Study-based evaluation of accuracy and usability of wearable devices in manual assembly

Barbara Tropschuh, Susanne Windecker & Gunther Reinhart

Published online: 14 Jul 2022.

Submit your article to this journal ⤤

Article views: 1383

View related articles ⤤

View Crossmark data ⤤

Taylor & Francis
Taylor & Francis Group

Study-based evaluation of accuracy and usability of wearable devices in manual assembly

Barbara Tropschuh, Susanne Windecker and Gunther Reinhart

Institute for Machine Tools and Industrial Management (*iwb*), Technical University of Munich, Garching, Germany

ABSTRACT

The fourth industrial revolution shapes today's private and industrial environments, implying increased digitalization, connectivity, and artificial intelligence. Wearable devices support digital communication by displaying information and monitoring health-related aspects by measuring vital signs. Even though various wearables for measuring vital signs are already used in private life, they have not yet found their way into the production environment. This could be due to poor data quality or a lack of acceptance among employees regarding the usability of wearables during work activities. This paper aims to evaluate the accuracy and usability of selected wearable devices in manual assembly. Therefore, two user studies were conducted in a rebuilt production environment. The first study focuses on the data accuracy of the heart rate measurement of different wearables during manual assembly. In the second study, the usability of the selected wearables is evaluated with the thinking-aloud method during a manual assembly task and questionnaires.

1. Introduction

In today's turbulent industrial environment, manufacturing companies are confronted with contrary developments regarding employees and workplaces. Due to the demographic change, companies face an aging workforce with declining physical and mental performance requirements (Gajewski et al., 2018). At the same time, the increasing number of product variants with low repetition rates and frequent changes in the work task lead to more complex workplaces (ElMaraghy et al., 2013). The adaptation to these evolving requirements poses various challenges, especially for elderly employees (DAK-Gesundheit, 2018). In this way, mental and physical overstrain can arise, leading to assembly errors in the short term and sick days in the long term (Rusnock & Borghetti, 2018). To detect this overstrain as well as to prevent work errors and sick days, wearable devices can be used to measure employees' vital signs to master their health in production, e.g. measuring the heart rate with a fitness tracker or an ear sensor (Storm, 2020). Due to the progress in electronics, microsystems, and information technology, new

devices with enhanced functions constantly enter the market (BMBF, 2018). These devices are already used extensively in private life. However, they have hardly been implemented in the production environment (Teucke et al., 2020).

This paper examines the accuracy and usability of wearable devices measuring vital signs in manual assembly. Section two reviews the literature on data accuracy and usability of wearables. Afterward, the accuracy and usability of different wearables are analyzed in two user studies in the Innovation Lab of the *iwb* (section three). Thereby, test persons conducted a specific assembly task in a replicated production environment. The results of the studies are described in section four. An interpretation and discussion of the results follow in section five before this paper is summarised.

2. Wearable devices in manual assembly

Wearable devices represent a subset of smart devices worn directly on or in the body. Sensors, cameras, and microphones are embedded in these devices to record personal and environmental data (European Commission, 2014). According to Guk et al. (2019), wearable devices are classified into portable, attachable, implantable, and ingestible devices (Guk et al., 2019). These devices can, for example, continuously monitor employees' vital signs, postures, or actions. In the approach of Dimitropoulos et al. (2021), a human-robot collaborative cell is described where the robot adapts to the ergonomic and process needs of the worker using data from wearable devices (Dimitropoulos et al., 2021). Gkournelos et al. (2018) also focus on the collaboration between robots and humans by developing a smartwatch application to directly interact with the robot, e.g. via voice or manual guidance (Gkournelos et al., 2018). By monitoring employees' vital signs, physical and mental overstrain can be detected (Teucke et al., 2020). Thus, corrective measures can be initiated to avoid long-term effects on the employees (Teucke et al., 2020). Peruzzini et al. (2017) present an overview of parameters for human factors monitoring in industry 4.0, e.g. heart rate or skin temperature, with suitable monitoring tools for user experience analysis (Peruzzini et al., 2017). Moreover, Peruzzini et al. (2020) use these human factors to assess workers' ergonomics performance and perceived comfort with eye-tracking systems and wearable biosensors (Peruzzini et al., 2020). Furthermore, various comparative studies analyzed the data accuracy of, e.g. step counts, sleep duration, or estimated energy expenditure (EEE) recorded by different wearable devices, e.g. Fitbit One and Zip, Jawbone UP, or Nike Fuelband (Case et al., 2015; Ferguson et al., 2015; Lee et al., 2014; Wallen et al., 2016). The results of these studies show that the wearable devices are consistently less accurate than the actual step count or the measures of research-grade devices (e.g. chest strap, ECG-electrodes). Research-grade devices are devices for which data quality and accuracy have already been scientifically verified. These devices are now used as comparison devices. The above results were obtained both under laboratory conditions and under real conditions. For instance, (Wang et al., 2017) conducted a comparative study to determine the heart rate (HR) accuracy of selected wrist-worn wearable devices. The heart rate data tracked by the Apple Watch, Mio Fuse, Fitbit Charge HR, and Basis Peak were compared to the Polar H7 chest strap data within a laboratory setting. The study revealed that none of the examined wearables achieved the accuracy of the chest strap (Wang et al., 2017). Besides the data validity of heart rate, the comparative study by Wallen et al. (2016) also

considered the accuracy of step count and the estimated energy expenditure (EEE) of wrist-worn devices (Wallen et al., 2016). Therefore, the authors compared the accuracy of the *Apple Watch, Fitbit Charge HR, Samsung Gear S*, and *Mio Alpha* against research-graded measures under laboratory conditions. The study's main findings inlcude that there is no device superior to the comparative devices and, on a variable level, the HR data is consistently more accurate than EEE and step count throughout all devices (Wallen et al., 2016). Regarding the accuracy of wearable devices, the research focuses on evaluating wrist- and hip-worn wearables of the category of portable devices. Moreover, the papers concentrate on the accuracy of wearable devices for use in private life and healthcare. Thus, no studies have considered the accuracy of the wearables during production-related activities. Furthermore, new devices with improved functions and better data quality enter the market, making it necessary to check whether the improved data quality is already suitable for use in production. Regarding the use of wearable devices in occupational environments, Mettler & Wulf (2019) analyze the affordances and constraints of wearable use from an employee's perspective (Mettler & Wulf, 2019). They analyze possible physiologic measurement systems (e.g. measuring physical activity with step count) and constraints like privacy or technological independence. Privacy concerns and the associated restrictions regarding personal freedom and individuality play an essential role in using wearables in a company environment (Mettler & Wulf, 2019). Luse & Burkmann (2020) focus on the use of RFID wearables in the workplace, analyzing privacy concerns (Luse & Burkman, 2020). The results show that being monitored has a negative impact on employee satisfaction. Therefore, higher transparency during implementation can help integrate such new devices into the corporate environment (Luse & Burkman, 2020). Usability is very closely linked to the acceptance of the employees. Khakurel et al. (2020)_describe usability issues related to wearable devices (Khakurel et al., 2020). Thereby, they identified three categories regarding usability challenges. These are device characteristics, deployment on the body, and the external devices used to synch with the wearables. Due to motion artifacts, device characteristics and wearing position can negatively affect data accuracy or connectivity (Khakurel et al., 2020).

This paper compares the data accuracy of a wrist-worn, ear-worn, and attachable device to a criterion measure (Polar H10 chest strap) under production-specific conditions. In addition to the accuracy aspects, a second user study applying the think-aloud method is conducted to externalize cognitive processes concerning selected wearable devices while performing assembly tasks. This provides insights into relevant usability topics in the production environment.

3. Research method

To measure the data accuracy and usability of selected wearable devices in manual assembly, two user studies were conducted at the Innovation Lab of the *iwb*. The test setup consisted of an assembly workstation with an axle carrier and a flow rack containing all the mounting parts as well as the required assembly tools. A digital worker information system displayed the right assembly locations on the axle carrier. The test persons assembled and disassembled mounting parts, such as hoses, clamps, beams, and bolts on a truck axle carrier at the rebuilt cycle-based

Figure 1. Study setup at the Innovation Lab of the *iwb*.

manual assembly workstation. Figure 1 shows the study set up at the Innovation Lab of the *iwb*. The overall overarching study concept for measuring physical and mental strain during manual assembly with smart devices was approved by the ethics committee of the Technical University of Munich (No. 388/20 S-KH). In addition, participation in the study was voluntary, and each test person signed an informed consent form. The participants in both studies were between 19 and 28 years old. Due to Covid-19 restrictions, it was not possible to include external industrial test persons in the study. Therefore, only students and members of the Institute for machine tools and industrial management participated. The wearable devices under evaluation were chosen according to the selection method in (Tropschuh et al., 2020), considering restricting criteria of the workplace and discussions with partners from the automotive industry to integrate a practical view on smart devices. Thereby, heart rate was selected as the measurement parameter, restrictions of the production environment were identified (e.g. no protruding buttons on watches), and a final selection of possible devices was conducted with the help of industry partners to consider real company criteria such as economic aspects. The selected devices are the cosinuss°One ear sensor, the Garmin vívosmart 4 fitness wristband and the skin patch movisens EcgMove 4 (see, Figure 2). For example, the Garmin vívosmart 4 fitness wristband was selected due to its pure form (no protruding buttons), the small size, and the moderate price compared to Apple watches. The Polar H10 chest strap is considered a reference device. The following sub-sections address the systematic approach of the two studies separately.

Figure 2. Selected wearable devices to be evaluated.

3.1. Accuracy of wearable devices in production

The first study focuses on the accuracy of the heart rate measurement of the selected wearable devices in the rebuilt production setup. The heart rate has been the most frequently used value for measuring and evaluating health status (Dias & Paulo Silva Cunha, 2018). For this reason, the current study uses this vital sign to compare the accuracy of the cosinuss°One ear sensor, the Garmin vívosmart 4 fitness wristband, and the skin patch movisens EcgMove 4 against the data from the Polar H10 chest strap. Due to its high data accuracy, the Polar H10 chest strap has already been used as a reference device in past research (e.g. Dooley et al., 2017; Gillinov et al., 2017; Wang et al., 2017). The sample size of 25 participants (9 women, 16 men) resulted mainly from the timing of the study between two covid lockdown phases. As 25 test subjects are sufficient for initial suggestions and statements on the subject, the data collection was ended. The 25 participants wore the four devices simultaneously during the user study while performing the assembly task. This assembly task is performed twice at the replica workstation with different cycle times (7.30 min, 5.45 min).

The data analysis is carried out in three steps: In the first step, descriptive statistics were used to identify the essential characteristics of the individual data sets. Additionally, Spearman's rank correlation analyses were calculated to determine the strength of the

monotonic relationship between the HR data of the devices. In the second step, the data sets of the evaluated wearable devices were mapped to the chest strap data according to the timestamps. Regarding the wearables, the fitness tracker measured every second if the wrist movement was not too much. The ear sensor tracked the heart rate every two to three seconds, and the skin patch, as well as the chest strap, were measured continuously. All data sets from the beginning of the assembly were considered, excluding data due to incorrect configuration of the device or if the device didn't measure the heart rate due to movements. To compare the heart rate, only the timestamps that are available for both devices were compared. The conformity of the wearables towards the reference device is assessed and quantified using Bland-Altman diagrams. These plots allow comparing an established measurement method (i.e. the Polar H10) with a new measurement method (i.e. the Garmin vívosmart 4, the cosinuss°One, and the movisens EcgMove 4) by identifying the range of variation using 95 % limits of agreement (LoA) (Bland & Altman, 1999). This way determined whether two measurement methods were of similar quality (Bland & Altman, 1999). In the third evaluation step, Wilcoxon's signed-rank tests were performed to determine whether the HR data recorded by the evaluated devices were statistically significantly different from the data measured by the reference device. Furthermore, the mean percentage error (MPE) and the mean absolute percentage error (MAPE) values were calculated as the average absolute value of the errors of each device relative to the criterion measures to measure how accurate the devices are compared to the reference device.

3.2. Usability of wearable devices in production

In the second user study, the usability of the selected wearable devices is evaluated. Fifteen new participants (aged 19 to 28 years, 7 women and 8 men) performed the same assembly tasks as in the previous study in four iterations. In each iteration, a different wearable device is worn by the participant and assessed with the think-aloud method as well as a follow-up paper-based questionnaire. The think-aloud session is intended to generate insights into relevant usability aspects and which features of the selected wearables are perceived positively or negatively by the participants in the assembly environment. To capture these feelings and preferences as part of their underlying cognitive processes, participants are encouraged to verbalize anything that comes to mind associated with the worn device while performing the assembly task. In contrast, the follow-up questionnaire for the respective device after each assembly round serves to quantitatively obtain specific information on the topics of intrusion, portability, ease of use, wearing comfort, and interchangeability. Each criterion expresses one attribute, except for portability, which describes two characteristics, i.e. shape and dimensions and weight. These attributes are measured on a 5-point Likert scale, with strongly agree and strongly disagree anchor points (e.g. The chest strap is comfortable to wear. | I fully agree → I do not agree at all). Additionally, two open-ended questions focus on which attributes and features the participants like best or worst about the device, e.g. What features did you like best about the Skin Patch?/What did you notice most positively when wearing the Skin Patch?. The questionnaire's general, cross-device part is answered at the end of the fourth iteration. Thereby, the preferred device for use in manual assembly is selected among the four evaluated devices, and the reasons for this selection
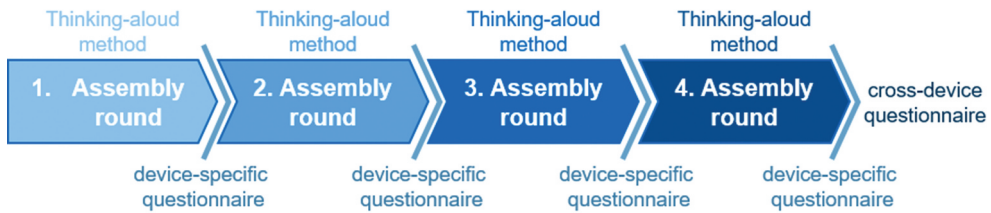
Figure 3. Protocol of the usability study.

should be provided. Participants conduct the study with a random order of wearables to eliminate bias and increase data quality across all devices. The first iteration might require participants to focus more on the underlying task even though a trial assembly run was carried out before the first iteration. Figure 3 shows the protocol of the usability study with the used methods.

The analysis of the think-aloud method starts with the review of the protocols (Tesch, 1990). This review serves to gain a comprehensive understanding of the nature of the preferred characteristics of wearable devices in general and in the assembly context, in particular. On this basis, common themes are determined and coded to derive the frequencies of occurrence of the aspects identified. For the follow-up questionnaires, the answers to the open-ended questions regarding the most positive and negative characteristics of the respective device are combined with those of the think-aloud protocols and jointly evaluated according to the identified common aspects. The answers to the closed questions are evaluated using descriptive statistics. This approach is employed for the per-device questions based on the predefined criteria and the cross-device question where the preferred device is selected.

4. Description of the results

4.1. Accuracy of wearable devices in production

The descriptive statistics reveal that the differences in the HR data between the cosinuss° One and the reference device are the greatest. In contrast, the differences between the movisens EcgMove 4 and the reference device are the smallest (see, Figure 4). The box plot was chosen to display the data distribution of the different devices graphically. The colored rectangle, the box, indicates the area that contains the middle 50 % of the data (Emerson & Strenio, 2000). It is limited by the lower and upper quartiles, and the line in each box represents the median of the values. The adjacent antennas at the top and bottom, also called whiskers, represent the mild outliers (Emerson & Strenio, 2000). The points outside the whiskers are extreme outliers, which either represent true extreme values or indicate device malfunctions. All relevant requirements underlying the Bland-Altman plot are fulfilled. The Bland-Altman diagram is a method for comparing two measurement methods with the graphical display of the data deviation (Martin Bland & Altman, 1986). The green line M indicates the mean value of the data difference between two measurement methods. By distributing the deviation points, the range of variation can be visualized, and it can be checked whether one measurement method measures
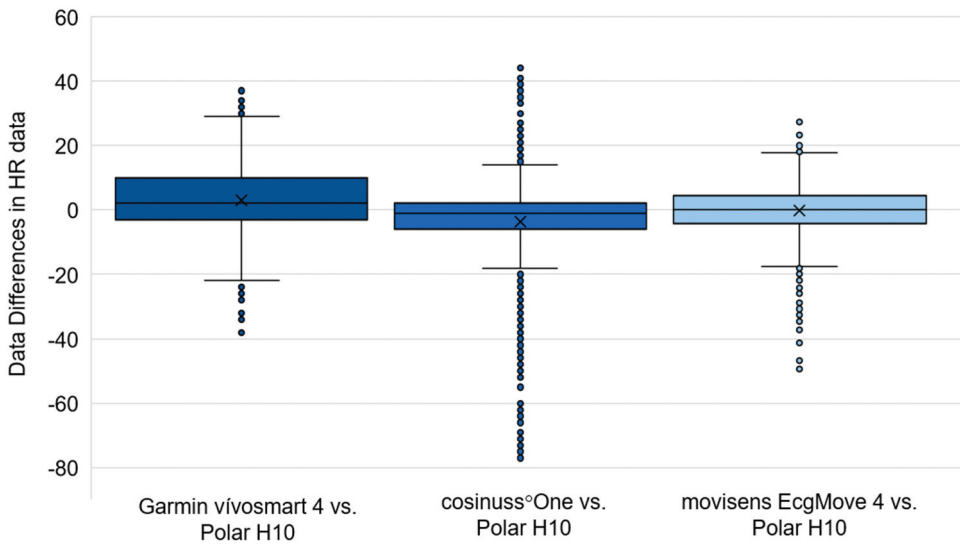
Figure 4. Comparison of the differences in HR data.

higher or lower than the other in principle (systematic measurement error). The Bland-Altman plots indicate that the Garmin vívosmart 4, on average, overestimates the HR data compared to the reference device. In contrast, the cosinuss°One and the movisens EcgMove 4 underestimate this data (see, Figure 5). Furthermore, visual inspection of the Bland-Altman plots for the Garmin vívosmart 4 and the cosinuss°One illustrates that
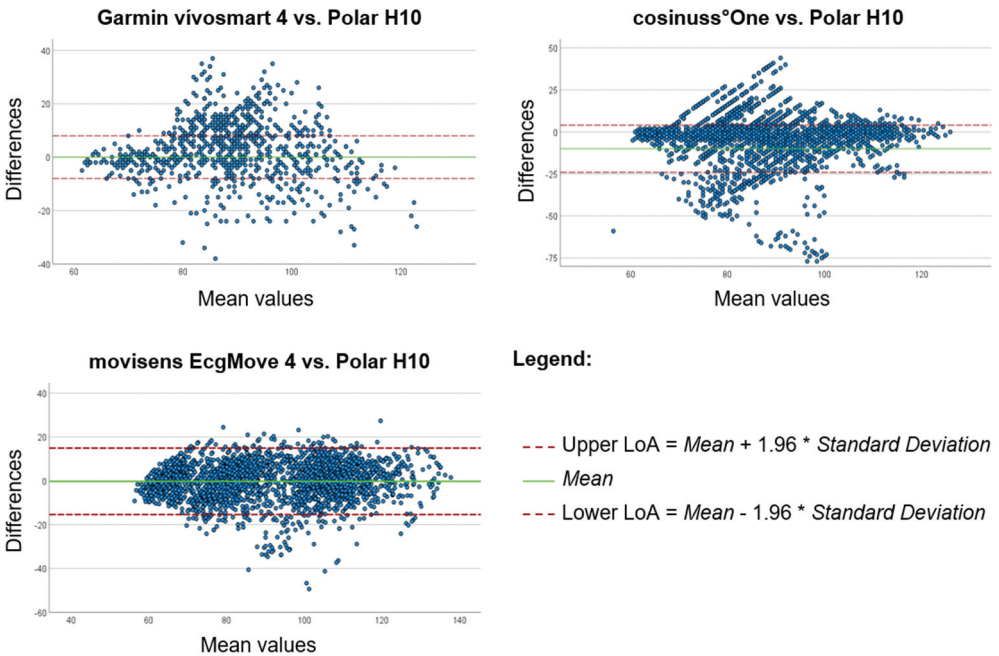

Figure 5. Bland-Altman plots of the devices to be evaluated.

more than one-third of the mean differences lie outside the limits of agreement (LoA), i.e. the lower and upper LoA. The LoA is calculated according to (Martin Bland & Altman, 1986) so that 95 % of the received values are contained. Consequently, these measurement methods do not represent acceptable deviations from the Polar H10. In contrast, almost all mean differences lie within the range of variation for the movisens EcgMove 4. Hence, this device represents an acceptable deviation and, thus, a valid alternative to the Polar H10.

Since a normal distribution of the data is not assumed, the following non-parametric statistical tests are performed. Spearman's rank correlation coefficient reveals a strong correlation for all devices under evaluation with the reference device (ear sensor and chest strap: p = 0.575; fitness tracker and chest strap: p = 0.612; skin patch and chest strap: p = 0.911). Calculating the Wilcoxon signed-rank test with a significance level of 5 %, the significant results are obtained for the Cosinuss chest strap combination (z = 14.95, p < 0.001) and the Garmin-chest strap combination (z = −8.05, p < 0.001). The Wilcoxon test was not significant for the combination of movisens and chest strap (z = −0.42, p = 0.672). This indicates that the cosinuss°One and the Garmin vívosmart 4 have a higher deviation in HR data compared to the polar chest strap, whereas the movisens EcgMove 4 has a very low deviation but is not statistically significant. Regarding the Garmin vívosmart 4, the mean percentage error (MPE) is 3.44 % and the mean absolute percentage error (MAPE) is 9.70 %. For the cosinuss°One, the MPE is 4.33 % and the MAPE is 9.69 %. The slightest error was detected for the movisens EcgMove 4 with an MPE of 0.46 % and a MAPE of 6.60 %.

4.2. Usability of wearable devices in production

Based on the qualitative part of the study, ten usability topics are identified by a theoretical deductive literature approach: wearing comfort, ease of use, restrictions in performing the assembly task, and fit represent the predominant themes across all devices. In contrast, hygiene, appearance, wearing with other equipment, sustainability, damage during assembly, and occupational safety are less prevalent.

For the Garmin vívosmart 4 fitness wristband, almost all participants stated that it is comfortable to wear and easy to use, like an ordinary watch. In addition, the lightweight and small dimensions of the wristband were positively emphasized. However, some participants reported that the rubber wristband felt uncomfortable on the skin, especially when sweating. In addition, it was difficult for almost all participants to navigate through the device without a prior explanation. The reasons are the small display, inconvenient haptic feedback, and undefined symbols. However, many participants indicated that they expected to get along well after a short learning period. Regarding restrictions in performing the assembly task, most participants did not feel restricted by the wristband. Nevertheless, occupational safety problems and damage during the assembly task could occur when the wristband gets caught or the employee gets distracted by the display's information. Some participants positively mentioned the appearance of the wristband as high quality.

Regarding the cosinuss°One ear sensor, the majority of the test persons could not imagine becoming accustomed to the device and wearing it for an extended period as the device felt uncomfortable and unfamiliar. In addition, almost all participants highlighted

that the device limits the hearing ability. This could also cause safety problems since warning signals might not be noticed. Some participants mentioned that the ear sensor does not fit properly and could slip out. However, many participants considered it positively that the ear sensor does not restrict movement during the assembly task. Nevertheless, a few participants felt that they performed the assembly task more slowly not to slip out the ear sensor. Fitting the sensor to the ear was simple for some participants as they found this sensor to fit like regular headphones. However, the simultaneous wearing of the ear sensor with other equipment (glasses, hearing aid, ear protection) was difficult and unpleasant.

Regarding the Polar H10 chest strap, most participants stated that the chest strap felt uncomfortable to wear and described the strap itself as very disturbing. Numerous participants experienced that the strap cut into their skin while at the same time did not remain firmly in one place. Almost all participants stated that the wearing comfort decreased while performing the assembly task as the chest strap slipped even more during movements. Expectations about becoming accustomed to the strap varied across participants. Besides, it was positively commented that the Polar H10 had no immediate impact on the execution of tasks. Considering the ease of use, most participants found the strap intuitive to apply, whereas adjusting the length of the strap was not easy when it was already worn.

In terms of the movisens EcgMove 4 skin patch, most participants did not consider the device as uncomfortable but unusual. Therefore, the majority stated that they became accustomed to wearing the skin patch within a few minutes. Most participants noticed positively that the device had no direct contact with the assembly task. In some movements, especially rotational movements in the upper body or with the left arm, a restriction was perceived as the sensor is attached slightly to the left side of the body. Regarding the ease of use, almost all participants noted that the adhesive patches could only be attached with detailed instructions. Furthermore, a considerable amount of force is necessary to reattach the sensor to the adhesive patches leading to pain in the chest while clipping it in. Concerning the fit of the skin patch, the majority of the test persons stated that they were afraid that the adhesive would not hold properly. In contrast to the other devices, sustainability was addressed several times as the sensor is attached with two disposable patches. New patches must be applied each time the sensor is worn.

Considering the best and worst rated device per criterion, the following can be observed (see, Figure 6): The wristband scores best on all criteria except ease of use. The chest strap ranks best in terms of ease of use. However, it is also the most negatively evaluated device for the criterion of comfort. The ear sensor performs the worst across all criteria, except for comfort. Meanwhile, the skin patch is not rated best or worst for any criterion, with average values determined in each case. Regarding the cross-device evaluation, the majority of participants prefer the wristband. Including the option of multiple responses, twelve out of fifteen participants opt for the wristband, three for the chest strap, and two for the skin patch. Meanwhile, no person chooses the ear sensor as the preferred device. The reason given by the test persons for choosing the wristband was that they were already used to this type of device.
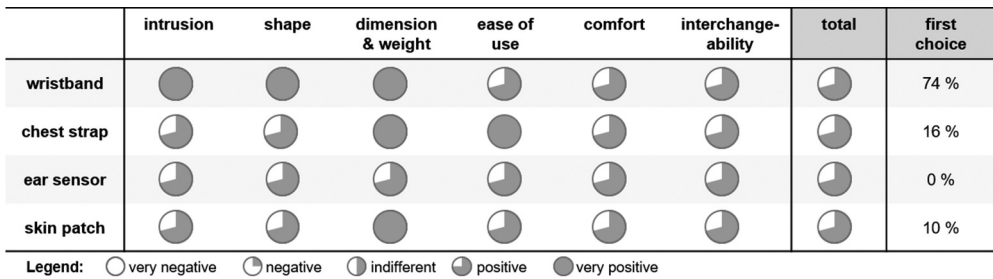
	intrusion	shape	dimension & weight	ease of use	comfort	interchange-ability	total	first choice
wristband	●	●	●	◔	◔	◔	◔	74 %
chest strap	◔	◔	●	●	◔	◔	◔	16 %
ear sensor	◔	◔	◔	◔	◔	◔	◔	0 %
skin patch	◔	◔	●	◔	◔	◔	◔	10 %

Legend: ○ very negative ◔ negative ◑ indifferent ◕ positive ● very positive

Figure 6. Overview of the evaluation of the selected devices by predefined criteria and preferred wearable devices of the participants (first choice).

5. Interpretation and discussion

The user study on the accuracy of selected wearable devices in manual assembly provides statistical evidence that the movisens EcgMove 4 represents a valid method for measuring the HR during production-related activities. By contrast, the Garmin vívosmart 4 and the cosinuss°One show high variability in HR data compared to the reference device. This implies that these two devices might be insufficiently accurate to measure the HR of production employees.

Regarding the usability of the selected wearable devices in manual assembly, the qualitative part of the user study reveals that wearing comfort, ease of use, restrictions in performing the assembly task and fit constitute the prevalent usability topics for all devices. The quantitative part of the study shows that among all devices evaluated, the Garmin wristband and the Cosinuss ear sensor were rated best and worst, respectively. This rating is also reflected in selecting the favorite device, where 80 % of the participants indicated that the Garmin wristband would be the preferred device to use in the production environment. This was because the participants were already familiar with this type of device, as it is similar to a conventional watch. This shows that employees tend to be more accepting of technologies they are already familiar with in their private lives than devices they have never been exposed to before. Furthermore, sustainability aspects like reusability of the devices and no single-use patches were especially important for the participants, as well as a self-explanatory user interface and a small size of wearables. These aspects, in particular, should be taken into account when designing next-generation wearables.

It is important to note that the current findings exhibit certain limitations. First, the results might be restricted because of the selection of participants (young, healthy, educated) and the small set of wearable devices (one device per category, i.e. wrist-worn, ear-worn, attachable). Furthermore, the analyzed data accuracy, usability, and comfort of the selected devices could be related to their brand and model. Therefore, generalizing the results of a single device to an entire product category is restricted. Second, the effects on the comparison between the Garmin vívosmart 4 and the reference device Polar H10 may be limited because the data of the Garmin device is only available in an irregular manner, without standardized measurement intervals (e.g. no measures for a minute). Therefore, it is possible that the actual time at which a value was tracked and the timestamp do not match. Third, the think-aloud method cannot ensure that every participant's thought is captured as only the verbalized aspects can be recorded.

6. Conclusion and outlook

This paper analyses the accuracy and usability of a few selected wearable devices in manual assembly. The results demonstrate statistical evidence that the movisens EcgMove 4 is valid for measuring HR data. At the same time, the Garmin vívosmart 4 and the cosinuss°One showed high variability relative to the reference device. Therefore, the latter two devices might not be accurate enough to collect HR data in manual assembly. The study on the usability of wearable devices combined qualitative (think-aloud) and quantitative (questionnaire) methods within a production-specific laboratory setting. The qualitative part of the user study revealed that the topics wearing comfort, ease of use, restrictions in performing the assembly task, and fit were the most critical topics across devices. Based on the quantitative part of the study, it became evident that the Garmin vívosmart 4 was rated best among all the devices evaluated.

Comparing the results of the two user studies, it can be seen that users most accept the fitness wristband, but it provides only insufficient data quality. Therefore, a compromise often must be made when selecting suitable wearable devices. For example, the skin patch achieved average evaluations regarding the usability aspects and, at the same time, reached high data quality.

The next step in our research is to verify the results in a field study with production employees of different ages and physical conditions. Furthermore, different devices could be included, e.g. implantable and ingestible devices and other devices of the already used wearable device categories. In addition, further studies will be conducted to validate the results with multiple devices per category (e.g. ear sensors or fitness trackers). On the one hand, this will allow generalizable, group-specific statements and the best devices for use in assembly to be identified. Moreover, further improved devices will enter the market, which should also be tested in terms of data quality and comfort. For practical application, restrictions and regulations regarding the privacy and tracking of employees need to be considered. Furthermore, it must be ensured that the collected data cannot be used for direct performance measurement of individual employees but are only available in an anonymous form, e.g. for planning and scheduling.

Disclosure statement

Funding

References

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods In Medical Research*, 8(2), 135. https://doi.org/10.1177/096228029900800204

BMBF. (2018). *Microelectronics from Germany - Driver of innovation for the digital economy: The German Federal Government's Framework Programme for Research and Innovation 2016-2020.* Federal Ministry of Education and Research (BMBF) Division for Electronics; Autonomous Electric Driving.

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA*, *313*(6), 625. https://doi.org/10.1001/jama.2014.17841

DAK-Gesundheit. (2018). *Betriebliches Gesundheitsmanagement.* Kordt, Martin: DAK Forschung.

Dias, D., & Paulo Silva Cunha, J. (2018). Wearable health devices-Vital sign monitoring, systems and technologies. *Sensors (Basel)*, 18.

Dimitropoulos, N., Togias, T., Zacharaki, N., Michalos, G., and Makris S. (2021). Seamless human–robot collaborative assembly using artificial intelligence and wearable devices. *Applied Sciences*, *11*(12), 5699. https://doi.org/10.3390/app11125699

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR Mhealth Uhealth*, *5*(3), e34. https://doi.org/10.2196/mhealth.7043

ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tseng, M. & Bernard, A. (2013). Product variety management. *Cirp Annals*, *62*(2), 629. https://doi.org/10.1016/j.cirp.2013.05.007

Emerson, J. D., & Strenio, J. (2000). *Boxplots and batch comparison, in understanding robust and exploratory data analysis* (pp. 58). Wiley.

European Commission, 2014. Article 29 data protection working party: Opinion 8/2014 on the on recent developments on the internet of things.

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *Int J Behav Nutr Phys Act*, *12*(1), 42. https://doi.org/10.1186/s12966-015-0201-9

Gajewski, P. D., Ferdinand, N. K., Kray, J., & Falkenstein, M. (2018). Understanding sources of adult age differences in task switching: Evidence from behavioral and ERP studies. *Neuroscience & Biobehavioral Reviews*, *92*, 255. https://doi.org/10.1016/j.neubiorev.2018.05.029

Gillinov, S., Etiwy, M., Wang, R., Blackburn, G., Phelan, D., Gillinov, A. M., Houghtaling, P., Javadikasgari, H., & Desai, M. Y. (2017). Variable accuracy of wearable heart rate monitors during Aerobic exercise. *Medicine and Science in Sports and Exercise*, *49*(8), 1697. https://doi.org/10.1249/MSS.0000000000001284

Gkournelos, C., Karagiannis, P., Kousi, N., Michalos, G., Koukas, S. & Makris, S. (2018). Application of wearable devices for supporting operators in human-robot cooperative assembly tasks. *Procedia CIRP*, *76*, 177. https://doi.org/10.1016/j.procir.2018.01.019

Guk, K., Han, G., Lim, J., Jeong, K., Kang, T., Lim, E.-K., & Jung, J. (2019). Evolution of wearable devices with real-time disease monitoring for personalized healthcare. *Nanomaterials (Basel)*, *9*(6), 813. https://doi.org/10.3390/nano9060813

Khakurel, J., Porras, J., Melkas, H., & Fu, B. (2020). A comprehensive framework of usability issues related to the wearable devices. In S. Paiva, & S. Paul, (Eds.), *Convergence of ICT and smart devices for emerging applications* (pp. 21). Springer International Publishing; Imprint Springer.

Lee, J.-M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1840. https://doi.org/10.1249/MSS.0000000000000287

Luse, A., & Burkman, J. (2020). Wearables in the workplace: Examination using a privacy boundary model. *Journal of the Midwest Association for Information Systems*, *2020*(2), 7. https://doi.org/10.17705/3jmwa.000058

Martin Bland, J., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, *327*(8476), 307. https://doi.org/10.1016/S0140-6736(86)90837-8

Mettler, T., & Wulf, J. (2019). Physiolytics at the workplace: Affordances and constraints of wearables use from an employee's perspective. *Information Systems Journal*, *29*(1), 245. https://doi.org/10.1111/isj.12205

Peruzzini, M., Grandi, F., & Pellicciari, M. (2017). Benchmarking of tools for user experience analysis in industry 4.0. *Procedia manufacturing*, *11*, 806. https://doi.org/10.1016/j.promfg.2017.07.182

Peruzzini, M., Grandi, F., & Pellicciari, M. (2020). Exploring the potential of operator 4.0 interface and monitoring. *Computers & Industrial Engineering*, *139*, 105600. https://doi.org/10.1016/j.cie.2018.12.047

Rusnock, C. F., & Borghetti, B. J. (2018). Workload profiles: A continuous measure of mental workload. *International Journal of Industrial Ergonomics*, *63*, 49. https://doi.org/10.1016/j.ergon.2016.09.003

Tesch, R. (1990). *Qualitative Research: Analysis Types and Software*. Falmer Press.

Teucke, M., Heuermann, A., Thoben, K.-D., & Freitag, M. (2020). Einsatz mobiler Computersysteme im Rahmen von Industrie 4.0 zur Bewältigung des demografischen Wandels. In *Handbuch Industrie 4.0* (pp. 1). Springer Berlin Heidelberg: Springer.

Tropschuh, B., Diez, M. A., Waldherr, S., & Reinhart, G. (2020). Psychische Beanspruchung in der manuellen Montage. *Vox sanguinis*, *115*(8), 703. https://doi.org/10.1111/vox.12917

Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisløff, U., & Coombes, J. S. (2016). Accuracy of heart rate watches: implications for weight management. *PLoS ONE*, *11*(5), e0154420. https://doi.org/10.1371/journal.pone.0154420

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology*, *2*(1), 104. https://doi.org/10.1001/jamacardio.2016.3340