Towards Understanding Multimodal Interaction for Visual Data Analysis

Marcel Ruoff[*]
Karlsruhe Institute of Technology

Alexander Maedche[†]
Karlsruhe Institute of Technology

Abstract

Multimodal interaction for visual data analysis and exploration provides new opportunities for empowering users to engage with data. However, it is not well understood which input modalities should be leveraged for certain information visualization (InfoVis) operations and how user would prefer to utilize them during data analysis and exploration. In order to close this research gap, we performed an user-elicitation study to examine how users utilize touch, speech, mid-air hand gestures and a combination of those for various InfoVis operations on large interactive displays. We believe this analysis will help us identify associated challenges and provide knowledge for the development of systems that provide multimodal interaction capabilities for visual data analysis and exploration.

Index Terms: Human-centered computing—Interaction techniques; Human-centered computing—Visual analytics

1 Introduction

Interaction in visual data analysis and exploration provides users the means to enter into a dialogue with the information visualization (InfoVis) and enables users to understand and gain insights into the underlying data. In recent years, InfoVis interactions have increasingly taken advantage of the new possibilities offered by advanced interaction technologies, such as touch, speech and mid-air hand gesture and the combination of those [3, 7].

However, a key challenge in designing multimodal interaction for visual data analysis and exploration is to understand how users want to interact with InfoVis systems using multimodal capabilities itself and how InfoVis operations, context and individual characteristics influence their preferences [8]. To date, multimodal interaction for InfoVis are mostly defined by system designers and not the users itself [3]. Although this may be chosen out of concern for reliable recognition [6], it is not helpful for determining which interactions and combination of modalities are intuitive to the users. Furthermore, the effects of different modalities and modality combinations on the interactivity and cognition of the users are not well understood [3].

Particularly in the context of large interactive displays, such as Microsoft's Surface Hub or Samsung's Flip, researchers strive to understand which modalities users would prefer for interacting based on their individual characteristics [8]. However, it is unclear how users want to interact with InfoVis systems on large interactive displays using touch, speech and mid-air hand gestures and how preferences depend on the InfoVis operation.

In order to address this research gap, we build on previous research [5] and report insights from a user-elicitation study with 30 participants. Specifically, for 15 typical InfoVis operations (e.g., filter, reconfigure) user preferences for touch, speech, mid-air hand gestures and a combination of these modalities were elicited as well as the most popular interactions were derived. After a more detailed analysis, we intend to contribute with a set of user-defined interactions for InfoVis systems on large interactive displays as well as modality preferences of users for the selected 15 operations.

2 Methodology

To better understand how users want to interact with multimodal InfoVis systems on large interactive displays using (1) touch, (2) speech, (3) mid-air hand gestures, and (4) a combination of these modalities, we conducted a lab-based user-elicitation study. Furthermore, we investigated how user preferences depend on the specific operation. In the following, we will describe the underlying methodology of the performed user-elicitation study in more detail.

2.1 Participants and Apparatus

The thirty participants in the user-elicitation study consisted of eight female and twenty-two male participants with a mean age of 22.8 years (SD = 5.94). In consistence with previous user-elicitation studies [1] our participants were students with a background in computer science, physics, engineering as well as management & economics. Five of the participants were left handed. We conducted the user-elicitation study in an experimental laboratory. Here we used a dedicated room with 23m^2. The room included a 65" LCD display which was mounted with a Logitech Brio 4k to simulate the input devices for touch, speech and mid-air hand gestures.

2.1.1 InfoVis Operations

In order to identify common InfoVis operations, we used the framework of Yi et al. [10] to categorize the operations with 11 InfoVis systems, such as [2, 4, 9]. Our selection criteria were that the systems provided an user interface either implementing touch, speech, mid-air hand gestures or a combination of them. Through this approach we identified the following InfoVis operations and extracted them for our user-elicitation study (Table 1).

Table 1: The list of InfoVis operations presented to participants grouped by category.

Interaction Intent [10]	**InfoVis Operation**
Select	Select
Filter	Filter
Explore	New Visualization
Explore	Question to Data
Explore	Scroll
Explore	Change Tab
Encode	Change Visualization Type
Reconfigure	Reconfigure
Abstract/Elaborate	Drill-Down
Abstract/Elaborate	Zoom In
Abstract/Elaborate	Zoom Out
Connect	Details
Connect	Externalize Insights
Others	Bookmark
Others	Undo

2.2 Procedure

The user-elicitation study consisted of two parts. First, participants elicited (1) touch, (2) speech, (3) mid-air hand gestures, and (4) a combination of these modalities for each InfoVis operation and

[*] e-mail: marcel.ruoff@kit.edu

[†] e-mail: alexander.maedche@kit.edu

rated them. Second, participants answered a series of interview questions about their elicited interactions and their preferences. The participants were asked to think out loud during the elicitation of the interactions and the sessions were recorded via video.

For the elicitation portion of the study, we told participants that they should suggest interactions for each modality which are most intuitive to them and would be preferred for a future multimodal InfoVis system independent of perceived recognition reliability. The experimenter then walked the participants through the 15 InfoVis operations, which were randomized for each session. As presented by Morris [5], for each InfoVis operation the experimenter stated the name and demonstrated the effect of the InfoVis operation as a video, then prompt the participant to suggest interactions that would cause this InfoVis operation to be executed (Figure 1).

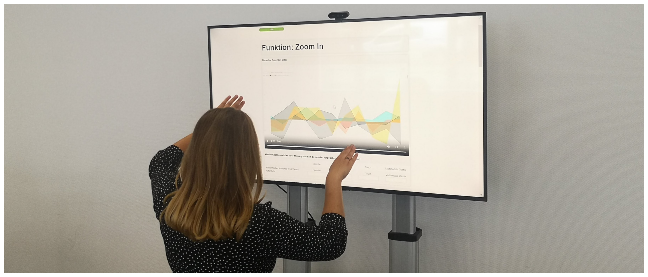

Figure 1: Elicitation of the InfoVis Operation 'Zoom In'

For each InfoVis operation, participants were able to suggest one interaction for (1) touch, (2) speech, (3) mid-air hand gestures, and (4) a combination of these modalities or could refuse to suggest an interaction if they think that no interaction would be possible or intuitive for a certain modality or combination. After each InfoVis operation the participants were asked to order the suggested interactions based on which interaction they would prefer most in a future system for that InfoVis operation and in which context they would use them. Afterwards we asked the participants to rate their interactions on two Likert scales, depicting *ease of use* and *goodness of mapping* of the interaction to the InfoVis operation.

With thirty participants, 15 InfoVis operations, and (1) touch, (2) speech, (3) mid-air hand gestures, and a combination of them, as well as 363 refusals by the participants because of the lack of an intuitive interaction, a total of (30 x 15 x 4) - 363 = 1.437 interactions were proposed by the participants. From these proposed interactions we derived 492 unique interactions for the subsequent analysis.

3 Initial Findings

Our findings suggest that users would be highly receptive to use multimodal InfoVis systems on large interactive displays. Our analysis of the agreement between participants based on the max-consensus metric [5] and the analysis of the preference regarding the modality used in a subsequent prototype confirms previous research that the combination of touch and speech is promising. Our results further show that especially speech is beneficial for interactions, which need to convey the intended interaction as well as supporting to access more specific information, e.g. by leveraging filters and configuration parameters. Touch, however, is great as a robust and basic interaction modality and is especially beneficial for interaction, which require the user to select targets for the intended interactions. Even though mid-air hand gestures are very well suited for simple interaction, which only need to convey the intended interaction, these interactions can also be provided using touch.

Furthermore, our analysis shows, that the agreement score should be complemented in multimodal user-elicitation studies by additional preference scores, such as our preference ranking. For example, based on the max-consensus metric designers and researchers could conclude, that using menus to create new visualizations, facilitated by touch or mid-air hand gestures, would be a intuitive interaction preferred by the users. However, when combining the max-consensus metric with our preference score, we were able to show, that speech is mainly preferred as a modality to create new visualizations and that we could focus in our future analysis of the results on synonyms for speech as well as touch for creating visualizations.

4 Conclusion and Future Work

We conducted a user-elicitation study to identify how users want to interact with multimodal InfoVis systems on large interactive displays using touch, speech and mid-air hand gestures and how their preferences depended on InfoVis operations. Overall, participants provided positive feedback regarding the usage of multimodal user interfaces for visual data analysis and exploration. Our results specifically show that touch and speech are promising input modalities for designing multimodal InfoVis systems on large interactive displays.

Besides further analyzing the results of our user-elicitation study, we are currently implementing a multimodal InfoVis system on large interactive displays. We especially plan to focus on touch and speech interaction, based on our results, and investigate how these modalities can be further enhanced to increase the effective use of these systems and the productivity of the users.

References

[1] S. K. Badam and N. Elmqvist. Visfer: Camera-based visual data transfer for cross-device visualization. *Information Visualization*, 18(1):68–93, jan 2019. doi: 10.1177/1473871617725907

[2] S. M. Drucker, D. Fisher, R. Sadana, J. Herron, and M. Schraefel. Touchviz: a case study comparing two interfaces for data analytics on tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2301–2310, 2013.

[3] B. Lee, P. Isenberg, N. H. Riche, and S. Carpendale. Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2689–2698, dec 2012. doi: 10.1109/TVCG.2012.204

[4] B. Lee, G. Smith, N. H. Riche, A. Karlson, and S. Carpendale. SketchInsight: Natural data exploration on interactive whiteboards leveraging pen and touch interaction. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, vol. 2015-July, pp. 199–206. IEEE, apr 2015. doi: 10.1109/PACIFICVIS.2015.7156378

[5] M. R. Morris. Web on the wall. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces - ITS '12*, p. 95. ACM Press, New York, New York, USA, 2012. doi: 10.1145/2396636.2396651

[6] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In *International gesture workshop*, pp. 409–420. Springer, 2003.

[7] A. Saktheeswaran, A. Srinivasan, and J. Stasko. Touch? Speech? or Touch and Speech? Investigating Multimodal Interaction for Visual Network Exploration and Analysis. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, jan 2020. doi: 10.1109/tvcg.2020.2970512

[8] F. Schüssel, F. Honold, M. Weber, F. Schüssel, F. Honold, ·. M. Weber, and M. Weber. Influencing factors on multimodal interaction during selection tasks. *J Multimodal User Interfaces*, 7:299–310, 2013. doi: 10.1007/s12193-012-0117-5

[9] A. Srinivasan and J. Stasko. Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 24(1):511, 2018. doi: 10.1109/TVCG.2017.2745219

[10] J. S. Yi, Y. A. Kang, and J. Stasko. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, nov 2007. doi: 10.1109/TVCG.2007.70515