Regression analysis of age effects during prolonged work with head-mounted displays

Matthias Willea,b,*, Alexander Mertensa, Lars Adolphb, Sabine Theisa, Sascha Wischniewskib, Christopher Schlicka

aInstitute of Industrial Engineering and Ergonomics of RWTH Aachen University, Bergdriesch 27, 52062 Aachen, Germany
bFederal Institute for Occupational Safety and Health Germany, Friedrich-Henkel-Weg 1-25, 44149 Dortmund, Germany

Abstract

The German Federal Institute for Occupational Safety and Health (BAuA) has been carrying out a project titled “head mounted displays – conditions for safe and strain optimized use”. One focus within the project was the influence of prolonged work with different head mounted displays (HMDs) in different assembling settings on performance, mental and physical strain while comparing the HMD to a wall mounted monitor and a tablet computer. A primary ANOVA analysis showed age effects regarding performance and some of the strain measurements. Present work contains a complete reanalysis of three laboratory studies using multiple regression to get valuable insights into the nature of age influences on performance and strain. In the first study, participants constructed Lego models. Here, the age of participants had a major influence on performance independent from display type. Regression analysis revealed that the predictors “age” and “technical affinity” led to highly significant regression models explaining 59% - 70% of variance in data. In a replication of this study with consumer products “age” was beneath display group again a valid predictor explaining 60% of variance. In the third study participants were assembling and disassembling a real car engine. Here “age” and “technical affinity” were rejected as predictor while display group explains about 10% of variance in task execution time. This illustrates that the “age” factor is highly task dependent and might have bigger effects on the fine motor skill demanding toy assembling task compared to the physically demanding work on a car engine. Within the paper, regression analyses of strain parameters are also discussed and regression models are presented and compared. This work helps to identify the influence of age within scientific research on new technologies like HMDs.

Keywords: Age effects; Regression; Performance; Strain; Head-mounted displays

* Corresponding author. Tel.: +49 241 8099479; fax: +49 241 8092131.
E-mail address: dr.matthias.wille@gmail.com
1. Introduction

The innovation circle of new technologies becomes shorter and shorter these days. These new technologies are often part of experimental setups, where special interface types, kind of interaction or applications are tested. The age of subjects in these experiments might have a wide range and it is often not clear, how age generally influence the attitude, perception and interaction with new technologies. On one hand a decline of visual and auditory abilities is known with increasing age [1]. On the other hand the performance of elderly people is not in first line determined by the chronological age [2] and the scatter of individual skills within elderly people is more dominant than the average decreases going along with age [3]. This makes general age effects hard to predict.

2. Background

The findings presented in this paper are based on a reanalysis of data from the project “head mounted displays – conditions for safe and strain optimized use” [4]. Therefore it is important to describe the scope, study design and main results of this project first.

2.1. The project

The project was carried out by the German Federal Institute for Occupational Safety and Health (BAuA) with subcontracted work by the Institute of Industrial Engineering and Ergonomics of RWTH Aachen University (IAW) and the Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE).

The main target of the project was to identify fields of application where workers can be supported by a head-mounted display (HMD) and to investigate the physical and mental strain while working with a HMD compared to other displays like a tablet-PC or a wall mounted monitor. Another important point of the project was to measure the effects of prolonged work over several hours with those systems and to investigate the development of strain over time with those technologies.

HMDs are small near-to-eye monitors attached to the persons head. This generic term describes a very diverse group of hardware: HMDs can be monocular (one display only in front of one eye) or binocular (a separate display in front of each eye). Those displays can be closed, covering the view into the surrounding (called “look around”) or they can be semitransparent, allowing seeing the environment shining through (called “see through”). Display resolution and construction method also differs strongly, same as weight and wearing comfort of the head carrier.

The focus of the project was to support manual workers, e.g. in maintenance or during assembling. This kind of work contains some aspects which can be perfectly supported by an HMD: If a lot of information is needed during the work process, if this information can be arranged sequentially (which is necessary as the HMD monitor is small), if mobility is required (so fixed installed monitors are no general solution) and if both hands are needed to work on the object (so using one hand to hold e.g. a tablet-PC is difficult). As the worker still has to navigate through and interact with the environment all HMDs used within the project were monocular (diverse see-through and look around types).

2.2. Study design

Beneath field studies and task analysis three laboratory studies were carried out within the project. These studies have in common that they investigated performance and strain during a sequential picture-based assembling task over several hours while the information was displayed on a HMD or for comparison on a tablet-PC or wall mounted monitor. Performance and strain were measured at different timestamps to get insight into the development of strain over time with this new technology. On the other hand the studies differ in design (within- or between-subject design), some dependent variables, task content for assembling and used displays. Table 1 gives an overview of all three studies and figure 1 shows the used HMDs.

Study A [5] followed a repeated measurement design, where each participant had three sessions of four hours each, while content was presented two times on the HMD and one time on the tablet-PC. A comparison between first and second HMD session revealed habituation effects and the comparison between second HMD session and...
tablet-PC the influence of display. Work content was assembling toy cars and parallel reacting to changes in a gauge at the display border (see fig. 2a/b). As no age effects showed up during the monitoring task, we exclude its description here. Dependent variable regarding performance is the number of processed instruction slides during the whole session. As subjective workload measures the rating scale of mental effort (RSME) [6] was prompted every 30 minutes during the session and the NASA-TLX [7] in the middle and at the end of the session. Hourly, participants completed the visual fatigue questionnaire (VFQ) [8] while technical affinity was enquired by a questionnaire [9] before the experiment started.

Study B [10] was a replication of study A. Work content and dependent variables were the same but lighter consumer hardware was used to investigate the effects of hardware on strain and performance [11]. Experiment time was shorter and a between subject design was used with 18 HMD and 18 tablet-PC users.

Study C [12, 13, 14] had a focus on physical strain: During the whole experiment the subjects activity of neck- and shoulder muscles was recorded via EMG. Visual acuity was measured before and after the experiment and visual strain by eye blink rate and duration captured by EOG during the whole experiment. Subjective workload measured by RSME and NASA-TLX were collected at the same timestamps as in study A. Work content was here to disassemble and assemble a real car engine while information was shown on a HMD or wall mounted monitor (see figure 2c). Dependent variable regarding performance was task finishing time. A between subject design was used with three groups (n=20 for each): One group with wall mounted monitor, one with HMD in look around mode and one with HMD in see-through mode. Age and gender were balanced among groups.

The original analysis of these three studies used an ANOVA while age of subjects was represented in two groups, done by median split for study A and B and ranging from 18–39 and 40–60 years in study C.

Table 1. Overview of study design, participants, apparatus and task.

| Study | n  | Age (yrs.) | Design | Duration | HMD          | Monitor                  | Task                                               |
|-------|----|------------|--------|----------|--------------|--------------------------|----------------------------------------------------|
| A     | 41 | 18–67      | Within subject | 4h       | MAVUS Industrial tablet-PC | Toy assembling + parallel monitoring task (reaction to stimuli) |
|       |    | M=36.7     |        |          |              |                          |                                                    |
|       |    | SD=15.3    |        |          |              |                          |                                                    |
| B     | 36 | 19–63      | Between subject | 2h       | Google Glass Consumer tablet-PC | Toy assembling + parallel monitoring task (reaction to stimuli) |
|       |    | M=37.2     |        |          |              |                          |                                                    |
|       |    | SD=14.6    |        |          |              |                          |                                                    |
| C     | 60 | 19–59      | Between subject | 4h       | Liteye 750 A Wall mounted monitor | Disassembling and assembling a real car engine |
|       |    | M=36.0     |        |          |              |                          |                                                    |
|       |    | SD=13.6    |        |          |              |                          |                                                    |

Fig. 1. Head mounted displays: (a) MAVUS-HMD; (b) Google Glass; (c) Liteye 750 A.
2.3. Main project findings and age effects

In study A the number of processed slides was significant lower if using a HMD and for the older participants, but no interdependency between those factors was identified, meaning elderly participants work slower with all displays in same amount. The subjective strain measured by the NASA-TLX showed a significantly higher score for the HMD. No main effect of age was found, but an interaction between display type and age reached statistical tendency, based on a higher increase in strain over time for older participants. The RSME value also showed significant higher values for the HMD, an increase over time and a main effect of age, which again indicates that elderly were more stressed. Visual fatigue questionnaire (VFQ) items remained in the lower third of the scale but mostly showed an expectable significant increase over time during the experiment. Many of the items proofed higher values for the HMD, especially “neck pain” and “headache” but age effects were only found in “mental fatigue” which had also a higher increase over time for the elderly.

Study B confirmed the findings of study A including lower performance with HMD and of elderly but again without interdependency among these factors. NASA-TLX showed again higher ratings for HMD and this time also significant higher ratings for older subjects, but no interdependency among these factors. RSME values missed significance for higher HMD values, but showed a tendency for higher ratings from elderly. VFQ showed again higher values for HMD, but no age effects were found. The items “neck pain” and “headache” showed no effect anymore, representing the lighter and more comfortable hardware setup.

In study C lower performance (longer task execution time) with the HMD was again significant, but no age effects were found there. Visual acuity tests revealed poorer values for the elderly in general, but independent from display type. Eye blink rate and duration showed no effect at all. VFQ values revealed higher values for HMD (especially in look-around mode) but no age effects. NASA-TLX and RSME showed no display or age effects. EMG analysis showed an increase over time for neck muscle activity (splenius capitus) and a higher increase for elderly.

3. Method

In order to investigate age effects observed in the previous results we conducted a post hoc correlation and simple as well as multiple regression analysis even if the studies were not intentionally designed for this kind of analysis. For (multiple) regression analysis the sample size is very small and the within subject design with repeated measurement used in study A is also not ideal for regression analysis. But despite these restrictions this detailed reanalysis can give a bit more insight in how age influence performance and strain while working with these new technologies.

During multiple regression stepwise backward method was used to identify relevant predictors with F-value $\geq 0.100$ as criterion for exclusion of predictors. Based on the previous analysis predictors were “age of participants”, “technical affinity score” measured by the TA-EG questionnaire and “diopter of display eye” as visual
acuity is important for these setups and a decline of visual acuity with age is known and documented in literature [3]. Diopter and TA-EG were not measured in study C, therefore only age and display group as predictor are possible in this part. All ps in correlation and regression are reported two-tailed. For study B and C - which followed a between subject design - the group variables are added as predictor. For study A all sessions are analyzed separately.

4. Results

First of all, correlation among predictors is important: Figure 3 show scatter plots of technical affinity and diopter of display eye with age for all participants of study A and B. It is interesting to notice that technical affinity has no relation with age ($r = .016, p = .914$). While diopter of the display eye had a significant relation with age ($r = .393, p < .01$). This reflects the well-known increase of long-sightedness with age. As especially for working with HMDs vision has a great impact, it will be difficult to separate these two conjunct factors “age” and “diopter of display eye”.

![Fig. 3. Participants in study A and B: (a) scatter plot from technical affinity and age; (b) scatter plot from diopter of display eye and age.](image)

Fig. 4. (a) Study A: scatter plot from processed slides within 4 hours and age for second HMD session and tablet-PC session (within subject); (b) Study B: scatter plot from processed slides within 25 minutes and age for HMD and tablet-PC group (between subject).
Figure 4 depicts the correlation between number of processed slides and age in studies A and B. It becomes clear that with increasing age the number of processed slides is getting smaller. In study A for HMD $r = -.715$ and for tablet-PC $r = -.662$; in study B for HMD $r = -.644$ and for tablet-PC $r = -.741$ (all $p < .01$). It is worth noticeable that even in this replication with complete different hardware the same effects were found. Age explains about 50% of variance in performance, independently from display type.

However, study C showed no age effects regarding performance: Not in the original analysis of variances between groups, nor in the reanalysis using correlation ($r = .068, p = .648$). This means the age effect found for task execution time is highly task specific and not a general effect of working with a HMD.

In a next step a multiple regression analysis was executed using stepwise backward method with the predictors mentioned above. Table 2 shows the final regression models for HMD- and tablet-PC-session in study A and for study B. The regressions can explain between 59 and 70 percent of variance. “Diopter of display eye” was excluded as predictor in both sessions of study A (HMD: $p = .148$; tablet-PC: $p = .180$) while it is a relevant predictor in study B. On the other hand “technical affinity score” was excluded as predictor in study B ($p = .833$), while it was a relevant predictor in both sessions of study A.

| Table 2. Coefficients of the final multiple regression models (stepwise backward) for number of processed slides in studies A and B. |
| --- |
| **Study A, 2nd HMD session; $R^2 = .589$, $F(2, 38) = 27.178, p < .001$** |
| Constant | 313.869 | 36.193 | 8.672 | .000 |
| Age | -2.743 | .389 | -.737 | -7.058 | .000 |
| Technical affinity score | -35.510 | 13.314 | -.278 | -2.667 | .011 |
| **Study A, Tablet-PC session; $R^2 = .702$, $F(2, 38) = 18.462, p < .001$** |
| Constant | 367.224 | 51.411 | 7.143 | .000 |
| Age | -3.241 | .552 | -.680 | -5.870 | .000 |
| Technical affinity score | -38.112 | 18.912 | -.234 | -2.015 | .051 |
| **Study B; $R^2 = .600$, $F(3, 32) = 16.008, p < .001$** |
| Constant | 34.008 | 2.703 | 12.583 | .000 |
| Display type | -4.894 | 1.736 | -.319 | -2.819 | .008 |
| Age | -3.10 | .626 | .593 | -4.997 | .000 |
| Diopter of display eye | -.702 | .408 | -.206 | -1.718 | .095 |

In study A the NASA-TLX score showed no correlation with age in the HMD-session ($r = .133, p = .409$) but in the tablet-PC-session ($r = .374, p < .05$). Furthermore, there is a highly significant correlation of NASA-TLX scores between HMD- and tablet-PC-session ($r = .665, p < .001$) reflecting the individual answering tendencies of subjects in this repeated measurement design. For the tablet-PC session a multiple regression analysis rejected none of the predictors and explains 31% variance [$R^2 = .308, F(3, 37) = 5.479, p = .003$] while age was the weakest of the predictors with $p = .088$. In study B no significant correlation of age and NASA-TLX score was found (HMD: $r = .444, p = .057$; tablet-PC:$r = .211, p = .416$) although there was a main age effect in the previous analysis of variances. A multiple regression analysis rejected “age” as predictor, while “display type” and “diopter of display eye” explain 31% of variance [$R^2 = .308, F(2, 33) = 7.340, p = .002$]. In study C again no significant correlation between age and NASA-TLX score can be found ($r = -.217 p = .096$). This tendency also got a new direction of lower score with increasing age, while in study A and B those weak effects that were found goes towards higher scores with increasing age. A multiple regression analysis rejected display types as predictors and ended in a simple regression with age, but not significant [$R^2 = .217, F(1, 58) = 2.864, p = .096$].

RSME score showed a highly significant correlation with age in study A (HMD: $r = .459 p = .003$; tablet-PC:$r = .438 p = .004$) and an even higher individual correlation among those both conditions ($r = .891, p < .001$). A multiple regression analysis rejected all predictors despite age in both displays sessions. In HMD-session age
explains 46% of variance \([R^2 = .459, F(1, 37) = 9.876, p = .003]\) and in the tablet-PC-session age explains 44% of variance \([R^2 = .438, F(1, 39) = 9.272, p = .004]\). But in study B no correlation between age and RSME-score can be found (HMD: \(r = .325, p = .189\); tablet-PC: \(r = .188, p = .469\)). In multiple regression analysis “age” and “display type” were rejected as predictors while “diopter of display eye” and “technical affinity” explains 48% of variance \([R^2 = .476, F(2, 32) = 4.691, p = .016]\). Again in study C no significant correlation between age and RSME score was found \((r = -.092, p = .483)\). Multiple regression rejected all predictors and no significant regression model was found.

Regarding the “mental fatigue” item of VFQ the difference between first and last inquiry were analysed to investigate the higher increase over time for elderly in study A. However, those values show no correlation with age (HMD: \(r = .041, p = .801\); tablet-PC: \(r = .154, p = .336\)). The higher increase over time in percentage of neck muscle activity for elderly found in study C was tested by building the difference between last and first work package for this item. But among display groups only the see-through HMD condition shows a tendency for correlation between age and the percentage of neck muscle activity \((r = .510, p = .052)\).

5. Discussion

Strong age effects in task performance were proven in study A and B: With increasing age less construction slides were completed within experiment time. This effect is independent from display type and it happens on HMD and tablet-PC in same amount. But it is also task specific, as no age effects were found in study C. So while the fine motor skill demanding task of assembling toy cars showed decreasing performance with increasing age the more physical demanding task of constructing a real car engine does not. The kind of work content presentation was comparable among studies using sequential picture based assembling instructions. Previous knowledge regarding toy assembling also showed no age effect. Most of the elderly had even more actual knowledge as they were used to assemble the well-known toys with their grandchildren. But the dual task paradigm with the additional monitoring task in studies A and B might have stressed older participants more and therefore explain task performance decrements in the assembling task. This will be also in line with the higher subjective strain ratings for elderly in study A.

However, in common subjective strain parameters NASA-TLX and RSME do not show clear and consistent age effects among studies and conditions. While in some cases correlation with age is strong (RSME in study A) in other cases it is not (RSME in study B, which was a replication of A with same work content). Furthermore, while direction in study A and B is towards increasing ratings with age, it is towards decreasing ratings with increasing age in study C (although only showing statistical tendency). Taken this all together it is not appropriate to speak of general age effects in subjective strain ratings.

Technical affinity score was treated as a between subject group factor (median split) in the previous analysis of variances in study A, but it had poor effects, mainly in triplet interdependencies. In this reanalysis it often turned out to be a valid predictor. Hereby the negative correlation indicates that higher scores lead to less finished assembling slides. This might be explained by motivation as more technical affine subjects might be more fascinated by the experimental setup and maybe try to do the work more accurate, which will take longer. However, technical affinity score is not linked to better signal detection rate in the monitoring task which speaks against this thesis.

“Diopter of display eye”, which correlates with age, was excluded as predictor in both sessions of study A and in study B it misses the rejection criterion only shortly with \(p = .095\) while age is a highly significant predictor \((p < .001)\) in that case. This might be interpreted as a hint that the age effects found here are not mainly based on the increase of long-sightedness with age but more on other effects which go along with age.

In common the age effects in performance found here are highly task specific and not linked to working with HMDs. Age effects in subjective strain measurements are not consistent and might be more based on individual answering tendencies within the small sample size especially as individual correlation across conditions are high. Nevertheless, investigation of age effects is important as the task specific age effects in study A and B show.
Acknowledgment

The results of this publication are applied within the research project “TECH4AGE”, which is funded by the German Federal Ministry of Education and Research (BMBF, Grant No. 16SV7111) supervised by the VDI/VDE Innovation + Technik GmbH.

More information about the project “head mounted displays – conditions for save and strain-optimized use” and a complete list of publications within the project can be found here: http://www.baua.de/en/Research/Research-Project/f2288.html

References

[1] F. Schieber, Changes in vision and hearing, in: J.E. Birren, K.W. Schaie (Eds.), Handbook of psychology and aging, Academic Press, San Diego, 2006.
[2] R. J. Shephard, Aging and productivity: some physiological issues, International Journal of Industrial Ergonomics, 25(5), 2000, 535–545.
[3] J. Ilmarinen, J. Tempel. Arbeitssicherheit 2010: was können wir tun, damit Sie gesund bleiben?: VSA-Verlag, 2002.
[4] M. Wille, S. Wischniewski, L. Adolph, S. Theis, B. Grauel, T. Alexander, T., Prolonged work with head mounted displays, in: Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program, 2014, pp. 221–224. ACM.
[5] M. Wille, B. Grauel, L. Adolph, Strain caused by head mounted displays, in: D. De Waard, K. Brookhuis, R. Wiczorek, F. Di Nocera, P. Barham, C. Weikert, A. Kluge, W. Gerbino, A. Toffetti, (Eds.) Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference, 2013. Available from http://hfes-europe.org
[6] F.R.H. Zijlstra, Efficiency in workbehaviour: an approach for modern tools. PhD thesis, University of Delft. Soesterberg, The Netherlands: Institute for Perception TNO, 1993. Online available: http://resolver.tudelft.nl/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f [April 2015].
[7] S.G. Hart, L.E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in: P.A. Hancock, N. Meshkati (Eds.), Human mental workload, 1988, pp. 139-183, Amsterdam: North-Holland.
[8] A.W. Bangor, Display technology and ambient illumination influences on visual fatigue at VDT Workstations. PhD thesis, Virginia Polytechnic Institute and State University, USA, 2000. Online available: http://scholar.lib.vt.edu/theses/available/etd-03072001-091123/unrestricted/vfatigue.pdf [April 2015].
[9] K. Karrer, C. Glaser, C. Clemens, C. Bruder, Technikhaffinität erfassen–der Fragebogen TA-EG. Der Mensch im Mittelpunkt technischer Systeme, 8, 2009, pp. 196–201. Online available: http://www.tu-berlin.de/fileadmin/f2s/8BWMMS/13.3-Karrer.pdf [April 2015].
[10] M. Wille, P. M. Scholl, S. Wischniewski, K. V. Laerhoven, Comparing google glass with tablet-pc as guidance system for assembling tasks, in: Wearable and Implantable Body Sensor Networks Workshops (BSN Workshops), 2014 11th International Conference on, 2014, pp. 38-41. IEEE.
[11] M. Wille, S. Wischniewski, Influence of head mounted display hardware on performance and strain, in: D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, H. Hoonhout (Eds.), Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference, 2014. Available from http://hfes-europe.org
[12] S. Theis, T. Alexander, M. Wille, M. Bertens, C. M. Schlick, Younger beginners, older retirees: Head-Mounted Displays and Demographic Change, Advances in The Ergonomics in Manufacturing: Managing the Enterprise of the Future, 13, 295, 2014.
[13] S. Theis, M. Wille, T. Alexander, The nexus of human factors in cyber-physical systems: ergonomics of eyewear for industrial applications, in: Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program, 2014, pp. 217–220. ACM.
[14] S. Theis, T. Alexander, M. Mayer, M. Wille, Considering ergonomic aspects of head-mounted displays for applications in industrial manufacturing, in: V. G. Duffy, (Eds.), Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management (DHM/HCI) 2013, Part II, Lecture Notes in Computer Science 8026, 2013, pp. 282–291. Berlin: Springer.