Valletto: A Multimodal Interface for Ubiquitous Visual Analytics

Jan-Frederik Kassel
Volkswagen Group
Data:Lab Munich
Munich, Germany
jan-frederik.kassel@volkswagen.de

Michael Rohs
University of Hannover
Human-Computer Interaction Group
Hannover, Germany
michael.rohs@hci.uni-hannover.de

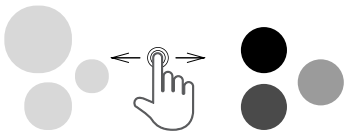
CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188445

Abstract
Modern technologies enable data analysis in scenarios where keyboard and mouse are not available. Research on multimodality in visual analytics is facing this challenge. But existing approaches consider exclusively static environments with large screens. Therefore, we envision *Valletto*, a prototypical tablet app which allows the user to generate and specify visualizations through a speech-based conversational interface, through multitouch gestures, and through a conventional GUI interface. We conducted an initial expert evaluation to gain information on the modality function mapping and for the integration of different modalities. Our aim is to discuss design and interaction considerations in a mobile context which fits the user's daily life.

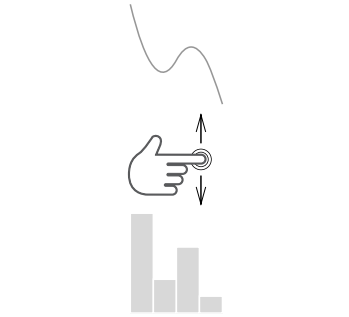
Author Keywords
Conversational Interface; User Experience; Visualization; Mobile Device; Multimodal Interaction; Ubiquitous Computing

ACM Classification Keywords
H.5.2 [User Interfaces]: Interaction styles, natural language

Introduction
Driven by mobile technologies and online infrastructure, like cloud services, mobile uses will get more and more common [10]: People increasingly like to access data of in-



Horizontal Swipes: Visual Encoding changes



Vertical Swipes: Changing the Type

Rotate Gesutre: Flips the axis

Figure 1: Three different executable gestures to change the current visualization's visual encoding.

terest [3, 4] whenever and wherever they want, e.g., while sitting in an autonomous vehicle on a way to a meeting. Hence, interfaces for visual analytics require a paradigm shift [7]. This new paradigm can base on the combination of conversational interfaces with other modalities of modern mobile devices. Conversational interfaces already showed their potential for visual analytics in desktop environments [14]. However, each modality is differently well suited for triggering typical visualization intents [2]. Furthermore, not only the functional range but also the user experience (UX) is critical. A concept with too many barriers, unclear functionality access, and insufficient integration in the user's daily life will probably fail [9].

In this paper, we approach the research gap of post-WIMP and multimodal user interfaces for visual analytics [14] in a mobile context. We introduce a prototypical tablet app, *Valletto*[1], which supports data-specific intents through speech and visual encoding intents through multitouch gestures. We illustrate design considerations and interaction possibilities in a mobile scenario as well as foster the discussion on the need of UX research in multimodal visual analytics.

Related Work

Research on multimodality for data analysis with visualization can be categorized into two groups [14]. The first group contains work on generating visualizations from scratch. Sun et al. [16] envisioned *Articulate*, a WIMP interface with an semi-automated natural language processing (NLP) procedure. In their approach, each visualization is individually shown in a pop-up window. Recently, Aurisano et al. [1] introduced *Articulate2* which implements the multi-window concept as well, but aims for an anticipating conversational interface. The challenge of ambiguity in the human language is discussed by Gao et al. [5]. They developed *DataTone*, a multimodal interface with two widgets, one to show the users how their queries are interpreted by the system and the other one to allow the users to adjust their query.

The second group focuses on ways to verbally manipulate existing visualizations. Setlur et al. [12] developed *Eviza*, a visual analytics tool for geospatial data. It shows the user three different visualizations (a map, a bar chart, and a scatter plot) for the given domain. The user is able to directly formulate his questions to these visualizations. Recently, Srinivasan and Stasko [15] introduced *Orko*, a multimodal post-WIMP user interface on a very large touchscreen for analyzing network data. It offers a solely, sequential or simultaneous use of touch and speech. In both *Eviza* [12] and *DataTone* [5], the user is able to adjust the interpreted query by using widgets. All discussed approaches are either WIMP [1, 5, 12, 16] or post-WIMP [15] multimodal interfaces for visual analytics on large and static screens. Our work belongs to the first category but explores visual analytics in a post-WIMP and mobile context.

Concept

Valletto is a prototypical tablet app that is structured into four tabs: The first shows for each available data attribute its corresponding histogram as well as standard statistical values, the second implements the multimodal visual analytics (see Figure 2), the third shows a history, and the fourth contains the user's meta information.

Multimodality

The user interactions on visualizations in a typical analysis process can be categorized into seven intents [19]. For the crucial mapping of modality to intent, we strictly consider the effectiveness and suitability in the visualization domain [2]. But certain modalities are hard to apply. In case of gaze, it is still unclear how to distinguish whether a user

[1] Valletto is a historical term for an assistant of an Italian sovereign.

Figure 2: A: Conversation between Valletto and the user with a live preview of the speech recording. B: Visualization for the current utterances with the ability to recognize gestures for manipulating the visual encoding. C: Recognizable intents and attributes of the data set where the color saturations indicate the system's beliefs (computed probability) as a feedback for the user.

is analyzing a visualization or intends to interact with the visualization. Spatial gestures are hardly applicable, when the user is already carrying a tablet device. Grammel et al. [6] have shown that (inexperienced) users can effortlessly formulate the visualization they need. Therefore, we chose speech and gestures to cover visualization-related intents [14]. Hence, more data-specific intents (generate, include / exclude data attributes, and filter) are executable via speech. In case of adjusting the visual encoding, we expect it is rather easier and faster to use gestures to search for a suitable visual encoding (like using a picture slider) than to formulate it verbally each time. Figure 1 shows the executable gestures on the visualization. As a result, the different modalities are used in complementary ways in the interface.

User Experience

From a UX perspective, it is necessary to achieve an accessible [9] and comprehensible concept for visual analytics. In our multimodal concept, the conversation with the system is essential. Beside a live preview on the user's utterance recognition, we included a dialogue in contrast to recent related work [5, 12, 15]. It helps to increase the transparency [8] on the already asked questions and to obviate redundancy in the analysis process. But the system does not only modify visualizations, it also gives verbal feedback (see Figure 3) in order to strengthen the user's engagement into the data analysis. Moreover, increasing the transparency of computations in the background fosters the UX. It makes the system more reasonable and comprehensible for the user [13]. Therefore, for each data-specific intent and for each data attribute, a visual indicator exists showing the system's utterance interpretation. It reflects the element's probability of being contained in the utterance. Furthermore, to show the user how many other accessible visual encoding variants are computed by the system, we use the familiar concept of picture sliders. This should motivate the user to use swipe gestures on the visualization to see a different one and helps to orientate themselves in the available visualization space.

Implementation

We implemented Valletto in React Native[2] and used the descriptive grammar Vega-lite [11] for visualizing the data. For NLP routines we used spaCy.io[3].

[2] https://facebook.github.io/react-native
[3] https://github.com/explosion/spaCy

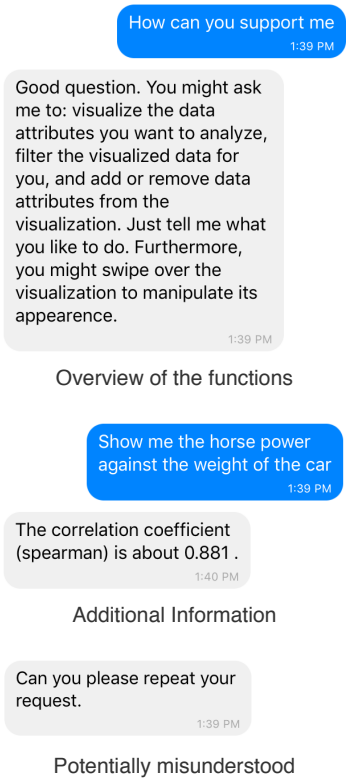


Figure 3: Exemplary answers of the system according to the user's utterance.

Exemplary Scenario

We consider the scenario of a manager analyzing reports while sitting in the back of a car on the way to a meeting. Many managers are very mobile persons, strictly limited in time and usually do not carry a notebook with them. Part of their daily business is the analysis of reports regarding the situation of the company (e.g. sales, productivity, cash flows). Usually, these reports are static documents and created by someone else. They contain visualizations as answers to previously provided questions based on quantitative data. Therefore, unprovided questions are not answered by the document. With Valletto, a manager can directly formulate questions regarding the data to the system instead of sending an email and waiting for the answer.

Evaluation

Initially, we conducted a hybrid expert evaluation composed of a cognitive walkthrough (CW) and a heuristic evaluation (HE). The CW requires a previously constructed list of tasks and the corresponding steps to fulfill them. We designed a sequence of tasks to simulate a typical visual analytics in the discussed scenario. It comprises both functionalities and modalities (e.g., complex visualization tasks), as well as UX-relevant actions (e.g. starting the app and logging in) in order to identify as many issues as possible. First, we introduced the expert to our evaluation procedure, the considered scenario, the designated user group and provided a sheet of paper with the heuristic of Weinschenk and Barker [17]. After each step the expert answered four questions to the interaction's nature [18]. With the CW, we are able to identify interaction issues immediately. After each fulfilled task, we asked the expert to note further recognized problems according to the provided heuristic. This helps us to discover more potential issues. Finally, we performed a severity rating on the findings.

Results and Discussion

We recruited four experts (computer scientist and engineers) working in both industry as well as research for more than 4 years on average. Their working background is in app and Web development with a focus on usability and UX and passing knowledge in multimodality. They discovered 14 potential issues (see Table 1) in total (some with multiple mentions) which are mainly rated as minor usability issues (see Figure 4). In general, the experts welcomed the use of speech in the given scenario, since the potential shaking when mobile has no influence on the interaction.

Nevertheless, the experts mentioned a transparency lack in the available gestures and their corresponding function. If users are not aware of the gestures, they might use speech instead of a gesture to perform a manipulation. This refers to the common issue in post-WIMP interfaces, since there is no graphical indicator for a function (e.g., button). A solution is a guided tour to educate the user how to use the system at the first start. Another – more robust – way is to extend the speech input to understand also rudimentary manipulations of the visual encoding, but aligned with feedback to inform the user that these actions can be easily executed by gestures. However, the experts think that the concept's interaction flow is easy to learn.

Furthermore, the reasoning panel (see Figure 2 section C) was discovered as a potential usability issue, although the experts welcomed the idea of more transparency on the reasoning. Some users might try to touch on the word, because it appears to be touchable, but it is not. A design change from border coloring to text coloring could be a workaround. A more effective solution would be making the panel touchable which can than be used as a visual grammar for generating visualizations e.g. selecting "generate" followed by selecting "origin" and "years". It would also

Heuristics	#Issues
Aesthetic integrity	3
Accommodation	1
Forgiveness	1
Fulfillment	2
Interpretation	2
Responsiveness	4
User support	1

Table 1: Results of the heuristic evaluation.

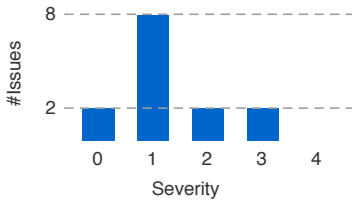


Figure 4: Severity ratings of the discovered usability issues.

keep the concept's UX in scenarios (transitions) in which speech is not easily applicable, e.g., during a walk through the production line or during a meeting.

It seems that slightly redundant functionalities in the modalities could potentially help to prevent usability issues and can foster the user's learning, although each intent has a most effective modality. However, this has to be validated in a separate study.

Limitations and Future Work

As we illustrated, the performance and the UX of multimodal ubiquitous visual analytics concepts are influenced by various factors. In our preliminary hybrid expert evaluation, we were able to identify and rate both interaction and design problems, but could not determine how well the system understands the user's needs. This is dependent on the robustness of the NLP routine and the performance of the speech to text function, which eventually influences the visualization and the UX. Therefore, we are going to investigate our improved concept by comparing it with common methods for data visualization as well as its behaviour in a field. The field study with potential users is a promising approach to eventually condense specific UX factors in the domain of (multimodal) visual analytics. Data about usage behaviour, frequently asked questions, typical visual encoding, and users' acceptance of modalities in the mobile context can be used to build up user models for visual analytics in order to foster the improvement of the UX.

We are planing to elaborate on the advantages of mobile devices for ubiquitous visual analytics. Location services can help speed up the user's analysis in case the data contains location information. Context-aware requests like "show me data related to my current position" can then be easily executed. Such requests can be helpful in a stocktaking scenario in which of a user walks through a warehouse. The tablet's sensors may also recognize mid-air gestures as a potentially additional modality.

Conclusion

Multimodality in visual analytics helps to analyze data of interests in scenarios, where mouse and keyboard are not available. Past work mainly discusses concepts for desktop or other static environments, but data analysis is increasingly becoming ubiquitous. Via Valletto, we provide an idea of a requisite UX design and use constraints of modalities for a ubiquitous visual analytics. In our expert evaluation, we received feedback from which we can conclude that it is beneficial to provide access to some functions via different modalities redundantly in order to prevent usability problems and support the learning process. Redundancy also simplifies scenario transitions, since the choice of modalities may be context-specific. We believe that our work contributes to the discussion of multimodal and post-WIMP user interfaces for ubiquitous visual analytics.

Acknowledgment

Any opinions, findings, and conclusions expressed in this paper do not necessarily reflect the views of the Volkswagen Group.

REFERENCES

1. Jillian Aurisano, Abhinav Kumar, Alberto Gonzalez, Jason Leigh, Barbara DiEugenio, and Andrew Johnson. 2016. Articulate2: Toward a Conversational Interface for Visual Data Exploration. In *IEEE Visualization 2016*.

2. Sriram Karthik Badam, Arjun Srinivasan, Niklas Elmqvist, and John Stasko. 2017. Affordances of Input Modalities for Visual Data Exploration in Immersive Environments. In *2nd Workshop on Immersive Analytics*.

3. Hsinchun Chen, Roger H. L. Chiang, and Veda C. Storey. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q.* 36, 4 (Dec. 2012), 1165–1188.
4. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding Quantified-selfers' Practices in Collecting and Exploring Personal Data. In *Proc. CHI '14*. ACM, 1143–1152.
5. Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proc. UIST '15*. ACM, 489–500.
6. Lars Grammel, Melanie Tory, and Margaret-Anne Storey. 2010. How Information Visualization Novices Construct Visualizations. *IEEE Trans. on Vis. and Comput. Graph.* 16, 6 (Nov 2010), 943–952.
7. Jan-Frederik Kassel and Michael Rohs. 2017. Immersive Navigation in Visualization Spaces through Swipe Gestures and Optimal Attribute Selection. In *2nd Workshop on Immersive Analytics*.
8. Robert J. Moore, Raphael Arar, Guang-Jie Ren, and Margaret H. Szymanski. 2017. Conversational UX Design. In *Proc. CHI EA '17*. ACM, 492–497.
9. Donald A Norman. 2010. *Living with complexity*. MIT press.
10. Jonathan C. Roberts, Panagiotis D. Ritsos, Sriram Karthik Badam, Dominique Brodbeck, Jessie Kennedy, and Niklas Elmqvist. 2014. Visualization beyond the Desktop–the Next Big Thing. *IEEE Comput. Graph. Appl.* 34, 6 (Nov 2014), 26–34.
11. Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. on Vis. and Comput. Graph.* 23, 1 (Jan. 2017), 341–350.
12. Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proc. UIST '16*. ACM, 365–377.
13. Rashmi Sinha and Kirsten Swearingen. 2002. The Role of Transparency in Recommender Systems. In *Proc. CHI EA '02*. ACM, 830–831.
14. Arjun Srinivasan and John Stasko. 2017. Natural Language Interfaces for Data Analysis with Visualization: Considering What Has and Could Be Asked. In *EuroVis 2017 - Short Papers*. Eurographics Association.
15. Arjun Srinivasan and John Stasko. 2018. Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks. *IEEE Trans. on Vis. and Comput. Graph.* 24, 1 (Jan 2018), 511–521.
16. Yiwen Sun, Jason Leigh, Andrew Johnson, and Dennis Chau. 2009. Articulate: a conversational interface for visual analytics. In *IEEE VAST '09*. IEEE, 233–234.
17. Susan Weinschenk and Dean T. Barker. 2000. *Designing Effective Speech Interfaces*. John Wiley & Sons, Inc., New York, NY, USA.
18. Cathleen Wharton, John Rieman, Clayton Lewis, and Peter Polson. 1994. Usability Inspection Methods. John Wiley & Sons, Inc., New York, NY, USA, Chapter The Cognitive Walkthrough Method: A Practitioner's Guide, 105–140.
19. Ji Soo Yi, Youn ah Kang, John Stasko, and Julie Jacko. 2007. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Trans. on Vis. and Comput. Graph.* 13, 6 (Nov. 2007), 1224–1231.