

VisConductor: Affect-Varying Widgets for Animated Data Storytelling in Gesture-Aware Augmented Video Presentation

TEMILOLUWA FEMI-GEGE, University of Waterloo, Canada

MATTHEW BREHMER, University of Waterloo, Canada

JIAN ZHAO, University of Waterloo, Canada

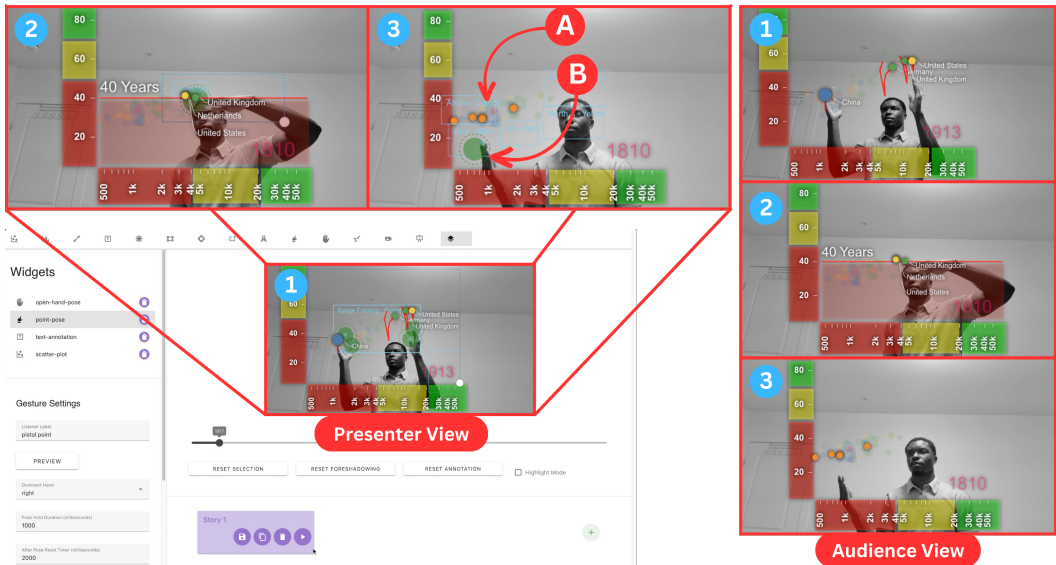


Fig. 1. A presenter uses VisConductor to deliver an augmented video presentation about life expectancy and income around the world. VisConductor provides a local presenter view (left) as well as a remote audience view (right); the former is equipped with additional widgets to ensure precise gestural control of the overlaid visual aid (in this example, an animated scatterplot). (A): A *Gesture Widget* outlines the boundary within which the presenter’s hand gesture should be placed to trigger an animation. (B): Upon detecting the presenter’s gesture, feedback in the form of a progressively expanding green circle appears, indicating the duration for which the gesture should be maintained for accurate detection.

Augmented video presentation tools provide a natural way for presenters to interact with their content, resulting in engaging experiences for remote audiences, such as when a presenter uses hand gestures to manipulate and direct attention to visual aids overlaid on their webcam feed. However, authoring and customizing these presentations can be challenging, particularly when presenting dynamic data visualization (i.e., animated charts). To this end, we introduce VisConductor, an authoring and presentation tool that equips presenters with the ability to configure gestures that control affect-varying visualization animation, foreshadow visualization transitions, direct attention to notable data points, and animate the disclosure of annotations. These gestures are integrated into configurable widgets, allowing presenters to trigger content transformations

Authors’ addresses: Temiloluwa Femi-Gege, tpfemige@uwaterloo.ca, University of Waterloo, Waterloo, Ontario, Canada; Matthew Brehmer, mbrehmer@uwaterloo.ca, University of Waterloo, Waterloo, Ontario, Canada; Jian Zhao, jianzhao@uwaterloo.ca, University of Waterloo, Waterloo, Ontario, Canada.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the ACM on Human-Computer Interaction*, <https://doi.org/10.1145/3698131>.

by executing gestures within widget boundaries, with feedback visible only to them. Altogether, our palette of widgets provides a level of flexibility appropriate for improvisational presentations and ad-hoc content transformations, such as when responding to audience engagement. To evaluate VisConductor, we conducted two studies focusing on presenters ($N = 11$) and audience members ($N = 11$). Our findings indicate that our approach taken with VisConductor can facilitate interactive and engaging remote presentations with dynamic visual aids. Reflecting on our findings, we also offer insights to inform the future of augmented video presentation tools.

CCS Concepts: • **Human-centered computing** → **Visualization systems and tools**; *Gestural input*; *Mixed / augmented reality*.

Additional Key Words and Phrases: Visualization, Augmented Video Presentation, Gestural Input, Animation

ACM Reference Format:

Temiloluwa Femi-Gege, Matthew Brehmer, and Jian Zhao. 2024. VisConductor: Affect-Varying Widgets for Animated Data Storytelling in Gesture-Aware Augmented Video Presentation. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 531 (December 2024), 22 pages. <https://doi.org/10.1145/3698131>

1 INTRODUCTION

The recent surge in remote work and education has triggered an evolution of presentation tools that enhance synchronous communication with audiences at a distance [7, 62], tools that convey information and accompanying visual aids in an engaging manner. Virtual camera applications (e.g., mmhmm [2], OBS [40]), teleconference applications (e.g., WebEx [16]), and operating systems (e.g., macOS's Presenter Overlay [6]) are now enabling speakers to easily composite their webcam video with their visual aids. This development means that remote audiences no longer need to divide their attention between the video of a presenter and their screen-shared visual aids. While screen-sharing visual aids and controlling them via mouse-based interaction will continue to remain common in informal teleconference meetings between colleagues, more formal types of presentations such as business pitches, class lectures, and conference talks are increasingly delivered online. It is in these contexts where communication is more effective when the audience can see the presenter and their non-verbal body language [30, 58, 63]. In this paper, we build upon recent work in augmented video presentation that goes beyond visual compositing, in which a presenter's hand gestures are used to manipulate and direct attention to overlaid visual aids [25, 37, 46]. The promise of gesture-aware augmented video presentation rests in the observation that gestures can be simultaneously operational and expressive: providing functional control of an interface while remaining an essential channel for non-verbal communication [58, 63]. Ultimately, gestures enrich one's communicative repertoire, allowing for a more holistic and nuanced exchange of ideas and affects, and whether used intentionally or spontaneously, they add depth and dimension to remote communication.

Presentations about data are common in organizational and educational contexts, with data visualization being a common form of visual aid for information delivery [19, 44]. However, authoring and delivering gesture-aware augmented video presentations about data during a live teleconference call is challenging [25], particularly if the content is dynamic (i.e., animated charts). It is essential to recognize the power of animation when storytelling with data [15], particularly as animating charts in a presentation can be critical for retaining audience attention and eliciting a range of affective responses [34]. However, animating a chart could involve one or more distinct types of transformation: scrubbing to control temporal playback, disclosing annotations, selecting or highlighting elements, or borrowing cinematic techniques such as foreshadowing to build anticipation [36]. In conventional slide-based presentation tools, presenters use peripheral input devices (e.g., a clicker, a mouse) to trigger predefined animated transitions or motion effects. Relying on a planned series of animation triggers limits the dynamism and spontaneity of a presentation, and

in the context of data visualization, this approach precludes satisfying ad-hoc interactions with a chart, such as in response to audience questions about aspects of the data. Returning to the prospect of controlling animated charts with gestures, we face the question of how these gestures would be specified, as well as how they would affect the audience experience.

In light of these challenges, we introduce *VisConductor*, a tool that allows presenters to author and perform augmented live video presentations featuring dynamic visualization controlled via hand gestures (Figure 1). *VisConductor* allows for the specification of various animated transformations to charts via a modular widget-based design, with widgets that simplify the authoring process and provide visual cues to presenters during presentation delivery. We deliberately restricted the gestural vocabulary associated with these widgets, so as to strike a balance between operational control and communicative expressiveness.

To evaluate *VisConductor*, we conducted two studies focusing on presenters ($N = 11$) and audience members ($N = 11$), respectively. In the presenter study, we asked participants to prepare a presentation using *VisConductor* and were given a dataset and story along with a sequence of speaking points, however, we allowed them the freedom to specify gestural widgets that would best serve them in delivering points of the story. In the audience study, we asked participants to watch a remote presentation given by one of the researchers. Our results suggest that *VisConductor* provides an effective authoring and presenting experience for presenters as well as an engaging experience for audience members. To summarize, our contributions include:

- The design and development of *VisConductor*, a tool for authoring and delivering live augmented video presentations featuring dynamic data visualization, which entails controlling animation playback, revealing annotations, and foreshadowing visual changes, all activated by gestures that are simultaneously operational and communicative.
- Empirical findings from two user studies evaluating *VisConductor*, considering the perspectives of presenters and audiences, respectively, which could inform the future design of augmented video presentation tools that allow presenters to speak about data to remote audiences during live teleconference calls.

2 BACKGROUND & RELATED WORK

Our research builds on an understanding of gestural communication as well as the use of gestural interaction and data visualization in presentation contexts.

2.1 The Communicative Use of Gestures

Gestures are a multifaceted tool in communication, often conveying subtleties and nuances that spoken words cannot capture [22]. They act as visual complements, sharpening the message's clarity and making it more accessible and memorable for the audience. This visual support not only heightens audience engagement but also fosters better retention of the information shared. For instance, Zeng et al. [63] documented the importance of gestures across diverse communication settings, from public speaking to daily interpersonal interaction. Their findings, along with those of Kang et al. [30] suggest that the likelihood of effective communication is greater when the gestures used to synchronize and resonate with the spoken content. While this synchronization often emerges naturally as a spontaneous reaction to a spoken narrative, speakers may also use gestures as a conscious strategy to emphasize and persuade. Finally, being particularly relevant to our focus on speaking about data, Tieu et al. [58] documented the amplifying effect of gestures when a speaker is presenting complex or abstract ideas.

According to taxonomies of communicative gestures [14, 31], *deictic* gestures such as pointing direct attention, while *iconic* gestures visually represent specific ideas or objects and *metaphoric*

gestures symbolize abstract concepts. Based on an analysis of 5,500 mid-air hand gestures, similarly, Aigner et al. classified gestures as follows: pointing, semaphoric, pantomimic, iconic, and manipulation [1]. In the context of presentations, Carstens et al. [14] noted that a presenter's one-handed gestures are often deictic (e.g. pointing), while two-handed gestures tend to be iconic or metaphoric. In our work, we primarily respond to a presenter's deictic gestures, though we also respond to a metaphorical gesture for the passage of time-based on the motion of an analog clock. We additionally build upon the realization that gestures can be simultaneously deictic and manipulative [25].

2.2 Gesture-Controlled Presentation Interfaces

Presenters frequently employ communicative gestures when speaking, however in conventional presentation settings in which a presenter and their audience are co-located, these gestures are often performed at a distance from the presenter's visual aids, particularly when using a projected display, resulting in potentially ambiguous deictic references [21]. This separation is greater in most online presentations, in which a video of the presenter and a screen-shared video feed of their visual aids are shown in separate windows or panels. Additionally, it is typical for presenters to position themselves such that their camera only captures them from the shoulders upwards, so any communicative gestures they make during their presentation go largely unseen by their audience. Hauber et al. [26] highlighted the importance of spatiality and social presence in videoconferencing by comparing 2D and 3D interfaces. Moreover, research on remote collaboration software indicates that superimposing a live video of users onto their shared content can significantly enhance collaboration [24, 33, 64]. Part of this effectiveness can be attributed to the preservation of eye contact in addition to the visibility of gestures, which closely mimics face-to-face interaction [60]. Overall, this research helps to explain the growing popularity of augmented videoconferencing applications in which visual aids are composited with a presenter's video.

In this work, we additionally distinguish between communicative and *operational* gestures in the context of delivering a presentation, with the latter representing those deliberately designed for controlling a presentation interface. Cao et al. [12] found that hand gestures were favored relative to a laser pointer for controlling a presentation interface. However, Fourney et al. [21] found that when granted hand gesture control of presentation materials, the range of interaction was often limited to slide navigation; people seldom used gestures to interact with slide content, and gestural interaction events were prone to unintentional triggering. As reflected upon by Hall et al. [25], this unintentional triggering could be a sign of interference between operational and communicative gestures. Presenters also struggle when a system fails to provide obvious feedback in response to the gestures being performed [8], and this is particularly pronounced when any operational gestures are separated in physical space from the visual aids they are intended to control.

Developments in real-time computer vision technologies such as Kinect [52], OpenPose [13], and MediaPipe [59] have recently yielded a collection of gesture-aware augmented video presentation tools that grant presenters functional control over visual aids composited over their video feed [25, 37, 38, 46]. Collectively, these tools demonstrate the potential of using gestural control that is spatially aligned with visual aids, thereby also serving a communicative purpose in that the audience's attention is no longer divided between a speaker and their content. However, even with these augmented video presentation tools, presenters still struggle to remember which gestures to perform at different points in a presentation, or to anticipate the effects that gestures will have on their visual aids [46]. VisConductor aims to address these challenges with gesture-controlled presentation tools with an authoring interface comprised of modular *Gesture Widgets*, along with visual feedback provided to the presenter as they execute these gestures at performance time.

2.3 Data Visualization in Presentation Contexts

Visualization is instrumental in translating data observations into impactful narratives [28, 32, 47, 50], with considerations that go beyond mere choices of representation (i.e., visual encoding).

The first major consideration relevant to our work is the strategic use of animation for enhancing audience engagement and comprehension [45]. While animated visualization can effectively demonstrate changes over time or evolving patterns, it can also contribute to the affective aspect of a narrative about data [3, 34]. This emotional connection is particularly vital in presentations, as it can promote audience engagement and information retention [20]. To this end, Lan et al. [34] identified the role of animation speed and easing functions on the affective impact of dynamic visualization. Similarly, Li et al. [36] demonstrated ways to employ visual foreshadowing in anticipation of animating a chart as a way to direct audience attention and keep them engaged.

While animation control can be realized using many data visualization libraries and tools, specialized packages such as *ganimate* for R [56] or interactive applications such as *Data Animator* [57], *DataClips* [4], and *Roslingifier* [51] surface convenient animation specification affordances. While we draw inspiration from the options that these specialized tools provide, their resulting specifications define automatic animation or playback triggered by conventional mouse and keyboard input, whereas we connect animation playback to real-time gestural input.

Another consideration is the where and how to employ highlighting and annotation, so as to direct the audience to the most crucial aspects of the data [54]. Annotations can assume many forms and be applied to various visual entities [42], from axes and coordinate spaces to data marks; they can directly reflect values in the form of mark stroke or fill highlights, mark labels and trend lines [55], or they might serve to provide additional context not captured in the underlying data.

2.4 Visualization in Augmented Video Presentations

In 2010, the late public health expert Hans Rosling produced a short documentary with the BBC entitled “*200 Countries, 200 Years, 4 Minutes*” [9]. In this video, Rosling stands in an empty room, appearing in the middle of the frame. Gradually, an animated scatterplot depicting the dynamics of public health metrics over time is composited in the foreground, seemingly controlled by his gesticulations. The video compositing and animation were both added in post-production, though the filming did require very deliberate choreography on Rosling’s part, so as to appear as though he was interacting with a gesture-aware interface. In doing so, Rosling motivated a line of work combining visualization in real-time gesture-aware augmented reality video, including this paper.

Hall et al. [25] recently demonstrated the detection of gestures to highlight values in composited bar, line, area, and pie charts. However, that work stopped short of using gestures for chart animation, such as showing how values change over time. To this end, *VisConductor* provides the means to specify the gestural control of dynamic representations of data, namely animated scatterplots and bar charts. These types of dynamic visualization differ from the alternating state-based animation of diagrams appearing in a demonstration of Saquib et al.’s gesture-controlled video performance system [46]; in contrast, our animations reflect values bound to a continuous time-series.

Beyond gestural control for animating time-oriented data visualization, *VisConductor* also provides a graphical authoring interface for specifying these gestures, incorporating familiar affordances from tools for web-based data storytelling (e.g., [4, 47, 51]), including those for communicating intended affects [34]. However, unlike these prior tools, *VisConductor*’s output medium is augmented video, with animation triggered by gestural input rather than by clicks or timers.

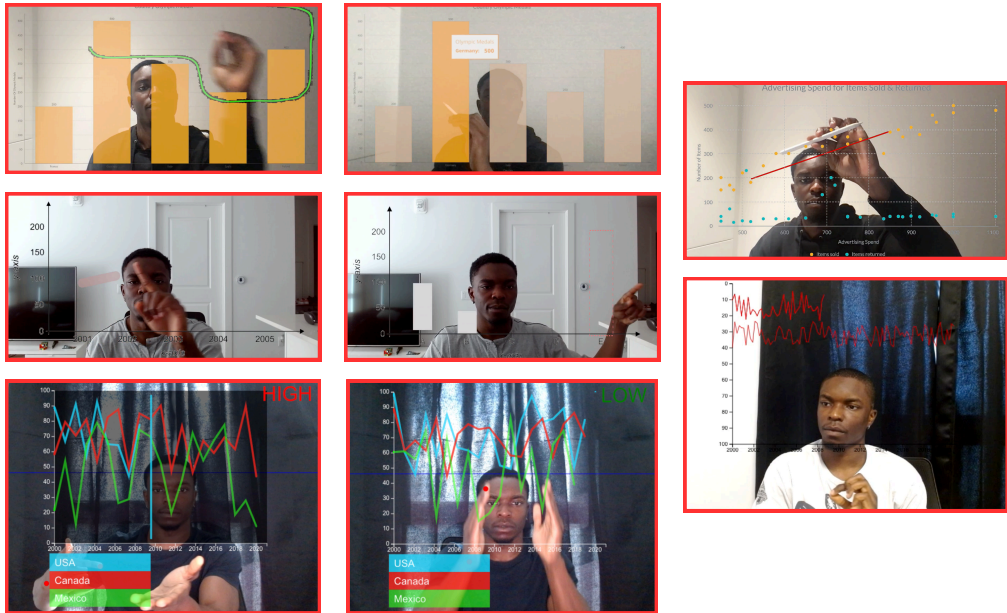


Fig. 2. Frames from early low-fidelity video prototypes of animated charts composited over a presenter gesticulating.

3 DESIGN CONSIDERATIONS

Our review of prior work yielded three design considerations. To validate and refine these considerations, we conducted two semi-structured interviews with a professional public speaking coach who has personally trained thousands of individuals over the course of their career [5]. Additionally, we solicited informal feedback via unstructured interviews with two university professors and two graduate students, with each having experience giving technical presentations to remote audiences via videoconferencing applications. In each case, we sought feedback on our design exploration with respect to the use of gestures to control dynamic visualization in augmented video presentations. Specifically, we solicited responses to low-fidelity video prototype demonstrations shown in Figure 2, inspired by the prior work summarized in Section 2.3. We produced chart animations in Adobe After Effects [29] and used OBS Studio [40] to composite them over a video of a presenter gesticulating, so as to simulate the interactive gestural control of chart animations.

D1: Recognize a small set of expressive and operational gestures. In response to early prototype demonstrations, the speaking coach reaffirmed the need to maintain a primary focal point of visual attention for the audience and to minimize the cognitive demand of remembering a large gestural vocabulary, echoing the findings of prior work [17, 63]. In other words, we require a set of gestures that are both easy to remember and easy to coordinate with speech. Moreover, the speaking coach urged us to consider gestures that would not distract the audience; rather they should reinforce deictic references and points of emphasis in the presenter’s oration, striking a balance between being expressive and operational [25]. These low-fidelity demonstrations therefore allowed us to identify and discard candidate gestures that were purely operational, as these were deemed most likely to distract audiences.

D2: Provide gestural control of visualization animation with intended affect. Prior work has underscored the role that animation plays in engaging presentations about data. In this context, animation is not limited to graphical transformations [27] or showing how patterns in data change over time [45]; it also encompasses the strategic reveal of annotation and highlighting [42], the use of affective motion to elicit appropriate emotional reactions to a narrative, and the use of visual foreshadowing to set audience expectations [36]. The speaking coach particularly supported our inclusion of foreshadowing in this consideration, as this tactic prepares the audience to receive new information. Given this multi-faceted characterization of animation, we sought distinct gestural interactions for controlling these aspects.

D3: Support authoring and delivery with dedicated views. Conversations with the speaking coach and the two professors also allowed us to refine the context of use for an augmented video presentation tool, namely professional environments in which there are numerous time pressures that demand that presentation content be prepared quickly. This context of use underscores the need for a predictable interface that allows for the skills developed using other tools to be more easily transferred. This sentiment was also captured in earlier work by Brehmer and Kosara [11], who found that not only do those who give presentations about data desire a graphical user interface that provides interactive affordances for authoring a presentation, modifying its structure, and previewing content, they also desire a secondary display visible only to them at the time of presentation delivery, following the conventions of popular commercial presentation tools (e.g., PowerPoint, Keynote). In the context of gesture-controlled augmented video presentations about data, we anticipate that such a secondary display could be used to provide visual feedback with respect to gesture recognition and animation state.

3.1 Usage Scenario

Given our three design considerations, we now describe a scenario in which a presenter uses gestures to control an augmented video presentation featuring dynamic data visualization; for a video of this scenario, please refer to our **supplemental material**. The presenter in our scenario is John, a public health official, and his presentation is about the shifts in monetary and healthcare metrics in countries around the world over the course of the last two centuries. We describe the authoring and delivery interface (D3) used by John in the next section; here, we describe the viewing experience of his remote audience.

John initially sets the scene by compositing a scatterplot over his webcam video. He proceeds to reveal a title with a deliberately-placed open hand gesture, as shown in Figure 5-A. This title text is revealed slowly, intended to mirror the profound and lingering impacts of war on public health, thereby establishing an emotional tone. He similarly directs attention to the elements of the scatterplot via a series of open hand pointing gestures, proceeding from the axes to the data points, highlighting each of them in turn, albeit with a neutral affect.

Next, he directs the audience's attention toward a few nations of interest. With pointing gestures aimed at specific data points, he triggers an animation in which these data points brighten while others dim (Figure 5-B). To herald changes in the data over time, John then uses a bimanual framing gesture to encircle a group of data points, highlighting them and revealing their projected trajectories (Figure 5-C), foreshadowing future public health trends in these countries.

Lastly, John animates the scatterplot to illustrate changes in the data over time while simultaneously eliciting an appropriate affective response from the audience. This animation is controlled by John's performance of a circular dialing gesture, evoking the passing of time as indicated by the movement of an analog clock. During this animation, some of the circular marks corresponding with countries smoothly and ephemerally morph into downward-facing arrows, reinforcing the

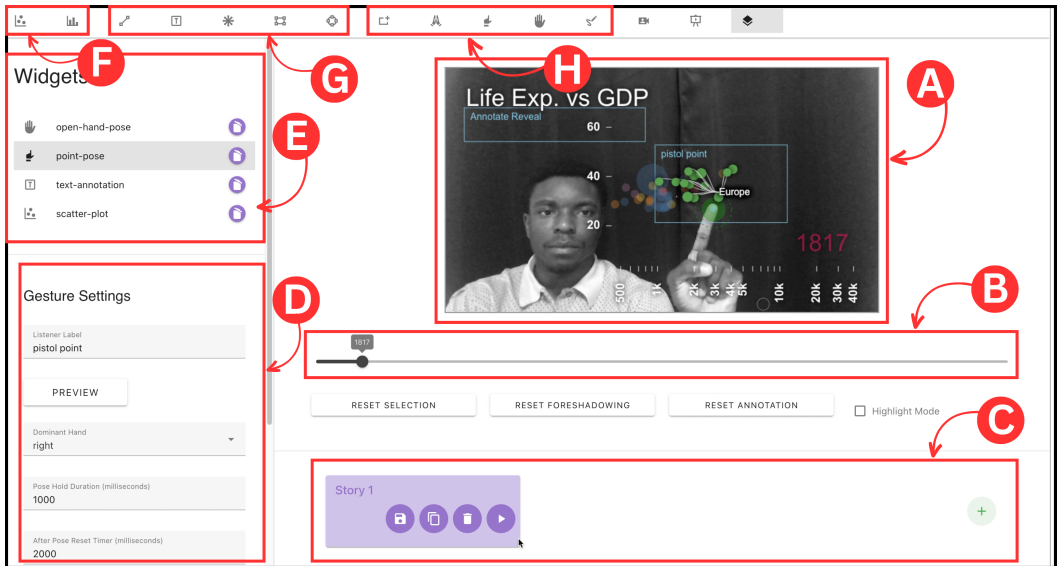


Fig. 3. The presenter interface of VisConductor consists of: (A) the *Presentation Preview*, (B) the *Timeline Slider*, (C) the *Storyline Tab*, (D) the *Widget Settings Tab*, (E) the *Widget List*, and a palette of (F) *Chart Widgets*, (G) *Annotation Widgets*, and (H) *Gesture Widgets*.

devastating impact of war (Figure 5-D). Once he reaches the most recent state of the data, John stops dialing to comment on the contemporary state of public health around the world.

4 VISCONDUCTOR SYSTEM

VisConductor is a browser-based application for authoring and delivering augmented video presentations about data. The name is an allusion to how an orchestra conductor’s hand gestures direct the performance of various instruments; in our case, hand gestures control various animation effects, including but not limited to visualization playback reflecting changes in data over time.

4.1 Presenter Interface

The presenter interface of VisConductor, shown in Figure 3, consists of several components. First, the *Presentation Preview* (A) provides a preview of what the remote audience sees—the presenter’s webcam feed composited with charts and annotations—additionally augmented with feedback visible only to the presenter (D3), indicating gestural activation cues and the progress of an animation. Next, the *Timeline Slider* (B) indicates the progress of an animation and allows for conventional cursor-based scrubbing to specific points in an animation. The *Storyline Tab* (C) lists distinct segments of the presentation, which are analogous to the convention of slides in a PowerPoint or Keynote presentation. The *Widget Settings Tab* (D) provides controls for specifying the parameters of widgets, while the *Widget List* (E) lists the widgets currently in use for the selected segmented, which may include one or more *Chart Widgets* (F), *Annotation Widgets* (G), and *Gesture Widgets* (H). The ordering of widgets within the *Widget List* reflects an ordering of layers, adopting a convention familiar to users of tools like Photoshop or Figma; the implication of this layering is that presenters can overlay multiple widgets within a single presentation segment.

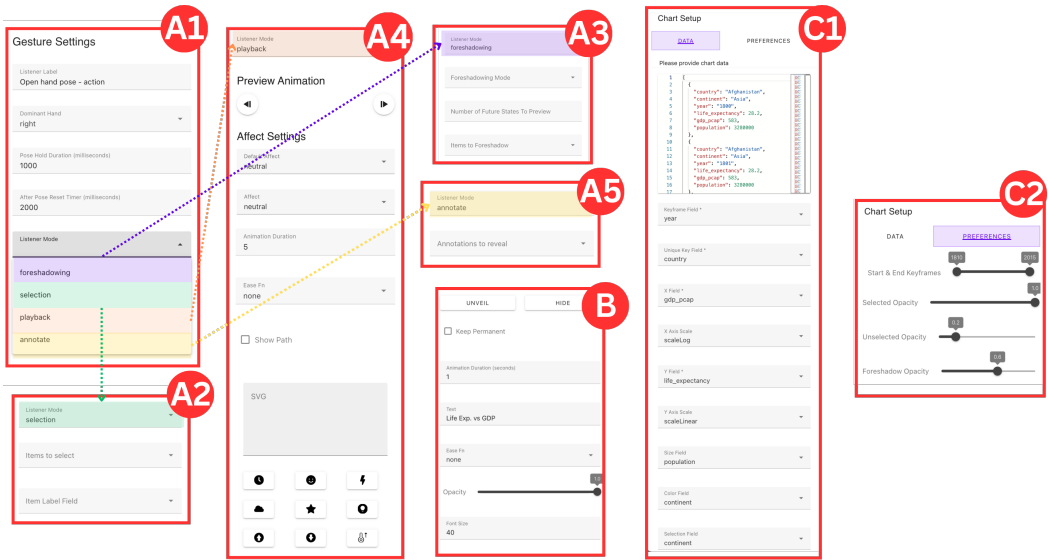


Fig. 4. An overview of widget parametrization in VisConductor: (A1) *Gesture Widget* settings for gesture type, recognition duration, which hand to recognize for one-handed gestures, and which operation the gesture will trigger: (A2) *Selection*, (A3) *Foreshadowing*, (A4) *Playback*, or (A5) *Annotation*; (B) *Annotation Widget* settings include those for specifying annotation type, color, opacity, reveal duration, and reveal easing; (C1) *Chart Widget* settings include those for ingesting tabular data and binding fields to position, size, and color scales used in the chart, while (C2) *Chart Widget* animation preferences allow for the selection of keyframes, mark opacities, and foreshadowing design.

4.2 Chart, Gesture, & Annotation Widgets

Every augmentation of a presenter’s webcam video in VisConductor is encapsulated as a modular widget, of which there are three types.

First, a *Chart Widget* (Figure 3-F) overlays a chart over the video feed. Currently, VisConductor supports two types of charts that are frequently used to tell stories about data over time (D2): an animated scatterplot and a bar chart race. In both cases, the position and size of data-bound graphical marks are animated to reflect values at a particular point in time. Figure 4-C1 illustrates the parameters available for configuring a *Chart Widget*, following familiar conventions for chart authoring exemplified in existing visualization tools (i.e., loading a tabular dataset and binding position, size, and color scales to field names).

Second, *Gesture Widgets* (Figure 3-H) add gestural activation to a specified region within the video frame; these could, for instance, trigger the selection and annotation of chart elements, reveal visual foreshadowing, or initiate the playback of animations corresponding to changes in data over time (D1). The parameters of a *Gesture Widget* (Figure 4-A1:A5) include those for binding a gesture to an existing *Chart Widget* or *Annotation Widget*, as well as those for specifying recognition duration, re-detection interval, and which hand the gesture should be recognized for.

Finally, *Annotation Widgets* (Figure 3-G) overlay text or shapes to the video frame. Figure 4-B shows the available parameters for specifying the visual style and revealing annotations, such as altering their color or opacity, which may have deliberate affective connotations. While text labels for data points can be retrieved from the underlying data and revealed when binding a *Gesture Widget* to a *Chart Widget*, *Annotation Widgets* can be used to freely place editorial text and graphical

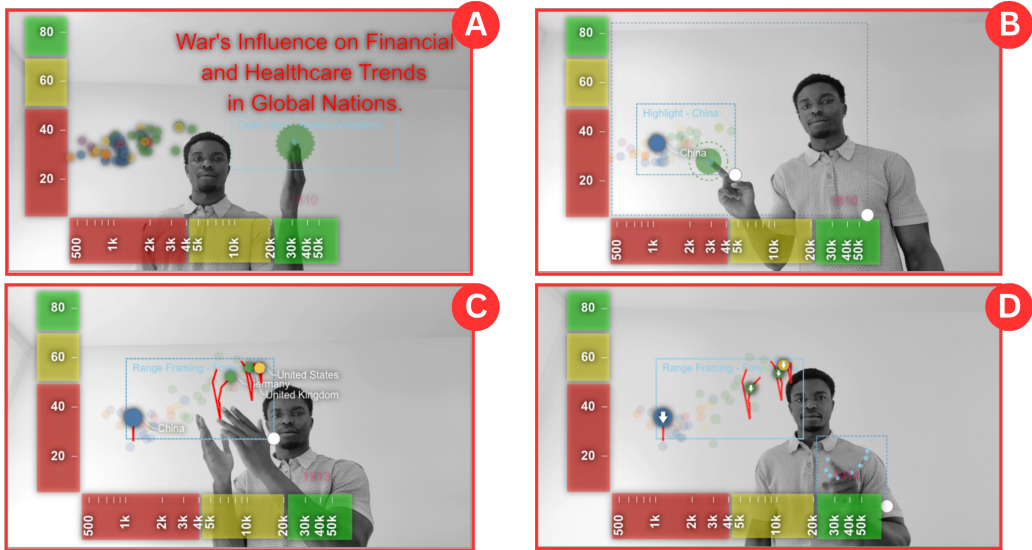


Fig. 5. Four moments in a presentation as seen from the Presenter Preview, exemplify visual feedback for four types of gesture: (A) an *Open Hand* gesture reveals a text annotation; (B) a *Pointing* gesture highlights specific data marks within the scatterplot; (C) a *Rectangular Framing* gesture foreshadows the trajectories of data marks appearing within that enclosed region; and (D) a continuous *Dialling* gesture controls the playback of an animation corresponding to changes in the data over time.

elements that do not have an explicit link to data, such as when directing attention to part of an axis or to a region within a chart where there is a conspicuous absence of data marks.

4.3 Gestural Vocabulary

In accordance with **D1**, VisConductor supports a small set of gestures that are simultaneously expressive and operational, differentiated according to whether they are single-handed or bimanual, stationary or dynamic, and which fingers are extended.

A stationary single-handed gesture can be used to draw attention to specific visual elements. Following Carstens' finding [14] that two gestures from this category are particularly common when making deictic references in presentations, we recognize an *Open Hand* gesture as well as a more directed *Pointing* gesture using the index finger. A widget for either of these gestures can be bound to an existing *Annotation Widget* or *Chart Widget*; Figure 5-A shows an example of using an *Open Hand* gesture to reveal a text annotation is shown in, while Figure 5-B shows an example of using a *Pointing* gesture to highlight data marks in a scatterplot.

Bimanual shaping gestures draw attention to ranges and boundaries [51]. If the palms of the two hands are facing each other with index fingers bent, we recognize this as *Rectangular Framing*, as it conveys a two-dimensional enclosure. If the fingers are not bent, we recognize this as one-dimensional *Range Framing*. When bound to a *Chart Widget*, either type of range gesture can be used to highlight data marks appearing within a two-dimensional enclosure or a one-dimensional span, respectively. Figure 5-C shows an example of a *Rectangular Framing* gesture used to trigger a visual foreshadowing of trajectories for data marks originating from the enclosed area.

Rounding out our set of gestures is a dynamic *Dialling* gesture inspired by radial scrolling on touch interfaces [39, 53]. This extension of *Pointing* controls the playback of animation corresponding to changes in data over time, using the visual metaphor of an analog clock (Figure 5-D; see video).

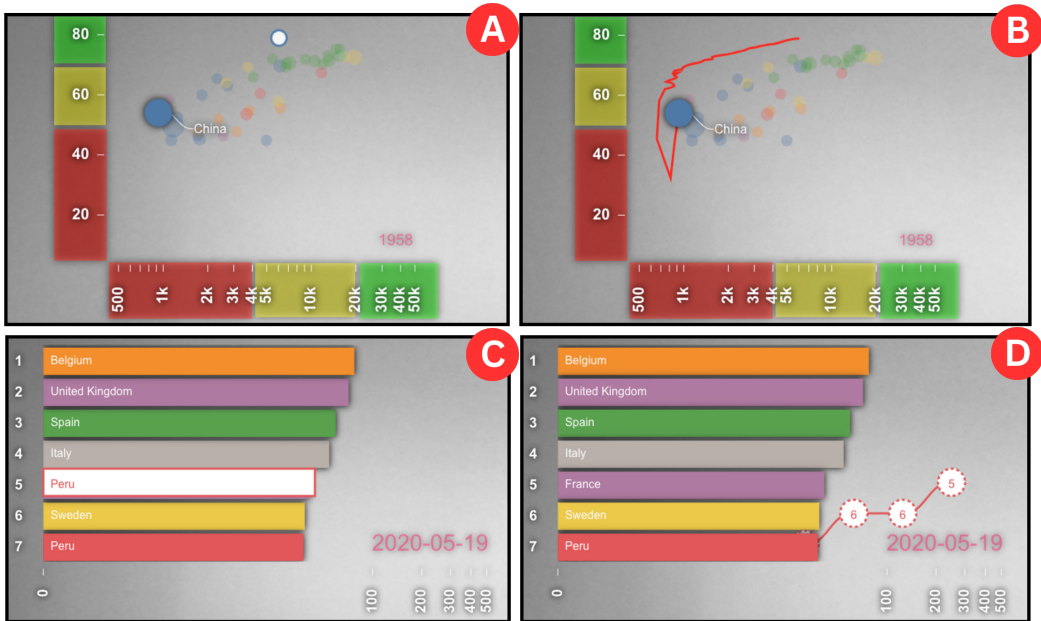


Fig. 6. Visual foreshadowing applied to the two *Chart Widgets*. (A) Scatterplot position foreshadowing, showing the initial and final position of the mark(s). (B) Scatterplot trajectory foreshadowing, showing the path of one or more data marks over time. (C) Bar chart race position foreshadowing, highlighting the future position of a bar. (D) Bar chart race trajectory foreshadowing, showing an ephemeral bump chart of the bar's future positions; e.g., Peru's rise from 6th to 5th place over three time steps.

4.4 Combining Widgets to Communicate Intended Affects

The *Playback* settings of a *Gesture Widget* (Figure 4-A4) allow presenters to establish different affects via animation (D2). In particular, easing controls provide a channel through which to convey a positive, neutral, or negative affect. Additionally, presenters can opt to ephemerally morph selected data mark shapes into semantically-suggestive SVG-based icons during playback. For example, in Figure 5-D and in the supplemental video, the presenter specifies an animation in which a scatterplot's data marks move downwards, reflecting the devastating loss of lives during World War I; next, they quickly return upwards, representing the economic recoveries that followed the war. The former movement is deliberately slow, while the latter is quick and bouncy, suggesting first a profound loss followed by a hopeful rebound. To further establish a negative affect, the circular data marks ephemerally morph into a downward arrow shape as the marks descend.

Controls for specifying *Foreshadowing* (Figure 4-A3) also provide an opportunity to elicit an appropriate affective response from the audience. Point foreshadowing juxtaposes a data mark's current state with a future one via an ephemeral pulsing animation, as illustrated in Figure 6-A and Figure 6-C. In contrast, trajectory foreshadowing ephemerally sketches out path that a data mark will take in a subsequent animation, with subtle variations in the appearance of this trajectory based on the type of *Chart Widget* it is bound to (Figure 6). In both cases, visual foreshadowing is intended to build anticipation for what is soon to be revealed via animation.

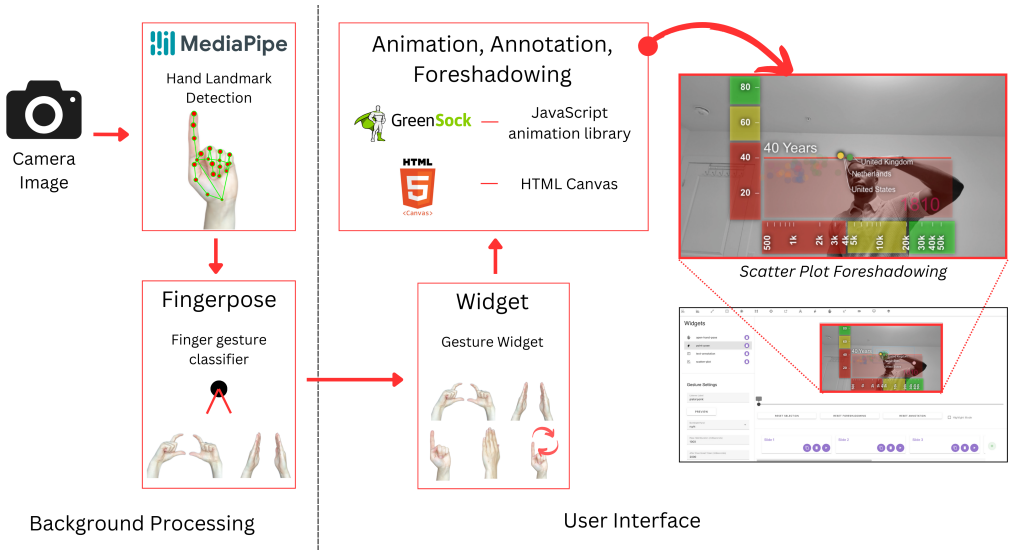


Fig. 7. The implementation of VisConductor: MediaPipe detects the presenter’s hand and finger positions, which are processed by Fingerpose to classify gestures that activate *Gesture Widgets*. These in turn trigger the pre-configured GreenSock-based animation, annotation, and visual foreshadowing in linked *Chart Widgets* and *Annotation Widgets*.

4.5 Presentation Delivery Experience

Like popular commercial slide presentation tools, VisConductor provides both an *audience view* and an *presenter view* (D3, Figure 1). The former can be shared with a remote audience either via the window sharing functionality offered by teleconferencing applications or via a virtual webcam application (e.g. OBS [40]). We enrich the latter with labelled blue bounding boxes for *Gesture Widgets*, visible only to the presenter, delineating the boundaries of gestural activation; we show additional examples of this visual feedback in Figure 5. Additionally, when performing a gesture within these bounds, we show expanding green bubbles in the presenter view, ephemeraly indicating the duration required for gesture recognition.

4.6 Implementation

VisConductor is a browser-based application built using Vue.js. As illustrated in Figure 7, the presenter’s webcam video is ingested by MediaPipe [59], which captures the presenter’s hand and finger positions. Fingerpose [49] uses these positions to detect predefined static and dynamic gestures. Depending on how the presenter has assigned *Gesture Widgets* to *Chart Widgets* and *Annotation Widgets*, these gestures will trigger selection, annotation, animation, or visual foreshadowing. Finally, we implemented our chart animations using GreenSock (GSAP) [23]. The source code of VisConductor can be found at <https://github.com/WatVis/visconductor>.

5 EVALUATION

We qualitatively evaluated VisConductor from two perspectives: our first study focused on a presenter’s experience of authoring and presenting using the tool, while our second study considered the audience’s experience of watching a presentation delivered via VisConductor. A precedent for this methodology is Hall et al.’s [25] evaluation of their own gesture-aware augmented video

environment for giving presentations about data. While a comparative evaluation between VisConductor and their environment might have been technically feasible, we opted against this given VisConductor's focus on animation and affect, the different chart types supported by both tools, as well as the latter's absence of both an interactive authoring interface and publicly-available source code. Hall et al. additionally compared the audience experience of gesture-aware augmented video against the convention of screen-sharing content when presenting data, finding that audiences were more engaged by the former; this finding obviated the need for us to perform a similar comparison. The study protocol was reviewed and approved by our institutional research ethics office.

5.1 Presenter Study

Our focus in this study was on the *utility* and *usability* of VisConductor with respect to presentation authoring and delivery.

Participants. We recruited 11 participants (3F, 8M) via mailing lists and word of mouth, ranging in age between 18 and 34. Their professional backgrounds included engineering, business, and computer science; eight held post-graduate degrees and three were college graduates. Regarding technological proficiency with presentation applications, five self-identified as *Advanced*, four as *Expert*, and two as *Intermediate*. Four reported having *Advanced* presentation skills, while seven reported having *Intermediate*-level skills. All participants reported experience giving in-person presentations, while three were familiar with giving online presentations. As for the frequency of delivering presentations, eight reported giving about one presentation a month and three reported doing so weekly. Their presentations incorporated various elements, including data visualization, text, images, diagrams, and multimedia elements such as videos.

Procedure. After signing a consent form and filling out a questionnaire regarding their backgrounds, we gave the participants a walkthrough of VisConductor, highlighting its features and encouraging questions. Next, we introduced them to a template presentation with five stories where each story in the presentation contained approximately one-minute worth of content to present, each seeded with an initial *Chart Widget*: five participants encountered the bar chart race presenting data on COVID deaths per 100,000 people across various countries between 2020 and 2022, while six experienced the animated scatterplot showing the relationship between countries' GDP per capita and average life expectancy. Next, participants proceeded to add and configure additional *Chart*, *Annotation*, and *Gesture Widgets* at their own pace, though we provided assistance when requested. We then asked participants to prepare a presentation using the tool based on a provided script of speaking points. Lastly, we conducted a semi-structured interview about their experience. The whole study lasted about one hour and each participant received \$20.

5.2 Audience Study

The primary objective of this study was to assess the tool's efficacy in presenting visualization with affect-varying animations, visual foreshadowing, and highlighting, with a specific focus on how VisConductor might affect audience engagement.

Participants. We recruited another 11 participants (4F, 7M), ranging in age between 18 and 44, having professional backgrounds that included engineering, computer science, and business. Five held post-graduate degrees and four were college students, while the remaining two had completed college. When assessing their data visualization literacy, six self-identified as *Advanced*, four as *Intermediate*, and one as a *Beginner*. Most (nine) frequently attended presentations on a daily (1), weekly (7), or monthly (1) basis. All of the participants had attended in-person presentations, while six had attended online presentations.

Procedure. After signing the consent form, participants completed a questionnaire about their past experiences. Then, participants watched a three-minute presentation conducted in VisConductor delivered via Zoom. While the presentation content varied (five saw a presentation featuring a bar chart race while six saw one featuring an animated scatterplot), the presentation featured all of VisConductor’s major features. After watching the presentation, participants were asked to summarize the content, which helped us gauge if they understood the presentation content and whether they had paid sufficient attention. Finally, we conducted a semi-structured interview and asked participants to complete a questionnaire regarding their experience. Each study session lasted about 30 minutes, and each participant received \$15.

5.3 Observations and Participants’ Reflections

Following an iterative qualitative analysis approach, we categorized findings from both studies according to common themes. Throughout this section, “P#” denotes a presenter study participant while “A#” denotes an audience study participant.

General utility. Presenters appreciated having gestural control of a presentation. P7 likened the experience to that of performing, stating that *“it gives the idea that I have more control over my slides ... almost like a performance for the audience.”* Further, P2 suggested that such a tool would make presenters commit more when giving a presentation: *“Working and doing physical gestures rather than just talking through the slides ... it might be engaging for them as well,”* a reflection that engagement is not only a concern for a remote audience. A7 specifically lamented the relative difficulty of animating charts in other tools, asking *“how much effort does this take to set up? ... What level of abstraction do I deal with or is it video editing?”* One potential answer is a statement by P10, who spoke to VisConductor’s capacity for configuring animation via gestural interaction in contrast to how animations are specified in other platforms: *“The ease of interaction is what’s the best point ... for [a] simple presentation tool like PowerPoint, it is harder to go within the presentation itself to point to stuff, play or annotate stuff ... it takes much more technical expertise.”*

Effectiveness of animation for communicating affect. A9 spoke about the benefit of controlling an audience’s visual attention, implying that with VisConductor, *“you can control what you want the audience to see ... especially if you try to convey something emotional,”* suggesting its potential for affective resonance. Meanwhile, A10 took specific note of the animations that ephemerally morphed data marks into icons that conveyed semantic and affective resonance: *“There was like an arrow or something that showed up when I was expecting the circle to drop ... I feel like these small things really help me understand the point you’re trying to convey.”* A9 also remarked that these animations, particularly when spatially superimposed with a presenter’s video, can reinforce the affective nature of their gestures: *“It helps if you want people to see your emotional gesture ... [it] makes it more clear to show which part of the presentation you want people to focus on, whether it’s the graphics or is it both you and the graphics.”*

Effectiveness of visual foreshadowing. While visual foreshadowing in narrative visualization has been studied independently [36], A7 remarked upon the effectiveness of this technique while also being able to see a presenter’s body language: *“Your face being behind the data, or being with the data, definitely emphasizes the importance of where to look ... especially the World War One part where you show the trend lines before the actual animation plays out ... that was super interesting.”* From a presenter’s perspective, P4 felt that foreshadowing helped both to highlight parts of a chart graph and to pique the audience’s curiosity, or as P3 put it, to keep the audience *“on their toes.”*

Effectiveness of annotation disclosure. A7 made specific mention of the interplay between a presenter’s gestures and the visual design of annotations, such as their color and placement: *“Your gestures explaining the axes and the different colors and how they move ... it was easy to understand ... I*

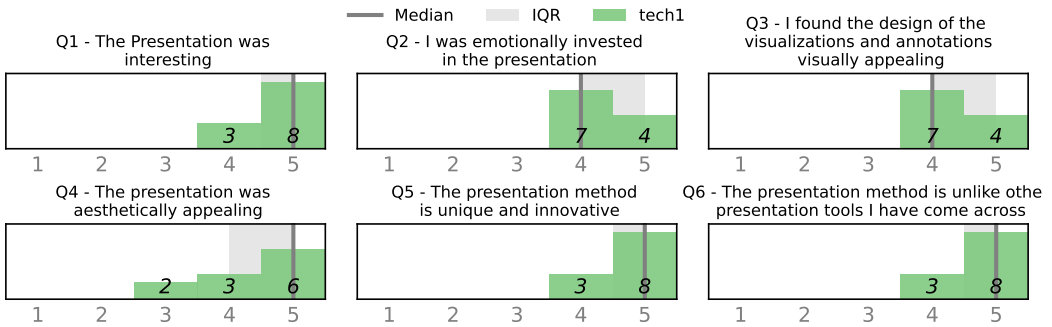


Fig. 8. Distribution of responses to User Engagement Scale [41] questions among participants in the audience study (1 = ‘strongly disagree’, 5 = ‘strongly agree’). The dashed line indicates the median and the shaded region denotes the interquartile range (IQR).

don’t think there were any gaps or lack of understanding because of too much information on the plot itself.” From a presenter’s perspective, P1 spoke about this interplay as well: “The gestures are good, I can just highlight which position of the graph I want to tell to the audience and we can set beforehand which part I will show so it will highlight accordingly.” P4 further attested to the utility of connecting gestures to the reveal of annotations, noting the synergistic effect of selection and foreshadowing in emphasizing specific chart elements: “[it] helps me highlight certain parts of the graph, just to show whatever is coming next.” Presenters also discussed the relative advantages of using gestures to reveal annotations over conventional presentation tools, with P10 stating that “in a simple presentation tool like PowerPoint, it is harder to go within the presentation itself to point to annotate stuff.”

Effectiveness with respect to audience engagement. Figure 8 shows the audience study participant responses to User Engagement Scale [41] questions. The predominant sentiment was positive, reflected in a median score of 4 or above across all questions, complementing the preceding remarks about specific aspects of the presentation experience. Overall, participants found the presentation they saw to be interesting (Q1) to the point of eliciting emotional investment (Q2), visually or aesthetically appealing (Q3-Q4), unique or innovative (Q5), and unlike presentations they had seen delivered using other tools or approaches (Q6).

Customizability and usability. P3 remarked that “authoring is pretty easy ... it’s really accommodative ... there are a lot of options to choose from.” More specifically, P9 appreciated the ability to morph chart elements into various SVGs: “all the visual effects are really good ... we can customize the icons ... the tool’s customizability allows presenters to create diverse and engaging animations that complement their data and message.” However, while VisConductor offers several animation parameters, P11 mentioned that it requires greater attention to design details: “it means that you need to put more details into the design part ... but once you get familiar with the whole system, it allows you to create the perfect slides.”

With respect to *Gesture Widgets* and their associated feedback, P3 was familiar with a widget-based approach based on experience with other tools: “it was easy for me as a first-time user to do it really fast ... It felt familiar because when we use various other apps, we see these options available as widgets.” Meanwhile, P1 appreciated VisConductor’s flexibility with respect to gesture positioning, stating: “everything is manually tunable, so I can position the boxes I can position the gestures.” Presenters also commented on the quality of feedback from the gestural widgets, with P5 finding the experience to be better than expected during presentation delivery, pointing out that “I expected

it to be a bit more confusing and hard to handle, but it was a really good experience, I could easily communicate, understand why it's not working or why it is working."

Challenges. One issue raised by presenter study participants was a need to accommodate more spontaneous interactions during a presentation. For instance, P6 expressed a desire for "*more dynamic control over charts*," emphasizing that there should not be a necessity to extensively "*pre-plan things*" in advance. We interpret such comments as a need for ways that presenters can adapt to unexpected events and varying levels of audience engagement during a presentation: be able to go off-script, ad-lib, or respond to interruptions and questions. This feedback underscores the need to balance planned automation and real-time manual control, to ensure that presenters can seamlessly navigate prepared material without feeling overly constrained by the tool. To this end, P3 suggested prompts visible only to presenters that indicate when specific gestures could be performed, allowing some gestures to be skipped. We acknowledge that by providing participants with a list of prepared speaking points, they might have interpreted this list to be a linear sequence, however, this was never explicitly stated as such.

Participants in the audience study raised a different concern, namely one relating to animation and gestural coherence. For instance, VisConductor allows *Gesture Widgets* to be freely placed, meaning that a misalignment of a pointing gesture with an animated referent, despite being valid from an implementation perspective, could confuse the audience and divide their attention. This potential disconnect, according to A8, could occur if gestures do "*not match the visual effects or properly communicate the message*," underscoring the need for coherent gesture-animation interactions, such as by enforcing position and / or distance thresholds when linking a *Gesture Widget* to an existing *Chart* or *Annotation Widget*.

6 DISCUSSION

We now reflect on the design of VisConductor and on findings from our evaluation with reference to design requirements **D1–D3**.

6.1 Design Implications for Gesture-Aware Augmented Video Communication

We pursued a modular widget-based approach in VisConductor, not only as a means to balance creative flexibility and usability, but also as a way to address the unique requirements associated with presenting data (**D2**): distinguishing between visualization and annotation and establishing affective associations with trends in the data. While some of our reflections below are unique to presenting data, we anticipate that aspects of our research are generally applicable to gesture-aware augmented video communication in a broader sense, particularly as our other two design considerations are not unique to presenting data, with **D1** calling for a small vocabulary of expressive and operational gestures, and **D3** calling for dedicated authoring and presentation interfaces that are distinct from what an audience sees.

Balance expressivity and operational control while imposing spatiotemporal constraints.

Prior work [25] in gesture-aware augmented video presentation warns of a potential clash between expressive and operational gestures, where the former serves as non-verbal communication with an audience and the latter controls aspects of an interface. In VisConductor, we opted for a small vocabulary of gestures that satisfy both criteria simultaneously (**D1**). Most of our gestures are deictic in nature, in which an extended finger or enclosed pair of hands are placed near a visual referent that will be in some way transformed, either animated, annotated, or contrasted with its future state via visual foreshadowing. A notable exception is the *Dialling* gesture, a metaphorical gesture for advancing in time; this gesture has no immediate deictic referent, though this absence of visual correspondence is appropriate here, as time is not represented visually, as it would be in

the horizontal axis of a line chart, for example. Otherwise, the spatial and temporal correspondence between a gesture and a dynamic visual aid helps to establish cause and effect and reinforces the agency of the presenter. Considering what participants in the audience told us with respect to the reveal of annotations and visual foreshadowing, this correspondence is critical for directing the audience's attention and building rapport.

We acknowledge the possibility that a spatial correspondence between a deictic gesture and its visual referent(s) may not always be possible, particularly if the relative positioning of the presenter and webcam puts some visual elements out of reach. In these cases, we envision two potential solutions. First, a presenter could alternatively opt to use a physical pointing device that is similarly tracked by the system. Second, we could extend *Gesture Widgets*. For instance, a *Pointing* gesture could draw a vector emanating from the extended finger, a vector that could highlight or annotate distant targets. Another possible extension is an adaptation of the bimanual *Rectangular Framing* gesture to recognize a triangular aperture formed by two hands [61]; we could visually demarcate this aperture and similarly highlight or annotate targets that fall within it.

Convey a spectrum of affects via gesture when telling stories with data. If the goals of data storytelling are to change viewers' opinions and beliefs, to persuade them, or to inform their decisions, establishing an appropriate affect for a narrative is critical [35]. Unfortunately, it can be challenging to convey semantic and affective associations with the abstract visual elements used in conventional chart types. As we demonstrated in VisConductor, a presenter can perform gestures to animate visualization elements, so as to reflect how values change over time (D2), where elements could, for instance, change in size, position, shape, or color. If we consider a linearly-paced and automatically-triggered animation to be a baseline, one suggestive of more neutral affect, the ability to parameterize the easing and pacing of an animation via gesture can suggest a broader and more nuanced range of affects. Consider our scenario (Section 3.1), in which the presenter drew attention to the cataclysmic effects of World War I on global health metrics followed by a triumphant rebound, using expressive gestures to reinforce this contrast. While prior visualization authoring tools have offered extensive options to parameterize animation (e.g., [51, 57]), including those for selecting specific affects [34], what is unique about VisConductor is that a viewer can directly perceive this parameterization in real-time by observing a presenter's gestures.

A presenter's body language is not the only way to communicate affect when animating data-bound visual elements. Recall that VisConductor additionally provides affordances to ephemerally morph the visual appearance of abstract shapes into semantically-charged ones, such as when we morphed circles in the scatterplot into downward-facing arrows when the values fell precipitously. Altogether, we argue that using an expressive gesture in tandem with the selective morphing of visual elements are complementary strategy for communicating an appropriate affect in the context of data storytelling. Prior research has highlighted the value of using semantically-rich or iconic representations of data in terms of memorability [10]. While we did not empirically assess memorability in our audience study, we are optimistic based on participants' comments and ratings relating to engagement that this combination of strategies would make for a memorable presentation of data.

Beyond the broad distinction between positive and negative affects, anticipation is a specific affective state that can be elicited when telling stories, particularly through the employ of non-chronological narratives. One promising technique for establishing this sense of anticipation in the context of data storytelling is that of visual foreshadowing as described by Li et al. [36]: ephemerally juxtaposing the final and initial states of visual elements and / or their intermediary states along a visible trajectory before an animation begins. In VisConductor, presenters invoke this technique via gesture. This gesture-driven foreshadowing was seen not only as a way to guide the audience's

attention, but also as a way to elicit curiosity and hypotheses before the presenter could explain a trend in the data during an animation.

Provide dedicated controls and views to support the flexible and dynamic nature of presentations. Prior work in augmented video communication offers alternative strategies in terms of presentation delivery support. Hall et al. [25] proposed a *what-you-see-is-what-I-see* (WYSIWIS) approach, in which there is no ambiguity with respect to what the presenter sees relative to what the audience sees. However, a downside of this approach is that a presenter receives no additional visual feedback or guidance as they perform gestural interactions. Alternatively, Liu et al.'s recent work on visual captions [38] reveals recommended visual aids in a dedicated interface adjacent to the video feed, from which a presenter can select recommended items, which in turn are composited in the WYSIWIS video. In VisConductor, authoring controls remain accessible throughout a presentation's delivery; while a presenter can share and monitor a WYSIWIS audience view (Figure 1-right), VisConductor has a dedicated presenter view (Figure 1-left) that additionally overlays gestural activation cues and feedback in addition to the visual aids (D3). Ultimately, we argue that WYSIWIS is insufficient for gesture-aware augmented video presentations about data, especially for those involving animated charts or gestures that have precisely specified spatial or temporal constraints.

A major benefit of gesture-aware augmented video presentation is the potential for presenters to deviate from a planned linear sequence. In conventional presentation tools, the animated reveal of visual aids on a slide follows a planned order, whereas with gestural activation, these reveals can be performed in different sequences at presentation delivery time, accommodating unforeseen events such as responding to questions or interruptions from the audience. While this non-linearity was realized in Hall et al.'s presentation environment [25], their gestural vocabulary and activation parameters could not be changed at presentation delivery time. In contrast, VisConductor provides presenters with full authoring control during a presentation, allowing them to modify existing widgets or add new ones on the fly. While we did not directly consider this scenario in our presenter study, we expect that having this interface on hand would reassure presenters. One promising approach would be to recommend new widgets during presentation delivery in a manner similar to that of Liu et al.'s visual captions [38], a generative approach that would require a real-time transcription of the presenter's monologue along with any audience questions or comments. For example, if a presenter mentions a particular entity in the data, the system could recommend a *Gesture Widget* tailored for highlighting the corresponding visual element. However, these widget recommendations would need to be delivered in an unobtrusive manner, such as by providing affordances to disable or dismiss recommendations during a rehearsed part of a presentation and enable them during unscripted segments, such as when fielding questions from the audience.

6.2 Limitations and Future Work

As a research prototype, we constrained the scope of VisConductor to two types of animated chart commonly encountered in data storytelling: the animated scatterplot has been a topic of study in the visualization research community for nearly two decades [45], while the animated bar chart race is a newer convention, one recently considered in Li et al.'s work on visual foreshadowing [36]. However, there are many other animated chart design conventions used in practice, from dynamic node-link representations of networks changing over time to various forms of animated thematic maps, not to mention the opportunities for animating bespoke works of visualization design. We acknowledge the possibility that some of our proposed *Gesture Widgets* may generalize in the context of other visualization design idioms, while others may require adaptation depending on the specific visual encoding design choices in use. That being said, a promising potential contribution

of future research would be a systematic investigation of the efficacy of common expressive and operational gestures when paired with different visual encoding channels (e.g. position, orientation, area, hue, etc.) and animated transitions [27], as opposed to specifying unique gestures for each type of chart. In the meantime, should VisConductor evolve into a publicly-available presentation tool, its adoption would depend on the addition of affordances for using other types of visual aids in presentations, from images and text to static (i.e., non-animated) charts.

It is also important to acknowledge the known difficulties of evaluating narrative visualization authoring tools [48] such as VisConductor. Despite these challenges and a modest number of participants in both of our studies ($N = 11$), we were nevertheless able to gather insightful quantitative and qualitative data. In the presenter study, we provided participants with a template presentation featuring several widgets, along with data and speaking points to seed their own presentation. While template reproduction or extension as well as open-ended exploration are commonly associated with evaluating visualization authoring tools [43], particularly when working with familiar constraints (i.e. participant time), these activities can only provide initial assessments of utility and usability. One aspect of usability that we did not assess involves the potential fatigue associated with performing gestures over the course of long presentations. However, we deem it unlikely that animated charts would be required throughout a long presentation; instead, we expect that they would be interleaved with other visual aids such as text or images that require no overt gestural interaction. A longitudinal deployment study of VisConductor that allows participants to use their own data at their own pace would certainly be illuminating, though VisConductor's functionality would need to be extended as described above to attract prospective early adopters.

In our audience study, we could not guarantee that participants had a preexisting interest in the topic of the presentations they saw, though they were able to provide us with a preliminary sense of whether presentations delivered using VisConductor could be engaging and memorable. Similarly, a brief study context could not shed light on whether presentations delivered via VisConductor might be used to combat so-called *Zoom Exhaustion and Fatigue* [18], a condition identified in remote learning environments that appears to be exacerbated by a lack of non-verbal cues. One possible next step would be to complement our initial results by partnering with post-secondary educators conducting online lectures in STEM fields where it is common to use charts and animations as visual aids. Across a diversity of presentation topics and durations, observing how students engage with lecturers and speaking to them about their experiences would be informative, particularly if the students are intrinsically motivated to learn about the subject matter being presented.

7 CONCLUSION

We introduced VisConductor, an augmented video presentation tool that combines the innate communicative power of hand gestures with spatially-superimposed animated data visualization. Recognizing that many gestures can be both communicative and operational, VisConductor incorporates gestural recognition to trigger animation playback, visual foreshadowing, and the reveal of annotation. Our evaluation, considering the perspectives of presenters and audience alike, provided insights into the utility of this medium for telling stories with dynamic representation data. Presenters appreciated having fine-grained control over animation and annotation, while audience members were captivated by the immersive and engaging nature of presentations delivered using VisConductor. As remote communication needs and use cases continue to proliferate, tools like VisConductor are poised to make presentations more than just an exchange of information, but rather an engaging and emotionally-resonant experience.

ACKNOWLEDGMENTS

We thank the participants for their time and effort. This research is supported by the NSERC Discovery Grant (RGPIN-2020-03966).

REFERENCES

- [1] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindbauer, Alexandra Ion, Shengdong Zhao, and Jeffrey Tzu Kwan Valino Koh. 2012. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI. *Microsoft Research Technical Report MSR-TR-2012-111 2*, MSR-TR-2012-111 (2012).
- [2] All Turtles. 2021. mmhmm. <https://www.mmhmm.app/>
- [3] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Jason Leboe-McGowan, and Pourang Irani. 2018. Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement. In *Proceedings of the ACM International Working Conference on Advanced Visual Interfaces (AVI)*. <https://doi.org/10.1145/3206505.3206552>
- [4] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andrés Monroy-Hernández, and Pourang Irani. 2017. Authoring data-driven videos with DataClips. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 23, 1 (2017). <https://doi.org/10.1109/TVCG.2016.2598647>
- [5] ANONYMOUS. [n. d.]. personal communication.
- [6] Apple. 2023. Use reactions, presenter overlay, and other effects when videoconferencing on Mac. <https://support.apple.com/en-us/HT212950>
- [7] Jose Maria Barrero, Nicholas Bloom, and Steven J Davis. 2023. The evolution of working from home. <https://siepr.stanford.edu/publications/working-paper/evolution-working-home> Working paper. The Stanford Institute for Economic Policy Research (SIEPR).
- [8] Thomas Baudel and Michel Beaudouin-Lafon. 1993. Charade: Remote control of objects using free-hand gestures. *Commun. ACM* 36, 7 (1993). <https://doi.org/10.1145/159544.159562>
- [9] BBC. 2010. Hans Rosling’s 200 Countries, 200 Years, 4 Minutes - The Joy of Stats. <https://www.youtube.com/watch?v=jbkSRLYSojo>
- [10] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 19, 12 (2013).
- [11] Matthew Brehmer and Robert Kosara. 2022. From jam session to recital: Synchronous communication and collaboration around data in organizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VIS)* 28, 1 (2022). <https://doi.org/10.1109/TVCG.2021.3114760>
- [12] Xiang Cao, Eyal Ofek, and David Vronay. 2005. Evaluation of alternative presentation control techniques. In *Extended Abstract Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. <https://doi.org/10.1145/1056808.1056888>
- [13] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. <https://doi.org/10.48550/arXiv.1812.08008> arXiv preprint.
- [14] Adelia Carstens. 2019. Advice on the use of gestures in presentation skills manuals: alignment between theory, research and instruction. *Image & Text* (2019). <https://doi.org/10.17159/2617-3255/2018/n33a8>
- [15] Tony Chu. 2016. Animation, pacing, and exposition. Presented at OpenVisConf. <https://youtu.be/Z4tB6qyxHJA>.
- [16] Cisco Systems. 2021. Webex. <https://www.webex.com/>
- [17] Stefania Cucurullo, Rita Francese, Sharefa Murad, Ignazio Passero, and Maurizio Tucci. 2012. A gestural approach to presentation exploiting motion capture metaphors. In *Proceedings of the ACM International Working Conference on Advanced Visual Interfaces (AVI)*. <https://doi.org/10.1145/2254556.2254584>
- [18] Géraldine Fauville, Mufan Luo, Anna Carolina Muller Queiroz, Jeremy N. Bailenson, and Jeffrey T. Hancock. 2021. Zoom exhaustion & fatigue scale. *Computers in Human Behavior Reports* (2021). <https://doi.org/10.1016/j.chbr.2021.100119>
- [19] Jean-Daniel Fekete, Jarke J. van Wijk, John T. Stasko, and Chris North. 2008. The value of information visualization. In *Information Visualization: Human-Centered Issues and Perspectives*, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.). Springer. https://doi.org/10.1007/978-3-540-70956-5_1
- [20] Danyel Fisher. 2010. Animation for visualization: Opportunities and drawbacks. In *Beautiful Visualization: Looking at Data through the Eyes of Experts*, Julie Steele and Noah Iliinsky (Eds.). O’Reilly.
- [21] Adam Fourney, Michael Terry, and Richard Mann. 2010. Gesturing in the wild: Understanding the effects and implications of gesture-based interaction for dynamic presentations. In *Proceedings of the Conference on Human-Computer Interaction (HCI)*. <https://doi.org/10.14236/ewic/HCI2010.29>
- [22] Susan Goldin-Meadow. 2023. *Thinking with Your Hands: The Surprising Science Behind How Gestures Shape Our Thoughts*. Basic Books. <https://books.google.ca/books?id=4XKSEAAQBAJ>
- [23] GreenSock. 2023. GSAP: JavaScript HTML5 animation. <https://greensock.com/gsap/>

- [24] Raja Gumienny, Lutz Gericke, Matthias Quasthoff, Christian Willems, and Christoph Meinel. 2011. Tele-Board: Enabling efficient collaboration in digital design spaces. In *Proceedings of the International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. <https://doi.org/10.1109/CSCWD.2011.5960054>
- [25] Brian D. Hall, L. R. Bartram, and Matthew Brehmer. 2022. Augmented chironomia for presenting data to remote audiences. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. <https://doi.org/10.1145/3526113.3545614>
- [26] Jörg Hauber, Holger Regenbrecht, Mark Billinghurst, and Andy Cockburn. 2006. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*. <https://doi.org/10.1145/1180875.1180937>
- [27] Jeffrey Heer and George G. Robertson. 2007. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 13, 6 (2007). <https://doi.org/10.1109/TVCG.2007.70539>
- [28] Jeffrey Heer and Ben Shneiderman. 2012. Interactive dynamics for visual analysis. *Commun. ACM* 5, 4 (2012). <https://doi.org/10.1145/2133806.2133821>
- [29] Adobe Systems Incorporated. 2023. Adobe After Effects. <https://www.adobe.com/products/aftereffects.html>. Accessed: 2023-04-23.
- [30] Seokmin Kang and Barbara Tversky. 2016. From hands to minds: Gestures promote understanding. *Cognitive Research: Principles and Implications* 1, 1 (2016). <https://doi.org/10.1186/s41235-016-0004-9>
- [31] Adam Kendon. 2004. *Gesture Visible Action as Utterance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511807572>
- [32] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The next step for visualization. *IEEE Computer* 46, 5 (2013). <https://doi.org/10.1109/MC.2013.36>
- [33] Andreas Kunz, Thomas Nescher, and Martin Küchler. 2010. CollaBoard: A Novel Interactive Electronic Whiteboard for Remote Collaboration with People on Content. In *Proceedings of the International Conference on Cyberworlds*. <https://doi.org/10.1109/CW.2010.17>
- [34] Xingyu Lan, Yang Shi, Yanqiu Wu, Xiaohan Jiao, and Nan Cao. 2022. Kineticharts: Augmenting affective expressiveness of charts in data stories with animation design. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VIS)* 28, 1 (2022). <https://doi.org/10.1109/TVCG.2021.3114775>
- [35] Elsie Lee-Robbins and Eytan Adar. 2023. Affective learning objectives for communicative visualizations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VIS)* 29, 1 (2023). <https://doi.org/10.1109/TVCG.2022.3209500>
- [36] Wenchao Li, Yun Wang, Haidong Zhang, and Huamin Qu. 2020. Improving engagement of animated visualization with visual foreshadowing. In *Short Paper Proceedings of the IEEE Visualization Conference (VIS)*. <https://doi.org/10.1109/VIS47514.2020.00035>
- [37] Jian Liao, Adnan Karim, S. Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-time speech-driven augmented presentation for AR live storytelling. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. <https://doi.org/10.1145/3526113.3545702>
- [38] Xingyu "Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang "Anthony" Chen, and Ruofei Du. 2023. Visual captions: Augmenting verbal communication with on-the-fly visuals. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. <https://doi.org/10.1145/3544548.3581566>
- [39] Tomer Moscovich and John F. Hughes. 2004. Navigating documents with the virtual scroll ring. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. <https://doi.org/10.1145/1029632.1029642>
- [40] OBS 2022. OBS Studio. <https://obsproject.com>.
- [41] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018). <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- [42] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyoung Choe. 2017. ChartAccent: Annotation for data-driven storytelling. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*. <https://doi.org/10.1109/PACIFICVIS.2017.8031599>
- [43] Donghao Ren, Bongshin Lee, Matthew Brehmer, and Nathalie Henry Riche. 2018. Reflecting on the evaluation of visualization authoring systems. In *Workshop Proceedings of Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. <https://doi.org/10.1109/BELIV.2018.8634297>
- [44] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale (Eds.). 2018. *Data-Driven Storytelling*. A K Peters/CRC Press.
- [45] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. 2008. Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 14, 6 (2008). <https://doi.org/10.1109/TVCG.2008.125>

- [46] Nazmus Saquib, Rubaiat Habib Kazi, Li-yi Wei, and Wilmot Li. 2019. Interactive body-driven graphics for augmented video performance. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. <https://doi.org/10.1145/3290605.3300852>
- [47] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring narrative visualizations with Ellipsis.... *Computer Graphics Forum (Proceedings of EuroVis)* 33, 3 (2014). <https://doi.org/10.1111/cgf.12392>
- [48] Arvind Satyanarayan, Bongshin Lee, Donghao Ren, Jeffrey Heer, John Stasko, John Thompson, Matthew Brehmer, and Zhicheng Liu. 2020. Critical reflections on visualization authoring systems. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 26, 1 (2020). <https://doi.org/10.1109/TVCG.2019.2934281>
- [49] Andreas Schallwig. 2023. Fingerpose. <https://github.com/andypotato/fingerpose>
- [50] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 16, 6 (2010). <https://doi.org/10.1109/TVCG.2010.179>
- [51] 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 (TVCG)* 29, 6 (2023). <https://doi.org/10.1109/TVCG.2022.3146329>
- [52] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. *Commun. ACM* 56, 1 (2013). <https://doi.org/10.1145/2398356.2398381>
- [53] Graham M. Smith and m. c. schraefel. 2004. The radial scroll tool: scrolling support for stylus- or touch-based document navigation. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. <https://doi.org/10.1145/1029632.1029641>
- [54] Charles D Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2018. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. In *Data-Driven Storytelling*, Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Cpendale (Eds.). A K Peters/CRC Press.
- [55] Hariharan Subramonyam and Eytan Adar. 2018. SmartCues: A multitouch query approach for details-on-demand through dynamically computed overlays. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 25, 1 (2018). <https://doi.org/10.1109/TVCG.2018.2865231>
- [56] Thomas Lin Pedersen. 2023. ganimate. <https://gganimate.com/>.
- [57] John Thompson, Zhicheng Liu, and John Stasko. 2021. Data Animator: Authoring expressive animated data graphics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. <https://doi.org/10.1145/3411764.3445747>
- [58] Lyn Tieu, Robert Pasternak, Philippe Schlenker, and Emmanuel Chemla. 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa* 2, 1 (2017). <https://doi.org/10.5334/gjgl.334>
- [59] Julien Valentin, Adarsh Kowdle, Suhil Nassar, Aameesh Makadia, Angjoo Kanazawa, Michael Ratasich, and Caroline Pantofaru. 2020. MediaPipe: A framework for building perception pipelines. <https://doi.org/10.48550/arXiv.1906.08172> arXiv preprint.
- [60] Rob van Eijk, André Kuijsters, Klaske Dijkstra, and Wijnand A. IJsselsteijn. 2010. Human sensitivity to eye contact in 2D and 3D videoconferencing. In *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEX)*. <https://doi.org/10.1109/QOMEX.2010.5518040>
- [61] Haijun Xia, Ken Hinckley, Michel Pahud, Xiao Tu, and Bill Buxton. 2017. WritLarge: Ink unleashed by unified scope, action, & zoom. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. <https://doi.org/10.1145/3025453.3025664>
- [62] Longqi Yang, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, et al. 2022. The effects of remote work on collaboration among information workers. *Nature Human Behavior* 6, 1 (2022). <https://doi.org/10.1038/s41562-021-01196-4>
- [63] Haipeng Zeng, Xingbo Wang, Yong Wang, Aoyu Wu, Ting-Chuen Pong, and Huamin Qu. 2023. GestureLens: Visual analysis of gestures in presentation videos. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 29, 8 (2023). <https://doi.org/10.1109/TVCG.2022.3169175>
- [64] Jakob Zillner, Christoph Rhemann, Shahram Izadi, and Michael Haller. 2014. 3D-board: A whole-body remote collaborative whiteboard. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. <https://doi.org/10.1145/2642918.2647393>

Received 2024-02-22; accepted 2024-05-30