

The Presenter Experience of Canvas Presentations

Leonhard Lichtschlag, Philipp Wacker, Martina Ziefle, and Jan Borchers

RWTH Aachen University, 52074 Aachen, Germany

{lichtschlag,wacker,ziefle,borchers}@cs.rwth-aachen.de

Abstract. Most presentations are given with supporting visuals and driven by specialized presentation software. Today, this software either follows the classic slideware metaphor, presenting a series of discrete screens—or it implements the more recent canvas presentation metaphor, using a zoomable free-form canvas to arrange information. Both paradigms were previously evaluated with presentation authors and audiences. In this paper, we extend our understanding to how they impact the presenter herself during delivery of a talk. In a lab study participants gave presentations with slideware and canvas tools, and we measured their emotional state through self-reporting. We find that a recommendation for a tool depends on the experience of the presenter or their spatial ability.

Keywords. Canvas presentations; slideware; zoomable user interfaces; lab study

1 Introduction

Visual presentation support is prevalent in talks in research, industry, education, and many other areas. Most software, e.g., Microsoft’s PowerPoint or Apple’s Keynote, employs the slide metaphor, which originates from the technical restrictions of physical slides used on overhead projectors. Yet today, presentation visuals are usually displayed using a computer connected to a video projector. This removes the necessity to show a series of slides, one at a time, and this format has been criticized repeatedly for the limitations it imposes on authors and presenters [16,21]. A recent alternative to the slide format are canvas presentations, which dismiss the slide metaphor in part or entirely. Instead, they place either the slides [7] or individual elements [11,17] on an infinite canvas. Authors then define viewports and transition paths across the canvas to define the presentation sequence, or present ad-hoc without a planned path.

It was previously studied how authors deal with the two paradigms when preparing for a talk [11,13], and how audiences perceive talks given in the two formats [12], but the effect on the presenter herself has not been investigated, yet. The canvas format should be especially helpful for navigations that deviate from the planned presentation delivery, e.g., in response to a question [2,7]. The presenter can quickly pan-and-zoom, thereby creating impromptu overviews and showing the macrostructure of the talk to the audience. However, we also hypothesize that the free format may be too demanding for a presenter who is preoccupied on talking. In this paper, we present a lab study with presenters who gave short talks in each format to investigate these issues. We measure the emotional state of the presenters during a presentation delivery in which several kinds of interruptions occur.

2 Related Work

Several tools have adapted the zoomable user interfaces paradigm [1,2] specifically for presentation support [7,11,17]. CounterPoint [7] broke new ground by positioning PowerPoint slides inside a zoomable user interface. It lets authors place slides at varying distances from a virtual camera and create a spatial layout of slides. Fly [11] and Prezi [17] have no notion of slides at all and present content elements (text, figures, etc.) directly on a canvas. Without the limitations of the slide frame, authors adapt their approach to presenting content less linearly and incorporate good presentation behaviors such as overviews [11,13]. Canvas presentations perform on par with slideware with regards to audiences' content recall and macrostructure understanding [12]. No studies investigated how presenters interact with canvas presentations.

Defining emotion is very difficult, and there are numerous attempts in the literature. A popular approach is to characterize emotion using a component model where expressions, bodily reactions, and the subjective experience have “long-standing status as modalities of emotion” [20]. There are different ways to combine these components; here we use a dimensional approach (e.g., [19]) with valence and arousal as the main components [20]. Valence dimension contrasts pleasure and displeasure, while the arousal describes intensity. The term *feelings* is defined in the component model as the subjective experience of an emotion and can occur separate from bodily reactions (e.g., [6]). As many people are anxious about public speaking, it is the presenter's feelings that we are interested in. Scherer [20] writes that these feelings can only be accessed through a person's self-reporting.

The *self-assessment manikin* (SAM) [4] (figure 1, right) is one way to elicit ratings from subjects. Each row represents change along one dimension of emotion as pictorial depictions of a person. In our study we used SAM without the dominance domain. While this method requires fewer ratings from a person, the two dimensional rating does provide little insight into the aspects that produced the emotion [20]. A more detailed technique is the *semantic differential* (SD) which measures the meaning of words [15] (figure 1). A semantic differential consists of point rating scales with bipolar word pairs on either end of the scale (e.g., good–bad). It has been used to measure attitude and feelings, for example in a classroom scenario [5]. A person rates a concept by indicating for each word pair where she places the concept between the limits.

3 Study Design

We conducted a lab study with every tester giving two presentations for about 7 min each (figure 1, left). One presentation was given with Apple's Keynote [10], representing slideware, and one with Prezi [17], representing the canvas condition (During presentation delivery, the differences between the different canvas tools are negligible and we do not expect our choice of software to impact the study. Similarly for slideware). In this format, we were able to have comparable talks for all users and to include interesting tasks in the presentations as well as simulated technological problems. In each presentation, we first asked the tester to give their presentation normally

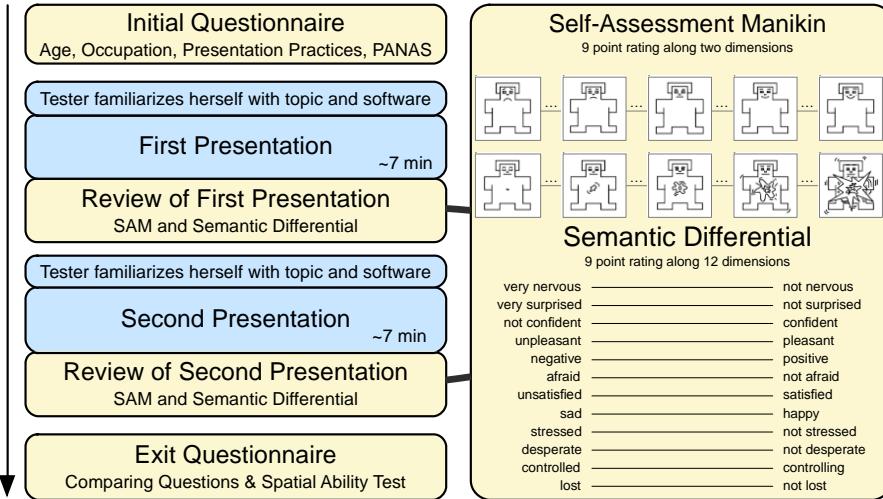


Fig. 1. *Left:* each participant presents two talks, after each she rates her feelings with the Self-Assessment Manikin (SAM) and the Semantic Differential (SD) described on the right. *Top Right:* One row changes from an unhappy person to a happy person (the valence dimension), one row from a person who has the eyes closed to a person experiencing intense feelings (arousal dimension). *Bottom Right:* SD dimensions used to quantify more precise feelings.

by *stepping forward* through the presentation. Then we interrupted them and asked to *move to a well defined position* in the presentation (e.g., going back to a specific previous slide in response to a question) and then to *skip forward* towards the end (e.g., due to time constraints). Finally, we asked to *search for a loosely defined position* (e.g., the presenter has to find the slide to a question). One of the two talks also included simulated errors: a *misinterpreted input* (simulated by a step backwards on a forward command), and an *unresponsive program* (simulated by no action on the first command). We counterbalanced the order of conditions.

Our setup excludes many confounding variables (e.g., varying documents, audiences, stakes, presentation occasions, length, etc.) to build a baseline understanding in the lab with high control. As such, the results of this setup are limited until a field study can investigate their generalizability (cf. the approach taken by [11,13]).

In both talks, testers presented the 2014 Soccer World Championship, a topic presented heavily in the media at the time, so that we could expect our testers to have at least some prior knowledge. The topic was also a good fit for a spatial layout by placing players on their actual position on the pitch. Therefore, the topic could be presented spatially in the canvas condition in a manner that was approachable to our testers (cf. [12] discussing of the problem of comparable documents between presentation software). Additionally, we introduced the participants to the documents and the tools, then they used the software to familiarize themselves with the matter. The documents were created by the authors; while is common to present foreign slides, this remains a limitation of the lab setup. All presentations were driven by an iPad carried

by the presenter. Using the same input device for all presentations lowered the possible differences in interacting with the different software. Both software animated transitions, canvas with the inherent flyover, slideware with a slide-in from right to left between each slide. During the presentation delivery, the input modalities are step forward, step backward, as well as zoom and pan. All materials and presentation documents to replicate the study are available at <http://hci.rwth-aachen.de/fly>.

During the presentation, the moderator acted as an interested audience member that smiles and acknowledges the information given. Two cameras recorded each talk, this increased the stakes for our testers as playacting the presentation in this manner gives similar results to the “real” situation [8,9]. Secondly, the cameras allowed us to watch the recording afterwards together with the presenter. Although memory of feelings lessens over time [18], the participant could relive the situation and assess her feelings, and we avoided interrupting the presentation. We used the self-assessment manikin with a nine-point rating scale [4] (figure 1) to measure the valence and arousal of the participant in the task situations. We also used the semantic differential [15] to ask her to rate twelve feelings on a nine-point rating scale to measure feedback on specific feelings. To find out which dimensions to use, seven weeks before our study, we had asked in an online survey which feelings presenters had experienced themselves or observed in others. We combined the reported emotions with directions from literature [7] and produced the dimensions in figure 1. Accounting for the possibility that underlying moods affect the feelings of the participants, we used the PANAS test, a reliable and valid method to measure mood over various periods of time [22]. Finally, in the exit questionnaire, we asked for informal feedback on the experience with the software, differences to their regular presentations, and how their feelings in the study related to feelings in real presentations. To see whether spatial ability influence the experience of presenting, we measured the participants spatial ability using a paper folding test [3]. We formulated these hypotheses:

- H1:** Feelings in canvas presentations are rated differently than feelings in slide presentations.
- H2:** Presentations with technical difficulties are rated differently than presentations without technical problems.
- H3:** Order of presentation and presence of errors does not influence ratings.
- H4:** Participants experience the same feelings during the study compared to a real world presentations.

4 Evaluation

We recruited 21 participants for the study with varying proficiency in presentation skills in general and technological skills in particular. The participants were 8 teachers, 7 students, 6 other professions, none familiar with the lab, aged 17–66 (mean=37.09, SD=16.02). To quantify the presentation experience we calculated a presentation age by subtracting the age at which a participant had given their first presentation from their current age. This *presentation age* (PAge) had a mean of 18.33 years and standard deviation of 11.73. The *technological expertise* (TE) was assessed

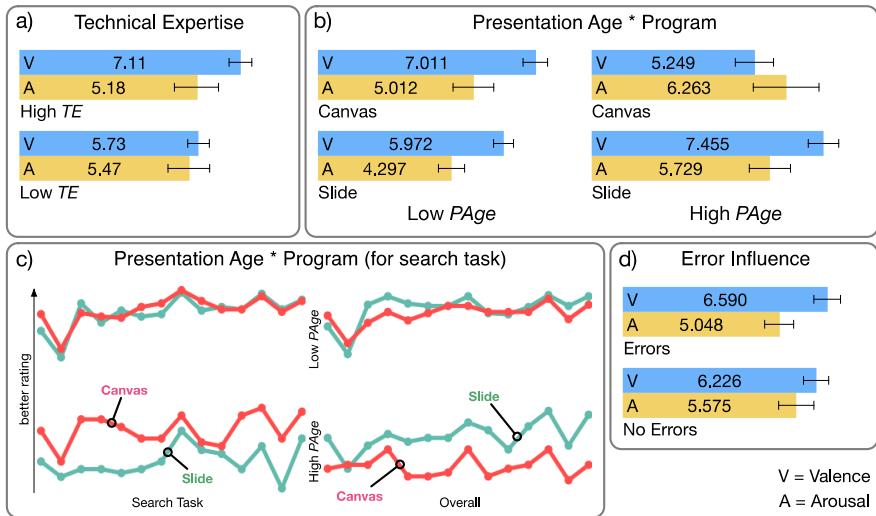


Fig. 2. a) SAM ratings for technical expertise show significant difference. b) Presentation experience influences emotional response. c) SD ratings show experienced presenters rated the search task differently from the trend. d) Arousal rating is influenced by presence of errors.

by calculating the mean between how often the participant uses canvas tools and slideware respectively (rated on a five-point scale where higher values mean more often). *TE* had a mean of 1.33 and a standard deviation of 0.56. We also asked participants how much they *liked* to present (*L*) on a five-point rating scale (1–5, 1=most enjoyment, mean=2.19, SD=0.93). Other gathered characteristics were *spatial ability* (*SA*, 0–20, number of correct solutions in the paper folding test; mean=12.76, SD=4.77) and mood (PANAS test, separated for *positive affect* (*PA*) (mean=31.14, SD=4.95) and *negative affect* (*NA*) (mean=12.57, SD=2.34)). A correlation of *PAge*, *SA*, *TE* and *L* showed that presentation age and spatial ability had a significant negative correlation. Hence we could not analyze them separately, and when we report on *PAge* below, *SA* can also be an explaining factor. For the evaluation we categorized the *PAge*, *TE* and *L* each into two groups.

As for the hypothesis of program influence (**H1**) we conducted two repeated-measures MANCOVAs with the valence/arousal ratings (SAM) and the semantic differential (SD) ratings as dependent variables respectively. *PAge*, *TE* and *L* were taken as between-subjects factors and positive affect (*PA*) and negative affect (*NA*) as covariates. For the valence/arousal ratings we found no main effect of the delivery method ($F(2,10)=3.00$, *ns*) but a significant between-subjects effect of *TE* ($F(2,10)=6.78$, $p<.05$) and a significant interaction effect of Program**PAge* ($F(2,10)=7.82$, $p<.01$).

Between-subjects, *TE* had a significant effect on the valence ratings ($F(1,11)=11.24$, $p<.01$) with more *TE* leading to higher ratings (figure 2a). The interaction effect of Program**PAge* was significant for the valence ratings ($F(1,11)=14.66$, $p<.01$). An analysis of the means showed that less experienced presenters gave higher valence ratings for the canvas presentation (5.97 to 7.01) while more experienced

presenters gave higher ratings for the slideware presentation (5.25 to 7.46) (figure 2b). The results from the analysis of SD ratings indicated a main effect of the delivery method ($F(11,1)=825.93, p<.05$), an interaction effect of Program**NA* ($F(11,1)=1481.85, p<.05$), an interaction effect of Program**PAge* ($F(11,1)=2249.68, p<.05$) and an interaction effect of Program**TE* ($F(11,1)=349.55, p<.05$). While the individual SD dimensions did not differ significantly between the programs, the overall trend was that the slideware presentation received more positive emotional response. The interaction effect of Program**PAge* was significant for *pleasantness* ($F(1,11)=8.4, p<.05$), *positivity* ($F(1,11)=10.11, p<.01$), *afraid* ($F(1,11)=14.67, p<.01$), *satisfaction* ($F(1,11)=20.38, p<.01$), *stress* ($F(1,11)=13.91, p<.01$), *desperation* ($F(1,11)=5.81, p<.05$), *controlled* ($F(1,11)=6.01, p<.05$), and *lost* ratings ($F(1,11)=16.83, p<.01$). More experienced presenters gave positive ratings for slideware on all these dimensions, while less experienced presenters showed only minor differences. The interaction effect of Program**TE* was significant for *surprise* ($F(1,11)=9.99, p<.01$), *unsatisfied* ($F(1,11)=9.21, p<.05$), and *lost* ratings ($F(1,11)=5.47, p<.05$). Presenters who had less *TE* gave higher ratings for slideware on all these dimensions while presenters with more *TE* showed only minor differences. In conclusion, we accept H1.

Exploring the data, we noted that the ratings for the search for a *loosely defined position* task showed a flipped behavior. An interaction effect Program**PAge* occurred once again and valence values were significantly different for this task ($F(1,11)=9.31, p<.05$). Further analysis showed a great difference between slideware and canvas presentations for experienced presenters, with canvas presentations having a better rating, while no difference was found for less experienced presenters (figure 2c).

To explore the error hypothesis (H2) we conducted two repeated-measures MANCOVAs with the valence/arousal ratings and the semantic differential (SD) ratings as dependent variables respectively. *PAge*, *TE* and *L* were taken as between-subjects factors and positive affect (*PA*) and negative affect (*NA*) as covariates. The analysis of valence/arousal ratings indicated a main effect of the error condition ($F(2,10)=5.55, p<.05$), an interaction effect of Error**PA* ($F(2,10)=6.62, p<.05$) and an interaction effect of Error**TE* ($F(2,10)=5.88, p<.05$). Arousal ratings were significantly different between the error conditions ($F(1,11)=7.90, p<.05$). Examination of means showed that arousal was rated higher in the no-error condition (figure 2d). The interaction effect of Error**PA* was significant for the arousal ratings ($F(1,11)=8.19, p<.05$), and the plot indicated that while the positive affect rating had no effect on the arousal rating in the no-error condition, it had a positive effect on the arousal ratings in the error condition. The interaction effect of Error**TE* was significant neither for valence nor for arousal ratings. The results from the analysis of SD ratings indicated between-subjects effects of *NA* on stress ratings and of *TE* on nervousness, pleasantness and positive-negative ratings. Further analysis showed that higher *NA* ratings correlated with more experienced stress, and that lower *TE* participants felt more nervous, more unpleasant, and more negative across both conditions. In conclusion, we accept H2.

Checking the quality of counterbalancing (H3), we compared presentations by order and found no differences between ratings for the programs (Canvas: $F(2,18)<1, ns$; Slide: $F(2,18)<1, ns$) or the errors (error: $F(2,18)=1.73, ns$; no-error: $F(2,18)<1, ns$). Thus, we accept H3.

As for **H4**, almost all (20) participants expressed that they felt similar to a real presentation. 14 mentioned that they felt less pressure since they had less stakes in the presentation, 10 mentioned an additional burden (e.g., the unfamiliarity of the topic), while 2 felt that the study was outright harder than their own presentations because of that. With this, we cautiously accept H4: the limitations of the lab study were manageable, and our setup was comparable to a real presentation.

5 Discussion and Summary

Our evaluation shows that presenters experience canvas and slide tools differently. More specifically, participants of our study that scored high on spatial ability or were less experienced preferred the canvas condition, while experienced or lower spatial ability presenters preferred classic slideware. Due to the strong overlap in our tester population between experience and lower spatial ability, we cannot attribute this effect to a single or a combination of these factors. We expected lower spatial ability to interact with the canvas condition due to its ZUI nature, but we could also explain that more experienced presenters are well versed in slideware and hence feel right at home. Interestingly, this difference is lessened in the *search for a loosely defined position*, a task that benefits particularly from the canvas format as the presenter can quickly zoom out to get an overview and pinpoint her target.

In summary, we have improved our understanding of canvas presentations and gained insight into who benefits from the format. Combined with the existing body of research [6,8,9,10] we now have an understanding of all the actors involved in presentations. One may think of author and presenter as the same person, but the results of [11,13] and this paper indicate opposing forces in the tools: The authoring work before the talk benefits from the freedom of expression of the canvas format. The same freedom of navigation while presenting comes with a drawback. Here, time and attention are limited and, as we have seen, a simpler linear format can be easier to handle for some presenters (cf. [7]). One could conclude to always limit the format during presentation delivery, but we have also shown that for some presenters this would be unfavorable. We suggest that during delivery, canvas tools should (1) allow the user to limit the navigation to the linear format until she needs the free format, and (2) offer her a simple way to get back to a safe place on the presentation path.

6 Limitations and Future Work

By design, our study was a lab study, and therefore is controlled situation that might not be representative of real-life use. A field study with presenters presenting their own presentations, with their own agendas on their own topic of expertise, and own audiences could corroborate the results of this paper. We were unable to attribute the interaction effect of program use to spatial ability or presentation age, as both independent variables correlated. A study with separated variables could investigate this issue further. Other limitations are the short nature of presentations and the novelty of

the canvas condition. We invite replication of this study and the materials used in this experiment can be downloaded at <http://hci.rwth-aachen.de/fly>.

References

1. Bederson, B.B., Hollan, J.D.: Pad++: a zooming graphical interface for exploring alternate interface physics. In: Proc. UIST 1994, pp. 17–26 (1994)
2. Bederson, B.B.: The promise of zoomable user interfaces. *Behaviour & Information Technology*, vol. 30 no. 6, 853–866 (2011)
3. Ekstrom, R.B., French, J.W., Harman, H.H.: Manual for Kit of Factor-referenced Cognitive Tests (1976)
4. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25 no. 1, 49–59 (1994)
5. Evans, G.T.: Use of the semantic differential technique to study attitudes during classroom lessons. *Interchange*, vol. 1 no. 4, 96–106 (1970)
6. Frijda, N.H.: The psychologists' point of view. *Handbook of emotions*, Lewis, M., Haviland- Jones, J.M., and Feldman Barrett, L. (eds.), vol. 3, 68–87 (2008)
7. Good, L., Bederson, B.B.: Zoomable user interfaces as a medium for slide show presentations. *Information Visualization*, vol. 1 no. 1, 35–49, (2002)
8. Higgins, R.L., Alonso, R.R., Pendleton, M.G.: The Validity of Role-Play Assessments of Assertiveness. *Behavior Therapy*, vol. 10 no. 5, 655–662 (1979)
9. Kern, J.M., Miller, C., Eggers, J.: Enhancing the Validity of Role-Play Tests: A Comparison of Three Role-Play Methodologies. *Behavior Therapy*, vol. 14 no. 4, 482–492 (1983)
10. Apple Keynote, <http://www.apple.com>
11. Lichtschlag, L., Karrer, T., Borchers, J.: Fly: a Tool to Author Planar Presentations. In: Proc. CHI 2009, pp. 547–556 (2009)
12. Lichtschlag, L., Hess, T., Karrer, T., Borchers, J.: Fly: Studying Recall, Macrostructure Understanding, and User Experience of Canvas Presentations. In: Proc. CHI 2012, pp. 1307–1310 (2012)
13. Lichtschlag, L., Hess, T., Karrer, T., Borchers, J.: Canvas Presentations in the Wild. In: Proc. CHI EA 2012, pp. 537–540 (2012)
14. Mauss, I.B., Robinson, M.D.: Measures of emotion: A review. *Cognition & Emotion*, vol. 23 no. 2 (2009)
15. Osgood, C.E., Suci, G.J., Tannenbaum, P.: The measurement of meaning (1957)
16. Parker, I.: Absolute PowerPoint: Can a software package edit our thoughts? *The New Yorker*, vol. 77 no. 13, 76–87 (2001)
17. Prezi, <http://prezi.com>
18. Robinson, M.D., Clore, G.L.: Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, vol. 128 no. 6, 934–960 (2002)
19. Russell, J.A., Mehrabian, A.: Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, vol. 11 no. 3, 273–294 (1977)
20. Scherer, K.R.: What are emotions? and how can they be measured?. *Social Science Information*, vol. 44 no. 4, 695–729 (2005)
21. Tufte, E.: *The Cognitive Style of PowerPoint*, Graphics Press, USA (2003)
22. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, vol. 54 no. 6, 1063–1070 (1988).