Comparing Visualizations and Text on Smart Watches while Running in Realistic Environment

by

Sarina Kashanj
B.Sc.,Azad University of Tehran, 2019

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

Comparing Visualizations and Text on Smart Watches while Running in Realistic Environment

by

Sarina Kashanj
B.Sc.,Azad University of Tehran, 2019

Supervisory Committee

Dr. Charles Perin , Supervisor
(Department of Computer Science)

Dr. Miguel Nacenta , Internal Member
(Department of Computer Science)

ABSTRACT

In today's digital age, smartwatches have become popular tools for millions of runners, providing metrics like pace, heart rate, and distance. Despite the widespread use of text-based displays on these devices, research suggests that visualizations could offer a more effective alternative. However, little is known about how visualizations perform in real-world running scenarios. This study addresses this gap by investigating how visualizations compare to text is aiding runners' performance and experience. Through a study involving 20 runners completing running tasks on an outdoor track, we found that visualizations significantly outperformed text, with runners completing tasks 1.5 to 8 times faster. Moreover, participants expressed a strong preference for visualizations and indicated a willingness to use them during their runs if available on their smartwatch. These findings highlight the potential of visualizations to enhance the usability and effectiveness of smartwatches for runners.

Contents

List of Figures

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to:

My supervisor, Dr. Charles Perin, for giving me the opportunity to undertake this research. Your guidance and dedication have been invaluable throughout this journey. You have not only imparted knowledge and expertise but also taught me the true essence of collaboration. Your support has been a light on my path, and I am truly grateful for the time and energy you have invested in my educational growth.

My brother, Dr. Sina Kashanj, and my sister-in-law, Dr. Yeganeh Saffar, for being my role models and unwavering supporters. Your strength, patience, and guidance have been a constant source of inspiration, and I am deeply grateful to always have you to look up to and count on.

My dad, whose boundless love and unwavering encouragement have given me the strength to overcome every challenge.

My mom, for always believing in me, teaching me the values of hard work, and encouraging me to pursue my dreams without hesitation.

My best friends, Sage Gervais and Luciana Bagia for being my pillars of support and sharing both the joys and struggles of this experience. Your companionship and understanding have made this journey more enjoyable.

To all those whose names may not be mentioned individually, your contributions have not gone unnoticed. Thank you all for being part of this accomplishment.

DEDICATION

To my Mina, who taught me the value of perseverance and hard work and who has always believed in me—I could not have done this without you.

To the women who fight fiercely for their dreams, who face every challenge with courage and determination, even when the path is difficult and there is no one to rely on but themselves. To those who stand strong through the storms, who refuse to give up on what they want, and who embody the true meaning of resilience.

Part of this thesis is based on the publication:
Kashanj, Sarina, et al. “Visualizations on Smart Watches While Running: It Actually Helps!.” IEEE VIS Short Paper, 23 July 2024. Web.

I acknowledge the use of AI-based technology to assist with writing and editing in this thesis. The research, content, and conclusions remain my own.

Chapter 1

Context and motivation

Running is a popular physical activity [1], undertaken by many individuals seeking a healthier lifestyle [2]. It is easy to perform, it has a social component, and it is relatively inexpensive, time efficient, and easily accessible [2], [3]. Running has important positive implications for mental health, particularly depression and anxiety disorders [4]. The high popularity and accessibility of running are seen as a strong contributor towards promoting and enhancing a physically active lifestyle within the population [3]. Running is the most-uploaded sport on the Strava sports-focused social network [5].

In recent years, the introduction of wearable technology has transformed how people manage and engage in their running routines. Among these technologies, smartwatches have become essential tools for runners, with about two-thirds of runners relying on them during their workouts to track various running-related data fields [6]. Smartwatches offer real-time tracking of essential running-related data fields, such as Distance covered, Heart Rate, and Pace. This data helps runners monitor their progress, set goals, and adjust their training accordingly. Smartwatches are an effective tool in motivating individuals to engage in exercise by enhancing the quality of exercise experiences [7].

Despite the advantages of using smartwatches to access running-related data fields, representing data on smartwatches presents significant challenges, primarily due to the small screen size [8]. Unlike larger screens, smartwatches have limited display space, making it difficult to display detailed information without reducing readability. Li et al. [8] also underscore users' difficulties when interpreting data on small screens, particularly in personal informatics systems, where clarity and conciseness are important.

Another challenge in presenting data on smartwatches is the dynamic nature of running. Runners are constantly in motion, and their environment is often unpredictable. Shifting the runner's attention from their environment to the display [9] breaks the running flow [10]–[12]. They must keep their eyes on the road to navigate obstacles and maintain their safety. Glancing at a smartwatch to check data can be risky, as it diverts their attention from their surroundings, increasing the potential for accidents or injuries. Studies have consistently shown that visualization can significantly enhance data comprehension [13]–[16] and speed of decision-making. Despite the proven benefits of visualization [17], [18], many smartwatches for runners still predominantly use text to present data (e.g., [19]–[23]). This discrepancy highlights a gap between what is known to be effective and what is commonly implemented in wearable technology. The reliance on text may be due to the technical limitations of small screens or the lack of emphasis on design innovation in wearable interfaces. However, it is important to bridge this gap to enhance the usability and effectiveness of smartwatches for runners. For runners, this means being able to understand their performance metrics at a glance without the need to interpret numerical data. This capability is particularly beneficial in scenarios where quick adjustments are necessary.

Glanceable feedback is likely to be useful. We know that glanceable feedback [24] on smartwatches can help for everyday activity tracking of health and activity levels; that it is possible to retrieve information from glanceable visualization on smartwatches [25]; and that real-time visualization can be useful for runners while running indoors on a treadmill, a low-risk environment [18].

However, we do not know whether this holds true in realistic running settings [26] where in runners might experience reduced cognition [27].

1.1 Research Questions

In this research, we aim to answer one primary question and one secondary question regarding the visualization of running data fields on smartwatches:

- **Primary Question: Does the use of visualization in showing the data fields on smartwatch interfaces help runners comprehend their data faster during physical activity?**

  The term "comprehending the data faster" refers to how quickly runners can understand and interpret the data displayed on their smartwatch interface while actively running. This question investigates whether visual representations of data (such as graphs and icons) enable faster comprehension compared to traditional textual data displays.

To make this broad question more manageable to answer, we broke it down into three sub-questions. Each sub-question addresses different types of tasks for different types of data fields that are meaningful to runners, and together, they help us answer the primary question.

Sub-questions:

Question 1.1: Does visualization help for tasks where the runner needs to know whether they are above, below, or on a specific target?

Question 1.2: Does visualization help with tasks where the runner needs to know their progress towards completing a set goal?

Question 1.3: Does visualization help for tasks where the runner needs to understand how much they need to make an adjustment to keep staying in their set goal?

- **Secondary Question: How do runners subjectively perceive the clarity, helpfulness, and usability of data visualization compared to textrepresentations on smartwatch interfaces?**

  The term "subjective perception" refers to the runners' personal experiences and preferences regarding how clear, helpful, and usable they find the different data representation styles. This question investigates whether runners prefer visual representations or traditional text displays and how these preferences impact their overall experience and satisfaction while using the smartwatch during physical activity.

1.2 Research approach

Our research approach consists of examining the effectiveness of visualization versus textual data representations.

- **Primary Question: Does the use of visualization in showing the data fields on smartwatch interfaces help runners comprehend their data faster during physical activity?**

To answer this question, we created two smartwatch interfaces, to compare visual and textual data representations. We then ran an experiment where we measured how quickly runners performed common running tasks. These tasks were designed to mimic real-world running scenarios to ensure the data collected was representative of actual use cases. Both watchfaces features a mix of visualization and text. When one watch face displays a data field using visualization, the other shows the same data field using text. This approach allows us to balance visualization and text in both watch faces and compare the results for each data field presented visually versus textually. We tested both watchfaces by defining the same tasks for each data field (Pace, Heart Rate, Distance) to compare how each representation type affects the time it takes for runners to understand their data and complete the tasks. By comparing the response times for both visual and textual representations, we could quantitatively compare the effectiveness of each representation.

Sub-question 1.1: Does visualization help for tasks where the runner needs to know whether they are above, below, or on a specific target?

To answer this sub-question, we defined a task where we asked participants to compare their current Pace/Heart Rate to the target Pace/Heart Rate shown on the watch face during their running activity.

Sub-question 1.2: Does visualization help with tasks where the runner needs to know their progress towards completing a set goal?

To answer this sub-question, we implemented tasks that required runners to recognize and interpret their Distance progress towards a set goal at the beginning, middle, and near the end of the run.

Sub-question 1.3: Does visualization help for tasks where the runner needs to understand how much they need to make an adjustment to keep staying in their set goal?

To answer this sub-question, we implemented tasks where we asked participants how much they needed to adjust their current Pace/Heart Rate compared to the target Pace/Heart Rate to stay on track with their set goal. This task required runners to recognize and interpret their data in relation to a set goal.

- **Secondary Question: How do runners subjectively perceive the clarity, helpfulness, and usability of data visualization compared to text**

presentations on smartwatch interfaces?

To answer this secondary question, we interviewed the participants to gather qualitative data on their subjective preferences and experiences with the two presentation styles. The interviews included a series of questions designed to rate their experience and preference for visualization versus text presentation. Participants were asked to rate the following statements on a scale from 1 to 5:

– The representation was clear and easy to understand.

– The representation was helpful in answering questions related to this data field.

– I would use this representation if it were supported by my watch.

1.3 Research contributions

The thesis makes two research contributions:

- **Contribution 1: A comparison of visualization vs text on smartwatches for runners in realistic settings**

  In our study, runners perform realistic tasks in a realistic environment. We derived the data fields, tasks, and representations from studying through a combination of an existing survey [6], our own survey, and our own experience.

  We devised five tasks involving three commonly used data fields (Distance, Pace, and Heart Rate). We presented each data field in both text and visual representations, as shown in Figure 1.1. We designed two watch faces: WF1 shows Elapsed Time and Heart Rate with text and Pace and Distance with visualization; WF2 shows Elapsed Time, Pace, and Distance with text and Heart Rate with a visualization.

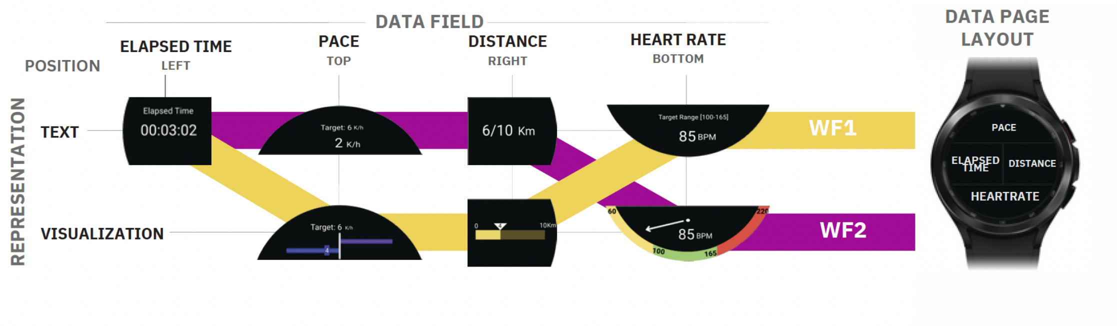


Figure 1.1: Watchface layout showing Elapsed Time (left), Pace (top), Distance (right), and Heart Rate (bottom). Pace, Distance, and Heart Rate are represented with text or visualization. WF1 shows Elapsed Time and Heart Rate with text and Pace and Distance with visualization; WF2 shows Elapsed Time, Pace, and Distance with text and Heart Rate with visualization

Our results show that runners

- i) complete the tasks much faster with visualization,
- ii) prefer visualization to text for those tasks,
- iii) would like to have access to such visualization while running.

- **Contribution 2: Methodology for evaluating smartwatches for runners in realistic settings.**
  Our methodology for evaluating smartwatches for runners stands out due to its emphasis on realistic settings, tasks, and the unique challenges encountered in study design and data collection.

  We conducted the evaluation on a 400m oval track, where runners often train. This ensured that our study closely replicated a natural running environment.

  One of the challenges we faced in conducting the test in a realistic environment was collecting data and communicating with runners while they were running and distant from the experimenter. To address this, we equipped participants with wireless earbuds for audio communication with the experimenter, ensuring clear instructions and timely data collection.

  Participants performed 26 tasks in total, including binary tasks (e.g.," Is your Heart Rate above or below the target?") and estimation tasks (e.g.," How much do you need to adjust your Pace?"). These tasks were designed to mimic real-world scenarios that runners encounter, making our findings directly applicable to everyday use.

To maintain consistency, we used a credible dataset for all participants, ensuring that the tasks were not based on their actual running data. This allowed for a controlled comparison between text and visual data presentations.

Data Collection and Analysis:

Structured Approach: Participants completed tasks while running and responded verbally to questions after a trigger sound. This method allowed us to accurately measure completion time.
Statistical Analysis: We used bootstrapped confidence intervals for data analysis, following best practices in human-computer interaction (HCI) and visualization research.

Subjective Feedback and User Preferences:

Surveys and Interviews: After each running trial, participants rated the clarity, usefulness, and usability of the data presentations using a 5-point Likert scale. This subjective feedback was important in understanding user preferences and identifying areas for improvement.

Subjective Feedback: Participants also participated in semi-structured interviews to share their experiences and provide qualitative feedback. This approach provided insights into how different smartwatch interfaces might impact user experience.

1.4 Thesis scope

This thesis focuses on information visualization. It focuses on how to display data that runners frequently use to track their workouts on the small screens of wearable devices.

1.5 Thesis overview

The thesis contains six additional chapters. These are described as follows:

- In chapter two, we present a literature review on information visualization on small screens, and the nature of visualization in motion for fitness trackers to provide an overview of the relevant research and insights in these areas.
- In chapter three, we look into the pre-study we conducted and describe the design of a survey to collect data from runners for a detailed examination of the data fields they most commonly use during their running activity.
- In chapter four, we look into the overall design principles and how we apply them to design the watch faces. We select data fields based on the results of the pre-study conducted for evaluation in this study and we explain the rationale behind the design choices.
- In chapter five, we discuss the two phases of the study and the setup for the evaluation test. We look into the different tasks we asked the participants to complete in order to answer the hypotheses, which lead to answering the research questions.
- In chapter six, we look into the analysis of the data gathered from the evaluation test and interpret them to answer the research questions.
- In chapter seven, we present a discussion of the findings along with limitations and future design recommendations.
- In chapter eight, we conclude the thesis with a summary.

Chapter 2

Background and Related Work

In this chapter, we provide an overview of the existing literature on wearable technology, human-computer interaction in wearables, and visualizationtechniques for fitness data, all of which are relevant to our research. We also present a brief history of wearable technology in running. The scope of this related work section includes studies and methodologies that have shaped the current understanding of these areas. By analyzing these works, we aim to identify gaps and highlight areas where further research is needed.

2.1 Wearable Technology in Running

Wearable technology for athletes can be traced back to the late 1970s and early 1980s when basic electronic devices like Heart Rate monitors were first introduced [28]. Polar Electro, a Finnish company, developed the first wireless Heart Rate monitor in 1977, specifically designed for athletes. This innovation marked the beginning of wearable technology tailored for sports and fitness, allowing athletes to monitor their Heart Rates during training and competition [28], [29].

Garmin registered the name "Forerunner" in 2001 but released the first watches in 2003. A major change came with the launch of the Nike+iPod Sports Kit in 2006. This device, developed by Nike in collaboration with Apple, allowed runners to track their Distance, Pace, and calories burned using a sensor placed in their shoes, which synced with an iPod [30], [31].

The purpose of this innovation was to enhance the running experience by providing real-time data and feedback to runners. This allowed them to track their performance more accurately and stay motivated through integrated music and audio feedback. The collaboration aimed to combine fitness and entertainment, making running more engaging and data-driven [32].

Wearable technology has revolutionized the way runners track and enhance their performance. Key works in this field, such as those by Bramble and Lieberman [33], have shown that endurance running plays an important role in human evolution, shaping our physical form and survival strategies. Modern research, including studies by Kilpatrick et al. [34], and Hespanhol Junior et al. [35], highlights the benefits of running for health and leisure, showing that a significant number of runners now use smartwatches to monitor metrics like Pace, Heart Rate, and Distance. Critical analyses by authors like Makikawa et al. [21] and Islam et al. [36] reveal that, while these devices provide valuable data, they primarily use text and numbers for data representation, which may not be optimal for quick comprehension during a run since these are symbolic representations that require cognitive effort as they require decoding each symbol one after the other [37]. Our research connects to this body of work by exploring how Visualization on smartwatches can improve the usability and effectiveness of these data presentations, ultimately aiming to enhance the runner's experience and performance.

2.2 Human-Computer Interaction and Wearables

The field of Human-Computer Interaction (HCI) in wearables has garnered significant attention, particularly in the context of enhancing user experience and interaction efficiency. Key works include Dunne and Raby's [38] exploration of the relationship between users and wearable devices, highlighting the importance of design in user engagement and satisfaction. Fogg's [39] concept of persuasive technology also plays an important role, emphasizing how wearables can influence user behavior positively through well-designed interfaces. Recent studies by Gouveia et al. [24] delve into glanceable feedback in wearables, demonstrating that brief, easily accessible information can significantly enhance user interaction and adherence to health-related goals.

Despite the progress made in HCI for wearables, several challenges remain. Dunne and Raby's work [38], while pioneering, often focused more on the theoretical aspects of design rather than practical implementations, leaving a gap in actionable guidelines for designers. Fogg's [39] persuasive technology principles are widely applicable but can sometimes lead to over-reliance on technology for behavior modification, potentially diminishing user autonomy. Gouveia et al.'s [24] research on glanceable feedback provides valuable insights but often lacks consideration of diverse user contexts, particularly in high-intensity activities like running. These gaps highlight the need for more context-specific studies that address wearable users' unique challenges and requirements in various environments.

Our research addresses these gaps by focusing specifically on runners, a group that requires both quick access to information and minimal distraction. By comparing visual and textual interfaces on smartwatches, our study aims to determine the most effective design principles for enhancing user interaction during running. Building on the foundational theories of Dunne and Raby [38], Fogg [39], and Gouveia [24], this research not only applies their insights to a specific context but also expands the understanding of HCI in dynamic, real-world scenarios. This contributes to the broader HCI literature by providing empirical data on the effectiveness of Visualization in wearable technology, particularly in the context of physical activities.

2.3 Visualization Techniques for Fitness Data

Several studies have explored visualization techniques for fitness data, aiming to enhance user engagement and comprehension. Gouveia et al. [24] investigated the design space of glanceable feedback for physical activity trackers, highlighting the importance of quick and easy-to-understand visual cues. Similarly, Schiewe et al. [18] demonstrated the effectiveness of real-time Visualization for runners on treadmills, showing that visual feedback could significantly improve performance and motivation. Amini et al. [26] explored in-situ Visualization for fitness tracking, emphasizing the potential of context-aware Visualization to provide meaningful insights during physical activities.

While these studies provide valuable insights into designing and implementing Visualization for fitness data, they often focus on controlled environments or specific user groups. Gouveia et al. [24] primarily examined daily activity tracking, which may not directly translate to the dynamic and intense running environment. Schiewe et al. [18] limited their research to treadmill running, an indoor and controlled setting, which may not reflect the challenges faced by outdoor runners. Amini et al. [26] proposed context-aware Visualization but did not extensively test them in real-world scenarios. This gap in research underscores the need for studies that evaluate visualization techniques in more varied and realistic running conditions.

Building on these foundational works, our research aims to address the gaps identified by testing Visualization in realistic running scenarios. By comparing visual and textual presentations of fitness data fields such as Distance, Pace, and Heart Rate, we seek to determine which method enhances runner performance and satisfaction in outdoor environments. Our study contributes to the literature by providing empirical evidence on the effectiveness of Visualization in real-world running, offering practical design recommendations for smartwatches used by runners.

2.4 Methodologies

The study of wearable technology, particularly in the context of fitness tracking, employs a variety of methodologies to gather data and derive insights. Common methods include user surveys, controlled experiments, field studies, and data analytics. User surveys, e.g., such as used by Janssen et al. [6] are often used to gather subjective feedback on usability and preferences. Controlled experiments, such as those conducted by Amini et al. [26], allow researchers to isolate specific variables and measure their impact on performance and user experience. Field studies, like those described by Schiewe et al. [18], involve testing devices in real-world scenarios to evaluate their effectiveness and user acceptance.

Each methodology has its strengths and limitations. User surveys provide rich qualitative data but may be subject to biases such as self-reporting inaccuracies and non-response bias by Janssen et al. [6]. Controlled experiments offer precise control over variables but may lack ecological validity, as they often occur in artificial settings by Amini et al. [26]. Field studies provide high ecological validity by capturing real-world usage, but they can be challenging to control and may introduce confounding variables by Schiewe et al. [18]. Data analytics offer quantitative insights but may miss the nuanced, qualitative aspects of user experience and motivation by Munson [40]. A balanced approach, incorporating multiple methodologies, can provide a comprehensive understanding of wearable technology in fitness tracking.

Our research aimed to address the gap in understanding the effectiveness of Visualization on smartwatches for runners in real-world conditions. Existing studies had primarily focused on controlled environments or theoretical designs, leaving a lack of evidence regarding how Visualization perform in practical, outdoor running scenarios.

To fill this gap, we conducted a study where runners used smartwatches displaying running metrics through both text-based and visual representations. We designed domain-specific tasks and scenarios that runners would encounter during typical runs, such as monitoring their Pace, Distance, and Heart Rate. By comparing the objective performance and subjective preference between these two presentation types, we are able to gather concrete data on how Visualization impact runners' experience and effectiveness in real-world settings. Our findings indicate that visual representations significantly improve task completion time and user satisfaction for the tasks we tested, demonstrating their potential to enhance data comprehension and decision-making for runners. This research contributes insights into the design and application of glanceable Visualization in wearable technology, which can lead to a practical framework for future developments in this field.

Chapter 3

Background Research, Survey

We conducted a preliminary study aimed to understand which data fields runners mostly use and how many data fields they track during their runs. This information was important because we wanted the watch faces to be designed with realistic data fields that runners actually use rather than random data fields. This study would help us design the watch faces and determine which and how many data fields to display.

Through an online survey, we collected data from runners, the primary demographic for this research. The survey targeted individuals who run at least once a week and use smartwatches to track their activity. It took 10 minutes to complete. Participants were invited to leave their email addresses for a chance to win a draw as a thank-you for their participation.

We recruited participants to answer the survey by distributing recruitment posters around the university campus and sharing them on social media. The survey was available for a month, and we received 200 responses. However, only 19 of these responses were complete and reliable enough to work with.

3.1 Survey Design and The Sections

The survey's main goal was to identify which data fields runners rely on and how these data fields are displayed. We wanted to ensure that the information we present in our study is realistic and relevant. Therefore, we designed the survey to meet this goal and to collect data on:

1. How their watch faces look while running,

2. How often they use the data fields provided in the survey,

3. How these data fields are visualized on the watch face they use.

Additionally, we included a section in the survey to gather information about runners' workout types to see if their preferred data fields change based on their type of run.

3.1.1 Consent Form and Demography

The survey started with a consent form stating that participation would be completely anonymous and highlighting the potential benefits of their participation. These benefits include contributing to the advancement of knowledge, particularly in the field of data visualization.

After consenting to participate, participants were asked to fill out a demographic section. This section asked about their age range, gender, running experience (how many years they have been actively running), and their expertise with smartwatches (how many years they have been using smartwatches while running).

For answering the survey questions, we provided multiple-choice answers. Also, we included an "Other" option so participants could provide their own answers if necessary.

3.1.2 Smartwatch Brand and its watch Faces

The brands we provided as answers to the multiple-choice question "What brand of smartwatch do you use while running?" were Apple Watch, Suunto, Garmin, Samsung, Fossil, Coros, and Huawei. For the "Other" option, two participants mentioned using Fitbit as their wearable device for running.

After selecting their smartwatch brand, participants were directed to questions specific to their chosen brand. All participants were asked the same questions, but the watch faces shown corresponded to their own watch brand.

3.1.3 Main Survey Questions

First Question: Which watch face do you use for each type of running?

We designed this question with a multiple-choice matrix. Figure 3.1 shows an example of how this section of the survey looks for Garmin watch faces, where the rows represent the watch faces and the columns represent the types of running. Participants could choose multiple types of running for each watch face.

We provided some watch faces that were available online for the specific brand that participants chose based on their responses to the previous section. We categorized the types of running into Base Run, Long Run, Tempo Run, Progression Run, Interval Training Workout, Fartlek Training, Hill Repeats, and Sprints. These types of running were listed in front of each watch face so participants could choose which watch face they use for different types of running.

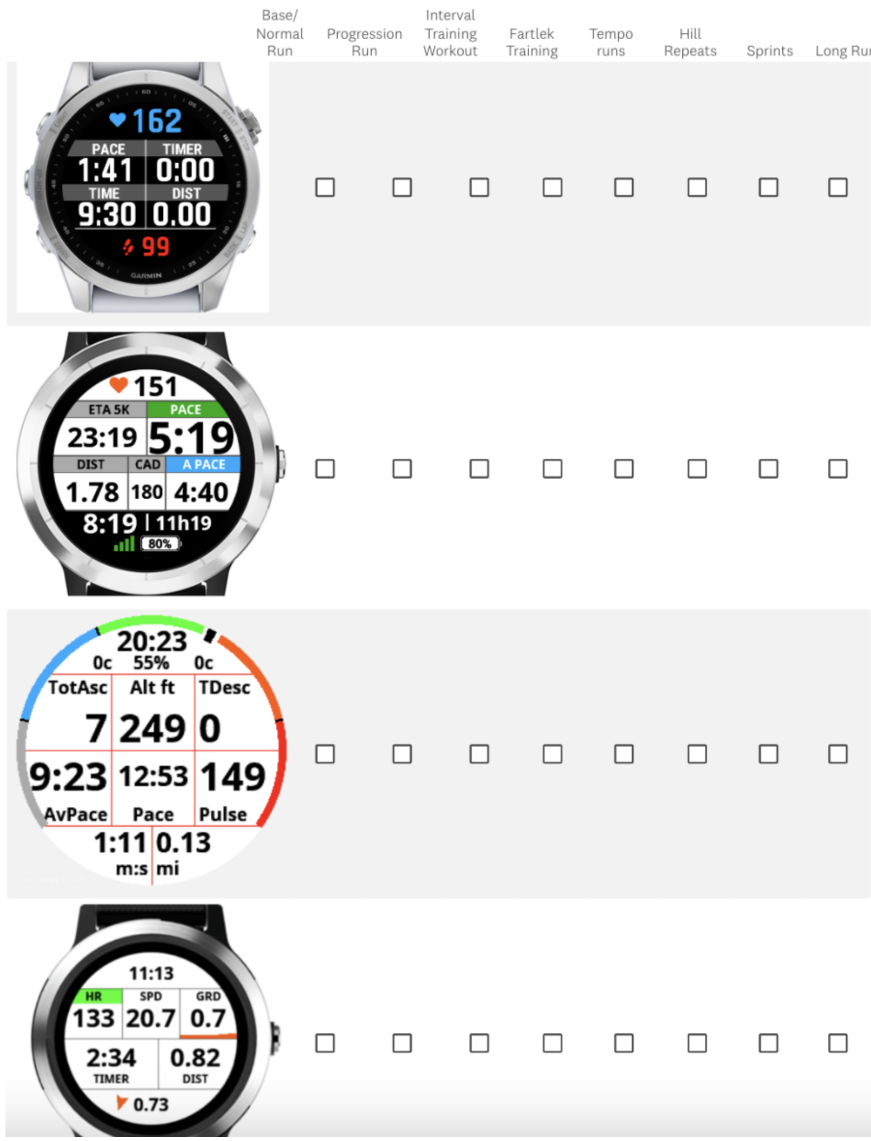


Figure 3.1: Selection of Garmin watch faces presented in the survey, enabling participants to identify and select the specific watch face utilized for various types of running.

If participants could not find their watch face among those we provided, they were directed to a section where we included three blank spots for watch faces. We asked participants if they had customized their watch face. If their specific watch face was not among the options we provided, they could either upload an image of their watch face, describe it, or share a link to a picture of their watch face.

Guidance was provided to participants in this section to gather more accurate data. The guidance suggested that participants either upload an image of their watch face (for example, start a run activity on their watch, display the watch face, take a picture with their phone, and upload it in the survey), describe the watch face following the example provided in Figure (see Figure 3.2 or share a link to an image of the watch face (for example, a link to an image found on the web).

Figure 3.2: example of how participants were asked to describe their watch face: At the top: Small heart icon, Heart Rate number (big font), bpm(small font size) in a row. In the middle: Heart Rate graph with low opacity. At the bottom: Elapsed time with the same font size as Heart Rate number.

Second Question: While Running, tell us how often you need to access this information on your smartwatch.

For this question, participants were asked how often they use the specific data fields we provided in the survey while running. The purpose of this question was to understand which data fields runners use most while running. This information would allow us to use the same data fields in our study and display them in the watch faces we were designing.

The data fields we included in the survey were: Time of day, Date, Weather, Power (watts), Elapsed Time, Lap Pace, Pace (Speed), Steps, Distance, Calories, Heart Rate, Activity Level, Navigation (Map), Body Battery, Barometric Altimeter, Accelerometer, Thermometer, Compass, Notifications, and Calls or Messages. This list was based on typical data fields found in running watches. Additionally, we provided an "Other" option for participants to specify any additional data fields they used that we might have missed. No additional data fields were suggested by the participants.

Participants were asked to rate their frequency of using each of these data fields using either of these options: Never (0%), Sometimes (30%), Often (70%), and Always (100%). Figure 3.3 Shows the multi select matrix question.

	Never (0%)	Sometimes (30%)	Often (70%)	Always (100%)
Time of day	○	○	○	○
Date	○	○	○	○
Weather	○	○	○	○
Power(watts)	○	○	○	○
Elapsed Time	○	○	○	○
Lap pace	○	○	○	○
Pace / Speed	○	○	○	○
Cadence/ Steps	○	○	○	○
Distance	○	○	○	○
Calories	○	○	○	○
Heartrate	○	○	○	○
Activity Level (progress to daily goals)	○	○	○	○
Navigation/Map	○	○	○	○
Body Battery	○	○	○	○
Barometric Altimeter	○	○	○	○
Accelerometer	○	○	○	○
Thermometer	○	○	○	○
Compass	○	○	○	○
Notifications	○	○	○	○
Calls/messages	○	○	○	○

Specify if there is other important information that you use while running and indicate the frequency as the table above.

Figure 3.3: We provided a list of data fields runners might use while running, including a blank space labeled "Other" for participants to specify any additional data fields we might have missed. We asked about the frequency of using each of these data fields with the options: Never (0%), Sometimes (30%), Often (70%), and Always (100%).

3.2 Analysis of the data

The data we collected from the survey were divided into four categories: demography, description of the watch face, type of running, and features of the smartwatch.

We received 200 responses. Of these responses, 70% were incomplete, so we focused only on the completed responses. We filtered these responses through a step-by-step process to ensure reliability: first, we extracted the completed responses. Second, we excluded responses where participants chose to upload an image, share a link, or describe their watch face but did not provide the required information. Next, we removed responses with irrelevant images. Finally, we eliminated duplicate responses from participants who took the survey multiple times to increase their chances in the draw offered as appreciation for participation. Ultimately, we were left with 19 authentic responses for analysis.

3.2.1 Demography

In total, we considered 19 authentic responses for the survey. Participants were categorized as follows: 21% were under 20 years old, 47% were aged 20-30 years, 21% were aged 30-40 years, 6% were aged 40-50 years, and 5% were over 50 years old. 26.3% were female, and 73.7% were male. Additionally, 63.8% of participants had been actively running and using a smartwatch for at least two years, while 36.2% had been actively running and using a smartwatch for more than two years. In terms of frequency, 52.6% of respondents ran 3-5 times a week, 36.8% ran twice a week, and 10.6% ran more than five times a week.

3.2.2 Watch Faces List

We made a table showing all the watch faces provided in the survey, and we gave an ID to each watch face. In the table, we dedicated the next column to the watchface to describe the data Fields (Feature) showing in that watchface, how it is displayed (icon, number, color), and the position where the Feature has been located. This way, it was easier to see how the most used features were presented in each watch. Also, by the watch face ID, we could track which watch face has been used for what running type, and the ID was the link of these two spread shits of data (Watch Faces List and the Running Type).

3.2.3 Running Type

The list of running types was Base Run, Long Run, Tempo Run, Progression Run, Interval Training Workout, Fartlek Training, Hill Repeats, and Sprints. The responses we collected from the survey for this section were scattered and lacked consistency. For example, one participant selected a specific watch face for a Base Run, while another selected the same watch face for Hill Repeats. There was no clear relationship between the watch faces and the eight types of running.

To identify any common patterns, we narrowed the eight types into two categories: regular runs (Base Run, Long Run, Tempo Run) and workout runs (Progression Run, Interval Training Workout, Fartlek Training, Hill Repeats, Sprints). However, we still lacked consistency, as one participant selected a watch face for a regular run while another selected the same watch face for a workout run.

Therefore, we decided to focus only on the data fields that runners mostly use, regardless of their running type, and not consider any specific running type while designing the watch faces for this research.

3.2.4 Data Fields (Features)

In the survey, we listed 20 features (Time of day, Date, Weather, Power (watts), Elapsed Time, Lap Pace, Pace (Speed), Cadence (Steps), Distance, Calories, Heart Rate, Activity Level, Navigation (Map), Body Battery, Barometric Altimeter, Accelerometer, Thermometer, Compass, Notifications, Calls (messages)). We also asked participants to add any additional features they use during running that were not provided. For each Feature, participants had the option of choosing between Never (0%), Sometimes (30%), Often (70%), and Always (100%).

To analyze the responses, we assigned a numerical value to each option: Never was converted to 0, Sometimes to 1, Often to 2, and Always to 3. By calculating the mean and median for each feature, we determined which features were used most often.

As shown in Figure 3.4; the features with the highest mean scores were Time of Day, Elapsed Time, Pace (Speed), Distance, and Heart Rate, all with a mean of 1.73. Heart Rate, Pace (Speed), and Distance also had the highest median score of 2 among all features.

Additionally, Elapsed Time had a high mean score of 1.63 but a low median of 1, and Activity Level had a low mean score of 1.42 but a high median of 2. These were important features to consider for the watch face design in this research.

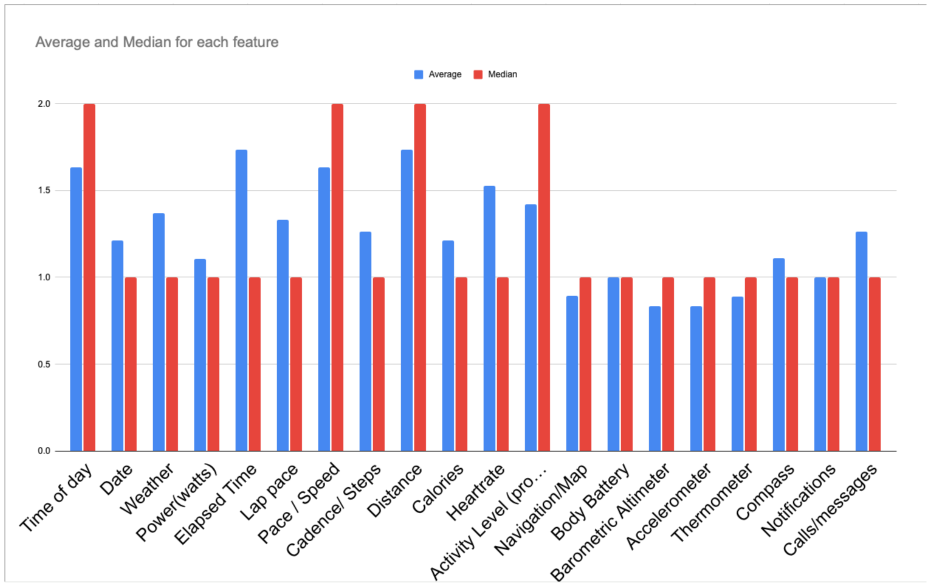


Figure 3.4: Average and median scores after converting the frequency of use into numerical values for each data field selected by participants.

3.3 Results of the survey

To choose which features to test in this research, we referred to the list of watch faces we created and checked how many times each of the 20 features provided in the survey were presented in these watch faces. This allowed us to understand which features were common across all watch faces. We then compared the most common features with the results from the data fields (features) list, focusing on those with the highest mean and median numbers. Ultimately, we selected the features that were common to both the watch faces list and the data field list.

As shown in Figure 3.5; In the watch face list, Elapsed Time appeared in 55% of all the watch faces, Pace appeared in 66%, and Distance appeared in 100%. These were the most common features across all the watch faces.

Comparing these features with the list of data fields, Pace and Distance both had high mean and median scores. Additionally, Elapsed Time, Heart Rate, and Time of Day were considered because Elapsed Time had the highest mean score and was present in more than 50% of all watch faces. Heart Rate and Time of Day also had high mean and median scores.

W ID	Features	Been Selected	Elapsed Time	Date	Weather	Power(watts)	Time of the day	Lap pace	Pace / Speed	Cadence/ Steps	Distance	Calories	Heartrate	Activity Level (pr	Navigation/Map
w1	3	7	X						X		X				
w2	9	1		X				X	X	X	X	X	X	X	X
w3	3	3	X						X		X				
w5	3	2						X	X		X				
w6	2	1	X								X				
w8	5	6	X				X		X		X				X
w17	5	1	X	X							X	X	X	X	
w18	7	1		X	X					X	X	X	X	X	
w19	5	2					X		X	X	X		X		
			55%	33%	11%		22%	22%	66%	33%	100%	33%	44%	33%	22%

Figure 3.5: In this table, the first column shows the most selected watchfaces, the second column indicates how many features are presented on each watchface, and the third column shows how many times each watchface was selected. An 'X' under a feature indicates that the feature is present on the corresponding watchface. The last row shows the percentage of times each feature was selected across all watchfaces. The features with the highest percentages are highlighted.

To decide how many features to include in our research, we examined the watch faces list to see how many features the most used watch faces had. Watch face number 1, chosen seven times, was the most selected watch face and had three features. Watch face number 8, chosen six times, was the next most used and had five features. All other watch faces were selected only once or twice. Therefore, we decided to use a median of four features and test four features for this study.

When choosing four features from the common list of data fields and watch faces, Pace, Elapsed Time, and Distance were the top three features with the highest percentage of representation in each watch face and the highest mean and median numbers in the data fields list.

For the fourth Feature, we had to choose between Heart Rate and Time of Day, as both had the same statistical values. We decided to go with Heart Rate because it was more related to running, and different designs could lead to different interpretations and results.

Therefore, we chose four features for our study: Pace, Elapsed time, Distance, and Heart Rate.

Chapter 4

Design and Representation of Watch Face Data Fields

This chapter explains the process of designing and implementing the data fields in the watch faces based on the pre-study results. We describe the overarching goals and principles used to design the watch faces, followed by an explanation of how these were applied throughout the design process.

4.1 Tools and Device

Before designing the watch faces for the evaluation test, we needed a device to test them. The watch we used for this study is the Fossil Gen6, which has a round screen resolution of 416 × 416 pixels and an AMOLED display.

We used Watch Face Studio for the design stage because it was compatible with the Wear OS watch we are using for the study. This allowed us to see the designs on the watch screen while we were designing the watchfaces.

This was important for the study, where initial findings required tweaks or entirely new data fields to address unforeseen issues and insights. It also allowed us to experiment with different layouts and visual elements.

Later in this chapter, we will discuss how we decided to present the participants' data for each data field in text and visualization.

4.2 Screen Sections

Based on the results from the survey, we identified four Data Fields for the study. Therefore, before starting to design for each data field, we needed to decide how to divide the watch screen into four sections.

The watch has a round screen, and we needed to divide it into four sections to place the four data fields. By examining the data screens of popular smartwatch brands used by runners like Garmin, Suunto, and Coros, we realized they commonly divide the screen into uneven sections depending on the number of data fields displayed. Figure 4.1 shows an example of a Coros smartwatch data screen. Our motivation for dividing the screen into three sections—top, middle, and bottom—and then splitting the middle into two sections is that it's a very standard layout on existing smartwatches. This layout aligns with the curvature of the watch, potentially making better use of space than a quadrant design, which can result in awkwardly shaped corners that are harder to utilize effectively.

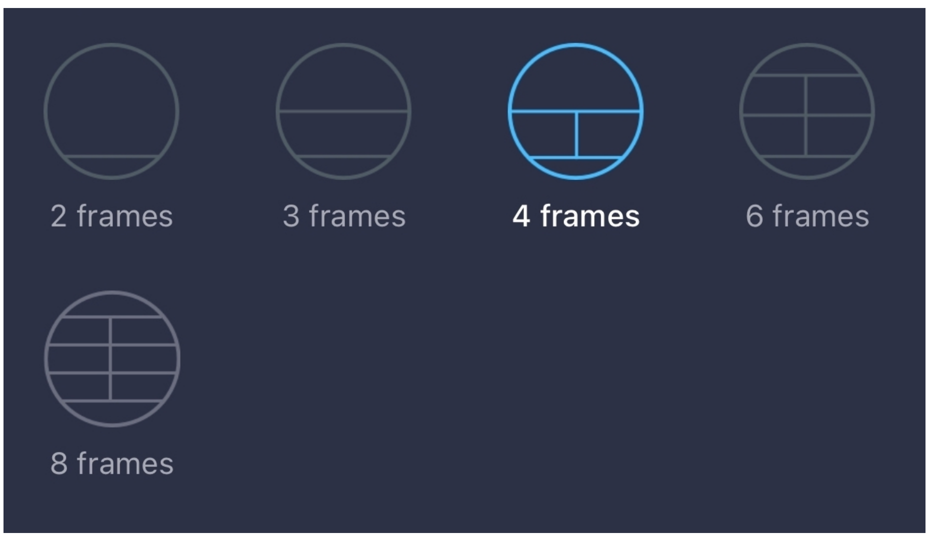
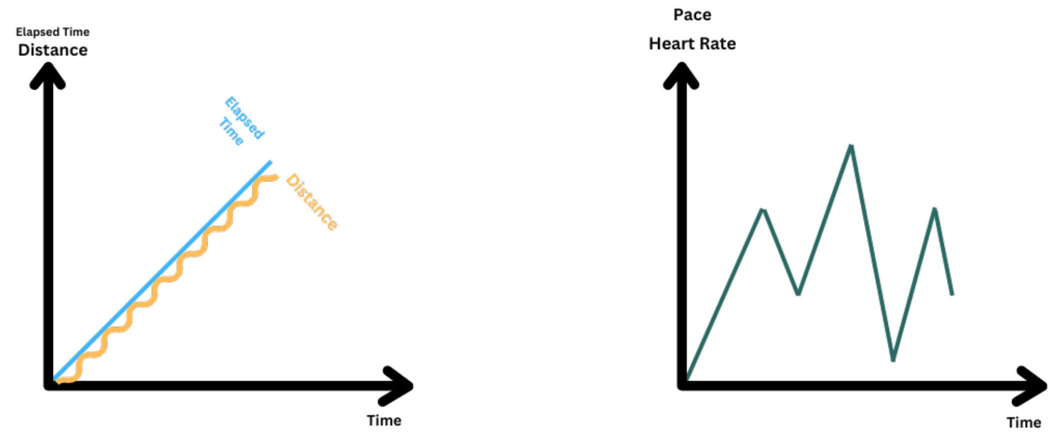


Figure 4.1: Example of a data screen for a Coros smartwatch, representing the sectioning of the screen based on the number of data fields shown.

Based on the discussions in the research meetings regarding placing each data field into sections, we realized that we could group the data fields into two categories. As shown in Figure 4.2, one group consists of data fields that continuously increase over time (Distance and Elapsed Time). The other group includes data fields that fluctuate unpredictably, either increasing or decreasing depending on the runner's performance (Pace and Heart Rate). Grouping similar types of information helped us place data fields like Elapsed Time and Distance next to each other, as both are measures of progress. As a result, we placed the other two data fields—Pace at the top and Heart Rate at the bottom of the watch face.



(a) Monotonic data fields: changes in Distance and Elapsed Time over Time.

(b) Fluctuating data fields: changes in Pace and Heart Rate over Time.

Figure 4.2: Grouping the data fields based on how they change over time.

4.3 Watchface Designs

To design the watch faces for this research, we began by considering how each data field was displayed on the watch faces chosen by participants in the survey.

For Pace and Elapsed Time, all the watch faces displayed these data fields using digital numbers without any visualization. However, there were two types of displays for Distance and Heart Rate: digital numbers and visualization.

Distance was visualized using icons combined with digital numbers or graphs to show progress. Heart Rate was visualized using icons combined with digital numbers, different colors to represent various Heart Rate ranges or animated heart-shaped icons with static graphs. Figure 4.3 shows some examples of how Distance and Heart Rate were displayed on the watch faces chosen by participants in the survey.

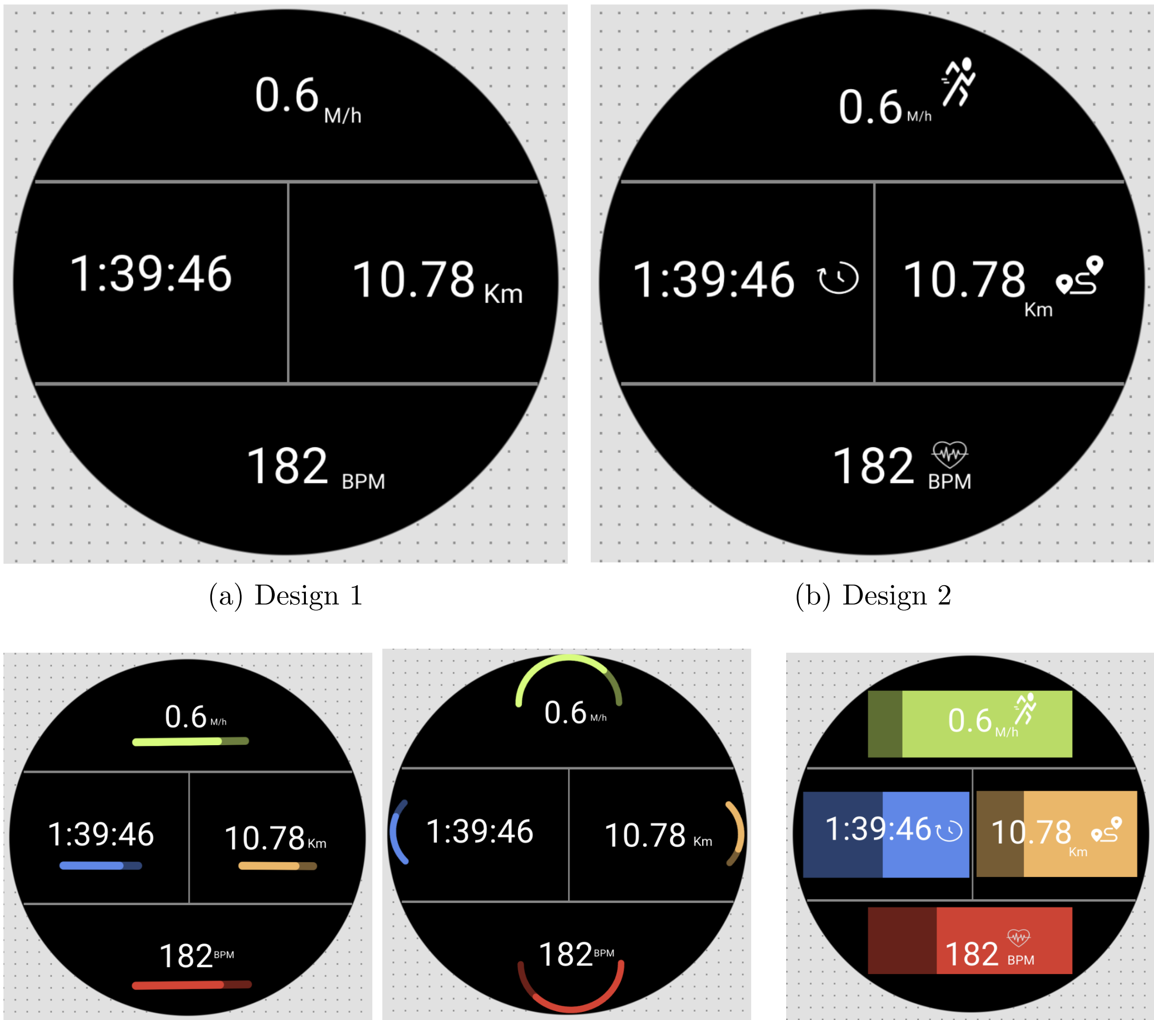
(a) Example of how Heart Rate was visualized in the smartwatches provided in the survey.

(b) Example of how Distance/steps were visualized in the smartwatches provided in the survey.

Figure 4.3: Example of how Distance and Heart Rate were visualized in the smartwatches provided in the survey.

Based on this information, we designed five different watchfaces. As shown in Figure 4.4, each watchfaces utilizes one of the following types: digital numbers, graphs to show progress, icons, and different colors to represent various ranges.

The first prototypes presented the four data fields (Pace, Elapsed Time, Distance, and Heart Rate). We began with Design 1, where the data fields were displayed only with text. In Design 2, we added an icon to each data field to indicate which data field each screen section represented. We incorporated color-coded progression bars for Designs 3, 4, and 5 to help runners visualize their progress. We used three different styles to present the progression bars: a line, a curve, and a section filled with different colors. Design 5 was a combination of text, a color-coded progression bar, and an icon for each data field. The use of color codes, icons, and progression bars aimed to help participants identify and understand their data more quickly and accurately, potentially improving their task completion speed compared to text-only displays



(a) Design 1 (b) Design 2

(c) Design 3 (d) Design 4 (e) Design 5

Figure 4.4: The first prototypes presented the four data fields (Pace, Elapsed Time, Distance, Heart Rate). In Design 1, the data fields were displayed only with text. In Design 2, we added an icon to each data field to indicate which data field each section of the screen represented. We added color-coded progression bars in Designs 3, 4, and 5. Design 5 also included an icon for each data field. The use of color codes and icons was intended to help participants identify which data field they were looking at, while the progression bars represented the data visually instead of textually.

To align with the research question of whether visualization helps runners understand their data faster and more accurately, we decided to develop two presentation types: text versus visualization. Therefore, instead of using one specific type for visualizing the data fields (icons, progression bars, colors) in one watch face and having multiple watch faces to test, we designed only two watch faces, WF1 and WF2. In these designs, each data field is presented either in a visual or text format, and where we used visualization, we applied different styles to visualize each data field. In the next section, we explain in detail how we visualized each data field.

(a) Watch Face 1 (b) Watch Face 2

Figure 4.5: Display of the final Design of the watch faces.

Based on several rounds of design ideation and discussions among the researchers, our design choices for visualizing each data field on the watch are grounded in both the survey results and the tasks defined to answer the research questions.

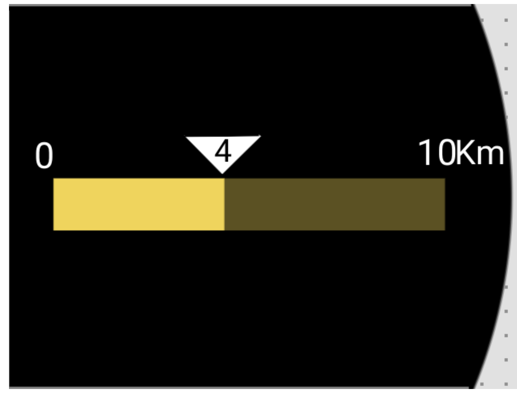
4.3.1 Tasks and Design choices for each data field

We defined five meaningful tasks based on the survey results and the selected data fields for this research. The two watch faces (WF1 and WF2) would allow us to study runners' ability to complete these identified tasks with either in the watchface presented with text or visualization.

- **Distance** is a fundamental data field that increases over time. Common runner questions associated with this data type include knowing the actual number and getting a sense of completion that tells them how far in their run they are.

  Therefore, as you can see in Figure 4.6, we visualized this data field with a progression bar that fills up as the runner progresses toward their goal Distance. We placed an arrow on top of the progression bar to indicate the runner's current position, with the exact Distance covered displayed inside the arrow. The text version of the watch face shows the completed Distance over the target Distance, allowing runners to compare how much of their total run they have completed.

(a) Distance displayed with only Text.



(b) Distance displayed with visualization.

Figure 4.6: Display of Distance.

We formulated the following task for Distance.

TDistance: Have you finished the first quarter of your run? Are you closer to the end or the beginning of your run? Do you have less than a quarter of your run left? These three questions ask runners to provide a binary response based on their estimated percentage of completed runs.

- **Pace** is important in workouts and races. Target Pace is often dynamic when performing high-intensity interval training. Common runner questions about Pace include knowing their current Pace, knowing whether they are on Pace or they should run slower or faster based on their activity goals, and by how much.

  Therefore, we visualized this data field to show runners their current Pace. As you can see in Figure 4.7 the watchface with text shows the current Pace with a large font and the target Pace with a small font. Figure 4.8 shows the watchface with visualization; a dual bar chart with a gray vertical line that is centered at the target Pace. A rectangular indicator shows the current Pace. It is placed on the left bar if the current Pace is lower than the target Pace at a Distance proportional to the Pace difference and on the right bar if it is higher. It also includes a small label that shows the exact Pace value.

Figure 4.7: Pace displayed with visualization.

Figure 4.8: Pace displayed with only text.

We formulated the following task for Pace:

TPaceAboveBelow: Is your Pace above or below the target Pace? This task asks runners to make a binary decision.

TPaceAdjust: Your Pace is below/above the target. How much do you need to adjust it? The two variations of this task (above or below instances) ask runners to estimate how much they need to adjust their Pace.

- **Heart Rate** is an important and complementary data field to Pace. It provides a way to measure effort, and in running, Heart Rate is typically utilized with respect to Heart Rate zones. For example, one might define a heart-rate zone for an 'easy pace' between 120 and 130 bpm and attempt to stay within that zone during an activity. Heart rate was a challenging data field to visualize because it needs to indicate the runner's current Heart Rate, what it means, and whether they need to adjust their speed. Among the visualization considered and chosen by runners in the survey, we decided to design something more informative than a simple animated heart icon but less complex than a static Heart Rate graph, which would be hard to understand at a glance while running. Ultimately, we designed the Heart Rate visualization as shown in Figure 4.9 the watchface with visualization shows a speedometer with three zones a common representation of running watch faces. The green zone is the target zone within the desired range. The yellow and red zones correspond to Heart Rate values below and above the target range. An arrow shows where the current Heart Rate value falls, and a label shows the exact value.

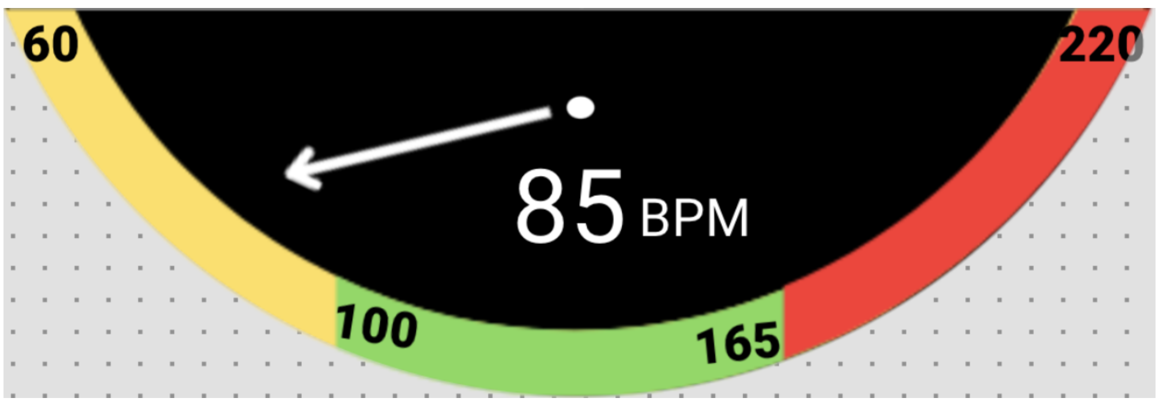


Figure 4.9: Heart Rate displayed with visualization.

As shown in Figure 4.10 the watchface with text shows the current Heart Rate value in a large font and the target range in a small font where runners can still see the target range and their current Heart Rate and can make a comparison to see if they are in the target range or how they need to adjust it.

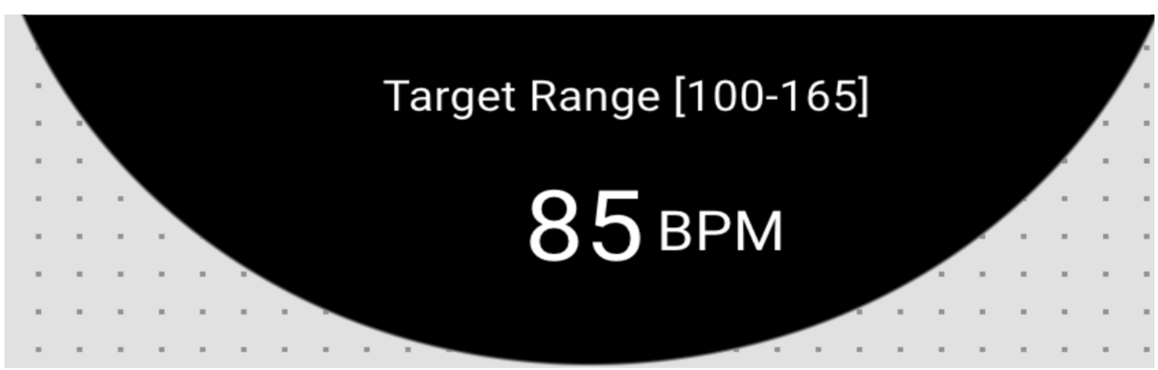


Figure 4.10: Heart Rate displayed with only text.

We formulated the following task for Heart Rate:

THeartRateInRange: Is your Heart Rate in the target range? This task asks runners to make a binary decision.

THeartRateAdjust: Your Heart Rate is below/above target; how much do you need to adjust it? The two variations of this task (above or below instances) ask runners to estimate how much they need to adjust their Heart Rate.

Together, these five tasks capture different data types and ask for different kinds of responses: binary decision based on proportion estimation (TDistance), binary decision based on a target value (Tpaceabovebelow) or a target range (Theartrateinrange), and estimation of deviation from either a target value (Tpaceadjust) or a target range (Theartrateadjust).

4.4 Design Goals

Our research's primary design goals when creating watch faces for runners were to present meaningful data and display it in ways that support the tasks. As shown in Figure 4.5, we finalized the two watch faces for this research to maintain a balance of text and visualization in both. We did not want to favor either watch face by presenting only text or only visualization, as this could affect the participant's responses to the tasks. Testing one watch face with all data fields presented as text and the other with all data fields as visualization might influence the results. Therefore, in WF1, Heart Rate is displayed with visualization, while Pace, Distance, and Elapsed Time are presented with text. In WF2, Pace and Distance are shown with visualization, while Elapsed Time and Heart Rate are presented with text.

4.4.1 Design Considerations

While these parameters were not the primary focus of our design goal, we still took them into account when designing the watch faces.

We aimed to make the watch faces clear and readable by displaying key metrics such as Pace, Elapsed Time, Distance, and Heart Rate using large fonts, simple graphics, and contrasting colors.

In both watch faces, target speed and current speed are positioned at the top, Elapsed Time and Distance are centered, and Heart Rate along with its target range is displayed at the bottom. We chose this layout because it follows a standard split-screen format (e.g., Coros), allowing runners to locate data fields they are already familiar with.

In both designs, we aimed to balance aesthetic appeal, minimalism, and the delivery of essential information in a straightforward manner without complicating the user interface.

Chapter 5

Study Design

In this chapter, we explain the methodology used to design an evaluation test. This test aims to determine whether visualizations help runners read and understand their data more quickly in a realistic setting.

The study is a quantitative study that objectively measures differences in performance and preference between the two representation types (Visualization and Text) for different data fields. We conducted an interview to collect additional qualitative feedback to understand why participants might prefer one type of display over another, beyond what can be gleaned from numerical ratings alone.

Our data analysis provides insights into the advantages and disadvantages of using Visualization to convey real-time performance metrics such as Distance, speed, and Heart Rate to runners and how these insights can be used to further improve future wearable systems for enhanced user understanding.

5.1 Participants

Since it was an in-person experiment, we used the university's departmental mailing lists and distributed university-approved posters on campus bulletin boards to recruit participants. Additionally, we used social media to reach out to people who are part of the running community and live in Victoria, BC. Our study evaluation criteria required that participants be at least 19 years of age and use smartwatches during running. Participants who responded to the invite were emailed to arrange the running trial based on their availability.

A total of 20 participants (13 males and 7 females) were recruited for the study on a first-come, first-serve basis. Participants were assigned a participant ID number based on the order of appointments. They ranged in age from 19 to 43, with a mean age of 24.6 and a median of 21.5. Among them, fifteen were undergraduates, three were graduate students, and one was pursuing a doctoral degree. Notably, the majority of participants (11 individuals) had backgrounds in Physical and Health Education, and 5 participants had a computer science background. All participants were running at least once a week, with some running up to 6 times a week. The majority (13 participants) were running four times a week.

5.2 Experiment Setup and Procedure

We divided the 20 participants into two groups. The first group contained participants from 1 to 10, and the second group contained participants from 11 to 20. Group 1 started their initial run with the first watch face (WF1), while Group 2 started their initial run with the second watch face (WF2).

Participants were asked to fill out a demographic survey form and confirm their consent to participate before the start of the experiment. The experimenter was present in the track field in the stadium throughout the session and provided a comprehensive explanation of the procedure to participants before commencing the experiment. Participants were encouraged to seek clarification by asking questions if needed.

The experiment had five steps. In step one, participants ran with the first watch face. In step two, they completed a follow-up questionnaire rating their experience with each feature of the watch face they used. Step three involved participants running with the second watch face. In step four, they completed another follow-up questionnaire rating their experience with each feature of the second watch face. In the final step, step five, participants participated in a short interview to compare and rate the two watch faces.

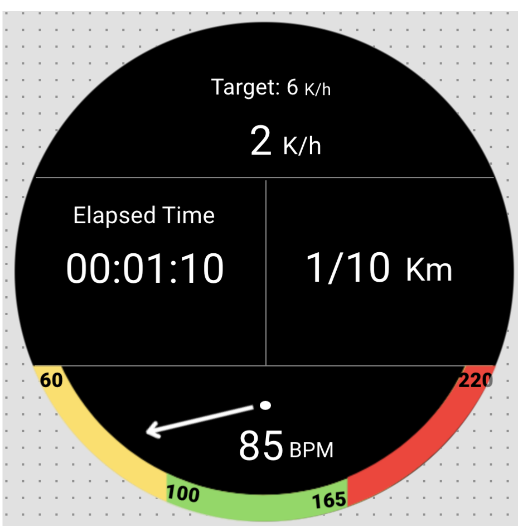
Following the completion of the initial run in the study (after step one), participants were afforded a brief respite of ten minutes. This interlude was provided to allow them to recover and to facilitate the completion of the initial questionnaire (step two).

5.2.1 Tasks and Its Relation to the Questions

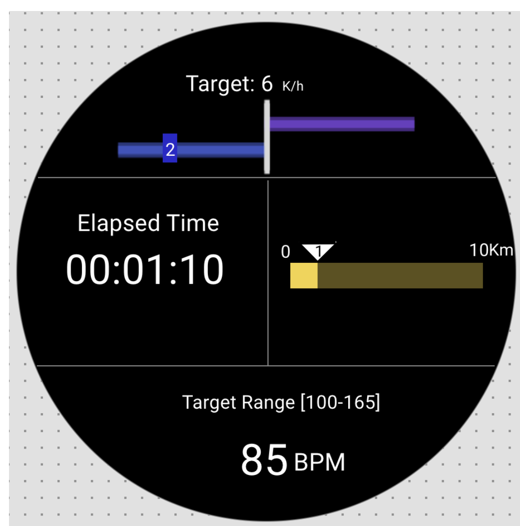
In Chapter 4, we discussed a series of tasks designed for the experiment that participants were to answer while running. Participants repeated the five tasks for a total of 13 questions with each watch face. TDistance (Questions 4, 8, and 11 mentioned in appendix) was repeated three times (each instance once), TPace (Questions 1, 6, and 13) and THeartRate (Questions 2, 7, and 10) were repeated three times each, and TPaceAdjust (Questions 3, and 9) and THRAdjust (Questions 5, and 12)were repeated twice each. Figure A.1 shows an example of how WF1 and WF2 was presented for Question 1 for TPace. (You can find all the Watch face Interfaces for all 13 questions in appendix section.)

The repetition helped us assess the consistency of participants' responses, ensuring that the Visualization (or Text) reliably showed the needed information. The purpose of these tasks was to evaluate how quickly and accurately participants could assess their current performance relative to a set target.

The tasks were not based on participants' actual running data; Instead, we generated a consistent dataset used by all participants to ensure consistency and maximize internal experimental validity.



(a) Question 1, WF1



(b) Question 1, WF2

Figure 5.1: Question 1: the task of realizing whether their speed is above or below the target where the left picture shows the watchface number 1 and the right picture shows watchface number 2.

The answers were of three kinds. Binary responses were used for TDist, TPace, and THR, where participants could answer ‘Yes’ or ‘No,’ ‘Beginning’ or ‘End,’ or ‘Above’ or ‘Below’. For estimating the adjustment (TPaceAdjust and THRAdjust), they could choose from three options: ‘A little’, ‘Somewhat’, or ‘A lot’. To avoid problems of subjective interpretation, we explained these terms at the beginning of the study. Although ‘Somewhat’ was an available option, the generated dataset presented the data so that the only correct answers for these tasks were either ‘A little’ or ‘A lot’—but participants were unaware of this.

5.2.2 Experiment Setup

The experiment took place in a controlled environment at Centennial Stadium (Figure 5.2), located at the University of Victoria. The stadium features a 400m synthetic rubberized track surface, which facilitates controlled communication for data collection.

The visualization system was built using Watch Face Studio, a graphic authoring tool for creating watch faces for the Wear OS smartwatch ecosystem. We tested the watch faces on a connected Fossil Gen 6 Wear OS smartwatch. During the test, the experimenter sent the watch faces displaying specific data for each task to the smartwatch via a wireless connection between the smartwatch and Watch Face Studio.

To maintain communication between the experimenter and the participant while they were running, we used wireless AirPods. The experimenter was present in the stadium, recording the call between the participants and the experimenter in which participants answered each question.

Participants were instructed to look at the watch only when prompted to complete a task. Each task instance was executed as follows:

1. The experimenter verbally asked the question.
2. Upon hearing a trigger sound at the end of the question, the participant looked at the watch.
3. The participant answered the question verbally.

Completion time was calculated as the difference between when the participant started providing the answer verbally and when the trigger sound was played.

Figure 5.2: Experiment setup centennial stadium track and field.

5.3 Pilot Studies

Before the main study, we conducted two pilot studies with two volunteer runners to assess the study design and procedure. During the pilots, we examined the clarity of the questions and the visualization of the data fields, the timing of each step of the experiment to ensure it did not exceed an hour, the connectivity between the experimenter and the participants during the run, voice recording, and the connection between the smartwatch and the system sending the watchface to the watch while the participant was running.

In the first pilot test, we also wanted to determine if participants were relying on their memory to answer the questions for the second watch face. Since the 13 questions and the unique data displayed for each question were the same for both watch faces, the only difference being the type of data presentation (Visualization vs. Text), there was a concern that participants might remember their answers from the first watch face and thus respond faster in the second run, regardless of the presentation type.

To address this, we asked both participants in the first pilot study if they were answering based on their memory. Both participants confirmed that they did not notice that the data was the same for both watch faces, so we decided not to change the order of the questions for the second run with the second watch face.

The only major change between the pilot test and the final evaluation test was in the tasks we described. After conducting the initial pilot study, we identified an issue with the task structure. Initially, we combined questions about Pace/Heart Rate adjustments with questions about whether these metrics were above or below target. For example, the original question format was: “Is your Pace above or below the target, and how do you need to adjust it?" This dual-question format proved challenging for participants during the first pilot study. They struggled to remember and accurately respond to the adjustment part of the question after focusing on whether their Pace/Heart Rate was above or below target.

We also realized that pilot participants spent more time observing their data to answer the first part of the question, making their immediate response to the adjustment part less valid. To address these issues, we decided to split each question into two separate questions. The first question asks whether their Pace or Heart Rate is above or below the target. The second question, introduced with additional information before the question, focuses only on how participants plan to adjust their Pace/Heart Rate (once you look at the screen, your Pace/Heart Rate would be Above/Below the target; how do you need to adjust it?).

This additional information aimed to ensure that participants' attention remains only on the adjustment process rather than splitting their focus between assessment and adjustment. Consequently, we finalized the task structure with five separate tasks instead of three.

Following these task revisions, we conducted a second pilot test with two different volunteer runners from the first pilot study. This second pilot aimed to verify clarity and minimize potential errors before proceeding to the main study. After confirming the clarity and not encountering any errors, we proceeded to the evaluation test without making further changes.

5.3.1 Test Procedure

The evaluation test had six phases and was completed by one participant at a time. The test started with the participants assigned to group 1 starting a 10-minute run while using the first watch face (WF1). Throughout the running trial, the experimenter asked the 13 questions and recorded the participants' responses.

The same thing was repeated for the participants assigned to group 2; They started the 10-minute run using the second watch face (Phase 1). After finishing the initial running trial, participants were given a 10-minute recovery break. During this time, they completed the first part of the survey, which asked them to evaluate their experience with the design of each data field of the first watch face (Phase 2). The survey used a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree).

	Strongly disagree 1	Somewhat disagree 2	Neither agree nor disagree 3	Somewhat agree 4	Strongly agree 5
The representation was clear and easy to understand					
The representation was helpful to answer the questions related to this data field					
I would use this representation if it was supported by my watch					

Figure 5.3: Rating table used for each data field.

After finishing the survey, participants began their second trial, using the other watch face for 10 minutes (Phase 3). The procedure was the same as Phase 1: the experimenter asked the same set of questions in the same order and recorded the participants' responses.

Following the second trial, participants were requested to complete the same survey, this time rating the design of each data field for the second watch face (Phase 4).

Finally, in the last phase, participants were asked to complete the final part of the survey. They were shown side-by-side representations of both watch faces and asked to rate each data field on a scale from 1 to 5, where 1 signified least preferred, 3 represented neutrality, and 5 indicated a strong preference(Phase 5).

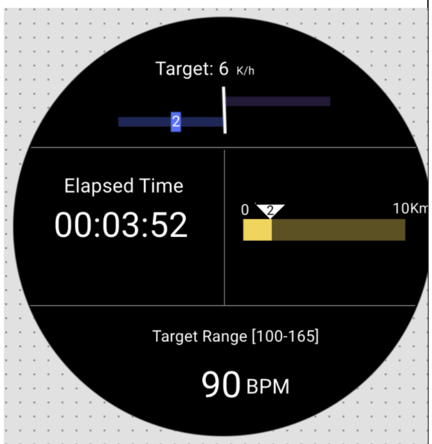
		
Distance		
Heart rate		
Speed		

Figure 5.4: The Final rating template for both watchfaces.

The session ended with a short interview. This discussion focused on participants' experiences with both watch faces and comparing the different designs (Text/Visualization), asking for feedback on possible improvements and any other comments they had (Phase 6).

5.4 Hypotheses

We formulated 5 hypotheses (one for each task) regarding completion time. We did not formulate hypotheses regarding accuracy because our pilot studies showed that accuracy was near perfect for all tasks with both representations. For the following reasons, we hypothesized that all tasks would be completed faster with Visualization than with Text.

- $Hypothesis_{PaceAboveBelow}$

  The average time to answer this question using Visualization would be faster than using Text.

  Rationale: There is no need to compare numbers. Participants can easily answer by indicating the position of the marker to see if it is on either the left or right side of the baseline.

- $Hypothesis_{HeartInRange}$

  The average time to answer this question using Visualization would be faster than using Text.

  Rationale: In Visualization, there is no need to compare numbers; The user can easily answer by indicating which color the pointer is pointing at. In contrast, Text requires reading three numbers and determining if one value is both higher than the minimum value and smaller than the maximum value.

- $Hypothesis_{PaceAdjust}$

  The average time to answer this question using Visualization would be faster than using Text.

  Rationale: Visualization requires estimating length (the vertical Distance between the marker and the baseline), while Text requires reading and estimating the proportion between two numbers.

- $Hypothesis_{Distance}$

  The average time to answer this question using Visualization would be faster than using Text.

  Rationale: Visualization supports visual estimation of proportions, while Text requires mental calculation.

- $Hypothesis_{HeartAdjust}$

  The average time to answer this question using Visualization would be faster than using Text.

  Rationale: Visualization requires estimating the angle between the arrow and a virtual vertical line, while Text requires reading three numbers and estimating the proportion between one value and a range.

5.5 Data Collection

We used four spreadsheets for data collection:

Demographic Information: Participants filled out this information on a paper given to them prior to starting their run. In the demographic survey, we asked about participants' age, gender, education level, and running habits (frequency per week). We did not provide multiple choices for answers and asked them to fill in the blanks with their own answers. This approach was chosen so we would know their exact responses and better understand who our participants are.

Completion Time and Accuracy: This recorded the time taken (in seconds and milliseconds) for each participant to answer each question and whether their answer was correct. Completion time was the difference between the start of the participant's verbal response and the trigger sound.

Data Field Ratings: As shown in Figure 5.3 participants rated each data field on a scale from 1 to 5, where 1 signified least preferred, 3 represented neutrality, and 5 indicated a strong preference for the watch face based on their experience running with either the Visualization or Text version.

Interview Responses: This captured participants' feedback on their experience, preferences, and any additional comments.

Chapter 6

Study Results

We analyzed completion times using Bootstrapped [41] confidence intervals (CIs) rather than $p-$values, following recommendations for statistical practices in HCI and visualization (e.g., [25], [42]–[44]). We interpret the statistical significance of the overlap of CI bars following [45]. We performed all analyses with 95% CIs computed with 1000 replicates and the BCa method [41], using the mean of the repetitions as the aggregated measure for each $\{participant, task, representation\}$ triplet.

6.1 Visualizing Quantitative Data

To better understand the data collected from testing the watch faces and to facilitate easier analysis, we use confidence intervals to visualize the average completion time for each task in different presentation types (Text/Visualization). There is also another representation of confidence intervals to show the differences in average completion time between the two presentation types. Additionally, we used data collected through a Likert scale to visualize the ratings for each data field in both presentation types (Text/Visualization).

6.1.1 Average Completion Time For Each Task

We developed a Python script to perform a bootstrap confidence interval (CI) analysis on average completion times for two user interface conditions: "Visualization" and "Text". This method resamples the data 1000 times to estimate the variability and the 95% confidence interval for each condition, rather than relying solely on traditional statistical tests. The results are then visualized using a horizontal bar chart, where each bar represents the mean completion time for one condition, with error bars showing the range of the confidence intervals. Additionally, I annotated the bars with the exact mean values and included visual markers to highlight them. Figure 6.1 is an example that illustrates how each representation of confidence intervals represents the mean estimate, and the black line shows the 95% confidence interval.

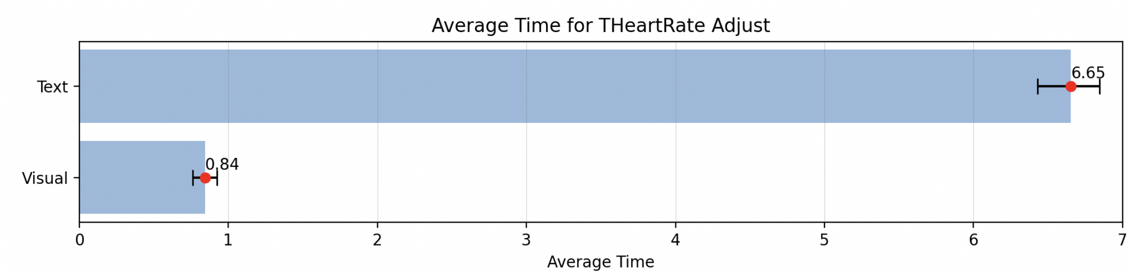


Figure 6.1: Bar charts representing average completion time for the Heart Rate Adjustment task in different presentation types (Text/Visualization) along with the 95% confidence interval

6.1.2 Differences In Completion Time

To determine whether there are significant differences in completion time between different presentation types (Text/Visualization) we calculated the average completion time for Text minus average completion time for Visualization. For each task, we visualized the differences in means with a horizontal bar chart with 95% CIs as shown in Figure 6.2.

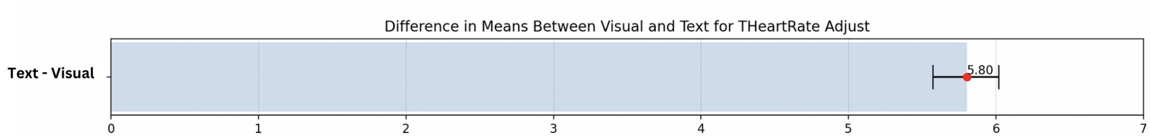


Figure 6.2: Bar chart representing differences in means and CI for completion time in the Heart Rate adjustment task

6.1.3 Likert Scale For Ratings

We used a Likert scale to collect survey responses and evaluate participant preferences for different presentation types (Text/Visualization) for each data field. These scales show how participants' responses are distributed, highlighting preferences and areas of disagreement. This helps us see patterns, such as which designs are most or least preferred. Figure 6.3, for example, shows participants' ratings for the different criteria asked in the survey for Heart Rate presented only in Text.

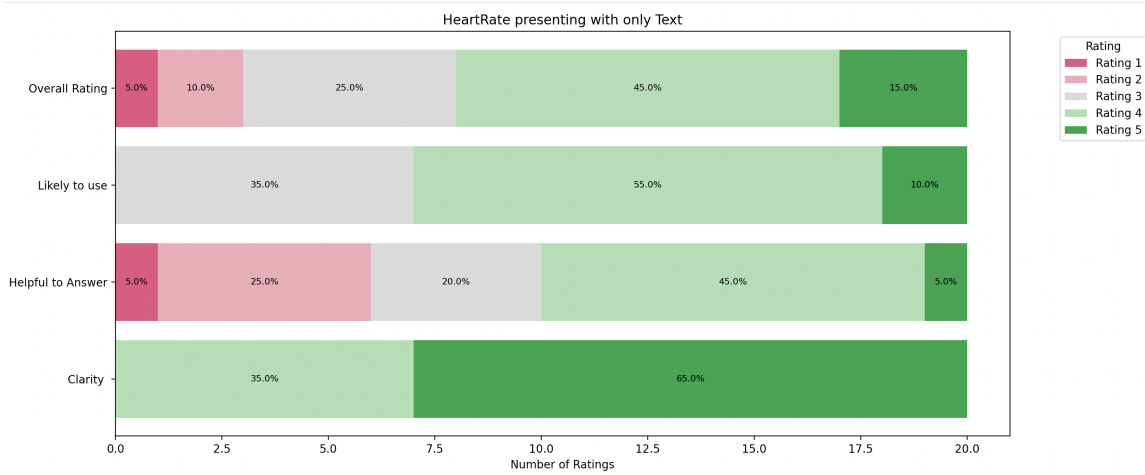


Figure 6.3: This visualization represents different rating for each criteria in the watch that Heart Rate was presented only by Text.

6.1.4 Confirmatory Analysis

For this study, the entirety of our recruited participants (20) were included in the analysis. In this section, we present the results for each hypothesis one by one.

We explain how pairwise comparisons show that participants were significantly faster with Visualization than with Text across all tasks, confirming our hypotheses.

- $Hypothesis_{PaceAboveBelow}$

  Analyzing the results shown in Figure 6.4, we can see that the average response time for the watch face with Visualization is much lower, about 1.01 seconds, as indicated by the red dot. In contrast, the Text-based display had an average response time of around 1.83 seconds. The error bars, representing the confidence intervals around the mean, indicate a clear difference between the two presentation types. The display with Visualization has a tighter confidence interval, meaning we are confident that the mean estimate (the red dot) is true.

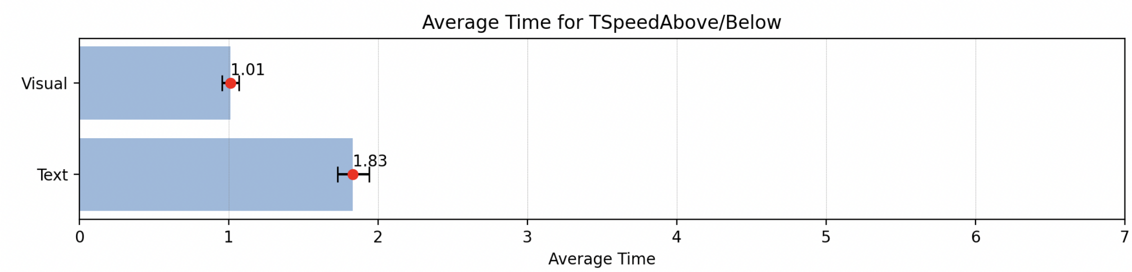


Figure 6.4: Mean time (seconds) with 95% CI for two type of presentation types (Text/Visualization) for the task TPace: Is your Pace above or below the Target Pace.

The data supports the hypothesis that using a visualization aid results in faster response times compared to using text alone. The results confirm that the Visualization helps with quicker responses, as participants can make decisions based on the graphical representation rather than processing numerical data.

In the comparison between Visualization and Text Display, we have evidence to strongly accept $Hypothesis_{PaceAboveBelow}$. As shown in figure 6.5 this conclusion is based on the observed pairwise difference of 0.82.

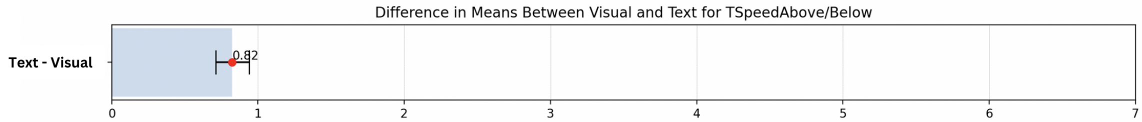


Figure 6.5: Difference of mean time (seconds) with 95% CI for two type of presentation types (Text/Visualization) for the task TPace: Is your Pace above or below the Target Pace.

- $Hypothesis_{HeartInRange}$

  The $Hypothesis_{HeartInRange}$ stated that Visualization would help runners respond faster to determine if their Heart Rate is in the target range. The results, shown in Figure 6.6, support this, with participants taking 2.40 seconds with Visualization compared to 3.88 seconds with Text and a lower confidence interval indicating quicker data comprehension. Figure 6.7 also shows participants understood Heart Rate information 1.48 seconds faster with Visualization, confirming faster completion times and supporting $Hypothesis_{HeartInRange}$.

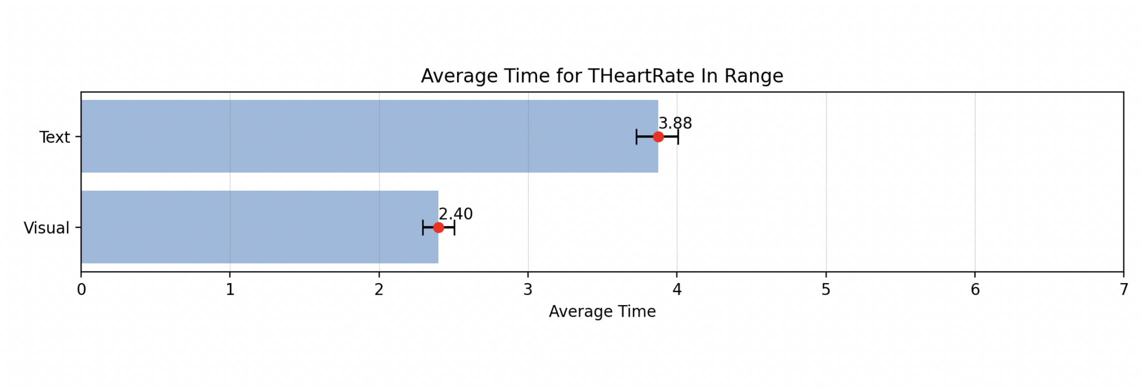


Figure 6.6: Mean time (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task THR: Yes or No, your Heart Rate is in the target range.

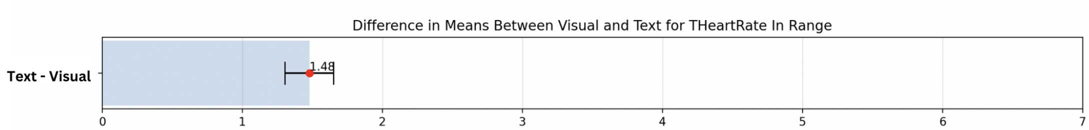


Figure 6.7: Difference of mean time duration (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task THR: Yes or No, your Heart Rate is in the target range.

- $Hypothesis_{PaceAdjust}$

  The $Hypothesis_{PaceAdjust}$ suggested that using Visualization would help runners respond faster to the task of adjusting their Pace. The results in Figure 6.8 support this, with participants taking 1.32 seconds with Visualization compared to 3.78 seconds with Text. Figure 6.9 further shows a 2.46 seconds quicker response with Visualization.

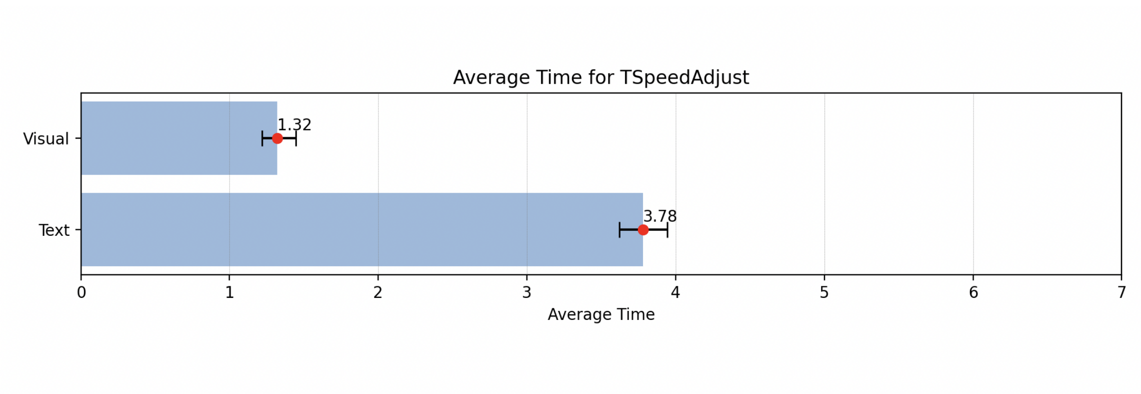


Figure 6.8: Mean time (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task TPaceAdjust: How much do you need to adjust your Pace?

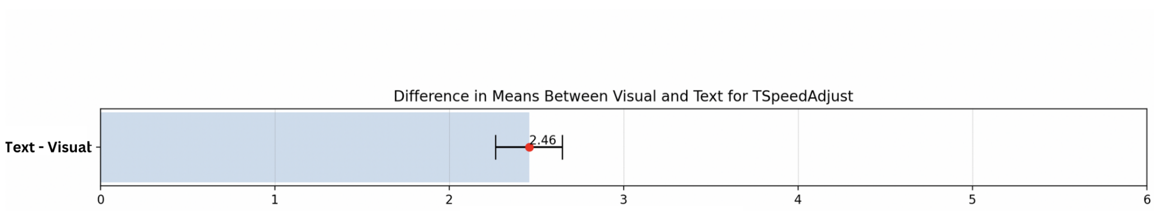


Figure 6.9: Difference of mean time duration (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task TPaceAdjust: How much do you need to adjust your Pace?

- $Hypothesis_{Distance}$

  Figure A.4 shows that using Visualization on smartwatches made participant respond much faster to Distance-related questions. On average, it took 1.52 seconds with Visualization compared to 4.83 seconds with Text. This supports the idea that Visualization cues, like showing progress toward a goal, can be understood much quicker than Text. As shown in Figure 6.11, On average, participants responded 3.31 seconds faster with Visualization than Text. This big difference, confirmed by confidence intervals, strongly supports our hypothesis.

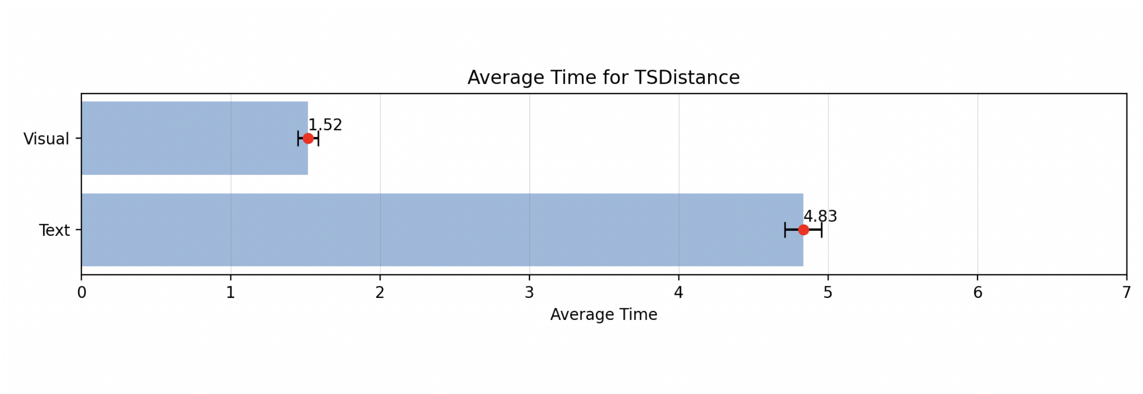


Figure 6.10: Mean time (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task TDistant: Have you completed a certain portion of the target Distance?

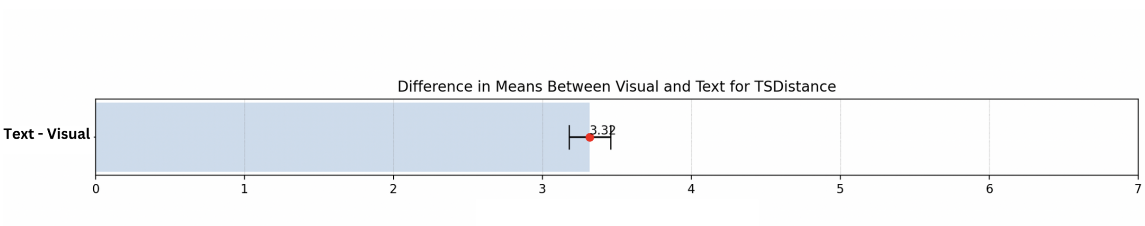


Figure 6.11: Difference of mean time duration (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task TDistant: Have you completed a certain portion of the target distance?

- $Hypothesis_{HeartAdjust}$

  The chart in Figure 6.12 shows that runners answered the Heart Rate adjustment question faster with Visualization than with Text. As shown in Figure 6.13, on average participants using Visualization took 0.84 seconds, compared to 6.65 seconds with Text. This supports the hypothesis that answering the Heart Rate adjustment question is faster with Visualization than with Text, as participants using Visualization took an average of 5.80 seconds less than those using Text. This difference is statistically significant, and the error bars confirm that this result is reliable.

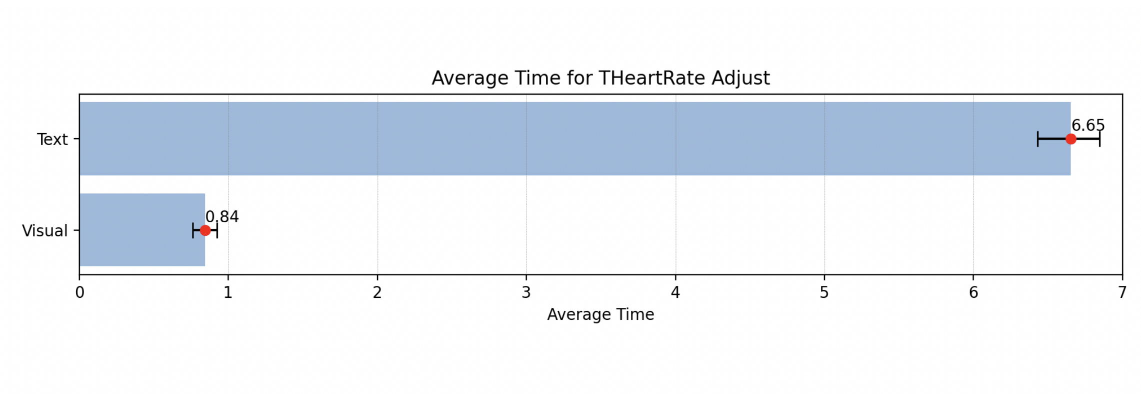


Figure 6.12: Mean time (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task THRAdjust: How much you need to adjust your Heart Rate.

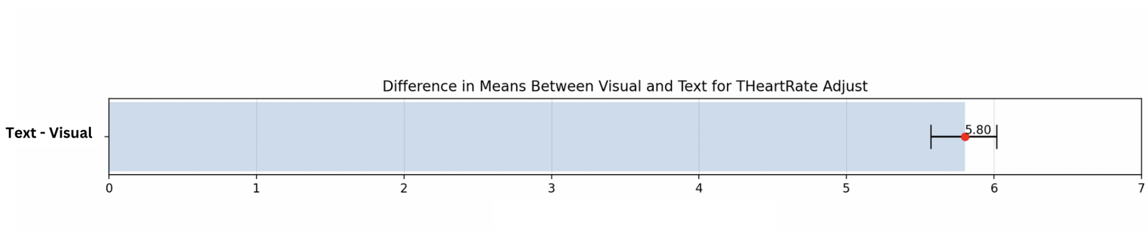


Figure 6.13: Difference of mean time duration (seconds) with 95% CI for two types of presentation types (Text/Visualization) for the task THRAdjust: How much you need to adjust your Heart Rate.

Figure 6.14 shows that Participants are consistently faster with the Visualization on the watch faces compared to Text across multiple tasks. This trend is clear from the mean differences, which favor Visualization. The data suggests that Text slows down response time compared to Visualization for tasks like interpreting Heart Rate and adjusting Pace and Distance. The advantage of Visualization is that they are more visible in tasks involving Heart Rate settings, with response times faster by nearly one to six seconds.

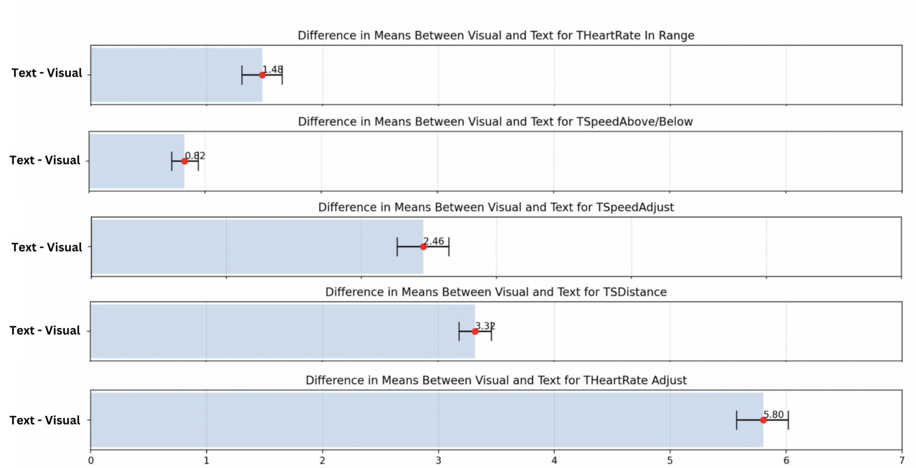


Figure 6.14: Mean Differences between Visualization and Text for each task.

6.1.5 Subjective Responses

Overall, Visualization was largely preferred for both Distance and Heart Rate, and slightly preferred for Pace. In this section, we examine the Likert charts showing participants' ratings for clarity, likelihood of use, and helpfulness in responding to tasks for both presentation types (Text/Visualization) in the watchfaces.

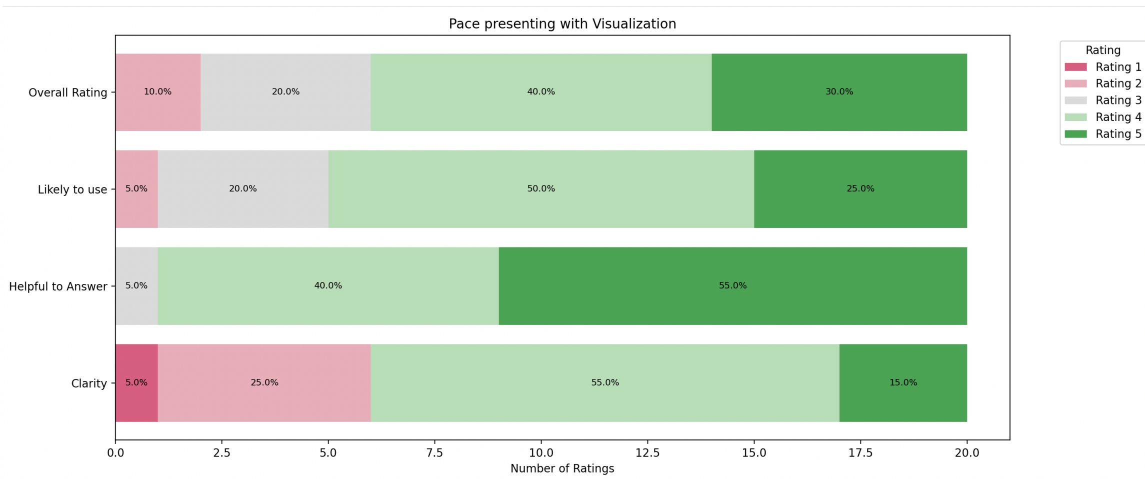


Figure 6.15: Subjective ratings for Pace presented with Text.

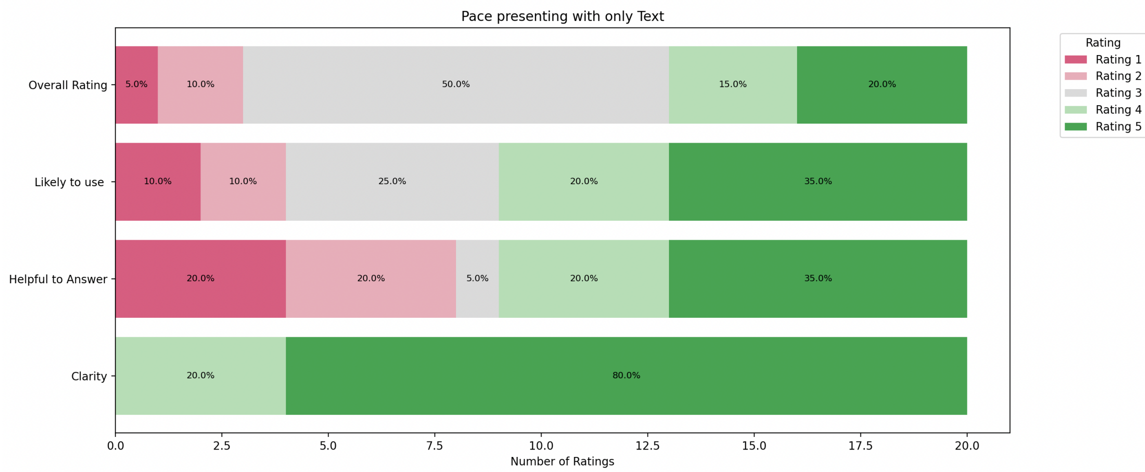


Figure 6.16: Subjective ratings for Pace presented with Visualization.

Figure 6.15 displays the Likert chart results for the Pace data field with Visualization, while Figure 6.16 shows the results for the Pace data field with Text. Comparing the rankings, the watch with Visualization was more helpful for answering questions, with 55% rating it 5 and 40% rating it 4. However, participants had difficulty understanding the design and clarity of the Visualization, with 25% rating it 2 and 5% rating it 1. In contrast, the Text version received no low ratings and had 80% rating it 5.

For the likelihood of use, 35% of participants rated the Text version a 5, while 25% rated the Visualization version a 5. However, combining ratings 4 and 5, the Visualization version had a total of 75% favorable ratings and only 5% rated it 2, compared to 55% favorable ratings and 20% least preferable ratings for the Text version. This indicates that the watch with Visualization was generally more preferred.

Overall, 70% of participants preferred the watch with Visualization, while only 35% preferred the Text version, even after their second run and overall rating.

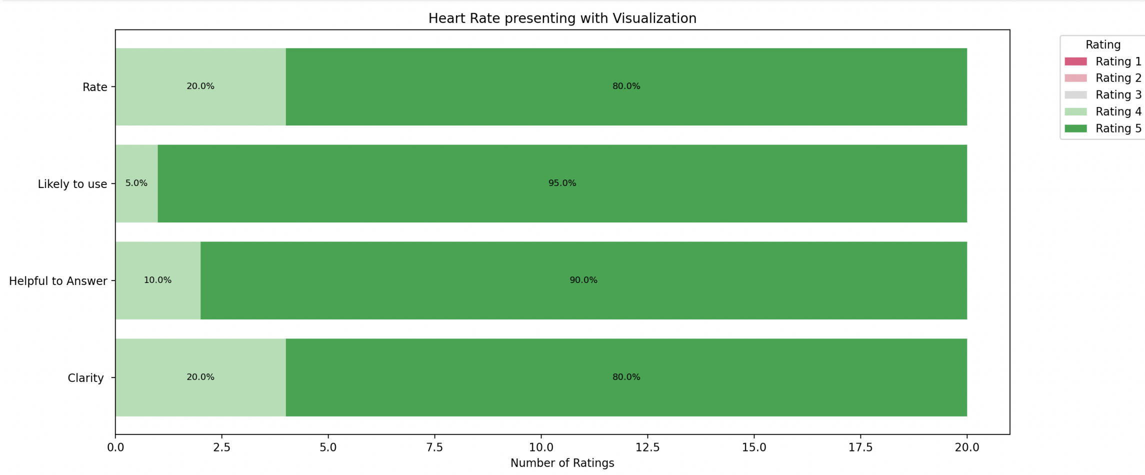


Figure 6.17: Subjective ratings for Heart Rate presented with Visualization.

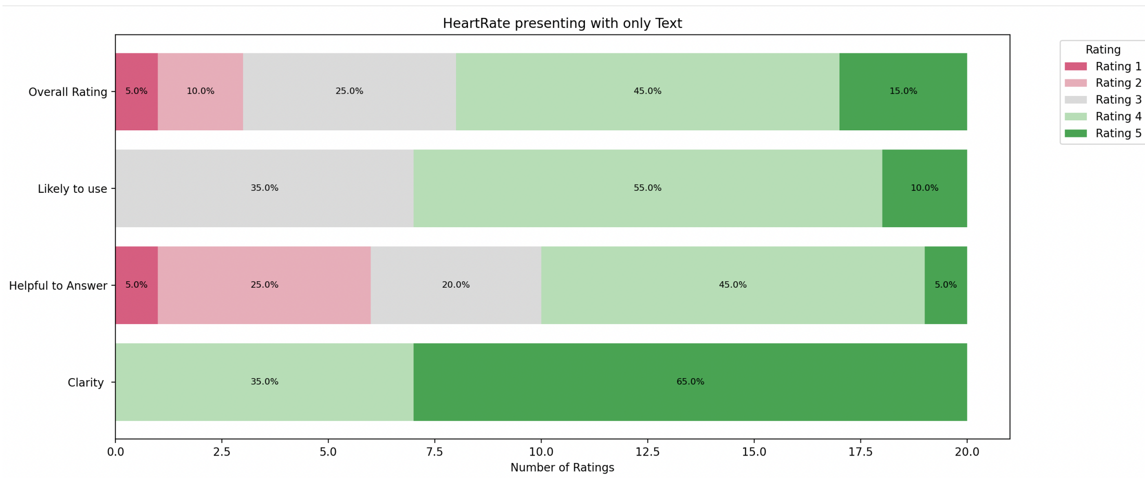


Figure 6.18: Subjective ratings for Heart Rate presented with Text.

Comparing the data visualization for Heart rate data field in both watchface, we can observe that the watchface using Visualization has been the most preferred in each category. The data visualized in this chart indicates a strong participant preference for the Visual presentation of Heart Rate information on a watch face. The majority of participants rated the clarity, helpfulness, and likelihood of using the Visualization very highly, with most of the ratings concentrated at 4 and 5.

While it seems participants had no difficulty understanding the watch face showing the Heart Rate with Text and the clarity for it has been ranked green, participants did not find the Text version as helpful to answer the questions for the watch face with presenting the Heart Rate by rating red for 35%. Also, it is noticeable that even though the watch face with the Visualization has 90% rated 5 for likely to use, seeing 0% ranked 1 or 2 shows that there have been no objections against using the watch presenting Heart Rate with Text.

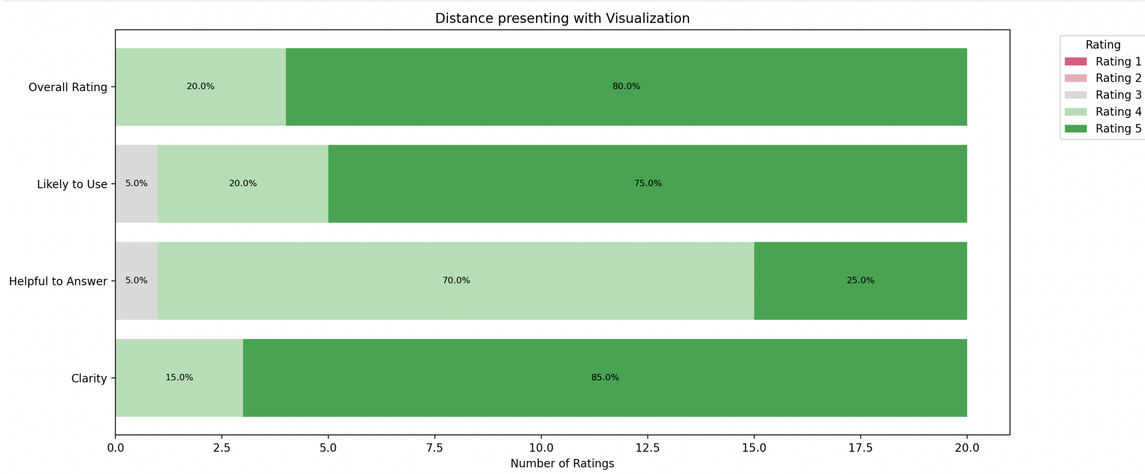


Figure 6.19: Subjective ratings for Distance presented with Visualization.

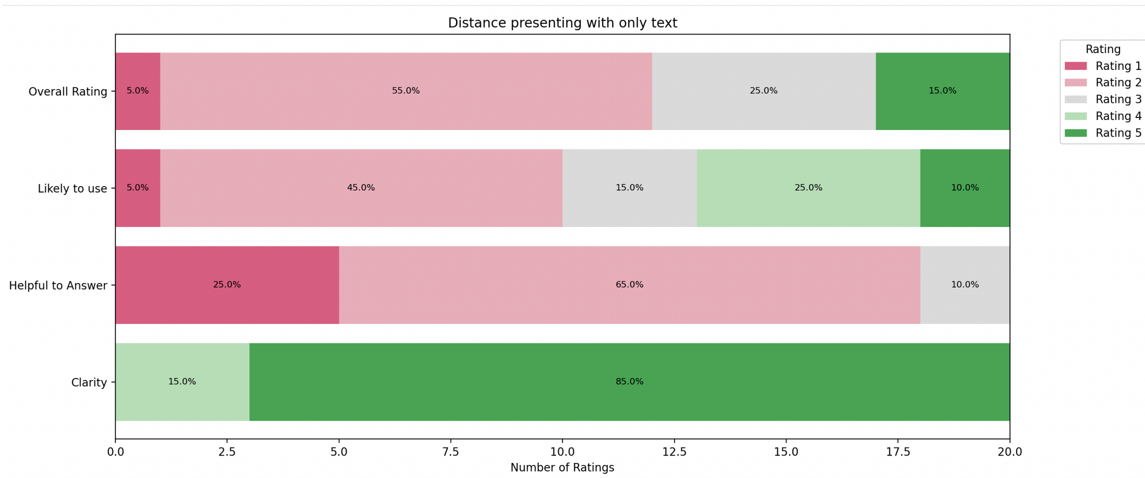


Figure 6.20: Subjective ratings for Distance presented with Text.

Comparing the data visualization for the Distance data field in both watchfaces, we observe that the watchface using Visualization was the most preferred in each category, with being rated only 4 and 5. In terms of clarity, both presentation types (Text/Visualization) were rated the same, with 85% rating 5 and 15% rating 4. However, none of the participants found the watchface presenting Distance with only Text helpful for answering the questions, with 25% rating it 1 and 65% rating it 2, resulting in a total of 90% least favorite ratings in this category. Only 35% rated it highly for likelihood of use, which decreased to 15% after the second run and in the overall rating.

Participants rated the Visualization highly across all categories, with most giving it the highest rating of 5 for clarity, helpfulness, the likelihood of use, and overall satisfaction. This suggests that the Visual presentation of the Distance data field is clear, useful, and likely to be used by users. The few ratings of 4 across categories indicate minor areas for improvement, but the absence of ratings 1 and 2 reflects a generally high level of satisfaction with the Visual representation of Distance.

Chapter 7

Discussion

In this chapter, we interpret what we have learned from the evaluation results. First, we examine the results of the quantitative phase of the study. We contextualize these results with quotes from participants during the interview to discuss the usefulness of visualization when running. Next, we discuss how context affects runners' representational needs. Later, we discuss the participants' feedback from the interviews. Finally, we address the limitations of our study and suggest future work that could further improve the presentation of data fields in smartwatches for runners.

7.1 Usefulness of Visualization While Running

Even with simple tasks that involve only a few numbers, visualization helped significantly. As shown in Figure 6.14, The difference in completion time between using Visualization and using Text is always a positive number. Participants were faster with Visualization than with Text for all tasks, and participants were often aware of it, as P1 said *"I quickly could see my progress visually, without having to check numerical values."* Several participants said that they struggled to complete TDistance with Text because of the cognitive load associated with performing mental math while running, as P10 said *"I don't like to do any sort of math or thinking so something that shows me where I am and where I need to be is pretty handy."*

While participants generally performed tasks more quickly using Visualization compared to Text in all cases, when it came to the specific task related to "Pace", their personal opinions and subjective preferences did not demonstrate a strong preference for Visualization over Text. In other words, even though Visualization helped participants complete tasks faster, they only slightly preferred Visualization to Text when dealing with the"Pace" data. Participants explained, for example, P14 said *"I prefer [the] numeric one, [it's] easier to just see the current Pace"*, and P3 said *"I am more comfortable with seeing a big number instead of bars and colors."*.

Considering all the illustrated Likert Scale, we can strongly state that the data fields for Heart Rate, as shown in Figure 6.17 and Distance, as shown in Figure 6.19, were the most preferred data fields when presented with Visualization. Participants were satisfied with the design in terms of comprehending it and finding it useful for running. While the data field for Pace was slightly favored towards the watchface with Visualization as shown in Figure 6.16, the results show that there is room for improvement in the design and visualization of this data field.

Overall, visualization are more than just an acceptable or adequate option; They are a strong and effective alternative to using numerical data for runners. Visual representations of information can be just as good, if not better, than relying on numbers alone when it comes to helping runners understand their data. As P14 said, *"I always find the visual representations are often easier to understand than numbers, especially when I am running or when I am doing a physical activity"*. This emphasizes the important role of glanceable visualization in helping runners quickly understand data during certain realistic tasks while running. This importance is likely even greater when runners have limited mental capacity [27], such as during intense activities like high-intensity interval training or racing. In challenging real-life environments, such as running on trails or in traffic, the ability to quickly glance at a smartwatch becomes crucial because spending too much time looking at the watch could lead to accidents, such as falling or getting injured.

7.2 Contextual Effectiveness of Data Representations for Runners

The findings of our study should be understood in the context of the specific tasks the participants were asked to perform. Although Visualization were effective for the tasks included in the study, there are many other tasks that are important for runners. For these broader tasks, numerical data might be more flexible or useful than visualization, allowing runners to accomplish a wider variety of tasks. Some participants recognized this, with P19 saying, for example, *"Visualization was good* *for comparing, but I prefer just numbers to see my speed."* P5 commented that *"Text and Visualization are very equally good to me, if both adapt my watch, based on my running plan I can choose either or."*

This outlines the usefulness of providing different ways to present data that runners can customize to their own specific running activities. Some activities, like high-intensity interval training or hill repeats, require careful monitoring of multiple metrics since these metrics directly measure performance and success. In contrast, other activities, such as a base run, may not require any data monitoring, as the primary goal is simply to spend time running without focusing on specific metrics.

How data is represented should depend on the specific context and task, which is particularly evident when considering how goals are achieved. Different tasks or situations may require different data representation types to help runners achieve their goals effectively. Echoing past work [18], three participants explained that they felt more motivated to achieve their goals and perform well with Visualization than with Text for TDistance. P5 said *"I was having a sense of accomplishment, and it was encouraging me to push myself further."*; When P16 said *"the visual designs were giving me this feeling to focus on the goal."*; and P8 *"the bar that was filling with color as I was running was giving me more motivation to fill it as soon as possible."*.

Understanding the influence of various visualization could help improve how runners set and achieve their goals and stay motivated. Future work should look into examining how different visualization designs impact goal-setting and motivation.

7.3 Qualitative Feedback

In the interview, we elicited further discussion from participants who deviated from the normal response patterns in terms of accuracy and response time. Although we did not formulate hypotheses regarding accuracy, as our pilot studies showed near-perfect accuracy for all tasks with both presentation types, we collected participants' responses to each question along with the completion time. We recorded four incorrect responses to the TDistance task when the data was presented using Text.

We investigated why participants gave incorrect responses to the Distance data field when presented as Text, compared to correct responses using Visualization. All four participants who answered incorrectly explained that they found it hard to do math while running. This made them take longer to respond and led to mistakes. In contrast, they found the Visual Distance data easy to understand at a glance, resulting in correct answers. As P3 said, *"trying to do math in my head while running threw me off. I was second-guessing myself. But when I looked at the visual data, I got it instantly."*; And P18 said *"I am not good at calculating numbers, especially when I'm doing an activity, but with the visualization, I didn't have to think much—it just made sense right away."*.

7.4 Limitations and Future Work

Our study has several limitations that should be considered when interpreting the results. Understanding these limitations helps contextualize our findings and guides future research directions.

The results of our study are specific to the set of tasks we tested and the demographic of runners we recruited. We expect that different tasks could lead to different outcomes. For instance, participants might be faster with Text than with Visualization if they are asked to retrieve an exact value and if the Visualization lacks precise labels. In such cases, participants would also likely be more accurate with Text. Future studies should establish a comprehensive list of meaningful tasks for runners based on various contexts, such as activity type, environment, and social setting, to investigate the usefulness of visualization more broadly.

Runners vary widely in their expertise and practice, which can influence their data needs and task performance. Some runners participate in casual runs, such as a turkey trot once a year, while others are weekend warriors, elite athletes, or professional runners. They may run on different terrains—tracks, roads, forests, or mountains—and in various conditions—sun, snow, or rain. Additionally, the Distance they run can range from 1 km to 200 km. Future research should consider these factors and study how the context affects tasks and representational needs. Conducting this study required us to rethink and adapt our typical approach to HCI and visualization studies. While HCI has a long history of conducting quantitative in-lab controlled experiments, it seldom conducts studies in realistic environments due to the increased difficulty. Our field lacks the infrastructure, knowledge, and practice history for such studies, making it challenging to find appropriate methods. This limitation presents both a challenge and an opportunity to develop and implement new methodological frameworks.

Our study used a synthetic dataset to ensure consistency across participants in task difficulty. However, this came at the cost of less realistic data, which might have distracted some participants who realized the data was inaccurate. This limitation highlights the trade-off between maintaining task consistency and using realistic data that participants might find more engaging and relevant.

We chose to ask questions and collect responses verbally, which introduced several challenges. Technological issues included how to send prompts, measure response times, and collect participant data. Bluetooth transmission likely introduced some noise in the response times collected. These factors threaten both external and internal validity, raising questions about the feasibility of fully automating this type of study, administering tasks, and accurately collecting participant data.

The factors mentioned above collectively impact our study's external and internal validity. External validity concerns the generalizability of our findings to other settings, populations, and times. The specific tasks, participant demographics, and controlled environment limit this generalizability. Internal validity refers to the extent to which we can attribute the observed effects to the interventions rather than other variables. The use of synthetic data, verbal data collection, and potential noise in response times are threats to internal validity.

Chapter 8

Conclusion

Our study evaluated the performance and preferences of 20 runners completing tasks on an outdoor track using Text and Visualization. The results showed that runners completed tasks more quickly with Visualization and expressed a clear preference for these visual tools, indicating their willingness to use them during runs.

Visual data displays have significant potential to enhance runners' understanding of their metrics, such as Heart Rate, Distance, and Pace. Runners often engage in various activities across diverse environments, each presenting unique challenges and data needs. Our research highlights the importance of moving beyond controlled environments, like treadmills, to study how visualization can assist runners in real-world, variable conditions.

The ability to quickly and accurately interpret visual data under physical and cognitive strain is valuable not only for runners but also in other scenarios where users operate in challenging conditions. Our findings suggest that well-designed visualization can reduce cognitive load and improve decision-making, thereby enhancing performance and overall experience for runners.

However, our study was limited to a specific set of tasks and a particular demographic of runners. Future research should explore a broader range of running activities, conditions, and participants to understand better how visualization can support the diverse needs of the running community. Additionally, studies conducted in more natural settings will provide deeper insights into the practical application of visualization in the environments where runners most need them.

Bibliography

[1] Emmanuel Stamatakis and Moushumi Chaudhury. "Temporal trends in adults' sports participation patterns in England between 1997 and 2006: the Health Survey for England". In: *British journal of sports medicine* 42.11 (2008), pp. 901–908.

[2] Laila Ottesen, Rikke Schou Jeppesen, and Birgitte R Krustrup. "The development of social capital through football and running: studying an intervention program for inactive women". In: *Scandinavian journal of medicine & science in sports* 20 (2010), pp. 118–131.

[3] Linda Ooms, Cindy Veenhof, and Dinny H de Bakker. "Effectiveness of Start to Run, a 6-week training program for novice runners, on increasing health-enhancing physical activity: a controlled study". In: *BMC public health* 13 (2013), pp. 1–12.

[4] Freya Oswald, Jennifer Campbell, Chloe Williamson, et al. "A scoping review of the relationship between running and mental health". English. In: *International Journal of Environmental Research and Public Health* 17.21 (Nov. 2020). ISSN: 1660-4601. DOI: 10.3390/ijerph17218059.

[5] Strava. *Strava Global Report 2023.* 2023. URL: https://downloads.ctfassets.net/9olkiac82a1q/iU421iiO6Hky1MVvYThmz/cf2f510407b1b16be35d1c323722f47a/Strava-Global-Report-2023-en-US.pdf.

[6] Mark Janssen, Ruben Walravens, Erik Thibaut, et al. "Understanding different types of recreational runners and how they use running-related technology". In: *International journal of environmental research and public health* 17.7 (2020), p. 2276.

[7] Jihyeon Oh and Daehwan Kim. "Workout with a Smartwatch: A Cross-Sectional Study of the Effects of Smartwatch Attributes on Flow Experience and Exercise

Intentions Depending on Exercise Involvement”. In: *Healthcare*. Vol. 11. 23. MDPI. 2023, p. 3074.

[8] Daniel Ashbrook, James Clawson, Kent Lyons, et al. “Quickdraw: The impact of mobility and on-body placement on device access time”. In: Apr. 2008, pp. 219–222. DOI: 10.1145/1357054.1357092.

[9] Florian ’Floyd’ Mueller, Chek Tien Tan, Rich Byrne, et al. “13 Game Lenses for Designing Diverse Interactive Jogging Systems”. In: *Proc. Annual Symposium on Computer-Human Interaction in Play*. New York: ACM, 2017, pp. 43–56. ISBN: 9781450348980. DOI: 10.1145/3116595.3116607. URL: https://doi.org/10.1145/3116595.3116607.

[10] Mads Møller Jensen and Florian ’Floyd’ Mueller. “Running with Technology: Where Are We Heading?” In: *Proc. 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*. Sydney, New South Wales, Australia: ACM, 2014, pp. 527–530. ISBN: 9781450306539. DOI: 10.1145/2686612.2686696. URL: https://doi.org/10.1145/2686612.2686696.

[11] Kristina Knaving, Paweł Woźniak, Morten Fjeld, et al. “Flow is Not Enough: Understanding the Needs of Advanced Amateur Runners to Design Motivation Technology”. In: *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York: ACM, 2015, pp. 2013–2022. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702542. URL: https://dl.acm.org/doi/10.1145/2702123.2702542 (visited on 01/27/2024).

[12] Lijie Yao, Anastasia Bezerianos, Romain Vuillemot, et al. “Visualization in Motion: A Research Agenda and Two Evaluations”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.10 (2022), pp. 3546–3562. DOI: 10.1109/TVCG.2022.3184993.

[13] F. Buttussi and L. Chittaro. “MOPET: A context-aware and user-adaptive wearable system for fitness training”. English. In: *Artificial Intelligence in Medicine* 42.2 (2008), pp. 153–163. ISSN: 0933-3657. DOI: 10.1016/j.artmed.2007.11.004.

[14] F. Buttussi and L. Chittaro. “Smarter phones for healthier lifestyles: An adaptive fitness game”. English. In: *IEEE Pervasive Computing* 9.4 (2010), pp. 51–57. ISSN: 1536-1268. DOI: 10.1109/MPRV.2010.52.

[15] Rodrigo de Oliveira and Nuria Oliver. "TripleBeat: enhancing exercise performance with persuasion". In: *Proc. International Conference on Human Computer Interaction with Mobile Devices and Services*. New York: ACM, 2008, pp. 255–264. ISBN: 978-1-59593-952-4. DOI: 10.1145/1409240.1409268. URL: https://dl.acm.org/doi/10.1145/1409240.1409268.

[16] Nuria Oliver and Fernando Flores-Mangas. "MPTrain: a mobile, music and physiology-based personal trainer". In: *Proc. International Conference on Human-computer interaction with mobile devices and services*. New York: ACM, 2006, pp. 21–28. ISBN: 978-1-59593-390-4. DOI: 10.1145/1152215.1152221. URL: https://dl.acm.org/doi/10.1145/1152215.1152221 (visited on 06/01/2023).

[17] Stina Nylander, Mattias Jacobsson, and Jakob Tholander. "Runright: real-time visual and audio feedback on running". In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. Toronto, Ontario, Canada: ACM, 2014, pp. 583–586. ISBN: 9781450324748. DOI: 10.1145/2559206.2574806. URL: https://doi.org/10.1145/2559206.2574806.

[18] Alexander Schiewe, Andrey Krekhov, Frederic Kerber, et al. "A study on real-time visualizations during sports activities on smartwatches". In: *Proc. International Conference on Mobile and Ubiquitous Multimedia*. 2020, pp. 18–31.

[19] Tian Hao, Guoliang Xing, and Gang Zhou. "RunBuddy: A Smartphone System for Running Rhythm Monitoring". In: *Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Osaka, Japan: ACM, 2015, pp. 133–144. ISBN: 9781450335744. DOI: 10.1145/2750858.2804293. URL: https://doi.org/10.1145/2750858.2804293.

[20] Alaul Islam, Anastasia Bezerianos, Bongshin Lee, et al. "Visualizing Information on Watch Faces: A Survey with Smartwatch Users". In: *2020 IEEE Visualization Conference (VIS)*. 2020, pp. 156–160. DOI: 10.1109/VIS47514.2020.00038.

[21] Masaaki Makikawa, Tadao Isaka, Takeo Iida, et al. "Jogging Support System with Portable Monitoring Device and Health Manage Software". In: *Studies in health technology and informatics* 107 Pt 2 (2004), pp. 1260–3.

[22] Fabrizio Mulas, Salvatore Carta, Paolo Pilloni, et al. "Everywhere Run: A Virtual Personal Trainer for Supporting People in Their Running Activity". In: *Proc. International Conference on Advances in Computer Entertainment Tech-*

nology. Lisbon, Portugal: ACM, 2011. ISBN: 9781450308274. DOI: 10.1145/2071423.2071510. URL: https://doi.org/10.1145/2071423.2071510.

[23] B. Smyth and P. Cunningham. "An Analysis of Case Representations for Marathon Race Prediction and Planning". English. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11156 LNAI (2018), pp. 369–384. ISSN: 0302-9743. DOI: 10.1007/978-3-030-01081-2_25.

[24] Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, et al. "Exploring the design space of glanceable feedback for physical activity trackers". In: *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. 2016, pp. 144–155.

[25] Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, et al. "Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 630–640. DOI: 10.1109/TVCG.2018.2865142.

[26] Fereshteh Amini, Khalad Hasan, Andrea Bunt, et al. "Data representations for in-situ exploration of health and fitness data". In: *Proc. EAI International Conference on Pervasive Computing Technologies for Healthcare*. Barcelona, Spain: ACM, 2017, pp. 163–172. ISBN: 9781450363631. DOI: 10.1145/3154862.3154879. URL: https://doi.org/10.1145/3154862.3154879.

[27] Mark M. Cullen, Daniel Schmitt, Michael C. Granatosky, et al. "Gaze-behaviors of runners in a natural, urban running environment". eng. In: *PloS one* 15.5 (2020), e0233158–e0233158. ISSN: 1932-6203.

[28] Accessed: 2024-09-02. URL: https://en.wikipedia.org/wiki/Polar_Electro.

[29] Accessed: 2024-09-02. URL: https://en.wikipedia.org/wiki/Heart_rate_monitor#:~:text=The%20first%20wireless%20EKG%20heart,heart%20monitors%20started%20in%201983..

[30] Accessed: 2024-09-02. URL: https://en.wikipedia.org/wiki/Nike%2BiPod.

[31] Accessed: 2024-09-02. URL: https://en.wikipedia.org/wiki/Garmin_Forerunner#cite_note-3.

[32] Accessed: 2024-09-02. URL: https://www.nike.com/ca/a/running-goals.

[33] Dennis Bramble and Daniel Lieberman. “Endurance Running and the Evolution of Homo”. In: *Nature* 432 (Dec. 2004), pp. 345–352. DOI: 10.1038/nature03052.

[34] Kelley Kilpatrick, Sharon Kaasalainen, Faith Donald, et al. “The effectiveness and cost-effectiveness of clinical nurse specialists in outpatient roles: a systematic review”. In: *Journal of Evaluation in Clinical Practice* 20.6 (2014), pp. 1106–1123.

[35] Luiz Carlos Hespanhol Junior, Julian David Pillay, Willem van Mechelen, et al. “Meta-analyses of the effects of habitual running on indices of health in physically inactive adults”. In: *Sports medicine* 45 (2015), pp. 1455–1468.

[36] Alaul Islam, Lijie Yao, Anastasia Bezerianos, et al. “Reflections on Visualization in Motion for Fitness Trackers”. In: *MobileHCI 2022 Workshop on New Trends in HCI and Sports*. 2022.

[37] Colin Ware. “Information Visualization: Perception for Design: Second Edition”. In: Apr. 2004.

[38] A. Dunne and F. Raby. *Design Noir: The Secret Life of Electronic Objects*. August, 2001. ISBN: 9783764365660. URL: https://books.google.ca/books?id=_49YTKJ16l4C.

[39] Brian J Fogg, Cathy Soohoo, David R Danielson, et al. “How do users evaluate the credibility of Web sites? A study with over 2,500 participants”. In: *Proceedings of the 2003 conference on Designing for user experiences*. 2003, pp. 1–15.

[40] Sean A Munson and Sunny Consolvo. “Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity”. In: *2012 6th international conference on pervasive computing technologies for healthcare (pervasivehealth) and workshops*. IEEE. 2012, pp. 25–32.

[41] Kris N. Kirby and Daniel Gerlanc. “BootES: An R package for bootstrap confidence intervals on effect sizes”. eng. In: *Behavior research methods* 45.4 (2013), pp. 905–927. ISSN: 1554-351X.

[42] Geoff Cumming. “The New Statistics: Why and How”. In: *Psychological Science* 25.1 (2014). PMID: 24220629, pp. 7–29. DOI: 10.1177/0956797613504966.

[43] Pierre Dragicevic. "Fair Statistical Communication in HCI". In: *Modern Statistical Methods for HCI*. Ed. by Judy Robertson and Maurits Kaptein. Cham: Springer International Publishing, 2016, pp. 291–330. ISBN: 978-3-319-26633-6. DOI: 10.1007/978-3-319-26633-6_13.

[44] Matthew Brehmer, Bongshin Lee, Petra Isenberg, et al. "A Comparative Evaluation of Animation and Small Multiples for Trend Visualization on Mobile Phones". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 364–374. DOI: 10.1109/TVCG.2019.2934397.

[45] Martin Krzywinski and Naomi Altman. "Points of significance: error bars". In: *Nature Methods* 10 (2013), pp. 921–922. DOI: 10.1038/nmeth.2659.

[46] Dennis M Bramble and Daniel E Lieberman. "Endurance running and the evolution of Homo". In: *nature* 432.7015 (2004), pp. 345–352.

[47] Daniel Boullosa, Jonathan Esteve-Lanao, Arturo Casado, et al. "Factors Affecting Training and Physical Performance in Recreational Endurance Runners". en. In: *Sports* 8.3 (2020), p. 35. ISSN: 2075-4663. DOI: 10.3390/sports8030035. URL: https://www.mdpi.com/2075-4663/8/3/35.

[48] Luiz Carlos Hespanhol Junior, Julian David Pillay, Willem van Mechelen, et al. "Meta-Analyses of the Effects of Habitual Running on Indices of Health in Physically Inactive Adults". In: *Sports Med* 45.10 (2015), pp. 1455–1468. ISSN: 1179-2035. DOI: 10.1007/s40279-015-0359-y.

[49] Marcus W. Kilpatrick, Mary E. Jung, and Jonathan P. Little. "High-intensity Interval Training: A Review of Physiological and Psychological Responses". In: *ACSM's Health & Fitness Journal* 18.5 (Oct. 2014), p. 11. ISSN: 1091-5397. DOI: 10.1249/FIT.0000000000000067. URL: https://journals.lww.com/acsm-healthfitness/fulltext/2014/09000/high_intensity_interval_training__a_review_of.5.aspx (visited on 01/15/2024).

[50] Pedro C Hallal, Lars Bo Andersen, Fiona C Bull, et al. "Global physical activity levels: surveillance progress, pitfalls, and prospects". In: *The lancet* 380.9838 (2012), pp. 247–257.

Appendix A

Additional Information

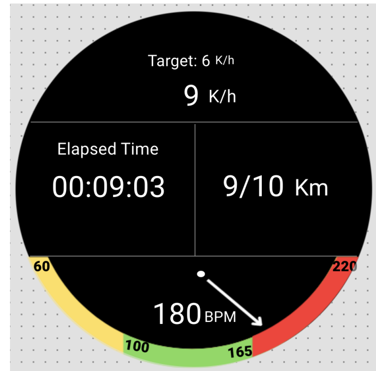
Questions:

- Is your speed above or below the Target speed? (Question Number 1,6,13)
- Yes or No, your heart rate is in the target range? (Question Number 2,7,10)
- Your Pace is below the target. How much do you need to adjust it? (Question Number 3,9)
- Yes/ No, You have finished the first quarter of your run? (Question Number 4,8,11)
- Your Heart Rate is above target, how much do you need to slow down?(Question Number 5,12)

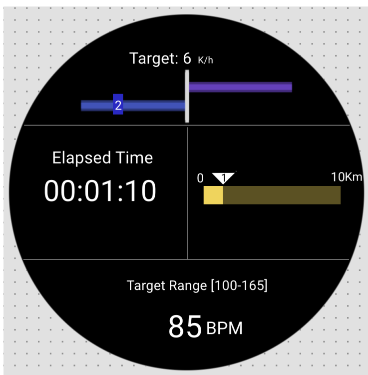
Watchfaces for each question:

(a) Question 1, WF1

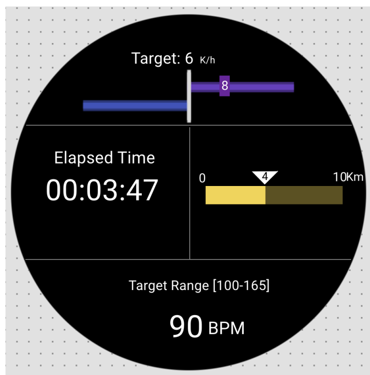
(b) Question 6, WF1



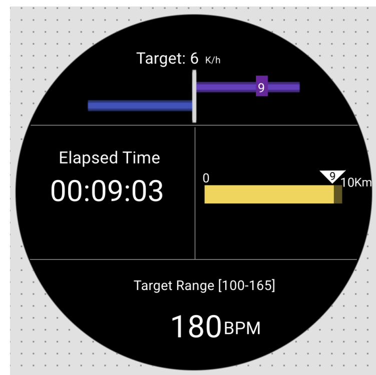
(c) Question 13, WF1



(d) Question1, WF2



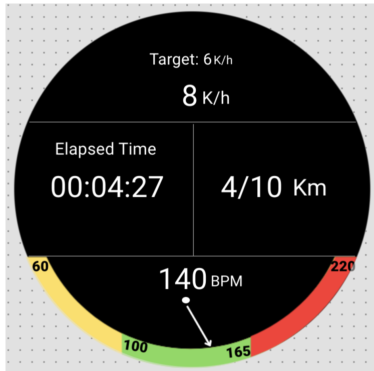
(e) Question6, WF2



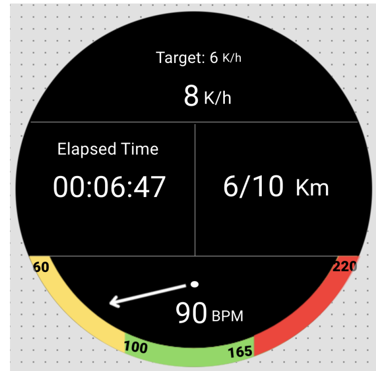
(f) Question13, WF2

Figure A.1: from left to right question 1,6, 13, task of realizing either their speed is above or below the target where top row showing the watchface number 1 and second row shows watchface number 2

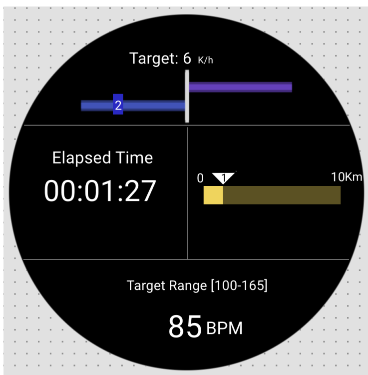
(a) Question 2, WF1



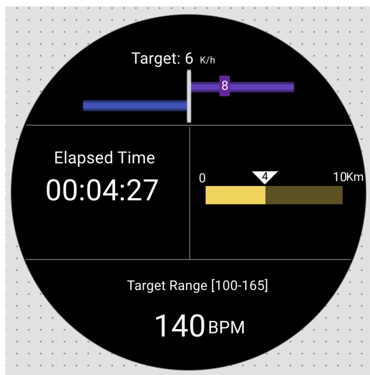
(b) Question 7, WF1



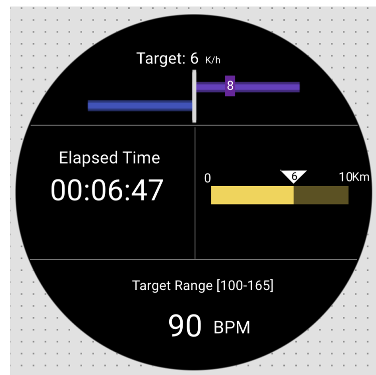
(c) Question 10, WF1



(d) Question2, WF2



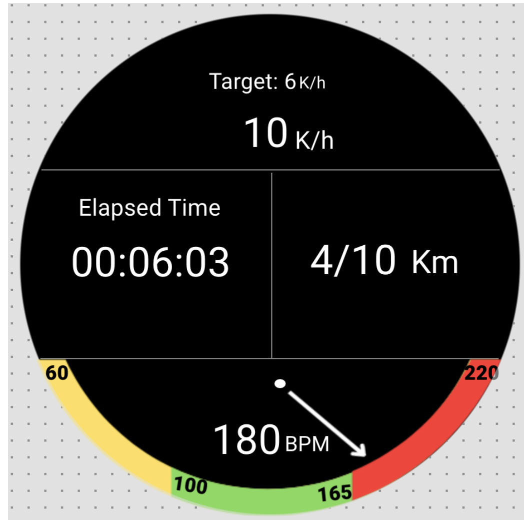
(e) Question7, WF2



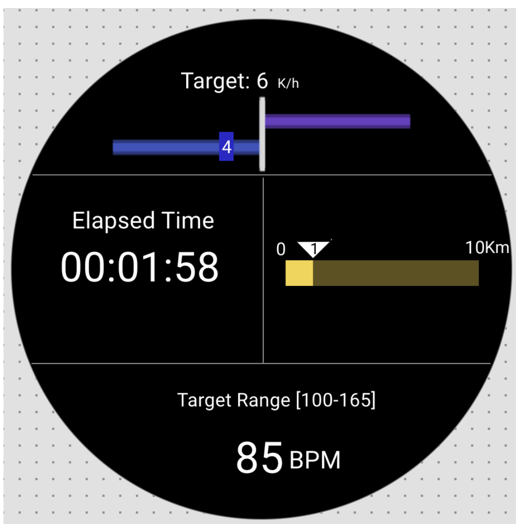
(f) Question10, WF2

Figure A.2: from left to right question 2,7, 10, task of realizing if their heart rate is in the target range, where top row showing the watchface number 1 and second row shows watchface number 2

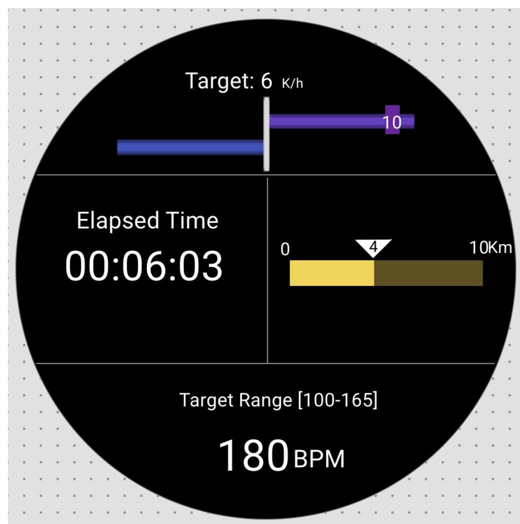
(a) Question 3, WF1



(b) Question 9, WF1



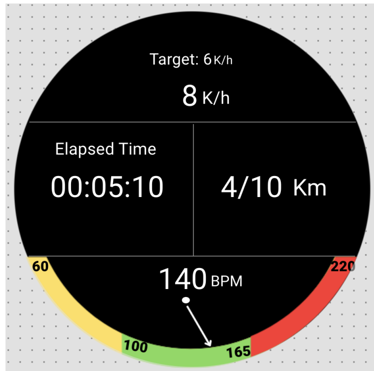
(c) Question3, WF2



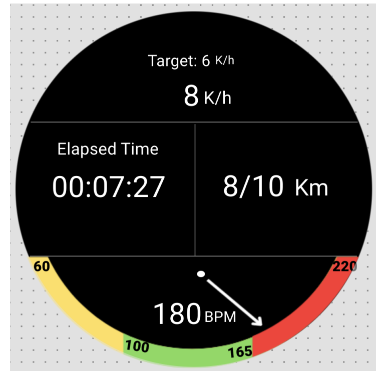
(d) Question9, WF2

Figure A.3: from left to right question 3,9 task of realizing How much the participants need to increase their speed to adjust to the target range, where top row showing the watchface number 1 and second row shows watchface number 2

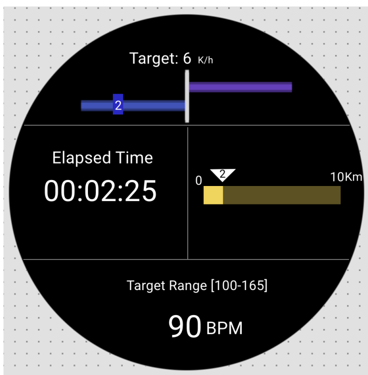
(a) Question 4, WF1



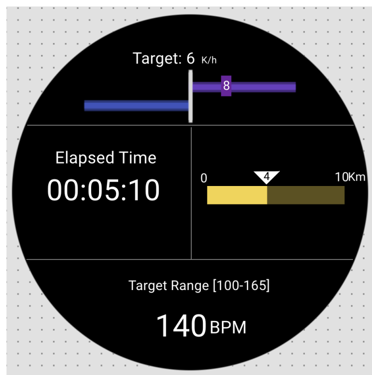
(b) Question 8, WF1



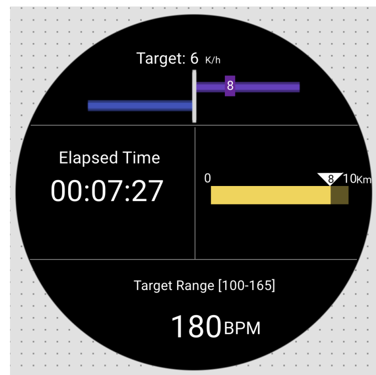
(c) Question 11, WF1



(d) Question4, WF2



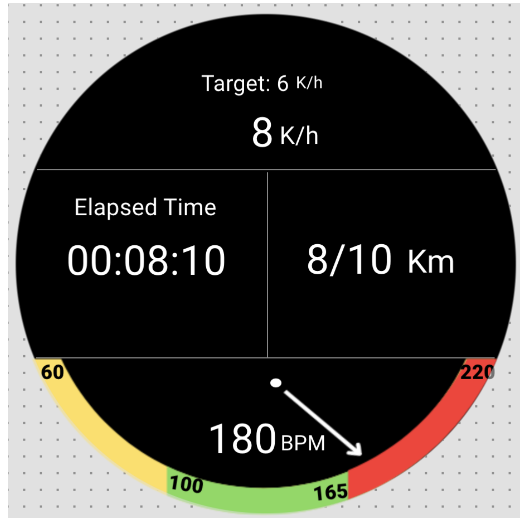
(e) Question8, WF2



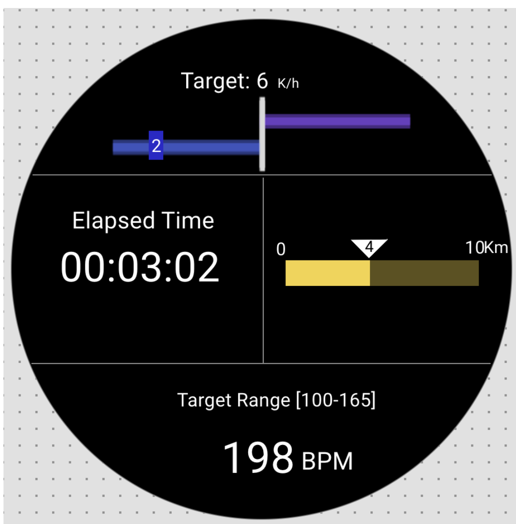
(f) Question11, WF2

Figure A.4: from left to right question 4,8, 11, task of testing participants‘ ability and the awareness to estimate distances covered during their run,, where top row showing the watchface number 1 and second row shows watchface number 2

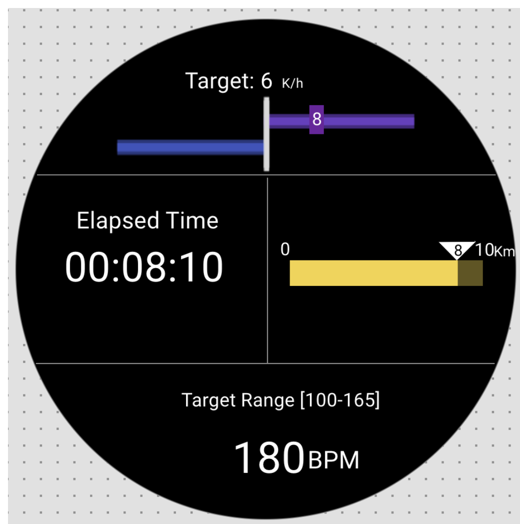
(a) Question 5, WF1



(b) Question 12, WF1



(c) Question5, WF2



(d) Question12, WF2

Figure A.5: from left to right question 5,12 task of realizing How much the participants need to slow down to adjust their heart rate to the target range, where top row showing the watchface number 1 and second row shows watchface number 2