Selection and Assessment of Activity Trackers for Enthusiastic Seniors

Viktoria Willner, Harald Rieser, Verena Venek and Cornelia Schneider
Mobile and Web-base Information Systems, Salzburg Research Forschungsgesellschaft m.b.H., Jakob-Haringer-Straße 5/3, 5020 Salzburg, Austria

Keywords: Activity Tracker, Seniors, Requirements, Usability, System Integration.

Abstract: A large percentage of older people do not achieve the recommended levels of health related activities. The planned CARIMO system addresses this problem and offers services and applications which motivate older users to do exercises and thus, improve their health. As additional incentive, an activity tracker is planned to be integrated in the CARIMO system. The paper describes the three-step process we defined to select a suitable device. First, available activity trackers were analysed according to a predefined criteria catalogue. Best ranked trackers were evaluated with a respect to usability and technical requirements. The Samsung Gear Fit2 offered a high range of functionality and was best ranked according to the usability evaluation. So, we finally decided to integrate the Samsung Gear Fit2 in the CARIMO system.

1 INTRODUCTION

In Europe, the demographic change and ageing is going to lead to a growing number of older people (Muenz, 2007). At the one hand this is a positive development of older people but on the other hand social care providers will have to face resource challenges. Therefore, maintaining older people as long as possible as healthy as possible is a main aim in the research field of Active and Ambient Assisted Living. One important aspect for healthy ageing is physical activity (WHO, 2003). According to the WHO (2003), physical activity improves balance, strength, coordination, flexibility, endurance, mental health, motor control and cognitive function. These improvements further reduce falls, decrease mortality and age-related morbidity, increase social inclusion, self-confidence and self-sufficiency. It is recommended that adults aged 65 years and above should at least perform 150 minutes of moderate-intensity aerobic physical activity throughout the week or at least 75 minutes of vigorous-intensity aerobic activity (WHO, 2010). A large percentage of the older population fails to achieve the recommended levels of activity (Vankipuram, 2012). According to Rasche (2015) even 70 % of older adults are inadequately active. The lack of information about individual capabilities and limitations as well as the lack of motivations due to the absence of a support structure are reasons for this high percentage number (Vankipuram, 2012; Rasche, 2015).

The CareInMovement project addresses these challenges. It aims at offering services and applications particularly for people aged 65 years and above motivating them to do more exercises and thus, improve their general health. Within the project, the CARIMO system is being developed which combines technologies and personal support. It provides personalized exercise programs, activity monitoring and motivational incentives. For monitoring purposes, an activity tracker should be used. Activity trackers like smartwatches and fitness bands are well suited for monitoring success of workouts, weekly routines or even the users' daily step counts. Although, tracker technology is a relatively new development, evidence supporting their potential health benefits has already become apparent (Angulo, 2016). The range of currently available fitness trackers is wide. But for CARIMO it is important to decide for one activity tracker which will be integrated into the system. Otherwise the system usage would be incomparable within the users and the effects not clear.

There are many ways of wearing the trackers. For example, the devices can be clipped onto clothing or worn as bracelets. Recorded data of the

Willner, V., Rieser, H., Venek, V. and Schneider, C.
Selection and Assessment of Activity Trackers for Enthusiastic Seniors.
DOI: 10.5220/0006256400250035
In *Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2017)*, pages 25-35
ISBN: 978-989-758-251-6

user's physical activity can be automatically uploaded to a computer or a mobile device (Collier, 2014; Butler, 2016). The functionality of available devices differs and ranges from simple pedometers to smart watches which include several sensors and applications. Most devices monitor the periods of the user's activity, inactivity, light, and deep sleep and feedback the recorded information (Nelson, 2015). Several trackers additionally include heart rate sensors (HR) displaying the exercising intensity to the user (Butler, 2016). Further, these data enables developers to provide adjusted fitness programs. An additional feature is the GPS functionality which allows position tracking. Such devices are able to provide information about distance, pace and average speed (Butler, 2016). Activity trackers are mostly designed for younger users and it has not yet been investigated if they are also suitable for older adults (Rasche, 2015).

This paper presents how an appropriate activity tracker for older adults can be selected. The structure of this paper is as follows. In the beginning the problem is raised and the CARIMO system is introduced. In chapter 2, the specific requirements are described. In chapter 3 the device selection process is introduced in detail. Chapter 4 presents the results of the device selection. In chapter 5, the results are discussed. Finally, in chapter 6 conclusion and outlook are given.

2 DEMANDS ON ACTIVITY TRACKERS FOR OLDER USERS

Currently, there are many fitness trackers on the market offering multiple features. The devices allow their users access to recorded data mostly in real time. Information is shown directly on a display or more detailed on a smartphone or web-based application. The functions are similar, but the applications, calculation algorithms and user interfaces differ (Kaewkannate, 2016). The specifics are even more interesting when integrating the devices into other systems. In order to obtain the greatest benefits of the specific devices, developers have to be sure that they consider the operation purpose and the target group they want to address. Further, developers need to understand any limitations the wearable might have (Butler, 2016). According to Ledger (2014), there are several well-known criteria that are essential for adoption and utilization of wearables. First, the value proposition of the device has to be clear for the user. Since most of the wearables are worn visible, aesthetics are critical. Nelson (2015) emphasized that wearables fulfill an aesthetic function and are not as a few years ago only used for utilitarian purposes. The initial impression and usability of a product are essential. Thus, the user experience should be intuitive from the beginning. The user interfaces should be kept simple and informative. Readability is a very important aspect (Nelson, 2015). Mercer (2016) approved that aesthetic and simplicity are criteria which are important not only for young users but also for older people. A common reason for liking or disliking a device is the design as well as the ease of use. The overall wearing comfort is additionally critical for adoption. Besides the general understanding of wearing, the comfort of wearing the device during common activities e.g. working, sleeping, doing sports, has to be considered (Ledger, 2014). The size should not exceed 50 x 50 x 20 mm and it should not be heavier than 100 g (Schneider, 2014). Most trackers are considered to be worn the whole day; therefore they need to be robust. Devices which require less behavior change of the users are more likely to be used for longer periods. Therefore, it is beneficial if charging or synchronizing is required less and it is not needed to take off the device while showering. In 2014, Schneider recommended a minimum battery life of 12 hours for older users with cognitive impairments. The ideal battery life was mentioned as more than 18 hours. It is recommended that the device is waterproof or at least water splash proof. For integration purposes, it is important that the data can be accessed by other services (Ledger, 2014).

3 DEVICE SELECTION PROCESS

The device selection process which was established to select an appropriate activity tracker for the CARIMO system consists of three steps which are conducted consecutively. First, devices available to the Austrian consumer market as of July 1, 2016 were identified and analyzed according to a set of criteria. Therefore well-established producers or distributors of electronics and organizations operating in the fitness or medical sector were investigated. Additionally, current test reports were used to get an overview about available fitness trackers (www.pcwelt.de, www.chip.de, www.testberichte.de). Devices which seemed to meet the criteria were analyzed in detail and registered including detailed specification. Within the second step, best ranked devices were bought for extensive usability tests. Next, technical characteristics of these trackers were tested under real world conditions. Based on the usability and technical test a decision was made.

3.1 Criteria Catalogue

For rating available activity trackers, a criteria catalogue has been developed based on consolidated demands which are described in chapter 2. As mentioned before, a clear value of the devices is important. Therefore, we are interested in the functionalities the devices offer. For CARIMO, which aims at motivating people to move, the device must have step counter functionality. In addition, heart rate and GPS are optional features. These features could enhance the motivation of the people to use the device. In order to provide suitable fitness exercises and to measure effects of the technology, the mentioned features could be beneficial for the developers. The device has to have a display to show information. Integrating a watch as well as calorie indication are optional features which could additionally motivate the users. Aesthetic and comfort are rather subjective measures, which are separately evaluated in the usability test. Anyway, it is noted in the criteria list if a device was available in different colors, if the wristband is changeable, which material is used initially, and size and weight of the device. The must-have requirement according to comfort is the adjustment of the device to different wrist sizes, because in the target group there is a wide range of users. Hence, it is important that people are not excluded by size limitations of the device. According to ‘less behavior change’, we defined that the battery life has to last at least five days (without using HR or GPS). The device has to operate alone, i.e. a permanent contact with a paired smartphone or tablet is not necessary, and it has to be waterproof or at least splash proof. For the system it is required that recorded data is accessible (application programming interface) and that data are transmitted either via Bluetooth or ANT. Technically, provision of activity recognition would be beneficial. The detailed implementation will be checked within the technical evaluation. Affordability and availability on the market are important project related requirements. For CareInMovement, the price of each device should not exceed € 200. The simplicity of the user interface as well as the robustness are not considered as part of the criteria catalogue, but evaluated within the usability test.

3.2 Usability Test

The usability test aims at comparing the devices according to simplicity, comfort and aesthetic. Since those three parameters are rather subjective measures, all devices were tested from people who represent the CARIMO target group. The usability test consisted of two parts. First, the initial experience was evaluated by four tasks the testers had to perform. The required time to perform each task was recorded. The tasks are indicators for the simplicity of the devices. For the first task asked the users had to charge the device. The second task was to put the device on their wrist or to clip it onto their clothes. Next, they had to turn the device on and finally, they had to find the information about

Table 1: usability questionnaire.

Questions	Answers
Q1: How would you assess wearing the device in general?	I did not notice the device at all, I hardly ever noticed the device, it was ok, it was rather irritating, it was very irritating
Q2: How did the band feel on the skin?	very comfortable, rather comfortable, rather uncomfortable, very uncomfortable
Q3: How would you describe the size of the device?	suitable, too big, too small
Q4: How would you describe the weight of the device?	suitable, too heavy, too light
Q5: How would you describe the size of the display?	suitable, too big, too small’
Q6: How would you describe the readability of the displayed information on the screen?	very well, well, it was ok, rather badly, very badly
Q7: How would you describe the usability of the device?	very easy, ok, too complicated
Q8: How often did you notice the vibrations or sounds of the device when it informed you about notifications?	always, mostly, sometimes, mostly not, never
Q9: Do you think that the device can be lost easily?	yes, no

walked steps. After performing the tasks, the testers wore each device for at least five hours and then filled out a questionnaire. The questionnaire consisted of nine questions (see table 1) aiming at evaluating the subjective impressions about comfort and aesthetic as well as usability.

Within the usability test, every participant tested every available device. For the evaluation, the results of the testers were combined. Arithmetical means of recorded task times and mean values of the questions are illustrated using spider charts.

3.2.1 Participants

Persons representing the CARIMO target group were asked to test the devices with respect to usability issues. Two women and two men aged between 55 and 77 volunteered. Three persons had minor visual impairments and wear glasses. One man and one woman are already smartphone-users; the others use feature phones. All participants had not yet possessed or used any fitness tracker before the test was conducted.

3.3 Technical Test

The technical test mainly focused on proving the product specifications, as well as data access and data privacy issues. As CARIMO intends to motivate users by informing about their fitness, the system has to collect activity and vital data from the users to monitor the behavior, and to measure the impact and optionally modify and improve their custom fitness plans. Though, privacy issues come into play. There are basically four methods to access data from a fitness tracker:

- The tracker has on-device APIs to access the data, and allows synchronization with custom services.
- The tracker provides an interface using Bluetooth or ANT+ to access the data directly from the device.
- The tracker synchronizes the data with a mobile application, and this application provides an interface for data access.
- The tracker synchronizes the data with a mobile application, this mobile application stores the data on a server, and the server provides an interface to access the data.

User data have to be stored and analyzed in the CARIMO system. Due to privacy issues, data captured by the devices and sent to third party services are a problem, since once the data is stored on third party servers, control of this data might be lost.

Within the technical test, data was recorded and it was proven how to access these data. Considering the criteria ‘less behavior change’, the battery life was tested in two different settings. First, the residual charge of fully charged devices after leaving them for five days in the standby mode was recorded. The second battery test tested the battery behavior during normal usage. The devices were worn by testers and they used the fitness trackers as such. Accuracy of recorded steps and heart rate were additionally assessed. Step accuracy was evaluated by comparing a standardized distance of 50 steps with the recorded data. The step accuracy was tested in a flat area, during uphill and downhill walking, respectively. To assess the accuracy and compare the monitored heart rate of the device, the values of a blood pressure monitor were used. In order to avoid bias by specifics of the test person, step and heart rate accuracy of all devices were tested by the same person. According to usability, the mirroring effects of the display were observed inside of buildings and outside (daylight). Furthermore, the adjustment of display brightness and availability of notifications were assessed. The data synchronization was important for the integration of the device in the CARIMO system. Therefore, a technician tested the integration of the device in practice due to possibility and simplicity. If available, the technician tested the functionality of the activity recognition of the device.

4 RESULTS

In the following, the results of the device selection are described in detail.

4.1 Market Survey

To assess commercially available wearable activity trackers, all devices available to the Austrian consumer market with 1st of July, 2016, were identified between July and August 2016. Finally, 33 devices (see figure 1) which met the criteria described in Section 3.1 were analyzed in detail. We identified and reviewed six devices by Fitbit, five devices by Garmin, two devices by each Withings, Mio, and Huawei, and one device by each Polar, LG, Runtastic, ihealthlabs, newgen medicals, Nike, Samsung, Asus, Epson, Jawbone, Xiaomi, Pebble, Tomtom, Suunto, Adidas, and Microsoft. The following figure shows the devices analyzed according to must-have criteria (see 3.1) and ordered by functionality (‘pedometer’, ‘pedometer + heart rate’, ‘pedometer + heart rate + GPS’).

	Device	display	size restrictions	battery life (days)	stand alone	water resistance	API	prize (€)	available
pedometer	Vivofit 2	✓	ok	356	✓	50 m	✓	100	✓
	Vivofit 3 *	✓	ok	356	✓	50 m	✓	120	✓
	Zip	✓	ok	180	✓	splash proof	✓	44	✓
	A300 *	✓	ok	28	✓	30 m	✓	100	✓
	One *	✓	ok	14	✓	splash proof	✓	100	✓
	Vivosmart	✓	ok	7	✓	40 m	✓	140	✓
	Activity	✓	ok	240	✓	50 m	✓	390	✓
	Alta	✓	ok	5	✓	no	✓	140	✓
	Newgen Medicals	✓	×	6	✓	1 m	✓	30	✓
	Talkband B2	✓	×	4,5	✓	1 m	✓	170	✓
	Orbit	✓	×	7	✓	100 m	×	120	✓
	Talkband B1	✓	ok	6	✓	1 m	✓	130	×
	Fuelband SE	✓	ok	6	✓	no	✓	139	×
	Lifeband Touch	✓	ok	5	✓	no	×	100	×
	iHealth Edge	✓	×		✓	no	✓	70	×
incl. heart rate	Fuse *	✓	ok	7	✓	30 m	✓	129	✓
	Vivosmart HR	✓	ok	5	✓	50 m	✓	150	✓
	Charge HR *	✓	ok	5	✓	splash proof	✓	150	✓
	Pulse Ox *	✓	ok	14	✓	no	✓	100	✓
	Blaze	✓	ok	5	✓	splash proof	✓	240	✓
	Vivowatch	✓	×	10	✓	1 m	✓	125	✓
	Pulsesense PS-500	✓	×	1,5	✓	30 m	✓	150	✓
	Alpha 2	✓	×	60	✓	30 m	✓	180	✓
	Band Pulse	×	×	30	✓	1 m	✓	19	✓
	Pebble 2	✓	×	7	✓	30 m	✓	130	×
	UP 3	×	ok	7	✓	no	✓	180	×
incl. GPS	Gear Fit 2 *	✓	ok	5	✓	40 m	✓	190	✓
	Surge *	✓	ok	7	✓	splash proof	✓	200	✓
	Band 2 *	✓	ok	2	✓	splash proof	✓	230	✓
	Spark Cardio	✓	ok	21	✓	40 m	✓	250	✓
	Ambit 3	✓	ok	5	✓	40 m	✓	299	✓
	Vivoactive HR	✓	ok	8	✓	40 m	✓	270	✓
	Micoach smart run	✓	×	14	✓	splash proof	✓	279	✓

Figure 1: Available activity tracker devices assessed between July and August, 2016.

Devices which did not conform to the must-have criteria are marked by grey. Size restrictions, water resistance and availability were the most common criteria for exclusion within basic activity trackers and devices including heart rate sensors. Price and battery life restrictions are more critical at devices with high functional range. After the market survey we decided to buy nine activity trackers; three basic activity trackers, three trackers with heart rate sensor (HR) and three trackers with HR and GPS (marked with *). One of the three selected basic activity trackers was the Garmin Vivofit 3. The device totally met the must-have criteria and compared to its precursor it offers activity recognition. The battery of the device has not to be charged (is a round cell battery) and should last one year. That would be sufficient for CARIMO’s eight months test phase. Additionally, the Polar A300 met the must-have criteria and its battery is chargeable. The third pedometer we decided to extensively test was the Fitbit One. It met the must-have criteria and in contrast to the other pedometers it can be clipped on clothing. For activity trackers with heart rate sensor, Mio Fuse, Fitbit Charge HR and Withings Pulse Ox were selected. The Mio Fuse was chosen because it met the must-have criteria and additionally offers ANT. The Fitbit Charge HR was selected because it met the must-have criteria and it provides activity recognition. Although the Withings Pulse Ox was worse ranked than the Vivosmart HR and is not waterproof, the device convinced by its longer battery life, inexpensive price and the possibility to wear it either on the wrist or clipped onto clothes. The three selected activity trackers with heart rate sensor and GPS functionality were Samsung Gear Fit 2, Fitbit Surge and Microsoft Band 2. Gear Fit 2 and Surge alone met the must-have criteria. The Band 2 convinced by its cheaper price compared to the other opportunities.

4.2 Usability Evaluation

Each participant tested the nine devices which were selected based on the market survey (see 4.1) without reading any user manual in advance. The average time spans the testers needed to charge every device ranged from 13 seconds (Pulse Ox) to 61 seconds (Fuse). The average time spans to place the devices ranged from 19.6 seconds (Surge) to 60.5 seconds (Vivofit 3). The practical test showed that most devices turned on automatically when they were charged, so task 3 could be skipped. To find the information about the step count participants needed 2.9 seconds on average using One and 91 seconds using the Surge.

Following, the results of the usability evaluation for each device are described in detail. From the questionnaire, the eighth question Q8 was not considered because within the five hours of testing notifications did not occur. The following table shows the weighted means of recorded task times and mean values of the questions for the basic trackers. The outcomes of this are the UI points.

Table 2: usability evaluation basic trackers.

	Vivofit 3	A300	One
Charging (task 1)		0.33	0.16
Placing (task 2)	0	0.5	0.66
Finding steps (task 4)	0.5	0	1
Comfort (Q1)	0.958	0.75	1
Feeling (Q2)	1	0.915	
Size (Q3)	1	0.75	0.5
Weight (Q4)	1	1	0.75
Display (Q5)	0.75	0.75	1
Readability (Q6)	0.75	0.625	0.937
Usability (Q7)	1	0	1
UI Points	**6.958**	**5.62**	**7.007**

The Garmin Vivofit 3 is not chargeable; hence, task 1 could not be performed. The placing of the device took 60.59 seconds on average. It took 17.65 seconds to find the step count information. Three testers reported that they did not notice the device while wearing at all; one person said that she hardly ever noticed the device. All testers said that the Vivofit 3 felt very comfortable on the skin. Weight and size were rated suitable from all participants. Three testers had the feeling that the display was suitable as well, one person thought that the display is too small. All testers thought that the readability is good. All said that the Vivofit 3 was very easy to use. Figure 2 illustrates the evaluation of the Vivofit3.

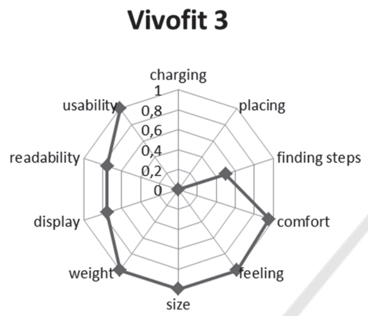

Figure 2: Evaluation Garmin Vivofit3.

The average time to charge the Polar A300 was 43.16 seconds and 31.98 seconds to place it on the wrist. The testers needed 48.93 seconds on average to find the step information. All testers said that they hardly noticed the device while wearing it. Three persons reported that the A300 was very comfortable to wear; one person said it was rather comfortable. This person thought that the device was too big; the others rated the size as suitable. The weight was suitable for all testers. The display was suitable for three persons; one person meant that it is too small. Three testers rated the display readability as good and one person as rather bad. All testers said that the usage of the A300 was too complicated.

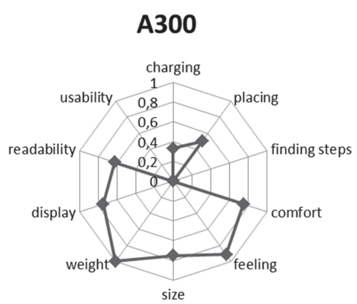

Figure 3: Evaluation Polar A300.

The average time to charge the Fitbit One was 59.9 seconds. All testers intuitively placed the fitness clip on their (trousers) waistband. It took 22.26 seconds on average. The step information was found averagely after 2.9 seconds. All testers reported that they did not notice the device on their waistband at all. The feeling on the skin could not be evaluated. Two testers rated the device as too small; one even said it is too light, the other one stated that the weight is ok. The other two participants reported that size and weight were suitable. All said that the display size was suitable and three persons thought that the readability is very good; one reported that it is good. All testers rated the usage as very easy.

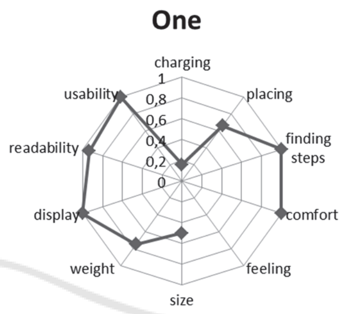

Figure 4: Evaluation Fitbit One.

Table 3 shows the usability evaluation of the activity trackers which include heart rate sensors.

Table 3: usability evaluation activity trackers with HR.

	Fuse	Charge HR	Pulse Ox
Charging (task 1)	0	0.83	0.83
Placing (task 2)	0.33	0.66	0.66
Finding steps (task 4)	0.16	0.83	0.66
Comfort (Q1)	0.375	0.562	0.625
Feeling (Q2)	0.577	0.66	0.83
Size (Q3)	0.25	0.5	0.25
Weight (Q4)	1	1	1
Display (Q5)	1	0.25	1
Readability (Q6)	0.375	0.937	0.187
Usability (Q7)	0.25	1	0.875
UI Points	**4.317**	**7.229**	**6.917**

The average time people need to charge the Mio Fuse was 61.08 seconds. It took them 43.14 seconds to place the wrist band and 26.18 seconds to find the step count information. After five hours of wearing the device, two persons said that wearing the device was ok and two persons reported that it was rather irritating. Three persons said that the device felt rather comfortable and one said that it felt rather uncomfortable on the skin. Three persons rated the

Fuse as too big; one thought the size was suitable. The weight and size of the display were suitable for each tester. Three participants stated that the readability was rather bad; one person rated it as good. This person reported that the Fuse was very easy to use; the others decided that it was too complicated. Figure 5 illustrates the evaluation of the Fuse.

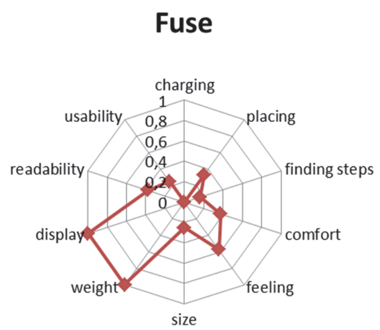

Figure 5: Evaluation Mio Fuse.

The testers needed 13.82 seconds on average to charge the Fitbit Charge HR. It took them 22.31 seconds to place the activity tracker and 7.62 seconds to find the step information. Three persons said it was ok to wear the tracker; one person reported that he hardly (ever) noticed the device. All testers stated that the wristband felt rather comfortable on the skin. Two persons said that the device was too big; the others said it was suitable. The weight was suitable for each tester. Three participants rated the display as too small; one decided it was suitable. Three participants confirmed that the readability was very good; one said it was good. The usage was rated to be very easy by all testers.

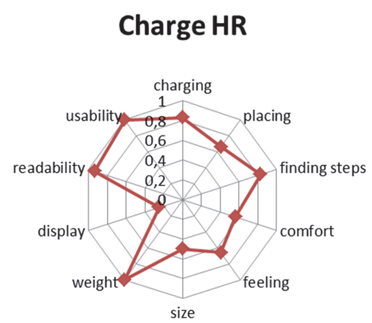

Figure 6: Evaluation Fitbit Charge HR.

The average time people need to charge the Withings Pulse Ox was 13.04 seconds. They needed 20.71 averagely to place the device on their wrist and 14.83 seconds to find the step count information. Two persons said they hardly (ever) noticed the device while wearing it; two persons said it was ok to wear it. The device felt very comfortable for two testers and rather comfortable for the other two testers. One person rated the size as suitable; the others decided that the Pulse Ox was too big. The weight and the size of the display were suitable for each participant. Three persons stated that the readability of the information on the display was rather bad; one person said it was very bad. Three participants confirmed that the usage was very easy and one participant decided it was ok.

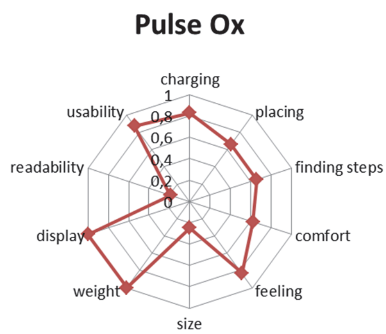

Figure 7: Evaluation Withings Pulse Ox.

Table 4 shows the usability evaluation of the activity trackers which include heart rate sensors and GPS functionality.

Table 4: evaluation activity trackers with HR and GPS.

	Gear Fit2	Surge	Band2
Charging (task 1)	0.83	0.83	0.66
Placing (task 2)	0.66	0.83	0.66
Finding steps (task 4)	0.66	0	0.83
Comfort (Q1)	0.875	0.562	0.375
Feeling (Q2)	0.943	0.747	0.332
Size (Q3)	1	0	0.5
Weight (Q4)	1	0.75	0.5
Display (Q5)	1	1	1
Readability (Q6)	1	0.687	0.75
Usability (Q7)	1	0.125	0.25
UI Points	**8.968**	**5.531**	**5.857**

The average time to charge the Samsung Gear Fit2 was 17.79 seconds. The users needed 20.38 seconds to place the wristband adequately. Within 11.92 seconds on average the testers found the information about the step count. Two persons said that they did not notice the device at all while wearing it and two persons said they hardly (ever) noticed the device. Three testers stated that the Gear Fit2 felt very comfortable on the skin; one said it felt rather comfortable. All testers decided that the size and the weight of the device, as well as the size of the display, were suitable. The participants agreed that the readability was very good and the usage was very easy. Figure 8 illustrates the evaluation of the Gear Fit2.

Figure 8: Evaluation Samsung Gear Fit2.

The average time to charge the Fitbit Surge was 18.13 seconds and it took 19.62 seconds to place the wristband. One person needed 91 seconds to find the step count; the other testers did not find this information at all. After five hours of wearing, two testers said they hardly (ever) noticed the device, one tester reported that it was ok to wear the device and one person said it was rather irritating. Two persons stated that the device felt very comfortable on the skin, one rated it as rather comfortable and one person said it felt rather uncomfortable. All testers agreed that the Surge was too big. One tester reported that it was also too heavy; the others rated the weight as suitable. The size of the display was suitable for everyone. One tester confirmed that the readability was very good; one said it was good and the other two testers thought it was ok. Three participants decided that the usage was too complicated and one participant said it was ok.

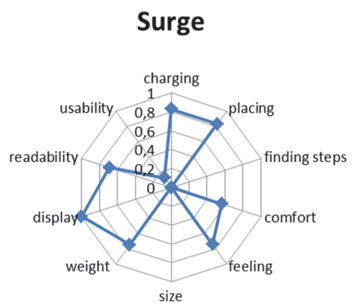

Figure 9: Evaluation Fitbit Surge.

The average time to charge the Microsoft Band 2 was 26.12 seconds. It took 20.39 seconds to place the activity tracker on the wrist and 9.24 seconds to find the information about the step count. Three persons reported that it was rather irritating to wear the device, one person mentioned that he hardly (ever) noticed the device while wearing it. He also said that it felt very comfortable on the skin. Two stated that the Band 2 felt very irritating on the skin and one person said it felt rather irritating. Two persons decided that the size and the weight were suitable and two persons rated the device as too big and too heavy. All agreed that the size of the display was suitable. One person decided that the readability was very good; two persons rated it as good and one person thought that it was ok. Two testers confirmed that the usage was ok and two testers rated that the usage of the device as too complicated.

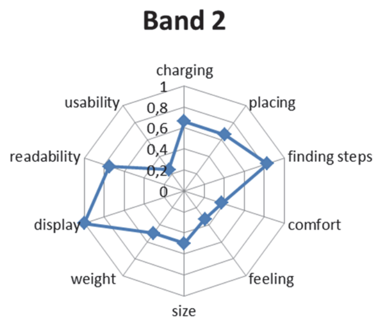

Figure 10: Evaluation Microsoft Band2.

Concerning the last question of the questionnaire (Q9), all testers agreed that solely the Fitbit One can easily be lost.

4.3 Technical Evaluation

Each of the tested devices provides a proprietary infrastructure for data exchange and data access. While there are standards for health data exchange and access (for example, ISO/IEEE 11073 Personal Health Data), those standards are not adopted for fitness and health trackers.

Data from the Fitbit devices are available via web API through the Fitbit developer API. Microsoft allows access to fitness data through the Microsoft Health APIs. Garmin, Withings and Polar provide access through the Garmin wellness API, the Withings API and the Polar Link APIs, respectively.

Only three devices provide an API on the mobile device: the Mio Fuse through Google Fit APIs and the Samsung Gear Fit 2 through the Samsung Health APIs. The Garmin Vivoactive HR is a special case, since Garmin provides an SDK for Android and IOS to communicate to the wearable, but most of the logic of an application has to be provided for the wearable itself using the Connect-IQ APIs.

The battery test verified the manufacturer information which was collected within the market survey. After five days operating in the standby mode the battery advice of the Fitbit One, Mio Fuse and Withings Pules Ox showed 100 %, the Fitbit Charge HR showed 50 %, the Polar A300 showed 30 %, and the Samsung Gear Fit2 and the Fitbit Surge showed 5 %. The Microsoft Band 2 automatically turned off on the second day. After five days of usage the battery advice of the One showed 100 %, the Pulse Ox showed 75 %, the A300 showed 50 %, , the Fuse showed 20 % and the Charge HR and the Surge showed 5 %. The battery of the Band 2 and Gear Fit2 did not accomplish five days usage.

Figure 11 shows the recorded steps of the activity trackers at a flat, uphill and downhill 50-steps distance.

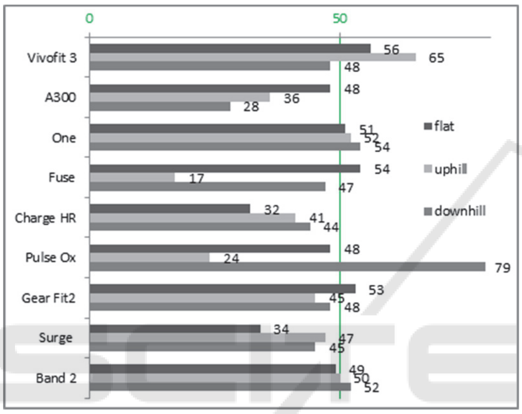

Figure 11: Step accuracy evaluation.

According to this figure, the Microsoft Band2, Fitbit One and Gear Fit2 were most precise. Mio Fuse and Withings Pulse Ox differed most between uphill and downhill distances and were rather precise at the flat distance. The Vivofit3 was more precise at the flat and downhill distance than at the uphill one. The A300 estimated well at the flat distance and worse uphill and downhill. Fitbit Charge HR and Surge were more precise at the uphill and downhill distances than at the flat one.

The pulse evaluation resulted in similar accuracies of all evaluated activity tackers (Fuse, Charge HR, Pulse Ox, Gear Fit2, Surge, Band2). The distinction between values acquired by the blood pressure monitor and the fitness trackers was between one and three heart beats. Excluding the Pulse Ox, all evaluated activity trackers measured the pulse directly on the wrist. The Pulse Ox had to be detached from its band. The pulse was measured on the fingertip.

Except One and Fuse, all trackers offered activity recognition. They distinguished between walking and running. The A300 additionally classifies five intensity levels. The Band2 included determination of cycling. Vivofit3, Charge HR and Surge distinguished cycling and swimming, and the Vivofit3 additionally cross training. The Gear Fit2 detects additionally cycling, rowing machine and cross training.

None of the devices' display mirrored inside buildings. The display of the Band2 mirrored outside, the others did not. Brightness adjustment was available only for Gear Fit2 and Band2. All devices offered visual notifications. Most of them also provided vibration notifications (A300, Fuse, Charge HR, Pulse Ox, Gear Fit2, Band2). A300, Fuse and Gear Fit2 additionally made sounds.

5 DISCUSSION

The Samsung Gear Fit2 was rated best at the usability test (8.968). Due to the charging plate, charging of the device was very easy. The wearing comfort, the feeling on the skin and the size of the device were rated very well. To place the tracker on the wrist is partially difficult. Due to this result, we have to emphasize the placing task at the initial system training. Finding the step count was not that easy as well. As the device enables own implementations, it is on us to focus on simplicity and implement plausible user interaction. Technically the tracker offered different activity recognition possibilities, brightness adjustment and any kind of notifications. The Gear Fit2 was one of the three most precise devices according to step accuracy. An API was provided on the device itself. The battery life was the only drawback. During normal usage it lasted three days. The battery life evaluation of devices with comparable functionalities (Surge and Band2) was similar. The Band2 automatically turned off after two days of usage. The Surge showed 5 % battery duration after five days. The usability was evaluated worse for both devices. Overall, the Band2 had 5.857 points in the usability evaluation. Mainly comfort and usability were rated badly. We think that the wearing comfort suffered because of the additional sensor on the closing part of the band. The device most precisely recorded walked steps within all slopes. The Surge received 5.531 usability points. The testers mainly criticized the size and the usability of the device. The middle section of the wrist band was rather big (32 x 55 x10 mm) and rigid. So it was difficult to put on a jacket or pullover. Three testers did not find the step information at all. We concluded that the mixture of physical buttons and the touch display were irritating for the test users. If the battery life would be the decisive factor, we would have chosen the Charge HR. The device was ranked secondly according to the usability test (7.229 points). The small display was criticized mostly. Low points in comfort and size could be argued with one size of the wristband (L) which was tested by all users. The Charge HR underestimated steps within all slopes. This could be a drawback for CARIMO because this might not be motivating for the users. The tracker did not provide GPS. Although, the One had got 7.007 usability points, we would not use this tracker for CARIMO, because, we such as our testers, think the device is easy to loose. Charging was not easy because one had to put it out of its tight silicone cover. The battery life did fine in all tests. However, the device only provided basic functionality without HR, GPS and activity recognition. In comparison to the other trackers, the device was worn at the waist; hence the step accuracy was very good. The spider chart of the Vivofit3 looked quite similar to the chart of the Gear Fit2. Since charging could not be evaluated, the device just got 6.958 usability points. The manufacturer information promised 356 days of battery life. This could be beneficial for CARIMO because the users could wear the device continuously. The only drawback within the usability evaluation was the placing of the device. Thus, we decided against the Vivofit3 for CARIMO because the Gear Fit2 was rated better and provided more functionality. The three last ranked trackers were Pulse Ox, A300 and Fuse. The Pulse Ox got 6.92 usability points; size, comfort and readability were criticized most. The battery life of the device was very good, although it provided also HR. However, pulse could not be measured directly on the wrist. It was measured on the fingertip. This required taking the device off each time the users wanted to measure the pulse. Another drawback of the device was the step accuracy. Thus, the Pulse Ox was rated not practicable for CARIMO. The A300 was rated with 5.62 usability points. The initial experience and usability were rated poorly. We saw that it was very difficult for the tester to figure out how to charge the device. In addition, the watch had to be disassembled out of the silicon wrist band each time to charge the device. This is particularly challenging for older users. The Fuse got 4.317 usability points and the technical evaluation showed poor results. All usability points except weight and display were rated poorly. It was difficult for the testers to find the steps. We observed that the handling with the display itself was challenging. The battery life was very well. The step accuracy was unprecise at the downhill distance.

After acquiring the test results, we decided to integrate the Samsung Gear Fit2 in CARIMO. It offers a high range of functionality including HR and GPS. So, we are not limited in implementing any planned CARIMO feature and it was rated best at the usability evaluation.

6 CONCLUSIONS

The variety of fitness trackers on the market is huge; however, it is challenging to select a proper one for a system like CARIMO. Based on the manufacturer information an explicit decision could not be made. Hence, we established a three step process to do so. First, we defined must-have and optional criteria based on literature research and project specific requirements. We did a market research to preselect trackers which basically met the criteria. The selected trackers could be classified in basic activity trackers, trackers with additional heart rate sensor (HR), and trackers with HR and GPS functionality. Based on the defined criteria, we further chose three devices of each group to test them extensively.

Each tracker was tested by four testers who represented the CARIMO target group according to usability, comfort and aesthetics issues. The first impression was evaluated by performing four initial tasks. The average time it took to conduct each task shed light on the simplicity of the usage. After wearing the devices for at least five hours, the participants answered questions about comfort, aesthetics and the overall usability. The average values of the tasks and questions were measured. For visualization the significant results were illustrated using spider charts. The highest possible score of the questionnaire resulted in 11.0. Additionally, technical characteristics were proven. The focus relied on battery life, accuracy of recorded data, data management and data accessibility.

The extensive tests allowed us to decide which activity tracker was suitable for the CARIMO system. Having the manufacturer information only, we could just decide by functionality. However, taking usability issues into account is important, particularly, considering selecting devices for older users. This cannot be assessed without testing the trackers by users representing the target group. Additionally, the technical characteristics given by the manufacturers have to be proven. These tests are helpful to gain further insights, e.g. about data accuracy, and can be used to test the manufacturer specifications e.g. related to battery life.

The on-going development of the CARIMO system will integrate the selected activity tracker; the Gear Fit2. The entire system will be tested and evaluated in a long term quasi experimental controlled field trial over eight months including 120 elderly users in Austria and Italy.

REFERENCES

Angulo, G., Brogan, D., Martini, A., Wang, J., Clevenger, L. A., 2016. Health Features of Activity Trackers: Motivation, Goal Achievement, and Usability. In: *Proceedings of Student-Faculty Research Day*, CSIS, Pace University.

Butler, M. S., Luebbers, P., E, 2016. Health and Fitness Wearables. *Wearable Technology and Mobile Innovations for Next-Generation Education*: p. 58.

Collier, R., 2014. Rapid growth forecast for digital health sector. In: *Canadian Medical Association Journal*, 186 (4), E143-E144.

Kaewkannate, K., Kim, S., 2016. A comparison of wearable fitness devices. In: *BMC public health*, 16 (1), 1.

Ledger, D., McCaffrey, D., 2014. Inside wearables: How the science of human behavior change offers the secret to long-term engagement. *Endeavour Partners*, 200(93), 1.

Mercer, K., Giangregorio, L., Schneider, E., Chilana, P., Li, M., Grindrod, K., 2016. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. In: *JMIR mHealth and uHealth*, 4(1).

Muenz, R., 2007. *Aging and demographic change in European societies: main trends and alternative policy options*. World Bank SP Discussion Paper No, 703.

Nelson, E. C., Verhagen, T., Noordzij, M. L., 2016. Health empowerment through activity trackers: An empirical smart wristband study. In: *Computers in Human Behavior*, 62, p. 364-374.

Rasche, P., Schäfer, K., Wille, M., Theis, S., Schlick, C., Mertens, A., 2015. Self monitoring–an age-related comparison. In: *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference*.

Schneider, C., Henneberger, S., 2014. Electronic Spatial Assistance for People with Dementia: Choosing the Right Device. In: *Technologies*, 2(2), p. 96-114.

Vankipuram, M., McMahon, S., Fleury, J., 2012. ReadySteady: app for accelerometer-based activity monitoring and wellness-motivation feedback system for older adults. In: *AMIA Annual Symposium Proceedings*, Vol. 2012, American Medical Informatics Association, p. 931.

World Health Organization. (2010). *Global recommendations on physical activity for health*.

World Health Organization. (2003). *Health and development through physical activity and sport*.