Visual Data Storytelling: A Case Study of Turning Big Data into Chinese Painting

Yanru Lyu[1(✉)], Tuck Fai Cheng[1,2], and Rungtai Lin[1]

[1] Graduate School of Creative Industry Design,
National Taiwan University of Arts, New Taipei City 22058, Taiwan
lyuyanru@gmail.com, rtlin@mail.ntua.edu.tw
[2] Graduate School of Applied Cosmetology, HungKuang University,
Taichung City 43302, Taiwan
Kennethl831@hk.edu.tw

Abstract. Visual data storytelling is a relatively new terminology that has emerged over the last decade. The use of information technology in data visualizing is becoming more common and accessible to users. The artist apply a variety of multimedia, symbols and metaphors to independently create a visual form that expresses data and communicates with real life. The visual data are the media that provide powerful and essential means of communication. If "visual data storytelling" is viewed as a process of communication, it follows naturally that how the artist's performances are conceived, developed, delivered and received, and how the audience is attracted, accurately understanding the data and affected by the data visualization are worth exploring. Therefore, this study is intended to propose a framework focusing on how the conception of the artist affects the creation process and how the creation process is understood by the audience. The artist's creation activities were analyzed through the framework of four steps using a case study of turning "big data" into "Chinese painting." The results showed that the approach could be applied to understanding data visualization and provides artists with an idea of how to concentrate their efforts at the creation stage, the easier to communicate with their audience. In addition, the research framework seems to provide a better way to explore the understanding of how data transforms into art forms, which is clearly worthy of further study.

Keywords: Big data · Data visualization · Storytelling · DIKW model

1 Introduction

In this data-driven era, there are a variety of purposes for data acquisition and application. The information made up of data can not only help people make more intelligent decisions, but also examine their view of the world from a more objective perspective [1]. Transforming data into a form that relies on the human visual system to perceive the embedded information, the resulting visual effect is as simple as discovering the underlying laws in movies [2]. On the one hand, through the research of visualization in computer science, data and facts can be presented rationally and objectively, thus making the data readable. On the other hand, there are parallel

P.-L. P. Rau (Ed.): HCII 2020, LNCS 12193, pp. 526–535, 2020.
https://doi.org/10.1007/978-3-030-49913-6_43

discourses regarding the artistic visualization aesthetics, which mainly focus on the exploration of human emotions and the transmission of inner views, but unfortunately rarely involves [1, 3–5]. As Norman [6, 7] argues in a successful design, visualization has evolved into as a functional art, which not only focuses on usability goals, but also needs to provide memorable, interesting, enjoyable, and engaging experiences. Therefore, more attention needs to be paid to artistic visualization in order to tell the audience something except what it is behind data.

Artistic visualization, as a type of visualization centered on emotional experience, refers to the artist's use of beauty in data by artist to arouse the perception of the audience. Different from the visualization just for usability goal, the goal of artistic visualization is usually to communicate a concern, rather than presenting data. The data is used as material and transformed into a form of beauty that makes relevant patterns visible to demonstrate the authenticity of the concerns [3, 4]. The information anxiety mentioned by Wurman is caused by the gap between what we understand and what we think we should understand [8], which can be bridged by the communication between the outer world and the audience in a meaning way [3]. Consequently, how to communicate is the research focus of artistic visualization.

As one of the most effective means of communication, stories have long served as a medium for conveying information, cultural values, and experiences. In this increasingly computerized world, with the development of technology and culture, there more complex narrative ways for the transmission of information [2, 8]. Visualization is not only about the exploration and analysis of data, but also about eliciting profound emotional and/or intellectual responses [1, 9]. There are countless real-life stories behind the data, which are displayed in ways that depend on what the audience wants to see or hope the audience to see. Therefore, visual data storytelling absorbs the ideas from both the artist/designer and the audience [1, 10].

If "visual data storytelling" is viewed as a process of communication, the following questions are worth exploring. How are the artist's performances conceived, developed, delivered and received? How are the audiences attracted, accurately understood the data, and affected by the visualization? Therefore, this study proposes a framework to explore how the conception of the artist affect the creation process and how the creation process is understood by the audience.

2 Theoretical Background

2.1 DIKW Model

The data-information-knowledge-wisdom (hereinafter referred to as DIKW) model is a common method to explain the human understanding in the perceptual and cognitive space [11]. Nathan Shedroff [8] gives an overview of understanding, that is, a continuum from data to wisdom. Ackoff, R. L. [12] defines data as symbols that has no value until they are processed into a useable form in a given context. Information consists of processed data with more usefulness that can provide answers to "who", "what", "where" and "when" questions. Knowledge is the application of data and information to provide answers to "how" questions. Understanding is the appreciation

of "why". Wisdom is the evaluation of understanding. The first four categories are related to the past, while the fifth category deals with the future. Zeleny [13] describes the components of the DIKW model for different purposes, including know-nothing (data), know-what (information), know-how (knowledge), know-why (wisdom), and know-for-sure (enlightenment) beyond wisdom. By integrating several models known as DIKW Hierarchies, Cairo [3] represents the gap that Wurman, R. S [8] describes between the data and the knowledge, and explains the process from reality to the human brain. The unstructured information from outer world can be encoded as data, which is shaped by the communicator. When the audience has a deep understanding of the acquired knowledge, relevant patterns can produce knowledge and even reach the realm of knowledge. Based on the previous studies combined with the DIKW model, the cognitive process from data to wisdom is shown in Fig. 1.

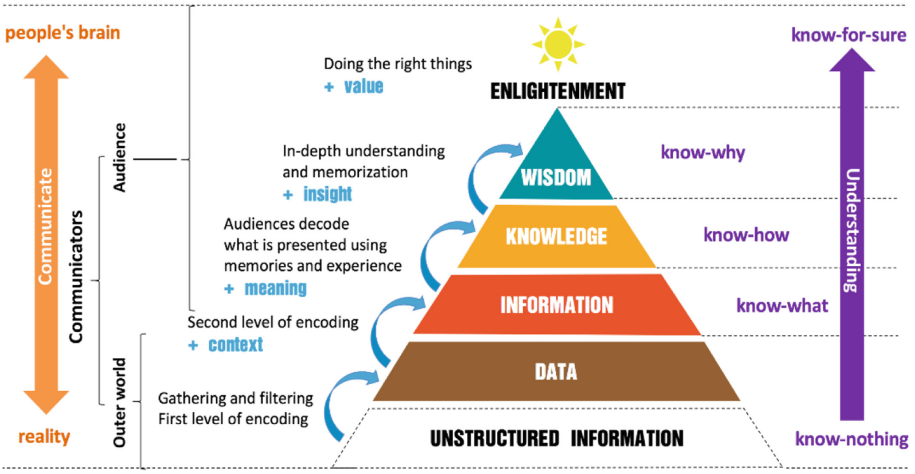

Fig. 1. Congnition from data to wisdom

2.2 Framework of Communication Research

Since the audience cannot gain personal insights directly from abstract data, the communication involving the artist as the communicator encoding and audience decoding is an important part of the discussion. There are six constitutive factors in Jakobson's [14] communication model, which are addresser, addressee, context, message, code, and contact. This model is used to analyze an act of communication in visualization. The artist (addresser) sends a message to the audience (addressee) through his/her visualization work. To be operative, the work (message) requires a story (context) and is mastered by the audience (addressee). The meaning of work must be based on his/her message and on the shared meaning system (code) that makes up the work. Finally, physical channels and psychological connections are established between the artist and the audience (contact), enabling both of them to enter and maintain communication. Of these six factors, each factor determines the corresponding different functions,

respectively emotive, conative, referential, poetic, phatic, and metalingual, which will be described in detail below. The emotive function focuses on the attitude or emotion of the artist. The conative function is a kind of orientation toward the audience. In the referential function, the work is related to the real data. The poetic function is the aesthetic expression of data art. For code and contact, the metalingual function is to confirm the code system, and the phatic function is the medium for communication.

Lin et al. [15–18] put forward a framework for communication research that contains three levels of problem, namely technical, semantic, and effectiveness. Firstly, the technical level requires the audience to receive information through his/her senses, that is, how accurately the artist can convey the information through his/her visualization work. Second, the semantic level requires the audience to be able to understand the meaning of the message without misinterpreting, misunderstanding, or even not understanding at all, namely the degree of accuracy with which the audience understands the original meaning of the message. Third, the level of effectiveness involves the audience taking corrective action in accordance with its original meaning, that is, how effectively affect conducts in the way expected. Based on the previous studies [14–20], a research framework combining communication theory with communication and DIKW models is proposed to explore the issue of turning data into story (see Fig. 2).

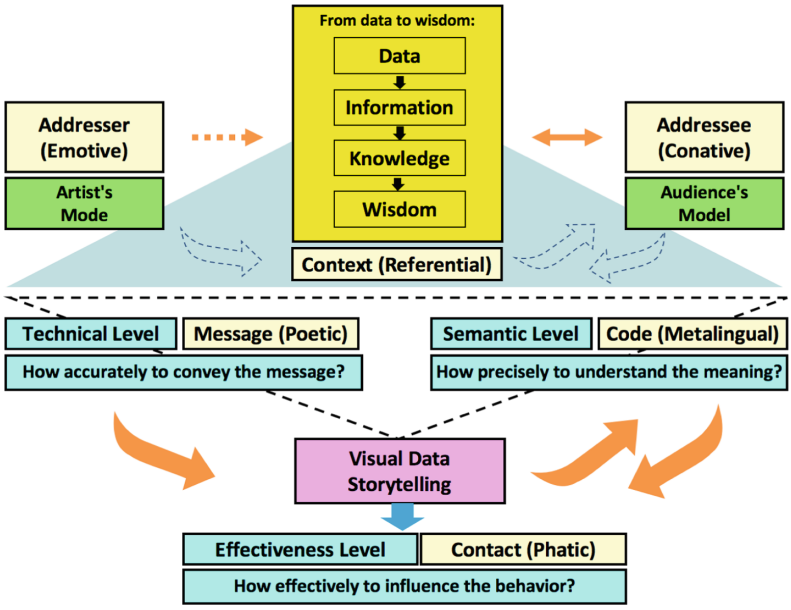

Fig. 2. Research framework for turning data into story

Visualization involves complex issues as interdisciplinary processes. On the basis of the above discussion, the framework in this study can be used for a continuous search for a deeper understanding of the process of visual data storytelling.

3 Methodology

3.1 Visual Data Storytelling Model

According to the studies of Lin [15] and the previous research framework (see Fig. 1), the visual data storytelling model includes three main stages, which are data processing, coding model and visualization works. The model focuses on how to extract the semantic features from data and then then transform these features into art forms, which can be divided into three stages: analyzation, translation and implementation, as shown in the upper part of Fig. 3.

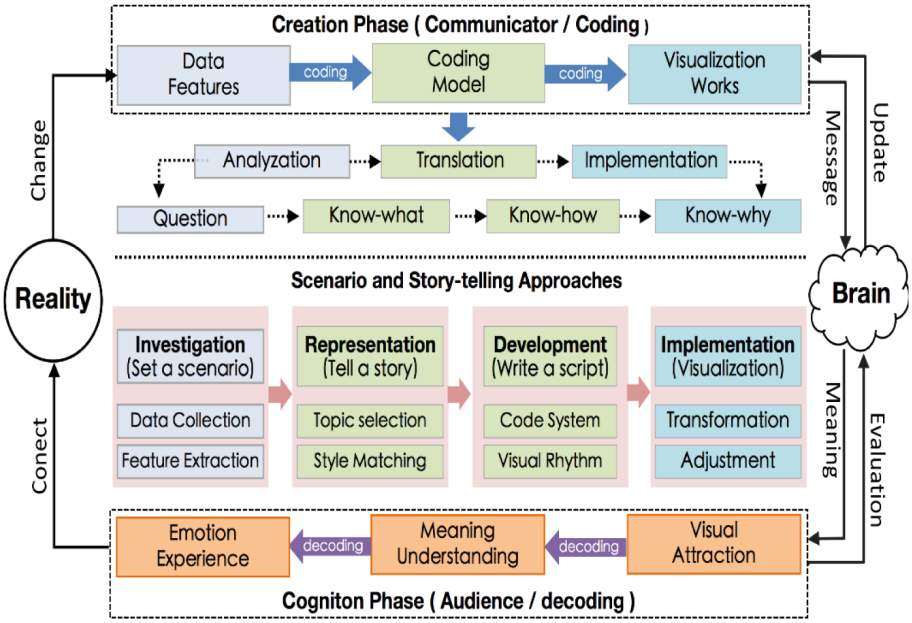

Fig. 3. Visual Data storytelling model

3.2 The Three Stages of the Research Method

As shown in Fig. 3, there are three stages of the visual data storytelling model, including the analyzation stage, translation stage and implementation stage. A further description is given below.

Analyzation Stage: This stage looks for hidden problems in the data. By observation and measurement, data are recorded. The data features are identified during processing, including the outer level of visual form, the mid level of meaning, and the inner level of emotion. Through this process, artists can use the methods and tools of big data analysis to obtain contextual information from useful data.

Translation Stage: The purpose of translation is to explain "what is the problem" and "how to cause the problem". The information from the data is translated into knowledge by associating it with the memory and experience of the audience. At this stage, the artist can relate this knowledge to the problems in modern society, and create an interaction between the reality and the audience's brain.

Implementation Stage: From this stage to the completion from data to wisdom. The implementation stage involves the expression of knowledge related to the data features, the artist's understanding of the meaning of information, his/her aesthetic sensibility, and his/her flexibility to adapt to various media. In this process, the artist's in-depth understanding of knowledge can add value to wisdom. This knowledge is combined with the artist's sense of visual communication to employ these three levels of data features in visualization.

3.3 The Four Steps of Visual Data Storytelling Process

Based on the creation model and the previous studies [10, 15, 21], the big data into story is used in scenario and story-telling approaches is used in scenario and story-telling approaches. In the practical visualization process, four steps are used to transfer data into story, namely, investigation (setting a scenario), representation (telling a story), development (writing a script), and implementation (visualization), as shown in the middle part of Fig. 3. The four steps of this process are further described as follows:

Investigation/Setting a Scenario: Data recorded by observation and measurement is the core of any visualization, which can be found in a lot of places, including experts in the related fields, a variety of online applications, or collections by people [1]. It should be explored after formatting and then used to identify data characteristics, the source of the problem, with a critical eye that includes technical, semantic and effectiveness levels. Through this process, the artist can obtain contextual information from the data.

Representation/Tell a Story: Through the structure of the story, this step aims to inform the audience what the question is about the data. Therefore, based on the previous scenario, the topic can be selected and narrowed down. In order to establish a connection between the story topic and the audience, an externally recognizable style should be used on the topic to make it easier for the audience to understand.

Development/Write a Script: As a step in developing concepts and prototypes, this step focuses on code system, which explores the transformation mechanism from data to visual form. During this step, the scenario and story may require modification to highlight the important point. In addition, this step provides a way to illustrate the story logic with visual rhythm.

Implementation/Visualization: The goal of this final step is to evoke an emotional response of the audience from data to wisdom through a comprehensive visual experience. Therefore, all data features should be listed in a matrix table to help artist examine the transformation effect. Moreover, the artist needs to evaluate the recognition of visualization from three levels, and then update the work based on the result of this evaluation.

Finally, it is necessary to always consider the audience and purpose of the visualization works. In traditional data visualization studies, there have been assessments of usability goals, such as effectiveness and efficiency. Alternatively, There are few approaches to experience goals in the visualization literature [22]. As shown in the lower part of Fig. 3, the process of audience cognition includes three phases of visual attraction, meaning understanding and emotion experience, which are not only the reference for the creation process, but also the three levels matrix of evaluation.

4 Case Study: Turning Big Data into Chinese Painting

Based on the previous discussion, the framework of visual data storytelling is applied to a case study that transforms big data into art forms and tell a story. The work includes more than 100,000 ecological data from 34 cities, including the capital cities of mainland China and Hong Kong, Macao and Taipei, between 2013 and 2019. Based on the features extracted from the data, the coding system on the artistic style of Chinese landscape painting is constructed, with the purpose of making use of the visually attractive style of landscape painting to make the audience pay attention to and reflect on natural ecological issues. In a practical process, four steps are used to transfer ecological data to Chinese paintings (see Fig. 4).

First of all, environmental issues have been the subject of extensive discussion for a long time, and we try to identify some specific problems. Based on the four steps of the visual data storytelling process, the scenario is set through the analysis for the four sets of ecological data, air quality, water quality, temperature and green coverage ratio. By using web crawlers to collect aerial data and manually collect other data from government reports, the statistical analysis method is used to explore the relationship between data, time and cities. The results of the analysis help to find interesting questions. For example, there are significant differences in the air quality index (AQI) among the cities in the south and north, while the cities in the same river basin have similar data trends. According to this data feature, in-depth exploration can be focused on the group of basin-based cities.

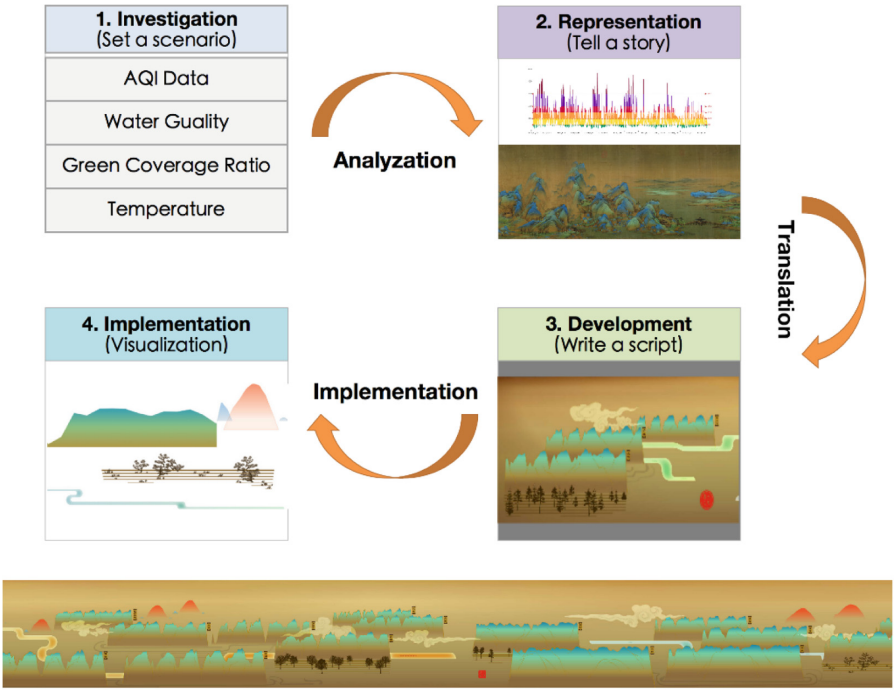

Fig. 4. The process of turning big data into Chinese painting

The four steps and the visualization creation process are described as follows.

Second, in the representation step of telling a story, the topic should be selected according to the data characteristics and what want the audience to know. In this case, although the data set shows that those northern cities are in a harsh environment, there has been a trend of improvement in the recent two years. Therefore, by choosing a theme full of positive hope, this visualization is expected to have an impact on the audience behavior in the future. In order to match the scenario and topic of looking forward to the environment with beautiful mountains and rivers, Chinese landscape painting is used as the main reference style.

The third step is to develop the story by writing a script. In order to communicate effectively, it is necessary to use universal encoding to transform abstract information into concrete perception, so that the audience can easily decode with their memories and experience. For instance, the perception of mountains is usually that the higher the altitude, the greener the color, and the better the air. Consequently, the higher the AQI value, the worse the air quality, which is encoded as a lower altitude and the yellowish color location, and vice versa. Similarly, following encoding rules, other data can be converted into shapes and colors. On the other hand, just like a movie story, visual rhythm can also highlight points and tell vivid stories. In order to make storytelling both poetic and pictorial, the screen slowly scrolls from left to right with background music before triggering the interaction. More label information is displayed rhythmically with interactive events that accompany mouse hover or click. In addition, the

story will be told what they want to know or want them to know according to the needs of the audience.

As the last step, the visualization combines aesthetic sensitivity with the process of dealing with data feature, the context of information and meaning of knowledge. No matter where the work is presented in different media in any field, the cognitive processes involved in the visual data storytelling model can serve as the basis for creation and evaluation for visualization. This work can be evaluated at the three levels of what to be seen, what to be understood, and what to be touched, and ultimately identify the important factors that influence audience perception and behavior.

According to visual data storytelling model and the four steps of the scenario and story-telling approaches, the visualization of natural ecological data into a long scroll of Chinese painting is realized from the three stages of data analyzation of data feature, translation based on code system, and implementation. The purpose is to complete the audience's three cognitive levels, namely form perception, semantic cognition and inner feeling. At first, the audience is attracted by the Chinese landscape painting. Then, through careful observation and interaction, the audience understands the theme of the overall picture, as well as the data types and changing trends of each visual symbol. Finally, the integration of cognition, memory, and experience of the visualization touches the audience's response to ecological phenomena, thereby affecting the behavior in natural ecology. The artist's creation process and the audience's cognition process constitute a complete process from "big data" to "Chinese painting", thereby stimulating the audience to use their behavior to draw more beautiful landscape paintings through the artistic visual works.

5 Conclusions and Suggestion

This study proposes a framework of visual data storytelling based on the DIKW model and communication theory, focusing on how the artist affects the creation process from data to visualization, and how the artwork is understood by audience from meaning to insight. Through a case study of transforming "big data" into "Chinese painting", the artist's creation process is analyzed from the four steps of the scenario and story-telling approaches. The results indicate that the approach could be applied to understand data visualization and provide artists with an idea of how to transform the data into art forms and how to tell stories during the creation stage, making it easier for them to communicate with the audience.

The research framework seems to offer a better way to explore the understanding of how "big data" can be transformed into art forms, which clearly deserves further study. Furthermore, detailed creation processes need to be developed in the future to provide artists with specified storytelling procedures for visualization. In addition, it suggests that the audience's cognitive experience be taken into account in the assessment, which is identified as three levels, namely technical, semantic and effectiveness levels.

Acknowledgments. The authors would like to express their sincere thanks to Professor Lin for his guidance in the research. In addition, special thanks to Long et al. for participating in the creation process of the study case in this paper.

References

1. Nathan, Y.: Visualize This: The FlowingData Guide to Design, Visualization, and Statistics. Wiley, Indianapolis (2011)
2. Gershon, N., Page, W.: What storytelling can do for information visualization. Commun. ACM **44**(8), 31–37 (2001)
3. Cairo, A.: The functional Art: An Introduction to Information Graphics and Visualization. New Riders Press, Berkeley (2012)
4. Kosara, R.: Visualization criticism-the missing link between information visualization and art. In: 11th International Conference Information Visualization, pp. 631–636. IEEE, Zurich (2007)
5. Judelman, G.: Aesthetics and inspiration for visualization design: bridging the gap between art and science. In: Paper presented at the proceedings. eighth international conference on information visualisation (2004)
6. Norman, D.A.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books, New York (2004)
7. Norman, D.A.: The Design of Everyday Things: Revised and, Expanded edn. Basic Books, New York (2013)
8. Wurman, R.S.: Information Anxiety 2. Que, Indianapolis (2000)
9. Kosara, R., Mackinlay, J.: Storytelling: the next step for visualization. Computer **46**(5), 44–50 (2013)
10. Edward, S., Heer, J.: Narrative visualization: telling stories with data. IEEE Trans. Visual Comput. Graphics **16**(6), 1139–1148 (2010)
11. Chen, M., et al.: Data, information, and knowledge in visualization. IEEE Comput. Graphics Appl. **29**(1), 12–19 (2008)
12. Ackoff, R.L.: From data to wisdom. J. Appl. Syst. Anal. **16**(1), 3–9 (1989)
13. Zeleny, M.: Human systems management: Integrating knowledge, management and systems. World Scientific (2005)
14. Jakobson, R.: Language in Literature. The Belknap Press of Harvard University Press, Cambridge (1987)
15. Lin, R.: Transforming Taiwan aboriginal cultural features into modern product design: a case study of a cross-cultural product design model. Int. J. Des. **1**(2), 45–53 (2007)
16. Lin, R., Lin, P.-H., Shiao, W.-S., Lin, S.-H.: Cultural aspect of interaction design beyond human-computer interaction. In: Aykin, N. (ed.) IDGD 2009. LNCS, vol. 5623, pp. 49–58. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02767-3_6
17. Lin, C.-L., Chen, J.-L., Chen, S.-J., Lin, R.: The cognition of turning poetry into painting. J. US-China Educ. Rev. **5**(8), 471–487 (2015)
18. Fiske, J.: Introduction to Communication Studies. Routledge, London (1990)
19. Barthes, R.: Elements of Semiology. Hill & Wang, New York (1968)
20. Silverman, K.: The Subject of Semiotics. Oxford University Press, New York (1983)
21. Gao, Y., Wu, J., Lee, S., Lin, R.: Communication between artist and audience: a case study of creation journey. In: Rau, P.-L.P. (ed.) HCII 2019. LNCS, vol. 11577, pp. 33–44. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22580-3_3
22. Saket, B., Endert, A., Stasko, J.: Beyond usability and performance: a review of user experience-focused evaluations in visualization. In: Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization, pp. 133–142 (2016)