the presence of the presenter's body, and the gestures it emotes can be used to emphasize and draw attention to ideas and content in the slides. Much like we did, participants envisioned presenters using dual-purpose gestures—both to control the system, and to communicate with their bodies to the audience. Further, participants imagined themselves using their bodies in coarse-grained ways to pace content, or to cognitively "chunk" their delivery to make things interpretable for their audience. We also saw that participants envisioned gestural controls that were decidedly *not* intended for the audience to see—either subtle gestures, or gestures intended to be hidden from the audience's view.

We view gesture interaction design for VR presentations as belonging on a continuum: some gestures are intended primarily for communicating with the audience, while others are intended primarily for controlling the system. In the middle, there exists a set of dual-purpose gestures, which can conceivably be used for both communication and control. In JollyGesture, we explored how an initial set of three generic types of gestures well-established in communication, can be extended to *also* control the presentation content. Our findings suggest that beyond this set (which already suits a number of intentions), there exists a rich vocabulary of communicative gestures that people already perform, or imagine performing, which could be augmented to add flourish to presentations. We derive from our observations that, where possible, dual-purpose gestures should be based on existing conversational gestures, as these have potential for more seamless integration within one's performance flow, compared to novel gestures which presenters may find less natural to perform to trigger effects (e.g. [14]).

We suspect that there will be additional work that is required to explore this space. Many dual-purpose gestures may be domain and topic specific, and given that VR provides presenters with a wider palette to design visual materials for communicating ideas, we expect that the space for "transitions" and "interactivity" to be correspondingly large. We also argue that control of digital content through dual-purpose gestures only is not desirable, as it can be seen as a forcing mechanism. This calls for VR presentations of the future to support gestures across the continuum that we have described, as different moments in a presentation will demand a presenter to gesture with different intentions.

7.2 Presentation Flow: Gesture Recognition

Participants made clear in using the JollyGesture probe that having to attend specifically to the presentation system itself interrupted the flow of their presentations. We expect this is representative of most presenters when the presentation experience goes beyond rehearsing: that delivering a presentation is mainly about delivering a good performance for an audience. We found that JollyGesture took away from this experience for many presenters: the added functionality also added undue complexity that needed to be managed by the presenter *during* the presentation.

Part of the complexity was due to the uncertainty around whether/when a gesture might be recognized, which could inadvertently trigger an animation in the system. And, while this could be resolved by better gesture recognition, P1 explained that the *ad libbing* that brings presentations "alive" comes between moments of interacting with the system. If the gesture recognition system is always "on", it is difficult to know when a gesture might inadvertently be recognized. To work around this issue, prior work has often relied on gestures that are unlikely to occur (un)intentionally during a presentation. However, such gestures could feel contrived or awkward to both the presenter and audience [14, 30]. In JollyGesture, we have chosen a relatively simple method for gesture recognition:

activation zones determine whether recognition is "on/off", and gestures are defined as a sequence of targets to trace over. While this "connect-the-dot" approach supports recognition of well-established communicative gestures, it still imposes constraints on how the gestures are executed and introduces visual distraction.

Overall, relying solely on constant gesture sensing to infer the presenter's intent is currently not a feasible approach. Meanwhile, excessive structuring and guidance for gestural input shifts the problem, potentially creating new challenges for presenters. We recommend that future systems that aim to support presentation delivery allow for both clutching mechanisms as well as backup techniques for moving forward and back through the slide deck. Clutching mechanisms would allow the presenter explicit control over whether a gesture should be recognized as a control gesture. Presumably, these could also be shown or hidden from the audience when they were performed. And, simple mechanisms to move forward and backward in the presentation slides would obviate the necessity for gross movements each time a slide was to be changed.

7.3 Authenticity, Reality, and Exaggeration

We used a 1:1 mapping between the presenter's movements and the movement of their avatar in VR, which does not technically differ much from other forms of XR where the presenter's real image would be rendered in lieu of the avatar. As such, we posit that our findings insofar would generalize to other forms of extended reality, *from the perspective of the presenters*.

One question we did not explore in this work remains a central issue for VR presentations—the issue of authenticity. Is this presenter live? Are they responding to me and my presence as an audience member? If all aspects of a VR presentation is data (the atrium, the audio, the slides, the avatar, the gestures, and so on), what makes it different from a pre-recorded presentation?

Our participants explained that they had a hard time imagining wanting to give presentations that would require them to move physically. At the same time, they delighted in the idea that their avatars could still use communicative gestures that would enhance the presentation for audience members. A handful suggested the idea of exaggerated gestures, where a slight movement with the arm might create larger "in-world" effects—for example, the avatar taking several steps toward a destination. Alternately, one could imagine a system that responds to button presses that play canned gestures and animations (i.e. saving the presenter from having to enact the gestures in an embodied way). Another logical extreme would simply be to author presentations where the avatar's movements and actions are also authored beforehand.

On the other hand, presenters may not want all actions to be captured and conveyed to an audience. For instance, perhaps sneezing or other involuntary body movements that a presenter finds undesirable could be "filtered out" by the system. In contrast, it might also be true that the lack of filtration, is the aspect of VR which makes the presenter authentic. Like all the other aspects that we discuss that presenters dislike, some of the undesirable acts from the perspective of presenters could be desirable to the audience. Hence, the importance of finding a fine balance between two different perspectives should be embraced in future works.

Our work does not provide us with deep insight into what would be desirable for audience members, but it does suggest that for presenters, some forms of augmentation may be desirable—certainly when the presenter's "authorial intent" is captured in the moment. To address this question will require additional future study.

7.4 Limitations and Future Work Opportunities

Participants in our user study were from a relatively homogeneous population: all had a background in Computer Science, most were graduate students around the same age. Future studies should strive for a broader audience. Similarly, the materials in the mock presentation were deliberately neutral (i.e. no participant had any deep knowledge of the content). It would be valuable to study how participants might use such a system in domain-specific contexts. Finally, the study focused primarily on the presenters' experiences. We did not explore whether the presentations that these presenters delivered were acceptable or "better" than non-instrumented versions of the presentations. This is an opportunity for future work.

8 CONCLUSION

We explore how dual-purpose gestures—gestures recognized by the system as commands to progress through presentation materials, and at the same time used by the presenter to communicate and reinforce ideas effectively to an audience—can be realized in a VR-based presentation system. Using our technology probe JollyGesture, participants were able to use dual-purpose gestures, and showed that the guidance system was helpful, albeit distracting. We also found that participants would, on their own, want to use these dual-purpose gestures, although these would *supplement* other gestures (communicative-only and control-only) rather than replace them entirely. We learned that participants found value in the idea, but many challenges remain to make systems usable for presenters. This detailed account identifies some of the most pressing challenges ahead for designers of VR presentation systems, and provides potential avenues for further exploration.

ACKNOWLEDGMENTS

This research was generously supported by Meta Reality Labs, the National Science and Engineering Research Council (RGPIN-2017-04883, and RGPIN-2018-05072), and the Faculty of Information (University of Toronto). We would like to thank Dr. Rubaiat Habib Kazi and Dr. Matthew Brehmer for allowing us to include snippets of their video in our supplementary material. We would also like to thank the members of the Dynamic Graphics Project for their valuable advice and assistance, as well as the participants of the user study for their contributions.

REFERENCES

[1] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindbauer, Alexandra Ion, Shengdong Zhao, and JTKV Koh. 2012. Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI. *Microsoft Research Technical Report* (2012), 10.

[2] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, St. Andrews Scotland, United Kingdom, 311–320. https://doi.org/10.1145/2501988.2502045

[3] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2016. Visual menu techniques. *ACM Computing Surveys (CSUR)* 49, 4 (2016), 1–41.

[4] Olivier Bau and Wendy E. Mackay. 2008. OctoPocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08*. ACM Press, Monterey, CA, USA, 37. https://doi.org/10.1145/1449715.1449724

[5] Thomas Baudel and Michel Beaudouin-Lafon. 1993. Charade: remote control of objects using free-hand gestures. *Commun. ACM* 36, 7 (July 1993), 28–35. https://doi.org/10.1145/159544.159562

[6] BBC. 2010. Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats - BBC Four. https://www.youtube.com/watch?v=jbkSRLYSojo

[7] Philip Cash and Anja Maier. 2016. Prototyping with your hands: the many roles of gesture in the communication of design concepts. *Journal of Engineering Design* 27, 1-3 (March 2016), 118–145. https://doi.org/10.1080/09544828.2015.1126702

[8] Ben J. Congdon, Gun Woo (Warren) Park, Jingyi Zhang, and Anthony Steed. 2023. Comparing Mixed Reality Agent Representations: Studies in the Lab and in the Wild. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology* (, Christchurch, New Zealand,) *(VRST '23)*. Association for Computing Machinery, New York, NY, USA, Article 26, 11 pages. https://doi.org/10.1145/3611659.3615719

[9] Josh Urban Davis, Paul Asente, and Xing-Dong Yang. 2023. Multimodal Direct Manipulation in Video Conferencing: Challenges and Opportunities. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1174–1193.

[10] Marion Dohen and Benjamin Roustan. 2017. Co-Production of Speech and Pointing Gestures in Clear and Perturbed Interactive Tasks: Multimodal Designation Strategies. In *Interspeech 2017*. ISCA, 166–170. https://doi.org/10.21437/Interspeech.2017-1329

[11] Katherine Fennedy, Jeremy Hartmann, Quentin Roy, Simon Tangi Perrault, and Daniel Vogel. 2021. OctoPocus in VR: Using a Dynamic Guide for 3D Mid-Air Gestures in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 27, 12 (Dec. 2021), 4425–4438. https://doi.org/10.1109/TVCG.2021.3101854

[12] Gaëlle Ferré. 2019. Gesture / speech alignment in weather reports. *Proceedings of the 6th Gesture and Speech in Interaction Conference* (2019). https://doi.org/10.17619/UNIPB/1-805 Publisher: UB-PAD - Paderborn University Library Version Number: 1.

[13] Celeste Groenewald, Craig Anslow, Junayed Islam, Chris Rooney, Peter Passmore, and William Wong. 2016. Understanding 3D mid-air hand gestures with interactive surfaces and displays: a systematic literature review. In *Proceedings of the 30th International BCS Human Computer Interaction Conference*. BCS Learning & Development.

[14] Brian D. Hall, Lyn Bartram, and Matthew Brehmer. 2022. Augmented Chironomia for Presenting Data to Remote Audiences. In *The 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–14. https://doi.org/10.1145/3526113.3545614

[15] Ping-Hsuan Han, Kuan-Wen Chen, Chen-Hsin Hsieh, Yu-Jie Huang, and Yi-Ping Hung. 2016. AR-Arm: Augmented Visualization for Guiding Arm Movement in the First-Person Perspective. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, Geneva Switzerland, 1–4. https://doi.org/10.1145/2875194.2875237

[16] Benedikt Hensen, Lukas Liß, and Ralf Klamma. 2021. ImPres: An Immersive 3D Presentation Framework for Mixed Reality Enhanced Learning. In *Advances in Web-Based Learning – ICWL 2021*, Wanlei Zhou and Yi Mu (Eds.). Vol. 13103. Springer International Publishing, Cham, 28–39. https://doi.org/10.1007/978-3-030-90785-3_3 Series Title: Lecture Notes in Computer Science.

[17] Nicholas Kong and Maneesh Agrawala. 2009. Perceptual interpretation of ink annotations on line charts. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. 233–236.

[18] Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical research in mid-air interaction: A systematic review. *International Journal of Human–Computer Interaction* 35, 18 (2019), 1747–1768.

[19] Gordon Kurtenbach and William Buxton. 1994. User learning and performance with marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 258–264.

[20] Vincent Lambert, Adrien Chaffangeon Caillet, Alix Goguey, Sylvain Malacria, and Laurence Nigay. 2023. Studying the Visual Representation of Microgestures. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023), 1–36.

[21] Wenjing Li, Fuxing Wang, Richard E. Mayer, and Huashan Liu. 2019. Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology* 111, 8 (Nov. 2019), 1382–1395. https://doi.org/10.1037/edu0000352

[22] Jian Liao, Adnan Karim, Shivesh Singh Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-Time Speech-Driven Augmented Presentation for AR Live Storytelling. In *The 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–12. https://doi.org/10.1145/3526113.3545702

[23] Nicolai Marquardt and Saul Greenberg. 2015. *Proxemic Interactions: From Theory to Practice*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-02208-1

[24] Bernhard Maurer, Alina Krischkowsky, and Manfred Tscheligi. 2017. Exploring Gaze and Hand Gestures for Non-Verbal In-Game Communication. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, 315–322. https://doi.org/10.1145/3130859.3131296

[25] Richard E. Mayer and C. Scott DaPra. 2012. An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied* 18, 3 (2012), 239–252. https://doi.org/10.1037/a0028616

[26] Andreea Muresan, Jess McIntosh, and Kasper Hornbæk. 2023. Using Feedforward to Reveal Interaction Possibilities in Virtual Reality. *ACM Transactions on Computer-Human Interaction* 30, 6 (2023), 1–47.

[27] Michael Nebeling, Shwetha Rajaram, Liwei Wu, Yifei Cheng, and Jaylin Herskovitz. 2021. XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–12. https://doi.org/10.1145/3411764.3445323

[28] Paul, Annie Murphy. 2021. *The extended mind: The power of thinking outside the brain*. Eamon Dolan Books.

[29] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 955–960.

[30] Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, and Wilmot Li. 2019. Interactive Body-Driven Graphics for Augmented Video Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland UK, 1–12. https://doi.org/10.1145/3290605.3300852

[31] Rajeev Sharma, Jiongyu Cai, Srivat Chakravarthy, Indrajit Poddar, and Yogesh Sethi. 2000. Exploiting speech/gesture co-occurrence for improving continuous gesture recognition in weather narration. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. IEEE Comput. Soc, Grenoble, France, 422–427. https://doi.org/10.1109/AFGR.2000.840669

[32] Minjeong Shin, Joohee Kim, Yunha Han, Lexing Xie, Mitchell Whitelaw, Bum Chul Kwon, Sungahn Ko, and Niklas Elmqvist. 2023. Roslingifier: Semi-Automated Storytelling for Animated Scatterplots. *IEEE Transactions on Visualization and Computer Graphics* 29, 6 (June 2023), 2980–2995. https://doi.org/10.1109/TVCG.2022.3146329

[33] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 179–188. https://doi.org/10.1145/2207676.2207702

[34] Mark Stefik, Daniel G Bobrow, Gregg Foster, Stan Lanning, and Deborah Tatar. 1987. WYSIWIS revised: early experiences with multiuser interfaces. *ACM Transactions on Office Information Systems* 5, 2 (April 1987).

[35] Studdert-Kennedy, Michael. 1994. Hand and Mind: What Gestures Reveal About Thought. In *Language and Speech (2, Vol. 37)*. 203–209.

[36] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@Home: Exploring Visual Guidance and Feedback Techniques for Physiotherapy Exercises. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 4123–4132. https://doi.org/10.1145/2702123.2702401

[37] Yi Tian and Marie-Luce Bourguet. 2016. Lecturers' Hand Gestures as Clues to Detect Pedagogical Significance in Video Lectures. In *Proceedings of the European Conference on Cognitive Ergonomics*. ACM, Nottingham United Kingdom, 1–3. https://doi.org/10.1145/2970930.2970933

[38] Gustav Verhulsdonck and Morie Jacquelyn Ford. 2009. Virtual chironomia: Developing standards for non-verbal communication in virtual worlds. *Journal For Virtual Worlds Research* 2, 3 (Oct. 2009), 10.

[39] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. 137–146.

[40] Petra Wagner, Zofia Malisz, and Stefan Kopp. 2014. Gesture and speech in interaction: An overview. *Speech Communication* 57 (Feb. 2014), 209–232. https://doi.org/10.1016/j.specom.2013.09.008

[41] Bodo Winter, Marcus Perlman, and Teenie Matlock. 2013. Using space to talk and gesture about numbers: Evidence from the TV News Archive. *Gesture* 13, 3 (Dec. 2013), 377–408. https://doi.org/10.1075/gest.13.3.06win

[42] Haijun Xia, Michael Glueck, Michelle Annett, Michael Wang, and Daniel Wigdor. 2022. Iteratively Designing Gesture Vocabularies: A Survey and Analysis of Best Practices in the HCI Literature. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–54.

A PRESENTATION SLIDES

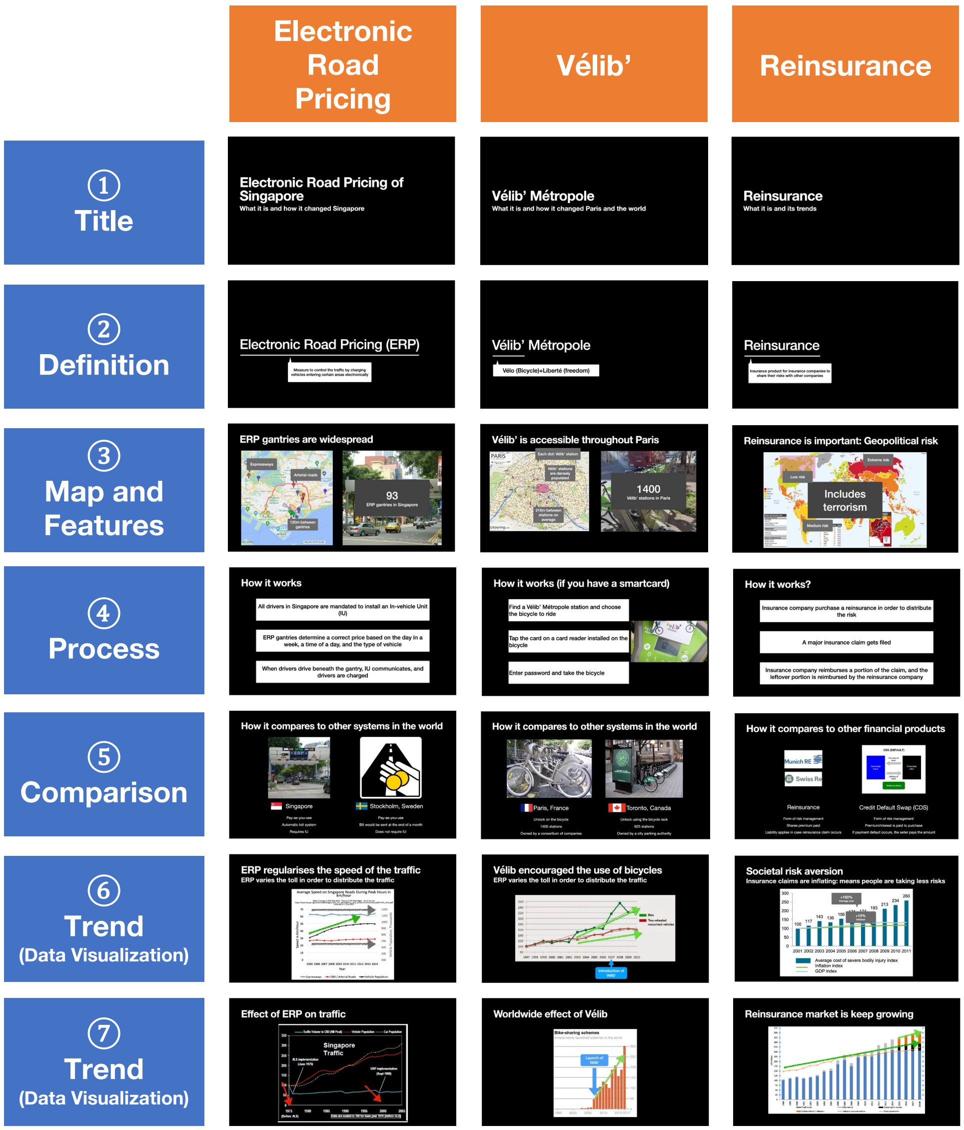

Figure 8: Slide decks created as part of the JollyGesture probe (and used in the study).