Younger Beginners, Older Retirees: Head-mounted Displays and Demographic Change

Sabine Theis[a], Thomas Alexander[a], Matthias Wille[b], Alexander Mertens[c] and Christopher M. Schlick[c]

[a]Institute for Communication, Information Processing and Ergonomics
Fraunhoferstraße 20
53343 Wachtberg

[b]Federal Institute for Occupational Safety and Health
Friedrich-Henkel-Weg 1-25
D-44149 Dortmund

[c]Institute of Industrial Engineering and Ergonomics of RWTH Aachen University
Bergdriesch 27
D-52062 Aachen

ABSTRACT

Demographic change and its consequences are tremendous challenges for industry and service enterprises (Dul et al., 2012). That implies the development of future working systems and technologies, which have to consider aspects of demographic change. Head-Mounted Displays (HMDs) may serve as intelligent solutions for multiple manufacturing scenarios by providing process knowledge already during fabrication and manufacturing. Since the early days, HMD-manufacturers have striven to improve technical and ergonomic characteristics of HMDs. Generic ergonomic aspects have already been investigated but the extent to which age influences workload and performance when using an HMD is merely known. Additionally, little is known about the acceptance of HMDs within different age groups. In this paper, we address these issues and present an empirical study (n = 40) analyzing the effect of HMDs on task execution time and workload (NASA Task Load Index). Two different age groups (18-39, 40-60 years) performed a manufacturing task supported by instructions displayed on an HMD. It is shown that elderly perform as good as young participants while there are significant differences concerning subjective effort. Young participants rate manufacturing tasks performed with an HMD to be more effortful as elderly. Regarding the acceptance of the HMD, we found that older participants are more likely to reject the HMD than younger ones.

Keywords: Wearable computing, demographic change, head-mounted displays, ergonomics

INTRODUCTION

Currently, population growth languishes in Europe and in other major industrial nations. Some publications even predict that population will decrease because low birth rates are no longer intercepted by higher immigration rates

(Hoßmann et al., 2008). This demographic change and its consequences form one of the biggest future challenges for industry and service enterprises (Dul et al., 2012). Demographic change has been defined as the shift within age structure of a population, which is influenced by birth rate, death rate, lifespan and migration. While the number of people under the age of 14 in Germany, for example, has decreased by 2.5 million, the number of people older than 65 years has increased by 5 million. The age structure of employees develops similarly and is influenced by immigration and emigration. While the average age of the West German population has changed in the years 1960-1996 from 35.6 to 40.1 years, the average age of the employable population changed from 38.9 to 39.8 years. The average age of employed people increased from 37.2 to 39.5 years (Buck et al., 2002). Women, elderly and unskilled younger people are considered as a compensation to a possible labor supply shortage (Tivig et al., 2013). But even if a shortage of labor supply does not occur, employees have at least to handle requirements of older workforces while keeping up productivity.

It is unquestionable that biological aging affects and reduces several physiological functions relevant for physical performance. Sensory performance declines as a side effect of biological aging. Different theories exist about the influences of age on productivity and performance. The deficient model states cognitive and physical resources as constantly declining. In fact, physiological parameters have their peak around the age of 20 (Nikolaus & Zahn, 1995). Afterwards they slightly decline until the age of 40. Beyond this age, physiological parameter curves are precipitous. In parallel, several cognitive abilities decrease as well. But age-dependent changes of cognitive abilities have to be differentiated according to Cattells two-factor model of intelligence about fluid and crystallized intelligence (Cattell, 1987). Fluid intelligence encompasses information processing and problem-solving, which are both independent from existing knowledge. It decreases in the course of life, has its effectiveness-peak around the age of 40 and then declines slowly onwards. On the other hand, crystallized intelligence encompasses knowledge and its application (Lehr, 1996). It stagnates or even increases from the age of 40. Nevertheless, performance and efficiency are not proven to be influenced by age-related deficiencies (Waldman & Aviolo, 1986). So, the deficient model was deeply shattered by the results of such large gerontological studies and the hypothesis of a differential aging became a likelier approach. Researchers ascribe performance as age-independent, due to reserve capacities which ensure that individuals' performance copes with given work requirements. Especially physiologically demanding environments lead to stronger decrease of age dependent physiological capacities (Dittmann-Kohli & Heijden, 1996). Furthermore, it has to be considered, that workplace design, besides age-dependent, lifestyle-dependent and psycho-social factors, form an influencing variable for productivity of elderly. Changing age structures and elderly workforces notably require an early ergonomic design of working facilities in order to foster employees being healthy up until old age.

Since the early sixties, head-mounted displays (HMDs) have become popular technologies and topics in the fields of military, medicine or training. Currently, industry and commercial companies are interested in applying virtual technologies and corresponding hard- and software (Friedrich et al., 2002; Schreiber & Zimmermann, 2011). Likewise, HMDs or digital data glasses attract the attention when it comes to production systems enhanced by ICT (Kagermann, 2013). They may serve as intelligent solutions providing process-relevant knowledge contemporaneous to the process itself. This facilitates a flexible, robust and high-quality processing within planning, production, maintenance and logistics. HMDs have excellent characteristics for this purpose: placed in front of the eye, they allow working with both hands while having work related information right beside. Thus, it becomes conceivable to entrust even inexperienced workers with complex tasks and consistently ensure high quality. An intelligent, adaptive system for example could monitor employees in order to provide visual, context-sensitive feedback displayed on a near-eye display. Furthermore, HMCs may support elderly by showing health-related content such as recommendations for correct working postures and object handling. But in which way they need to be supported and if HMDs might have negative side-effects, is merely known. Before delving into specific application scenarios, designing the interaction processes or visualizing information, we examine to which extent HMD hardware pose an ergonomic impact on elderly compared to younger people. While ergonomic aspects are closely related to the experimental task, we focus on the manufacturing sector as a paragon for cyber physical systems with highly demanding socio-technical interfaces. The presented study therefore investigates ergonomic effects of head-mounted displays among different age groups. So far, ergonomic research related to the application of data glasses or head-mounted displays exposed drawbacks primarily concerning visual parameters, performance, simulator sickness and workload . But due to remarkable changes of hardware characteristics and different experimental tasks, results are hardly transferable to modern HMDs within an industrial context. In order to investigate whether HMDs are able to approach implications of demographic change, we formulate the following research questions:

RQ1: How do industry-relevant parameters of young workers behave, compared to those of elderly if HMDs are applied as assistive technology?

Related to research questions one (RQ1) we are going to test the following hypothesis:

$H1_1$: *There are differences concerning performance among different age groups performing manufacturing task supported by an HMD.*

$H2_1$: *There are differences concerning workload among different age groups performing manufacturing tasks supported by an HMD.*

$H3_1$: *There are differences concerning acceptance among different age groups performing manufacturing task supported by an HMD.*

In order to answer above-mentioned research question and test related hypothesis, we perform a secondary data analysis of data from a previous experiment (Theis et al., 2014) which was funded by the Federal Institute for Occupational Safety and Health in Germany. We restrict the original dataset to those participants who used an HMD in order to focus on the influence of age and head-mounted displays on ergonomic issues.

METHOD

Participants

Each of the participants (N=40) was assigned to one of two age groups (AG1: 18-39 yrs., AG2: 39-60 yrs.). Both groups were almost balanced in terms of gender. Because of the random assignment the group sizes were slightly unequal (AG1: n=22, AG2: n=18). 13 participants wore glasses; none of them wore contact lenses. Concerning their experiences 23 participants stated that they frequently perform easy manufacturing tasks on their own; 7 participants indicated that they have a lot of experiences with manufacturing tasks and 10 claimed to have no experience. Just one person has used a head-mounted display before. Exclusion criteria encompassed: chronic and acute musculoskeletal pain or disease, heave peripheral vision defects, dermatologic diseases, cognitive and communicative disorders, neurological and psychiatric disorders.

Table 1: Sample characteristics (N=40)

		AG 1	AG 2	Tot al
Dominant eye	Right	17	13	30
	Left	5	5	10
Gender	Male	12	11	23
	Female	10	7	7
Glasses	No	18	9	27
	Yes	4	9	13
Manufacturing experience	none	4	6	10
	moderat	16	7	23

	e			
	extensive	2	5	7
Total		22	18	40

Apparatus

Within our experimental setup, we applied the monocular HMD Lite-Eye LE A750 (see fig. 1). The display consists of a small metal display unit, which can be adjusted in front of the eye. The display unit is fixed to a head-support. A slider in front of the display facilitates switching between see-through and look-around mode. Display and head-support together exhibit a total weight of 120g. The OLED-display has a resolution of 800 x 600 px and a 30 degree field of view. It provides colored images with a brightness of 30 cd/m^2. In the black-and-white-mode brightness increases to 1800 cd/m2 and in the monochromatic-green mode the display provides a brightness of 3000 cd/m2. Diopter compensation is possible between +2 and -6. Color and contrast adjustment can be made by a hardware interface attached to the cable of the HMD. In order to minimize and exclude the cable's influences on user behavior during the experiment it was fed from the ceiling and clamped to participants belt loop.

Figure 1. Working instructions were displayed on a monocular HMD Lite-Eye LE A750.

Subjective perceived workload was quantified by means of the NASA Task Load Index (NASA-TLX, Hart, 2006). Workload in general is defined as the effort devoted by a person to perform a task. The NASA-TLX questionnaire consists of six rating scales in total, in which the mental, physical and temporal demand as well as own performance, effort and frustration are evaluated. The rating consists of two steps: First, the task is evaluated in relation to all six items each consisting of a twenty-stage bipolar rating scale with verbal descriptors on each end. Then by means of a pairwise comparison the participant defines the importance of the six subscales over all possible pair combinations. Hence, based on the individual importance of subscales, weights for each subscale are derived and multiplied with the subscale result. These values are summed and divided by the sum of weights. In this way results disembogue in a total weighted value. In addition to the total weighted value, each item can be analyzed separately, allowing diagnostic statements about specific workload problems.

Experimental task

The experimental task consisted of 3.5-hour manufacturing work. We divided it into four working segments of 48 min. in order to mimic field conditions. After the first and third working segment we admitted a 10-minute break. The second working segment was followed by a break of 20 min. Precise timing was ensured by the software, which displayed the working instructions on the HMD and on the supervisor's desktop screen. On the supervisors side the software interface displayed the remaining time until the next break, the actual and next working instruction and control elements. The examiner was able to insert a break within the process or navigate between working instructions. The participants screen displayed the instruction only (see fig. 2). After the participant finished a working step, he provided feedback and the examiner continued with the next working instruction. After 48 minutes the software switched automatically to a break-instruction and – after the break – back to the working instruction.

The experimental task itself consisted of manufacturing tasks on an Opel Omega B X20SE engine, carburetor, starter and generator. Performance strongly depends on the task, and task execution time accumulates to a difference between participants over time. Therefore, a primary and a secondary task were introduced. During the primary task the participants worked on the engine. Remaining motor vehicle parts served the secondary task. During both tasks, participants measured subparts of the engine and noted the results into a documentation sheet. A differentiation between primary and secondary task had the advantage that each participant started with the same task at the beginning of each working segment. Primary tasks were evaluated and arranged in a way that it took about 30 minutes to finish them. Subsequently, secondary tasks were processed until the 48 minutes of a working segment are over. After the break the next primary task started. In case that a primary task could not been not finished within a working segment the software displayed all primary tasks in sequence and the participants' performance was not considered in the analysis.

Figure 2. A working instruction the 3D model of the manufacturing object, textual instructions, graphical representation to the tools and a textual position indicator.

Serial working instructions were displayed on the participants' HMD and on the examiner's desktop screen. A 3D-model of the vehicle part served as digital representation of corresponding real world object. Within the 3D model parts to be processed were highlighted in red, textual instructions below the visualization briefly described how to proceed and the required tools were visualized on the right side. It was a non-augmented-reality visualization. The instructor passed the tool and received it together with the dissembled part. A textual indication of the manufacturing part's location was given on the working instruction and on each side of the engine.

Procedure

Each participant was randomly assigned to one HMD-condition and took part in the experiment only once. The initial procedure involved a variety of physiological calibration methods listed in Theis et al. (2014) and the total procedure took accordingly between six and seven hours. Only the following steps are relevant within the scope of present research questions: (1) Welcome and demographic questioning (age, prior experience, etc.), (2) brief introduction to the experimental procedure, hardware and tools, (3) question answering, (4) examination of the

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2103-6

ocular dominance, (5) identifying participant's body height and adjusting working environment to it, (4) experimental task, (5) NASA-TLX questionnaire and (6) interview.

Study design

The analysis of this study is based on a between-subject design in order to exclude learning effects. Originally the data consisted of a third group, which was supported by a conventional screen (Theis et al., 2013). This part of the experiment has been left out for this analysis.

RESULTS

Performance. In total, N=17 cases were analyzed for age group AG1 (18-39 yrs.). Missing values relate to participants who required more than 48 minutes for the primary task during one working segment. The percentage on the task completion time for AG1 [D(17)=0.18, p > .05] and the task completion time AG2 (D=0.20 , p < .05) were both significantly normal distributed. For the assumption of equality of variances we conduct Levene's test. For the percentage on the task completion time, the variances were equal for AG1 and AG2, [F(1, 29)=0.0, p>.05]. An univariate analysis of variance between age groups (one-way independent ANOVA) revealed no significant differences in performance, [F(1, 29)=0.57, p<.05]. In order to investigate performance over time, we need to consider task differences, even if they were small. Therefore, we normalized performance values by computing the quotient of the task execution time (TET) and mean task execution time of working segment (mean [TET]). Mean task execution time for the first working segment (task 1) is about 1962 seconds (SD=341 sec.) while the mean task execution time for the last working segment (task 11) is about 2113 sec. (SD= 289). Based on normalized task execution times (TET/mean[TET]) a repeated measures ANOVA shows, that age-dependent performance differs not significantly over time [F(1,1)=0.2, p>.05].

Workload. As the NASA-TLX questionnaire provides multiple interrelated variables, we conducted a multivariate analysis of variance. One way to check the assumption of multivariate normality is a univariate normality tests for each dependent variable. A Kolmogrov-Smirnoff test investigating normality, revealed inconsistencies concerning normal distributions for the NASA-TLX scales within each sample (see Table 2). As the Q-Q plots showed data that only slightly deviates from normality we emanate from multivariate normality and conduct a MANOVA. Influences of non-normality on power have to be considered. The assumption of equality of covariance matrices was tested by Box's test. While Box's test is susceptible to deviations form multivariate normality and does not support normality of the effort scale data within AG2. With p<.05 the homogeneity of covariance matrices was violated. Using Pillai's trace, there was a significant effect of age on the NASA-TLX scales [F(7, 32) = 2,33, p < .05]. However, as post-hoc analysis separate non-parametric univariate analysis on the outcome variables revealed non-significant effects on the NASA-TLX total score [F(1, 38) = 0.676, p > 0.05] on mental demand [F(1, 38) = 1.062, p > 0.05] on physical demand [F(1, 38) = 4.030, p > 0.05] on temporal demand [F(1, 38) = 0.010, p > 0.05]; on overall performance [F(1, 38) = 0.882, p > 0.05] and on frustration [F(1, 38) = 0.283, p > 0.05]. Instead, we found significant difference for the effect of HMD user's age on effort [F(1, 38)=5,12, p<.05] as well.

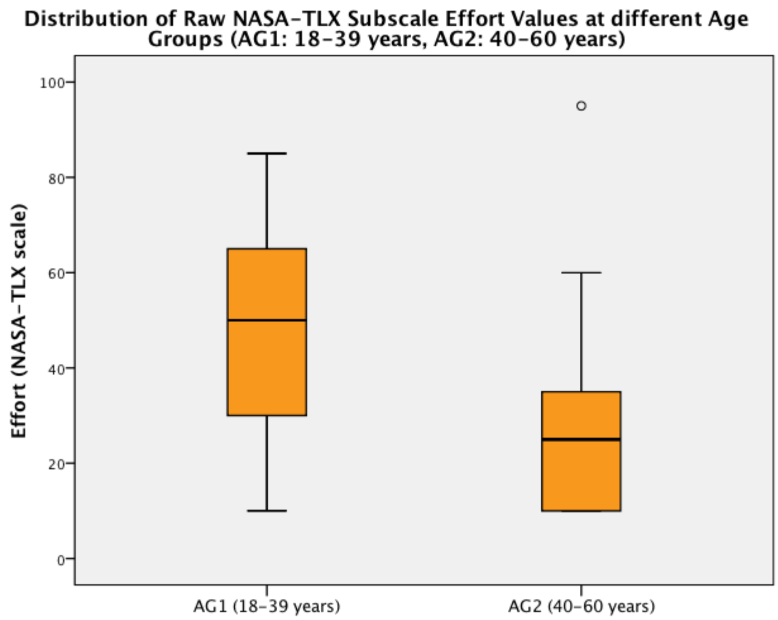

Figure 3. Subjective perceived effort of HMD users for the two age groups.

Acceptance. In order to investigate the acceptance of HMD characteristics of different age groups we coded answers with qualitative characters. If answers encompassed only two or three categories we ran a quantitative analysis, otherwise results were reported qualitatively. The question targeting at acceptance of the HMD was question Q2 (how do you judge the monocular display of information?), the others aimed at finding starting points for further research and intended to separate problems with the hardware from task related issues.

The following items were answered in a quantitative way (3 or less answer categories): (Q1) did you come along well with the task and the visualization? [Observed answers: yes, no, a bit], (Q2) how do you rate the monocular display of information? [Observed answers: positive, negative, neutral], (Q3) were you able to see the displayed information sharply? [Observed answers: yes, no, partly] Considering the categorical nature of the answers' data, we conduct a chi-square test in order to investigate differences between groups concerning issues contributing to acceptance. A chi-square test of independence was performed to examine the relation between age and Q1. The relation between these variables was not significant, $X^2(2, N= 35) = 4.264$, $p>.05$. 94.75% of age group one got along well with the task, as well as 68.8% of age group two. The question was posed in order to differentiate between problems caused by the task or by the hardware. Unlike the first question a significant relation was found between age and Q2, $X^2(2, N=37) = 11.84$, $p<.05$. While 55% of age group one judged the monocular display as positive, most participants of age group 2 (76.5%) judge the monocular display as negative. A chi-square test of independence was performed to examine the relation between age and Q3. The relation between these variables was significant, $X^2(2, N= 36) = 7.27$, $p<.05$. While no one of age group 1 had problems seeing sharply, ca. 31.1% of age group two encountered blurred vision, 56.2% of group two reports that they partially could see only blurred. Only 12.5% of age group 2 stated that they had no problems concerning blurred vision.

The qualitative questions investigated wearing comfort (Q4), altered perception (Q5) and gained insight into participant's ideas about hardware's potential for improvement (Q6), potential application areas (Q7) or further suggestions concerning HMD applications (Q8). Qualitative feedback revealed rather HMD specific characteristics than differences between groups.

Question Q4 (How do you judge wearing comfort of the HMD?) was answered by N=22 participants of AG1. They describe the wearing comfort predominantly as negative. People who stated the comfort as not negative indicated that it was 'ok' to use the device or that it was better than expected. Only one person stated that the HMD was good attachable to the head. Negative Feedback of AG1 can be clustered to one of the following categories: generic, concerning the head support, pressure, weight (distribution) and pain. While only one person reported neck pain, most participants indicated the head-support, pressure and weight as the most irritating characteristics of the display. All three aspects are related to each other. Participants of AG1 moaned that the head-support became unfastened, that it is too though and that bails are or could only awkwardly be placed. Pressure was especially exerted on the forehead and on the ears. They described the weight as unequally distributed and to heavy.

Within AG2, N=18 participants answered question Q4. The answers were similar to those of participants from AG1. They also described the head support as the most disturbing element with similar characteristics: could not be fastened adequately, too loose, and too tough, not absorbing sweat and had to be adjusted often during the working session. The weight was as well described as unequally distributed and too heavy. Pressure was less frequently mentioned then within AG1.

Vision and perception (Q5: Did you recognize altered perception and how would you describe it?) was described by N=8 participants of AG1. Headache was the most frequent word in this category. One could assume some kind of relationship between headache and vision parameters because it was not mentioned as comfort issue. Participants indicated that headache disappeared after some time and especially after the breaks, but it returned in the course of the experimental task. One participant even said that the headache got worse each time it came back. Breaks were perceived as helpful in reducing visionary problems and headache.

N=12 participants of AG2 answered the question Q5; two of them indicated that they had no problems. Participants here described similar issues. They reported primarily headache, blurred vision and a strong stress on eyes. One person encountered neck pain as well. Participants reported different length of familiarization phases and that negative symptoms rapidly disappeared with each break.

Q6, Q7 and Q8 did not differ between groups as well. The most often mentioned improvement was a wireless connection and an improved head-support. But also improvements of interaction techniques and advanced display adjustments in terms of distance to the eye were suggested. Some even suggested integrating the display into glasses, as done by current models like Meta, Google Glass or Epson Moverio BT-200.

DISCUSSION

The aim of this research was to quantify workload, performance and acceptance differences between younger and older participants supported by an HMD during manufacturing. Additionally it describes personal attitude and recommendations for the design and the use of HMDs. This study provides one of the first investigations of head-mounted displays considering implications of demographic changes.

Based on present results we reject $H1_{0,}$ which assumes **performance** differences among different age groups wearing an HMD as assistive technology. Accordingly we accept hypothesis $H1_1$ predicating no significant differences in performance among different age groups wearing an HMD. This finding supports the theory of differential aging rather than the deficient model of aging. Differential aging postulates performance as not necessarily declining with physiological or psychological parameters due to compensational abilities. We want to emphasize that differences in performance cannot be assigned to either age or HMD – and this is true for each of the investigated parameter. It results solely from a combination of both given specific task and hardware.

In general, current study reveals differences of subjective perceived **workload** among different age groups. This leads us to acceptance of hypothesis $H2_1$, even if solely the NASA-TLX subscale “effort” differed. Neither the total NASA-TLX scale nor remaining subscales distinguished between groups. However, elderly subjectively perceive lower efforts than younger participants. These findings correspond to the results of Wille et al. (2013).

Furthermore, significant differences appeared between both age groups concerning the **acceptance** of the display. During our subjective ratings, we asked participants how they liked the monocular display. While 55% of age group one rated the monocular display as positive, 76.5% rated it as negative. Based on these findings we accept hypothesis $H3_1$ that there are significant differences concerning acceptance among different age groups performing manufacturing task supported by an HMD. As a consequence, the feedback emerging from the interview becomes even more important. Although no differences between groups could be observed, it provides valuable suggestions for future improvement of the HMD.

It cannot be concluded that HMDs necessarily support elderly better than other devices. Comparisons between HMDs and mobile assistive technology devices are described in greater detail by Wille et al. (2013) and Theis et al., (2014). Furthermore, it has to be taken into account, that the participants of our experiments were new to the task. They were strongly dependent on the instructions and the display; otherwise they could not have performed it. That makes our results only transferable to situations where workers are exposed to new tasks. It should be investigated in future how the performance of experienced elderly changes if an HMD is applied during familiar tasks. But given the implications of demographic change, changing conditions together with new tasks are likely to constitute an integral part of working life. That makes our results highly relevant for employers planning to support elderly workers by a HMD.

CONCLUSION

In summary findings suggest that elderly perceive manufacturing work with an HMD less effortful while their performance remains comparable with the performance of younger people. At the same time, elderly's acceptance of HMDs remains a challenge, while younger participants are more open-minded here. Head-mounted displays though relate to demographic change in a way that they do not influence elderly's performance differently than younger ones and they might constitute an eligible mean to support workers during complex and even physically challenging tasks. Certainly, age-dependent differences and general hardware design issues need to be addressed.

REFERENCES