Ditto: Exploring data in large display environments through speech+mid-air gesture interactions

Jillian Aurisano*
University of Cincinnati

Abeer Alsaiari
Taibah University

Abhinav Kumar, Barbara Di Eugenio, Andrew Johnson
University of Illinois, Chicago

Jason Leigh
University of Hawaii, Manoa

Moira Zellner
Northeastern University

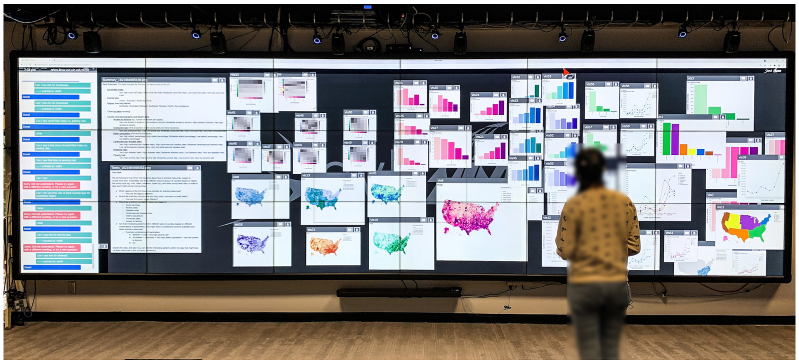

Figure 1: Exploring COVID-19 data using Ditto, a novel technique for data exploration using speech and mid-air pointing gestures.

Abstract

During visual data exploration, analysts often approach a dataset incrementally, segmenting data into meaningful partitions and representing these parts in multiple, related views. Large displays can support data exploration by providing space to juxtapose or organize related views. We present an interaction technique for creating many, related views of data for data exploration through synchronous speech and mid-air pointing gestures. We present our design goals, which leverage the combined affordances of speech, pointing gestures and large displays, with the aim of supporting exploratory tasks and transitions. We implemented our design in a large display environment with gesture tracking and a speech input system, along with a touch system for freely positioning visualizations. We implemented this technique in an application called Ditto, and evaluated through a user study where participants explored a COVID-19 dataset. We found that they used both modalities to interact and were able to efficiently create coherent sets of views which could be arranged on the large display.

1 Introduction

Visual data exploration is an iterative process, where an analyst alternates between open-ended exploration, targeted queries and iterative steps from current points of interest to new ones, based on observations or evolving exploratory goals [3, 15]. In the process, they may produce many visualizations which partition a dataset into meaningful pieces and which feature diverse selections of data values and data attributes. Large display environments may support visual data exploration by providing abundant space for displaying and juxtaposing multiple views produced during the exploratory process [4, 5, 10]. However, interaction in large display environments presents significant challenges, and recent research has considered ways to enable interaction using modalities beyond mouse and keyboard inputs, such as touch [1], proxemic interaction [7], and external devices [2, 6, 8, 11, 12].

We present a novel interaction technique for view creation in data exploration on large displays using synchronous speech and mid-air pointing gestures. Our technique enables analysts to create views of data around multiple, evolving points of interest and uses a design which targets the **interplay between synchronous speaking and pointing in communicating data exploration intentions**, and **leverages abundant display space and human spatial organizational capabilities**. We implemented this technique in an application called **'Ditto'**, which combines input systems for speech and mid-air pointing gestures, an input interpreter which translates spoken inputs and gesture targets into visualization responses, and a display application where many views can be freely positioned using touch interactions on a tiled-display wall. We performed a study with recruited participants, who explored COVID-19 data. We examined participant requests and how they used speech and mid-air pointing gestures together to express intentions. We found that they used both modalities, created coherent sets of views that could be positioned into meaningful configurations on the large display.

2 Ditto: Design, Implementation, Evaluation

In designing our technique, we developed the following design goals, driven by prior work in data exploration and interaction on large displays [4, 5, 10]. We will described these design goals in reference to a hypothetical city crime dataset, consisting in crime instances, classified by crime type (eg. theft, burglary...), month and year.

Design Goal 1: Multiple interests, coherent sets of visualizations. First, users of our technique can pose direct requests around one or many interests (data values and/or data attributes) through speech to create coherent sets of visualizations. These multiple interests may include an enumerated list of data value interests (such as wanting to see thefts and burglaries by year), or an enumerated listed of data attribute interests (such as wanting to see thefts by year and month and neighborhood). Our spoken command interpreter prioritizes responding with a set of views, rather than resolving the list into a single visualization, allowing the user to efficiently create multiple views of the data and utilize the large display space.

Design Goal 2: User-directed positioning into meaningful groups.. Second, users of our technique can freely position these

*e-mail: jillian.aurisano@uc.edu

visualizations on the large display, using touch interactions, into configurations that support their reasoning and exploration process, such as grouping multiple views showing a common attribute (Eg. year) with distinct filter criteria (Theft in Jan, Theft in May). This design goal is in line with prior work on how users of large display environments organize content on large displays [4, 10].

Design Goal 3: Enable exploratory transitions through reference to prior views or sets of views. Third, in our technique users can express exploratory transitions from one set of interests to another using synchronous speech and mid-air pointing interactions. In these *referential copy+pivot interactions*, users indicate a prior visualization and express that they want to see this view of the data but with some change, such as a new filter criteria or a new data attribute. For instance, a user could reference a view showing thefts by year and ask to see this visualization for burglaries. Thus, instead of re-articulating a complex request, users can use prior visualizations as a shortcut. We enable users to point to multiple, adjacent visualizations on the display and express a copy+pivot action that applies collectively to all the indicated views. Rather than having to reference each view individually, they can reference and pivot many visualizations at once. The 'pivot point' for this transition can be a common feature shared by the referenced views (such as a common filter criteria or data attribute). This takes advantage user spatial positioning decisions, where adjacent and related views can be pointed at in one interaction. This design goal is based on prior research large display interactions [5] and on exploratory transitions [9].

In our implemented system, Ditto, we capture spoken requests through a speech-to-text interface on a phone. A tracking system from a Kinect mounted to the center of the display is interpreted to detect pointing gestures, as depicted in Fig. 2. Our command interpreter translates these combined inputs into a response, consisting in one or several visualizations, which are encoded as VegaLite specification objects. Our interpreter uses a simple command grammar, which primarily captures user interests (Eg. data values and data attributes and other keywords). This is in line with other recent natural language for visualization research efforts [13], in which the initial focus is on novel interaction challenges prior to realizing robust NL interpretation, which has been extensively studied in recent years [14] in other application areas. Responses are displayed within Sage2, a tiled-display wall software [12], on a 24 x 6.75 feet (7.3 x 2 meters), 37 MPixel display.

We evaluated Ditto with 8 participants (plus 4 participants in a pilot study), who explored a COVID-19 dataset. Participants were provided with training, given a detailed data description and exploration prompt. We captured a total of 307 visualization queries with responses from Ditto. Of these, 55 percent (168) were 'direct' queries, where participants only used speech to express their intentions (DG1), and 45 percent (139) were 'referential' (DG3), involving simultaneous speech and mid-air pointing gestures. Of the direct requests, 21 percent (36) were 'cast-a-net' queries, where participants expressed several points of interest, and Ditto provided several views (DG1+3). Participants referenced recent views and views from early in the session (DG3), and created coherent groupings of visualizations to reflect their exploratory process (DG2).

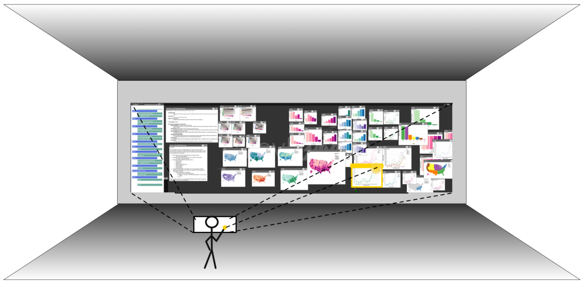

Figure 2: An illustration mapping mid-air pointing movements to an on-screen pointer.

Acknowledgments

This work was supported by NSF award IIS 1445751 and CNS 1625941

References

[1] M. Agarwal, A. Srinivasan, and J. Stasko. Viswall: Visual data exploration using direct combination on large touch displays. In *2019 IEEE Visualization Conference (VIS)*, pp. 26–30. IEEE, 2019.

[2] A. Alsaiari, J. Aurisano, and A. E. Johnson. Evaluating strategies of exploratory visual data analysis in multi device environments. In *EuroVis (Short Papers)*, pp. 91–95, 2020.

[3] S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst. Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE transactions on visualization and computer graphics*, 25(1):22–31, 2018.

[4] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 55–64. ACM, 2010.

[5] J. Aurisano, A. Kumar, A. Alsaiari, B. D. Eugenio, and A. Johnson. Many at once: Capturing intentions to create and use many views at once in large display environments. In *Computer Graphics Forum*, vol. 39, pp. 229–240. Wiley Online Library, 2020.

[6] S. K. Badam, E. Fisher, and N. Elmqvist. Munin: A peer-to-peer middleware for ubiquitous analytics and visualization spaces. *IEEE Transactions on Visualization and Computer Graphics*, 21(2):215–228, 2015.

[7] T. Ballendat, N. Marquardt, and S. Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*, pp. 121–130. ACM, 2010.

[8] T. Horak, S. K. Badam, N. Elmqvist, and R. Dachselt. When david meets goliath: Combining smartwatches with a large vertical display for visual data exploration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 19. ACM, 2018.

[9] D. Jung-Lin Lee, V. Setlur, M. Tory, K. Karahalios, and A. Parameswaran. Deconstructing categorization in visualization recommendation: A taxonomy and comparative study. *arXiv e-prints*, pp. arXiv–2102, 2021.

[10] S. Knudsen, M. R. Jakobsen, and K. Hornbæk. An exploratory study of how abundant display space may support data analysis. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pp. 558–567. ACM, 2012.

[11] R. Langner, U. Kister, and R. Dachselt. Multiple coordinated views at large displays for multiple users: Empirical findings on user behavior, movements, and distances. *IEEE transactions on visualization and computer graphics*, 25(1):608–618, 2018.

[12] T. Marrinan, J. Aurisano, A. Nishimoto, K. Bharadwaj, V. Mateevitsi, L. Renambot, L. Long, A. Johnson, and J. Leigh. Sage2: A new approach for data intensive collaboration using scalable resolution shared displays. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2014 International Conference on*, pp. 177–186. IEEE, 2014.

[13] A. Srinivasan, B. Lee, N. Henry Riche, S. M. Drucker, and K. Hinckley. Inchorus: Designing consistent multimodal interactions for data visualization on tablet devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[14] A. Srinivasan and J. Stasko. Natural language interfaces for data analysis with visualization: Considering what has and could be asked. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*, pp. 55–59. Eurographics Association, 2017.

[15] J. W. Tukey et al. *Exploratory data analysis*, vol. 2. Reading, Mass., 1977.