

Did We Miss Something? Correspondence Analysis of Usability Data

Stanislaw Zabramski, Wolfgang Stuerzlinger

► **To cite this version:**

Stanislaw Zabramski, Wolfgang Stuerzlinger. Did We Miss Something? Correspondence Analysis of Usability Data. David Hutchison; Takeo Kanade; Madhu Sudan; Demetri Terzopoulos; Doug Tygar; Moshe Y. Vardi; Gerhard Weikum; Paula Kotzé; Gary Marsden; Gitte Lindgaard; Janet Wesson; Marco Winckler; Josef Kittler; Jon M. Kleinberg; Friedemann Mattern; John C. Mitchell; Moni Naor; Oscar Nierstrasz; C. Pandu Rangan; Bernhard Steffen. 14th International Conference on Human-Computer Interaction (INTERACT), Sep 2013, Cape Town, South Africa. Springer, Lecture Notes in Computer Science, LNCS-8120 (Part IV), pp.272-279, 2013, Human-Computer Interaction – INTERACT 2013. <10.1007/978-3-642-40498-6_20>. <hal-01510508>

HAL Id: hal-01510508

<https://hal.inria.fr/hal-01510508>

Submitted on 19 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Did We Miss Something? Correspondence Analysis of Usability Data

Stanislaw Zabramski¹ and Wolfgang Stuerzlinger²

¹ Department of Informatics and Media, Uppsala University, 751 05 Uppsala, Sweden
stanislaw.zabramski@im.uu.se

² Department of Computer Science and Engineering, York University, Toronto, Ontario,
M3J 1P3, Canada
www.cse.yorku.ca/~wolfgang

Abstract. We have applied a multivariate exploratory technique called Correspondence Analysis (CA) to create and analyze a model of the dataset of experiment results. The dataset originates from a comparative usability study of tracing with the use of mouse, pen, and touch input and contains both categorical and continuous data – i.e. results of questionnaires and task measurements. CA allowed to visually and numerically assess the main variables in the dataset and how they interact with each other. In our study, pen input had the best measured performance and was preferred by the users. Touch input was the least accurate of all input methods tested but it was preferred by users over mouse especially in the conditions lacking of visual feedback of drawing. CA helped to detect that secondary effect even though it cannot be explained by the performance results alone. The importance of the influence of user’s previous experience is also noted. We conclude that CA helped to identify all major phenomena known from previous studies but also was sensitive to minor and secondary effects, what makes it a well suited method to quickly evaluate usability data.

Keywords: shape, freehand, tracing, drawing, mouse, pen, stylus, touch, evaluation, comparison, error, measurement, subjective.

1 Introduction

Usability of any technical solution can be perceived from the pragmatic point of view that focuses only on “getting things done”. However, there is a variety of users' beliefs, preconceptions, preferences, perceptions, and physical and psychological responses that affect the whole experience before, during and after use of the technique tested. Therefore the subjective and objective approaches should be used in usability evaluations and the results should be integrated into one consistent picture [1]. Even though subjective perceptions of usability are generally not correlated with objective measurements [2] their comparison is done rarely in practice [3]. The reason for that

might be that it is hard and time consuming to get a holistic view on the experiment from results obtained from a number of different methods of data analysis suitable for particular types of data.

Correspondence Analysis (CA) is a multivariate exploratory technique similar to Principal Component Analysis (PCA). In this study we apply CA and analyze objective measurements of users' performance i.e. task time and error, together with subjective evaluations of users' satisfaction in shape-tracing task.

Artistic drawing involving computer as a tool may be negatively influenced both by inaccuracies in the input method and by spatio-temporal constraints imposed on the user e.g. by task formulation. Unfortunately, there is no suitable model to describe spatially and temporally unconstrained freehand drawing when the path is unknown a priori and typical user errors are not characterized mathematically. Therefore, an exploratory study based on a standardized method [4] was performed to comparatively evaluate three popular input methods (mouse, pen, and touch input) in a simple shape-tracing task [5]. Previous research points to a positive influence of visual feedback. Therefore, the visual feedback of drawing was controlled in analyzed study through the presence or absence of solid black ink trace.

We use CA in an attempt to get a quick and simplified view on the experimental data-set and use it to evaluate if drawing performance can be a predictive criterion of users preferences [6] towards input methods tested with controlled visual feedback.

2 Procedures and Methods

We used CA to analyze the data-set of results from experimental study that aimed to evaluate the influence of mouse, touch and pen input on user performance in an *unconstrained* free-hand tracing task [5]. It also included the user experiences and preferences expressed towards these devices.

2.1 Experiment Design

The experiment featured a mixed design. Sixteen student volunteers (4 females) with average age 24 (SD = 2.3) were selected through convenience sampling from the local university campus. All except one were right-handed and had normal or corrected to normal vision. Half of the participants were randomly assigned to the visible drawing line condition and the other half had no feedback of drawing. Each participant used the mouse, pen and touch input on a Windows7 based HP Touchsmart TM2-1090eo Tablet PC to draw over a semi-randomly generated contour shape displayed on the screen once with every input device. Devices were assigned in a counterbalanced order in a within subject design. They were required to use the same starting point and direction to draw over the same nonsense closed contour shape presented, using a single continuous and uninterrupted stroke. Timing data was collected for each task. User error was calculated via a method that computes the average deviation of the user-generated shape from the original version [4, 5].

2.2 Questionnaires

All the participants filled pre-test questionnaires and provided basic screening information about demographics as well as their familiarity with the investigated computer input methods in the last 3 years.

After the experiment, post-test questionnaires were used to collect information about participant experience and opinions regarding three following aspects:

- input device of preference (statement S1)
- perceived ease of use (statements S2 and S4)
- perceived learnability (statements S3 and S5)

To avoid potential distortions originating from acquiescence bias the scale was balanced by offering equal number of positive and negative statements regarding the most important aspects [7]. The users' ratings of negative statements were later inverted and the final ratings then computed as the median of the responses to the paired questions (S2&4, S3&5). The questionnaire included:

- S1 - I think that I would like to use this input device frequently for sketching.
- S2 - I thought that the input device was easy to use.
- S3 - I would imagine that most people would learn to sketch with this input device very quickly.
- S4 - I found the input device very cumbersome to use.
- S5 - I imagine I would need to practice a lot before I could sketch efficiently with this input device.

Participants were asked to evaluate five statements on a symmetric Likert-like psychometric scale. The choices offered were: strongly disagree, somewhat disagree, indifferent, somewhat agree, strongly agree. Every statement was evaluated for each input method separately.

2.3 Correspondence Analysis

CA identifies a set of synthetic variables (factors) that summarize the original data set. In this study, CA is expected to reveal frequency-based associations between the categorical data from the questionnaires and the data from the measurements [8]. CA creates a model for data with fewer factors by decomposition of inertia describing the degree of variation or eigenvalue. Similarly to PCA, the calculation gives the factors and correlations (loadings) with each factor for each variable. The optimal number of factors can be decided with the scree-test or the Kaiser criterion. Then significance tests can be used to indicate that the occurrence of certain values is significantly more common than chance.

CA also enables the visualization of associations between variables in question, to simplify their identification, and to isolate relations between them. That can be done through two-dimensional configuration bi-plots, which show the relative proximity of data points and depict degrees of correlation and variation (inertia). Each dimension of a plot explains a certain percentage of the variation in the data. Normally, the first two dimensions capture a large percentage of the variation [8].

3 Results

3.1 Pre-test Questionnaires

The participants reported their average daily experience: 8.2h (SD = 5) of with computers in general, 3.7h (SD = 4.7) with a computer mouse, 0.8h (SD = 1) with pen devices (incl. mobile phones), and 3.3h (SD = 5.3) with touch devices (incl. mobile phones). These values have been called *experience* and were used in the analyses.

3.2 Time and Error

The ANOVA of task time and user error data showed no main effect for the visibility of visual feedback during drawing, nor any interaction between feedback visibility and input device used [5]. There was no significant difference between the visibility of feedback conditions either for time (Cohen's $d = 0.3$) or error (Cohen's $d = 0.14$). Power analysis revealed that in order for an effect of this size to be detected (90% chance) as significant at the 5% level, a sample of 238 and 1111 participants would be required for time and error respectively.

However, there was a main effect of input device on task time and a post-hoc analysis identified significant differences between touch input and mouse. The mean times were: mouse = 40 sec. (SD = 16.6 sec.), pen = 30.1 sec. (SD = 15.3 sec.), and touch = 21.5 sec. (SD = 9.3 sec.). User error data indicated also a main effect of input device. A post-hoc analysis showed that mouse and pen are not significantly different. The mean errors were: mouse = 5.84 px (SD = 2.21 px), pen-input = 4.98 px (SD = 1.92 px), touch-input = 7.95 px (SD = 2.94 px), with the grand mean error = 6.25 px.

3.3 Post-test Questionnaires

All the questionnaire responses were coded between 1 and 5 reflecting the level of expressed agreement or disagreement on an ordinal scale. Some of the scales were inverted to deal with acquiescence bias. We illustrate this inversion by adding a “-“ sign in front of the statement number (e.g. -S4). The paired statements, S2&-S4 and S3&-S5, were compared to check for acquiescence bias.

Polychoric correlation coefficients (pcc) were calculated and used in the analyses. To investigate the significance of potential differences between groups of responses a Wilcoxon matched-pairs test was used.

We found that the overall replies to statements S2 and -S4 for all input methods and all feedback visibility conditions have strong positive correlation (pcc=0.68) with insignificant differences between their medians. Similarly, in case of statements S3 and -S5, strong positive correlation was observed (pcc=0.74), but this time with significant differences between their medians ($p < 0.05$).

Figure 1 presents the summary of responses to post-test questionnaires.

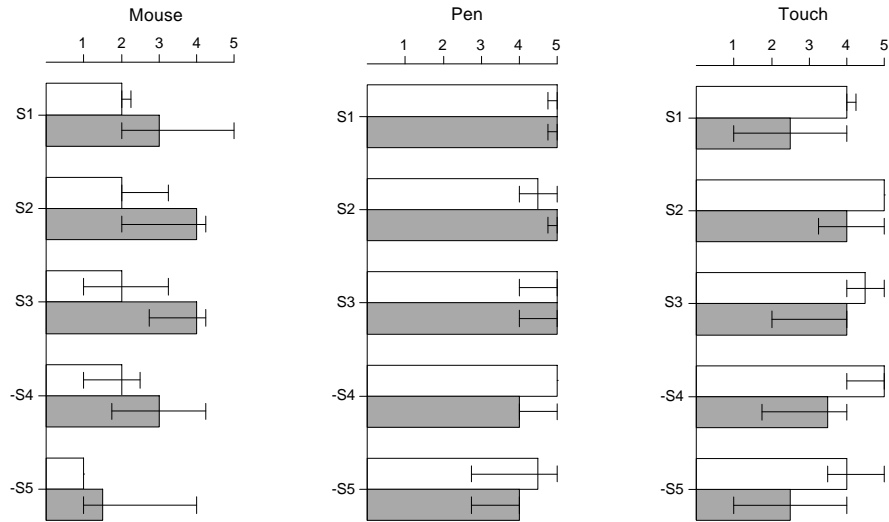


Fig. 1. Median responses to post-test statements for each input method and visibility of feedback mode (*white bars for feedback “off”, grey bars for feedback “on”*). Whiskers represent the interquartile range.

In following analyses the median value of replies to S2 and -S4 is referred to as *ease of use*, and the median value of replies to S3 and -S5 as *learnability*. Replies to S1 are termed *preference*. The correlation matrix of responses of all these three aspects illustrated positive correlations to each other ($pcc > 0.5$).

3.4 Correspondence Analysis

For extracting common explanatory factors a CA was calculated and analyzed. The first analysis compared the experience in using various input methods against the measurements and the post-test results, such as preferred input device and perceived ease of use.

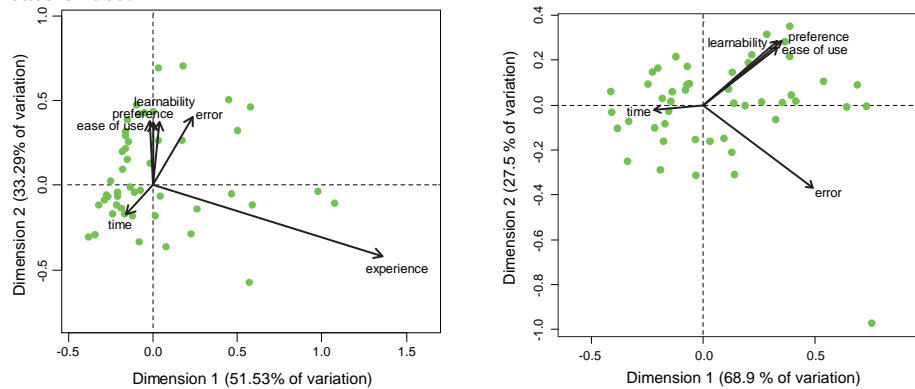


Fig. 2. Bi-plots from Correspondence Analyses with (*left*) and without users’ *experience* as variable in the dataset (*right*).

The bi-plot generated from the CA (see Figure 2 - left) shows that *experience* has a strong influence on all other variables. It shows potential strong relation to task time, but neither the correlation nor regression analysis gave any significant or consistent results. The *experience* factor was removed from data for the following analyses to better contrast the remaining variables (see Figure 2 - right).

Some differences between the visibility of the drawing feedback conditions have been noticed in the post-test responses within each input device (see Figure 1). The bi-plot generated from the CA confirmed that observation (see Figure 3).

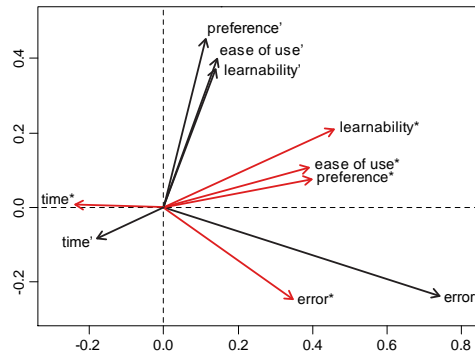


Fig. 3. Bi-plot from Correlation Analysis. The datasets with (marked with apostrophes) and without visible drawing feedback (marked with asterisks) were considered separately.

A Mann-Whitney-Wilcoxon test identified significant differences for *preference* and *ease of use* only for touch input ($p < 0.05$). In case of *learnability* that effect was not so strong for touch but a significant difference has been found in case of mouse ($p < 0.05$). Table 1 summarizes these results.

Table 1. Input methods ordered by users preference in each visual feedback condition for each aspect. “ \geq ” marks insignificant differences and “ $>$ ” marks significant differences ($p < 0.05$) after Mann-Whitney-Wilcoxon test.

Aspect	Visual feedback	Input methods
preference	on	Pen > Mouse \geq Touch
	off	Pen \geq Touch > Mouse
	on and off	Pen > Touch \geq Mouse
ease of use	on	Pen > Touch \geq Mouse
	off	Touch \geq Pen > Mouse
	on and off	Pen \geq Touch \geq Mouse
learnability	on	Pen > Touch \geq Mouse
	off	Pen \geq Touch > Mouse
	on and off	Pen > Touch \geq Mouse

4 Discussion

The multidimensional dataset generated by an experiment is challenging to analyze. Thus, it is essential to use a suitable analysis method, such as CA, which can identify the most influential variables.

Previous experience strongly varied between participants due to the lack of homogeneity in the participant's group. As the most influential variable, it biased both the measurements and subjective factors, but unfortunately not in a way that can be easily interpreted in a meaningful way. E.g. experience did not correlate with error or task time. Therefore it was excluded from the analyses.

CA suggested to contrast time with error (see Figure 2 - right). This is the evidence that the well-known speed-accuracy trade-off [5] took place in this study.

Subjective evaluations did not show meaningful differences between the aspects of user satisfaction addressed in questionnaires and strongly correlated to each other. The pen supports better performance and overall results for *device of preference* identify the pen as the best choice for frequent sketching. It is followed by touch and mouse with a non-significant difference between the two. However, this difference changed to significant when no feedback of drawing was provided (see Table 1) making touch more preferred when the feedback was absent. That trend is also present in the case of perceived *learnability* and even more visible for perceived *ease of use*. None of these however, correspond to significant differences of performance [5].

5 Conclusion

Correspondence Analysis is a frequently used technique e.g. in linguistics, sociology, or biology. When applied in HCI, CA helped to quickly analyze the results and identified all major phenomena in the studied experiment that mostly have confirmation in previous research. But, CA was also sensitive to secondary effects that might be harder to notice when the results are spread among multiple methods of data analysis.

In the experiment analyzed here the lack of visual feedback significantly improved the overall users' attitude towards touch as an input method for drawing (see Table 1 and Figure 1) – while having no significant effect on performance. The result is opposite for *learnability* for the mouse. The pen stays an obvious choice for drawing task both from performance and preference point of view. However, we think that more attention should be directed towards user biases favoring touch input over mouse observed in cases lacking of visual feedback. Also, the huge influence of user's familiarity with these input methods is worth of being investigated even more. Potentially, technical properties of the input methods, such as directness, occlusion effect, or latency may also contribute to the observed mismatch between the performance results and subjective evaluations.

Obviously, we have to assume that the results of the experimental study used here are true only for the particular shape investigated. Nevertheless, CA appears to be an available [9] and convenient method that might be appropriate also for exploratory analyses of multi-dimensional data-sets with results from other kinds of studies in-

volving mixed data-types, also with much higher sample size - what makes it worth being promoted more widely in HCI community.

Acknowledgments

I thank Dimitrios Gkouskos and Sarathkumar Neelakannan for their respective input.

References

1. ISO: 9241-210 Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. ISO 9241 (2010).
2. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring Usability: Are Effectiveness , Efficiency , and Satisfaction Really Correlated? Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00. pp. 345–352. ACM Press, New York, New York, USA (2000).
3. Barkhuus, L., Rode, J.A.: From Mice to Men – 24 years of Evaluation in CHI. alt.chi 2007 (2007).
4. ISO: 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. ISO 9241 (1999).
5. Zabramski, S.: Careless touch: A comparative evaluation of mouse, pen- and touch-input in shape tracing task. OZCHI 2011. pp. 329–332. ACM, Canberra, Australia (2011).
6. Nielsen, J., Levy, J.: Measuring usability: preference vs. performance. Communications of the ACM. 37, 66–75 (1994).
7. Schriesheim, C.A., Hill, K.D.: Controlling Acquiescence Response Bias by Item Reversals: The Effect on Questionnaire Validity. Educational and Psychological Measurement. 41, 1101–1114 (1981).
8. Greenacre, M.J.: Correspondence Analysis in Practice. Chapman & Hal / CRC (2007).
9. Nenadic, O., Greenacre, M.: Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. Journal of Statistical Software. 20, 1–13 (2007).