Databiting: Enabling Lightweight, Transient, and In-Situ Exploration of Personal Fitness Data on Smartwatches

by

Bradley Rey

B.Sc., The University of Manitoba 2020

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The College of Graduate Studies

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

December 2024

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, the dissertation entitled:

DATABITING: ENABLING LIGHTWEIGHT, TRANSIENT, AND IN-SITU EXPLORATION OF PERSONAL FITNESS DATA ON SMARTWATCHES

submitted by BRADLEY REY in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Pourang Irani, Irving K. Barber Faculty of Science, UBC Okanagan
Supervisor

Karon MacLean, Computer Science, UBC Vancouver
Supervisory Committee Member

Ramon Lawrence, Irving K. Barber Faculty of Science, UBC Okanagan
Supervisory Committee Member

Megan Smith, Creative and Critical Studies, UBC Okanagan
University Examiner

Petra Isenberg, INRIA
External Examiner

Abstract

Smartwatches have become increasingly popular for health and fitness tracking. Worn on the body, and equipped with powerful sensors, smartwatches enable comprehensive and ubiquitous health data collection. Exploring this collected data on the smartwatch is, however, less comprehensive and is largely unrealized. The small screen size and many daily scenarios of use are often considered challenges which limit exploratory capability. Yet, the immediate availability of the smartwatch remains to offer potential for actionable and in-situ insights derived from the collected data. As such, this dissertation introduces *databiting*, a novel concept for lightweight and transient data exploration that supports in-situ exploration.

Before realizing *databiting*, we conducted an empirical study with 18 individuals to better understand their smartwatch data exploration needs. We captured and analyzed 205 personal health data queries to understand where, when, and why these queries were desired. This analysis combined qualitative insights with quantitative metrics, uncovering patterns that inform the design and implementation of *databiting* on the smartwatch. We further identified key dimensions such as interrogatives, data sources, aggregations, and filtering mechanisms, which provide a foundation for enhancing natural language processing capabilities.

We then examined preferences of voice assistant responses to personal health data queries, as an additional output modality to visualization. Insights were gathered on perceived quality, behavior, comprehensibility, and efficiency to a variety of answer structures, providing guidelines to support *databiting* through an effective and complementary output modality.

Using these findings, we then developed DataWatch, a novel smartwatch application which facilitates in-situ exploration of fitness data through lightweight and transient touch and speech interactions. DataWatch supports Single Value, Browse,

and Compare exploration types. Furthermore, users can explore their data during different workout phases—before, during, and after a workout—enhancing one's ability to gain insights and make informed decisions while in-situ.

Overall, this research demonstrates the feasibility and benefits of *databiting* on smartwatches, providing a foundation for future advancements in personalized data exploration on wearable devices. DataWatch exemplifies how smartwatch health applications can begin to evolve within their means to meet user expectations, provide increased actionable insights, and motivate deeper engagement with health data.

Lay Summary

This thesis focuses on enhancing how people interact with their personal health data on smartwatches. We introduce the concept of *databiting*, with the goal of allowing exploration of data in both a quick and easy manner. Through a learned understanding of peoples' smartwatch data exploration needs, we developed DataWatch. DataWatch is an Apple Watch application allowing people to question and explore their past collected fitness data before, during, and after a workout. By enabling further exploration on the smartwatch, from that of the current visualizations, DataWatch provides the potential for increased actionable insights and motivation in exploring one's data. Our findings not only demonstrate that people desire more interactive and insightful health data exploration on their smartwatches, but also validate that such exploration directly on the smartwatch is both feasible and beneficial. Finally, results guide future advancements for increasingly personalized and capable health data exploration on the smartwatch.

Preface

Publications Included in Thesis

Sections of this thesis have been published in conference proceedings or journals. We note, both the ACM and IEEE (publishers for which the following work has been published to) provide confirmation that published material can be used in a dissertation[1][2]. The following is a list of publications in which portions of this thesis work appeared, organized by chapter. For each of the publications listed below, I was a either a lead or co-lead for the identification and design of the research, building of study methodologies and research instruments, conducting of study participation with participants, analysis of data, and preparation of manuscripts submitted. Moreover, each contributing publication was the result of collaborations, both local and abroad. For each, I further highlight the nature of any collaborative work that took place below.

Portions of Chapter 3

Bradley Rey, Bongshin Lee, Eun Kyoung Choe and Pourang Irani. 2024. Databiting: Lightweight, Transient, and Insight Rich Exploration of Personal Data, in IEEE Computer Graphics and Applications, vol. 44, no. 2, pp. 65-72, March-April 2024. DOI: `https://doi.org/10.1109/MCG.2024.3353888`

The article was a collaborative effort among myself, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. The collaboration took shape through joint meetings focused on concept creation and development, as well as later editorial work.

[1] `https://authors.acm.org/author-resources/author-rights`

[2] `https://www.ieee.org/content/dam/ieee-org/ieee/web/org/pubs/permissions_faq.pdf`

Portions of Chapter 4

Bradley Rey, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. 2023. Investigating In-Situ Personal Health Data Queries on Smartwatches. Proceedings of the ACM Interactive Mobile Wearable Ubiquitous Technologies, 6, 4, Article 179, December 2022, 19 pages. DOI: `https://doi.org/10.1145/3569481`

The article was a collaborative effort among myself, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. The collaboration took shape through combined efforts surrounding ideation, study design and editorial work. Furthermore, Charles-Olivier Dufresne-Camaro supported efforts in the data analysis/coding process highlighted within the chapter.

Portions of Chapter 5

Bradley Rey, Charles-Olivier Dufresne-Camaro, and Pourang Irani. 2023. Towards Efficient Interaction for Personal Health Data Queries on Smartwatches. In Proceedings of the 25th International Conference on Mobile Human-Computer Interaction, Article 18, 1–7. `https://doi.org/10.1145/3565066.3608700`

The article was a collaborative effort among myself, Charles-Olivier Dufresne-Camaro, and Pourang Irani. Charles-Olivier Dufresne-Camaro supported efforts in the data analysis/coding process highlighted within the chapter and provided editorial support.

Portions of Chapter 6

Bradley Rey, Yumiko Sakamoto, Jaisie Sin and Pourang Irani. 2024. Understanding User Preferences of Voice Assistant Answer Structures for Personal Health Data Queries. In Proceedings of the 25th International Conference on Conversational User Interfaces, Article 18, 1–7. `https://doi.org/10.1145/3565066.3608700`

The article was a collaborative effort among myself, Yumiko Sakamoto, Jaisie Sin, and Pourang Irani. Yumiko Sakamoto collaborated on study design, supported data analysis procedures, and editorial work. Jaisie Sin collaborated on study design and editorial work.

Thesis Publications in Progress

Sections of this thesis contain work of a publication in progress, and is to be submitted for peer-review at a later date.

Portions of Chapter 7

Bradley Rey et al. DataWatch: Fostering Exploration of Physical Activity Data on the Smartwatch Leveraging Speech and Touch Interaction for In-Situ Insight

All work in this chapter is solely my own. However, discussions surrounding the work put forward in Chapter 3 and Chapter 4 influenced aspects of the DataWatch application and study methodology. As such, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani have provided indirect collaboration.

Other Publications

Below is a list of additional works that were published during my time as a Ph.D. student, for which I contributed varying roles:

1. Shariff AM Faleel, Yishuo Liu, Roya A Cody, **Bradley Rey**, Linghao Du, Jiangyue Yu, Da-Yuan Huang, Pourang Irani, and Wei Li. 2023. T-Force: Exploring the Use of Typing Force for Three State Virtual Keyboards. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 723, 1–15. `https://doi.org/10.1145/3544548.3580915`

2. **Bradley Rey**, Kening Zhu, Simon Tangi Perrault, Sandra Bardot, Ali Neshati, and Pourang Irani. 2022. Understanding and Adapting Bezel-to-Bezel Interactions for Circular Smartwatches in Mobile and Encumbered Scenarios. Proc. ACM Human Computer Interaction 6, MHCI, Article 201 (September 2022), 28 pages. `https://doi.org/10.1145/3546736`

3. Sandra Bardot, **Bradley Rey**, Lucas Audette, Kevin Fan, Da-Yuan Huang, Jun Li, Wei Li, and Pourang Irani. 2022. One Ring to Rule Them All: An

Empirical Understanding of Day-to-Day Smartring Usage Through In-Situ Diary Study. Proc. ACM Interactive Mobile Wearable Ubiquitous Technologies 6, 3, Article 100 (September 2022), 20 pages. https://doi.org/10.1145/3550315

4. Anuradha Herath, **Bradley Rey**, Sandra Bardot, Sawyer Rempel, Lucas Audette, Huizhe Zheng, Jun Li, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2022. Expanding Touch Interaction Capabilities for Smartrings: An Exploration of Continual Slide and Microroll Gestures. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 292, 1–7. https://doi.org/10.1145/3491101.3519714

5. Fouad Alallah, Shariff Faleel, Yumiko Sakamoto, **Bradley Rey**, and Pourang Irani. 2022. SSCA: Situated Space-time Cube Analytics. EuroVis 2022 - Short Papers. https://doi.org/10.2312/evs.20221088

6. Ali Neshati, Fouad Alallah, **Bradley Rey**, Yumiko Sakamoto, Marcos Serrano, and Pourang Irani. 2021. SF-LG: Space-Filling Line Graphs for Visualizing Interrelated Time-series Data on Smartwatches. In Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21). Association for Computing Machinery, New York, NY, USA, Article 5, 1–13. https://doi.org/10.1145/3447526.3472040

7. Sandra Bardot, Surya Rawat, Duy Thai Nguyen, Sawyer Rempel, Huizhe Zheng, **Bradley Rey**, Jun Li, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2021. ARO: Exploring the Design of Smart-Ring Interactions for Encumbered Hands. In Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21). Association for Computing Machinery, New York, NY, USA, Article 12, 1–11. https://doi.org/10.1145/3447526.3472037

8. Shariff AM Faleel*, Bibhushan Raj Joshi* and **Bradley Rey*** (*equal author contribution). Writely: Force Feedback for Non-Dominant Hand Writing

Training. 2021. IEEE World Haptics Conference (WHC), Montreal, QC, Canada, 2021, pp. 340-340. `https://doi.org/10.1109/WHC49131.2021.9517209`

9. Ali Neshati, **Bradley Rey**, Ahmed Shariff Mohommed Faleel, Sandra Bardot, Celine Latulipe, and Pourang Irani. 2021. BezelGlide: Interacting with Graphs on Smartwatches with Minimal Screen Occlusion. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 501, 1–13. `https://doi.org/10.1145/3411764.3445201`

10. Sandra Bardot, Sawyer Rempel, **Bradley Rey**, Ali Neshati, Yumiko Sakamoto, Carlo Menon, and Pourang Irani. 2020. Eyes-free graph legibility: using skin-dragging to provide a tactile graph visualization on the arm. In Proceedings of the 11th Augmented Human International Conference (AH '20). Association for Computing Machinery, New York, NY, USA, Article 1, 1–8. `https://doi.org/10.1145/3396339.3396344`

Use of Generative AI

Throughout, the use of generative AI (e.g., ChatGPT, Gemini, etc.) has been limited to only that of editorial support. Specifically, ChatGPT has been used in various places throughout to critique my writing in a holistic manner, rather than providing individual grammatical edits. Prompts such as "How well do you feel these paragraphs work together in providing [the desired argument]?" and "What are some areas of improvement that could be incorporated into the shared text?" were used. No other generative AI models/tools were used. All figures are original creations, not created through the use of generative AI.

Ethics Approval

Research undertaken within this thesis spanned both the University of Manitoba, where I began my degree, and the University of British Columbia Okanagan, where I completed my degree. Within both institutions, ethics approval was obtained before conducting studies with participants. Specifically the following certificate numbers were obtained:

– **Chapter 4** - University of Manitoba, Research Ethics Board 1, Certificate Number HE2021-0070

– **Chapter 6** - University of British Columbia Okanagan, Behavioural Research Ethics Board, Certificate Number H23-00807

– **Chapter 7** - University of British Columbia Okanagan, Behavioural Research Ethics Board, Certificate Number H23-00805

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my gratitude and appreciation to my supervisor, Dr. Pourang Irani. I will be forever grateful for your belief in me, the opportunities given, guidance, conversations (both academic and non-academic), encouragement, and support provided throughout this journey. Your initial decision to hire me while an undergraduate student, and everything that has come since, has forever shaped my and my family's future. This is something I will never be able to repay, and can only hope to carry forward. Thank you.

I would like to thank my internal committee members, Dr. Karon MacLean and Dr. Ramon Lawrence, for their time and effort in meeting with us throughout and providing insights and feedback. Transferring from the University of Manitoba to the University of British Columbia affected the normal program deadlines, and both Dr. MacLean and Dr. Lawrence accepted the internal committee role without hesitation. I also want to thank my external committee members, Dr. Petra Isenberg and Dr. Megan Smith for their time and effort in evaluating this thesis.

A special thank you to all my external collaborators and lab-members (of both the HCI Lab at the University of Manitoba and OVI Lab at the University of British Columbia). In particular, I would like to thank Dr. Yumiko Sakamoto, Dr. Bong-shin Lee, Dr. Eun Kyoung Choe, Dr. Sandra Bardot, Charles-Olivier Dufresne-Camaro, Anuradha Herath, Shariff Faleel, Dr. Maxime Daniel, Dr. Fouad Alallah, Samar Sallam, Preeti Vyas, Dr. Ali Neshati, Dr. Jaisie Sin, Dr. Simon Perrault, and Dr. Kening Zhu. To my collaborators, working with you has taught me everything I know about conducting research and has shown me the level of excellence required to succeed. To my lab-members, I am grateful for the countless hours spent discussing research problems, brainstorming solutions, providing research support, and the fun times along they way. Meeting all of you has shaped this journey for the better, and I hope to continue our relationships beyond this degree.

To my family, thank you for instilling in me the values needed to complete such a degree. Without your nurturing and encouragement this would not have been possible. To my partner, Danica, *we* did it. Thank you for believing in me, supporting and encouraging this degree, moving across the country with me, patiently listening while I ramble about ideas (even when you barely knew what I was talking about), reminding me to celebrate the highs, and for comforting me during the lows. This has been a shared journey to say the least, and I am now more than ever excited for our future.

Finally, I would like to extend my gratitude to all those who have directly or indirectly contributed to my research and are not named above. Particularly, I want to thank the participants who generously gave their time and effort, and the funding agencies that made this work possible. I hope my work, as well as future motivated work, can carry forward a positive impact within many lives.

Chapter 1

Introduction

> *"If the smartwatch can't eventually get smarter and more useful, it risks becoming a footnote."* - Walt Mossberg, The Verge

The potential for smartwatches to become truly influential personal health tools lies in their ever-growing capabilities. Built with a wide range of sensors, processing power, network connectivity, and various interaction modalities (i.e., visual, auditory, and haptic output as well as touch, speech, and gesture input), smartwatches are well-equipped to allow for collection and monitoring of personal health data in real-time. So much so, the smartwatch's current ability to ubiquitously collect fitness and other health metrics is a major contributing factor to its widespread adoption [49, 52].

Data collection is just the first step. It is only once this data is explored and analyzed that insights can be drawn and meaningful actions taken. The smartwatch, worn on the wrist, is seamlessly integrated into daily activities such as sleeping, jogging, working, or simply while on-the-go. This level of integration makes the smartwatch an ideal tool for continuous data collection. Moreover, and importantly, this integration offers capability for convenient and accessible exploration [148, 192]. In fact, individuals are becoming increasingly dependent on the smartwatch, even more so than their other devices, for analyzing their collected data throughout their daily activities [34]. This in-situ exploration on the smartwatch is crucial, as it empowers individuals to quickly and easily gain actionable insights from their health data, enabling them to adjust their behavior in real-time, to optimize their activities, and achieve their goals [40, 62].

The smartwatch's full potential as a personal health tool remains largely unrealized, however, due to inadequate capabilities for data exploration [37, 70, 131]. The largest barrier for exploration on the smartwatch is often seen as its small screen size, limiting typical touch interactions and visual output [136, 155]. As such, guidelines for exploring data specifically on the smartwatch often focus on efficiently designing micro and glanceable visualizations [3, 13, 15, 26, 68, 85, 135, 136] (i.e., small screen visualizations that can be viewed and understood within five seconds [13, 68]). While beneficial in their own right, these visualizations still only allow for limited and discrete exploratory capabilities. This in turn, remains to hinder utility and access to a broader range of desired insights [37, 70, 131].

By allowing users to further explore their data directly on the smartwatch, individuals can quickly and easily gain actionable insights beyond what is currently possible. To illustrate this necessity we provide a storied example, illustrated in Figure 1.1: Consider a hiker, Sam, who checks their smartwatch during a hike. Current systems only display a handful of metrics, such as the current pace. Sam, while viewing their pace, is uncertain if they are slower than usual. This lack of information richness and overall usefulness of the interface ultimately hampers full engagement and a larger range of benefits that could be available to smartwatch users [70, 145]. Through enhanced exploration, Sam could tap on the pace number, and ask *"How does today's pace compare to my last six hikes?"*. A graph appears, revealing that their pace is indeed slower than their average. This quick in-situ exploration on the smartwatch provides a deeper and relational understanding, enabling Sam to immediately adjust their pace towards achieving their fitness goals.

1.1 Research Challenges

While the exploration in the above storied example may seem simple, there are a handful of underlying research challenges that need to be solved before enabling such exploration directly on the smartwatch.

To further highlight the complexity of the problem, and to better situate this work, we take a moment to demonstrate capability of such querying of data on currently marketed devices. To our knowledge, certain Apple devices through interaction with their voice assistant Siri, are the only commercial devices which can

Figure 1.1: To the left of the grey dashed line is one example of the current state of exploration for mobile data visualizations. To the right, databiting is enabled for physical activity related information. This has the potential to provide greater influence and insight of the current activity taking place. Notably, more in-depth long-term data exploration is left for a later time.

answer some personal health data questions; other commercial devices, and their operating systems, often only recognize key words and simply provide a prompt to open a respective health application. As such, we asked Siri a range of personal health data questions[3]. These questions and responses can be seen in Table 1.1. Notably, aside from questions pertaining to current and recently captured metrics, responses are unsuccessful or unsupported. Furthermore, when responses are unsuccessful, they often provide outright incorrect information. Thus, it is clear that the ability to accurately and efficiently process and respond to a diverse range of health data queries on the smartwatch necessitates further efforts.

[3]This is a new feature (as of December 2023) that only works with iPhones and iPads running iOS and iPadOS greater than 17.2 and Apple Watch Series 9 and Ultra 2 running watchOS greater than 10.2. Queries were asked on an iPhone 14 Pro running iOS 17.4.

Table 1.1: Recorded personal health data question and answers using Siri on an iPhone 14 Pro running iOS 17.3. We demonstrate this capability on an iPhone, rather than an Apple Watch, as functionality is further limited on the Apple Watch. The question asked, the voice assistant's response, supplementary information displayed on screen, and the success of the interaction are provided.

Question	**Vocal Response**	**Supplementary Information**	**Successful**
What is my current step count?	8739 steps	8739 Steps Today	Yes
How far have I walked this week?	7.2 km	7.2 km 2.4 km Daily Average Walking + Running Distance May 26 - June 1, 2024	Yes
What is my move ring at?	You've burned 210 out of your 500 calorie goal	Move Ring Today 10:02 am	Yes
What's my average distance walked in my last seven walks?	You've walked or run 2 km today	2 km Walking + Running Distance Today	No (Incorrect Information)
How much time did I spend walking in September 2023?	124.8 km	124.8 km 4.1 km Daily Average Walking + Running Distance September 2023	No (Incorrect Information)
What was my average running pace in March?	I can't help you with that		No (Unable to Respond)
How far have I walked since January 1st?	3.7 km	3.7 km Walking + Running Distance January 1, 2024	No (Incorrect Information)

The provided examples in Table 6.3 highlight a limitation in current solutions, and are underscored by a variety of research challenges that remain to be explored. First, we must understand *what* health data queries people feel are important. Furthermore, we must recognize *when*, *where*, and *why* these queries arise to afford context and understanding to the handling of the query. Then, specific understanding is needed of the natural language used which dictates *how* we can process such queries. Moreover, exploring *how* the smartwatch responds to these queries is equally as important. Only through addressing these challenges can we begin to realize the potential for broader, lightweight, and transient health data interaction on the smartwatch. Below, we discuss these research challenges in further detail and, throughout, the structure of the remainder of the thesis to follow. This outline can also be viewed graphically in Figure 1.2.

1.1.1 Conceptualization

Can we conceptualize data exploration that represents a lightweight and transient bridge between short- and long-form data exploration?

The first focus of this thesis is on the conceptualization of our targeted exploration capability of the smartwatch. Traditional research in data exploration often concentrates on either short, glanceable, or longer, more detailed, exploration. In Chapter 3, we position our work through a classification of what we term *databiting*—an intermediate level of data exploration, complementing short- and long-form exploration, that remains to provide richer insight while being transient and lightweight. We conceptualize *databiting* through metaphor, highlight the envisioned benefits of such exploration, and discuss associated research challenges. Chapter 3 approaches this conceptualization through a broader perspective of data exploration, hinting at personal health data exploration on wearable devices where appropriate.

1.1.2 What

What personal health data queries are of interest to people for exploration on their smartwatches throughout their daily lives?

In order to move forward in enabling *databiting* on the smartwatch, Chapter 4 looks to first understand *what* specific personal health data queries people have for exploration on their smartwatch, throughout their daily lives. Given the wide range of data available and the diverse contexts in which smartwatches are used, people likely have varied and specific desired exploration needs. Currently, our knowledge in this area is limited. Research is comprised of lab-based or survey studies [3, 154], which do not account for in-the-wild experiences, suffering from potential recall bias [75]. Furthermore, most prior work does not share concretely reported queries, rather they highlight broader insights into exploration practices. To address the *what* challenge, Chapter 4 details a week-long in-the-wild study designed to elicit queries from participants. We utilized a custom-built smartwatch application which enabled participants to quickly and easily record a desired personal health data query and other contextual information. Throughout the study, we collected 205 personal health data queries which have been made publicly available to support and facilitate future research in this area.

1.1.3 When, Where, Why

How does being in-situ influence the desired personal health data queries of smartwatch users?

Our next focus is on understanding the context of the personal health data queries by examining *when*, *where*, and *why* these queries are desired. Understanding *when* and *where* a query arises can help to give it context. Traditional, desktop, data exploration often happens in one static setting. In contrast, smartwatch data exploration, which can happen anytime and anywhere, can occur under varied contexts and while in-situ. This in turn, may influence both the questions posed and even the answers provided. Furthermore, the *why* challenge looks at increasing our knowledge of the broader exploratory insights that are desired. Further to Chapter 4, we analyze additional contextual information captured during our study to explore these dimensions. We find six insight categories to which our collected queries exist within. Compared with previous work, we highlight one entirely new Preemptive and Proactive insight category, as well as expanded Current

Status/Value and Contextual insight categories.

1.1.4 How

How can we expand the expressivity of data exploration on the smartwatch, through both multi-modal input and output modalities, within the confines of the smartwatch's capabilities and data exploration needs?

With a concrete understanding of *what* personal health data queries are desired, we then focus on *how* to allow for such exploration. Importantly exploration must remain lightweight and transient to support *databiting* on the smartwatch. Given the limited screen size and input capabilities of smartwatches, designing an intuitive and efficient user interface is key. As such, this research challenge focuses on *how* to support both both input and output modalities.

Through analysis of the personal health data queries collected, how can we process the components within a query to efficiently handle naturally spoken queries on the smartwatch?

Without the screen real estate for menus, toggles, sliders, etc., touch interactions can be difficult and limited on a smartwatch [61]. Speech, on the other hand, allows for the conveyance of more complex queries, a low barrier in expressing intent, flexibility in phrasing and querying [8, 42, 182, 184], and does not require as much screen real estate. Through the use of speech and touch, benefits of both can be utilized to enhance smartwatch data exploration. As touch is an already established input modality, we focus our work in Chapter 5 on examining how the previously collected queries are phrased. We analyze the components of speech within the queries, including interrogatives, data sources, aggregations, and filtering mechanisms. By understanding these elements, we aim to later develop a part-of-speech tagging framework for processing and responding to such queries directly on the smartwatch.

How should voice assistant answers be structured to support auditory responses to personal health data queries?

Output modalities are equally important towards enabling *databiting* on the smartwatch. While research exists on micro and glanceable smartwatch visualizations [3, 13, 15, 26, 68, 85, 86, 135, 136], Chapter 6 focuses on exploring auditory responses to personal health data questions. Previous work with voice assistants for question and answering focuses only on general knowledge queries (e.g., ”What is the weather today?”) [2, 71, 127]. By exploring voiced responses for personal health data queries, we aim to offer an output modality that leverages the strengths of auditory feedback, which can be beneficial in times where the visual system is overloaded [20, 147], as often seen on-the-go.

Through the use of a custom-built pseudo voice assistant, implemented using Google’s Web Speech API, Chapter 6 provides an online survey study where participants directly interacted with the developed voice assistant. Participants would ask a guided personal health data question and receive a voiced response in one of three defined answer structures (minimal, keyword, and full sentence). Responses were ranked by participants for quality, behaviour, comprehensibility, and efficiency. Notably, full sentence answers were better perceived across all ranked metrics. The insights gained from this chapter can be used to further enhance the *databiting* experience on smartwatches, ensuring that voiced responses are effective and appropriate.

1.1.5 Databiting Validation

Finally in Chapter 7, we integrate the knowledge gained in the preceding chapters to build a smartwatch workout application, DataWatch, that supports not only workout tracking, but also enables *databiting*. Participants installed DataWatch on their own smartwatch, using it throughout the week-long study within their day-to-day lives. We then validate and observe the application’s use, and any subsequent *databiting* that takes place. Through the findings of our study, both quantitative and qualitative, we showcase the concept of databiting as a valid means for data exploration on the smartwatch, pushing the boundaries of how the smartwatch is perceived for data exploration. This final step brings together all aspects of the research conducted throughout this thesis to demonstrate and validate the practical feasibility and benefits of *databiting* on smartwatches for lightweight and transient personal health data exploration anytime and anywhere.

1.2 Research Scope

The scope of this thesis follows a funnelled approach. In Chapter 3, we introduce the concept of *databiting* from a broad perspective. We conceptualize lightweight and transient exploration of any data on any device, while also highlighting benefits and challenges present. As the thesis progresses, the focus funnels towards a more specific context: personal health data on smartwatches. This focus is motivated by the immediate impact that can be realized through advancing capabilities of an already mature and widely adopted tool like the smartwatch. Furthermore, the smartwatch is a device that may have the most to benefit from a lightweight and transient form of data exploration, due to its immediate access to a range of personal data being collected, limited screen real estate inhibiting lengthy exploration, and typical and accessible on-the-go usage.

While Chapter 4, Chapter 5, and Chapter 6 focus on exploration of a wide range of personal health data, our final chapter, Chapter 7, specifically focuses on exploration of personal fitness/workout data on the smartwatch. This narrowed focus is justified by the immediate availability and practicality of fitness data on the smartwatch compared to any and all types of personal health data (e.g., women's health, mood, or nutritional information). Moreover, our elicitation study in Chapter 4 revealed that personal fitness and workout data were the most queried by participants, highlighting the high level interest in exploring such data.

By starting broad and funnelling inwards, we ensure that knowledge gained can transcend the immediate context of this thesis and as such apply to various types of data and devices. Importantly, however, by concluding with a focus on exploration of personal fitness data on the smartwatch, we create opportunity for our work to provide immediate impact.

1.3 Thesis Contributions

This thesis contributes the following high-level contributions, while lower-level, chapter-specific, contributions are further discussed within each chapter and where applicable:

C1: An introduction of the term *databiting*, conceptualized as lightweight and transient data exploration, that bridges the gap between quick, glanceable data visualizations and more extensive, detailed data analysis. Further to providing a conceptualization of *databiting*, we also discuss potential benefits and highlight future, and necessary, research directions. This foundation sets the stage for developing more intuitive and effective ways for people to engage with their personal data, both on smartwatches and other mobile platforms.

C2: An elicited, and now publicly available, dataset of 205 personal health data queries desired for exploration on a smartwatch throughout daily life. This dataset, coupled with a thorough qualitative and quantitative analysis, offers valuable insights into what personal health data query needs people have, including when, where and why these queries arise.

C3: Interaction requirements for lightweight and transient exploration of personal health data on smartwatches. We identify key dimensions of our captured queries, such as interrogatives, data sources, aggregations, and filtering mechanisms, which are crucial for enhancing input methods, particularly natural language processing capabilities. Additionally, we gather insights on user preferences for different output structures from voice assistants, focusing on perceived quality, behavior, comprehensibility, and efficiency. These findings provide essential guidelines for optimizing both input and output interactions, ensuring that data exploration on smartwatches can be supported in being lightweight and transient.

C4: Development and validation of DataWatch, an Apple Watch application that demonstrates the practical implementation of *databiting*. By integrating multimodal interactions, specifically touch and speech, DataWatch enhances in-situ exploration of past workout data, demonstrating the feasibility and user benefits of providing lightweight and transient data exploration capabilities directly on the smartwatch.

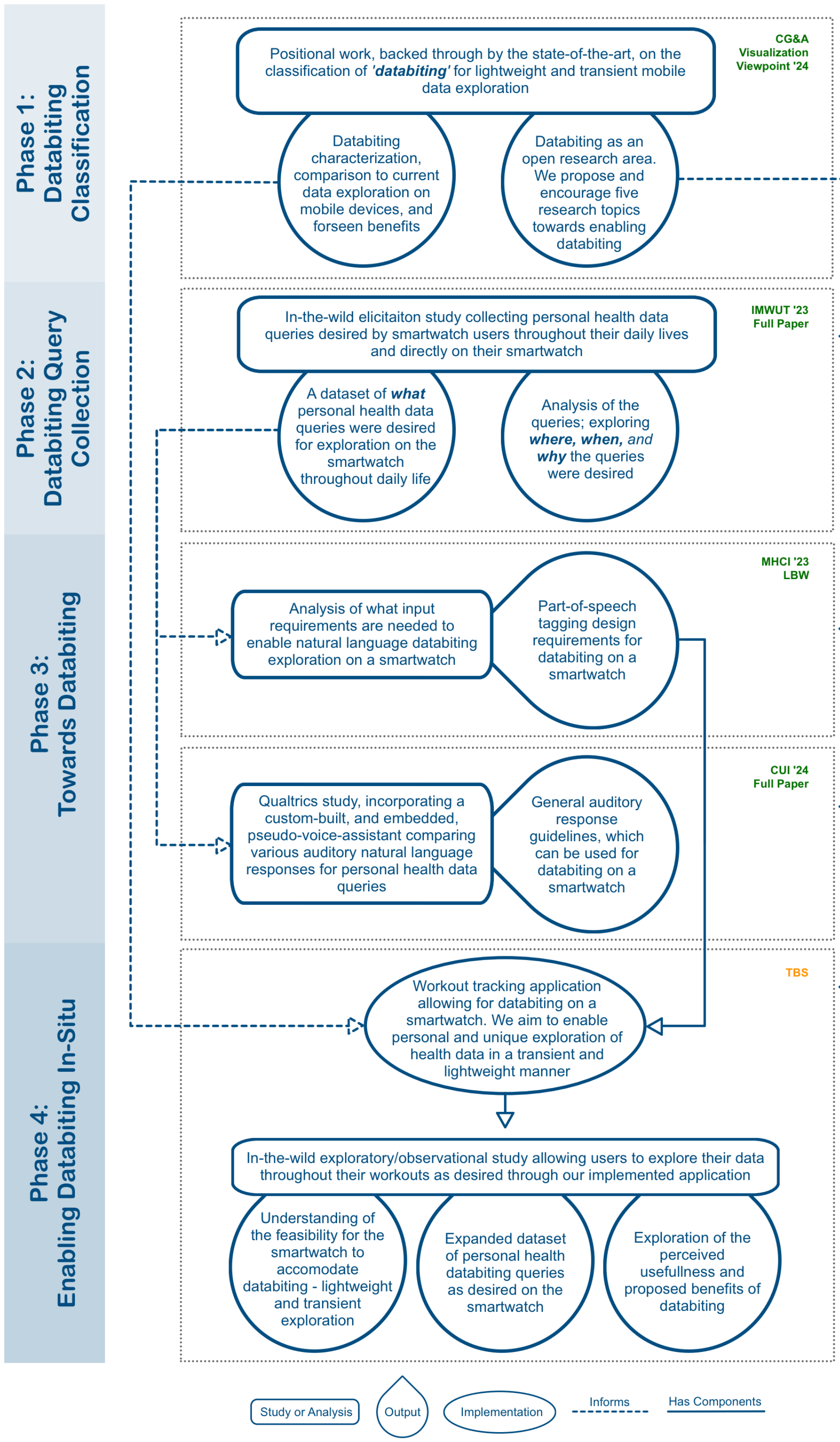


Figure 1.2: Outline of the components of this thesis, and how each interacts.

Chapter 2

Related Work

Throughout this chapter, we highlight research pertinent to the broader topics within this thesis. These include personal informatics, wearable computing, general smartwatch usage, and smartwatch data exploration. In each chapter to follow, chapter-specific related work is provided to better situate and justify topics and methods explored throughout.

2.1 Personal Informatics

Personal informatics (PI), an interdisciplinary field encompassing human-computer interaction and visualization research, emphasizes the collection, comprehension, and utilization of personal data [59, 107]. The core principle of PI is that of utilizing data to gain self-knowledge and enhance overall well-being. Lupton [118] describes this as *"practices in which people knowingly and purposely collect information about themselves, which they then review and consider applying in their lives.*"

The Quantified Self Exploring personal health data to gather insight into one's personal health journey has long been a goal of "Quantified Selfers" (QS) [118]. Thanks to the increasing capability and prevalence of mobile and wearable devices, a broader audience is now experiencing PI. These devices, equipped with advanced sensors for data collection, empower individuals to track a wide range of health metrics, from physical activity and sleep patterns to heart rate and stress levels. As a result of the wider spread adoption of wearables, PI is transitioning from a niche, QS, interest to now mainstream practice.

While collecting personal data is essential, true value emerges when individuals can explore and interact with their data presented in meaningful ways. Quantified Selfers conduct exploration for various reasons, such as maximizing performance [28, 197], reflecting on overall health [10, 28, 36, 40, 107], reinforcing prior knowledge and understanding [36, 58], viewing trends [28, 58], satisfying simple curiosity [28, 197], and comparing data [3, 28–30].

These reasons for exploration afford explicit queries which can be further categorized [3, 28, 29, 57, 98, 107]. These categories include questions about current status (e.g., "What is my current step count?"), historical data (e.g., "What was my average daily step count last month?"), goals (e.g., "How many more calories do I need to burn to reach my daily goal?"), discrepancies (e.g., "Did I exercise as much this week as I expected to?"), context (e.g., "How has the time of my walk affected my pace?"), and factors (e.g., "Has losing weight and getting more sleep been helping my mood?").

Modelling Personal Informatics Research has modeled the lived personal informatics journeys that people take when exploring their collected data for the variety of reasons above. Li et al. [107] first proposed a model to help understand how people use personal informatics tools. This model comprises of five stages, which include *preparation, collection, integration, reflection,* and *action*. The main outcome of this model is actionable insight that allows a person to create positive behaviour change from their reflection of the collected data. This model has been iterated upon many times. In fact, one year later, Li et al. [108] further divided the reflection stage into maintenance and discovery stages to better highlight the types of reflection that can be done.

Other researchers have further iterated upon this model. As the above model only highlights reflection as a separate and subsequent action to collection, Choe et al. [28] discuss how reflection is not always a separate stage, and often occurs in tandem with the data collection itself; this is especially true when using modern fitness trackers and smartwatches. Finally, Epstein et al. [59] further added *deciding* to track and *selection* of a tracking tool to the beginning of Li et al.'s model, while also adding *lapsing* and *resuming* to the end. This highlights the increasingly *lived* experiences of people, by suggesting there exists a repeated and looping nature

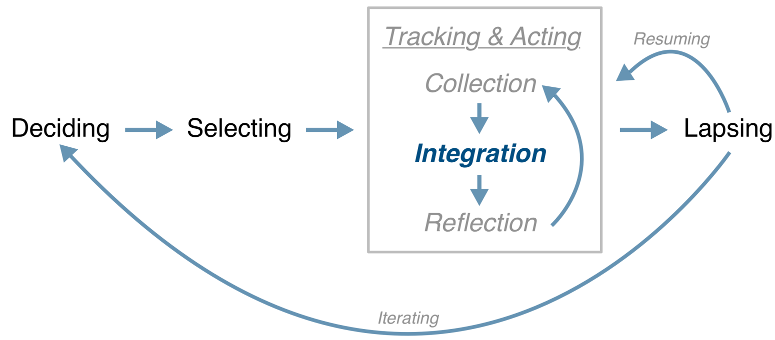


Figure 2.1: The lived personal informatics model by Epstein et al. [59] as illustrated by Moore et al. [128].

across many of the stages. See Figure 2.1 for a visual representation of this model.

Interestingly, these stages have not been equally studied [60]; of importance to this thesis is the integration stage, which is one of the more understudied. Integration involves preparing for reflection by combining and transforming data [107, 128] and is one of the more complex stages to facilitate. This complexity arises from the uniquely varied and large number of data exploration needs required, and the difficulty in providing tools that afford such potential exploration. Research has attempted to curb this complexity through narrowed design spaces and algorithms that detect a greater number of insights [91, 92]. However, these solutions are often fixed and do not involve user interaction and/or control which can have further negative effects [10, 92]. It is crucial that we develop more dynamic and user-centered integration tools. These tools should enable individuals to generate unique insights through data transformation across tracking and action stages, providing greater reflection and actionable insights beyond predefined metrics.

2.2 Wearable Computing

Thad Starner [185] defined wearable computers as *"any body-worn computer that is designed to provide useful services while the user is performing other tasks."*

Today, examples of wearable computers include smartwatches, smartrings, chest strap sensors, smart soles for shoes, body-worn inertial measurement units, smart glasses, and brain-computer interfaces [166]. As of March 2024, wearable sales were up 10.5% from that of the previous year, largely driven by the sale of new smart glasses, and sales are projected to continue to grow at around 4% per year up to around 650 million units sold in 2028 [112]. As it stands, one of the main services offered by wearables is to ubiquitously capture and track personal health metrics from the wearer [34, 69], which can then be later explored to support personal informatics and healthier living. These fitness metrics, and algorithms associated with their use, often track physical activity information (i.e., step count, distance, pace, time, cadence, etc.), sleep quality (i.e., hours slept, sleep rhythm, etc.), caloric burn, heart rate, energy expenditure, and oxygen saturation [43, 170].

Attributes of Wearables It is important to note that wearable computers offer different and adapted services from that of typical desktop computers and other mobile technologies (e.g., smartphone or tablet). With this in mind, the goal should not be to replace these traditional devices; rather, we can complement them by leveraging attributes that are unique to a wearable device. From a wearer's perspective, Steve Mann described six key attributes of wearable computing that make wearables, and their services, unique to that of other devices [123]:

1. *Nonrestrictive to the wearer: Wearable devices should not impede the wearer's ability to physically perform other tasks.* The physical miniaturization of sensors and hardware in today's wearable and on-body devices, has greatly allowed for for unrestricted use to become a reality [170]. Particularly, the compact size of smartwatches, and wrist-worn location, ensures they do not physically hinder daily activities, making them great devices to have on us at all times. Importantly, this is achieved while still offering a wide range of sensors, robust processing power, network connectivity, and various interaction modalities (i.e., visual, auditory, and haptic output as well as touch, speech, and gesture input). Together, this makes the smartwatch more versatile than many other wearables, such as a chest-worn heart rate monitor,

which may require to be even smaller than the smartwatch or worn in a less accessible location.

2. *Non-monopolizing the wearer's attention: Wearable devices, and their services, should be designed under the assumption that computing will be a secondary activity, not the primary focus of attention.* For example, tasks such as hiking, jogging, working, or walking to and from locations should remain the primary activity and focus, while interaction with a wearable such as the smartwatch is second [14]. In fact, not monopolizing interaction is imperative. If interaction with a wearable, or any mobile device for that matter, requires too much attention and focus during a primary task, then peoples' interaction performance will decrease [164], they will miss elements within their surroundings [82], and will change aspects of their primary activity, such as how they walk [44]. Consequently, any interaction with the wearable device during these times must be quick and lightweight such as to not distract from the primary task at hand.

3. *Observable by the wearer: The output medium of a wearable device should be constantly perceptible to the wearer.* Many wearables offer haptic feedback to convey pertinent information [9, 81, 147]. Smartwatches, afforded by the inclusion of a screen, also offer visual feedback to allow observation. However, to truly support non-monopolized interaction, it is crucial to provide alternative feedback options for when the visual system is overloaded [19]. In such cases, auditory or natural language output can be an effective means to allow the wearer to receive observable feedback without needing to visually engage with the wearable [151].

4. *Controllable by the user: Wearers should be able to take control of the device whenever they wish.* The smartwatch is wrist-worn making it easily accessible. This body-worn location makes it a perfect device for user input and engagement in many daily scenarios of use [68, 148]. In fact in certain daily scenarios, such as at the gym, use of the smartwatch can eclipse that of other devices such as the smartphone [34]. This access and convenience

allows users to interact with their device on the go, whether they are working out, commuting, or undertaking other activities. Importantly, however, interaction with the smartwatch often comes at the cost of monopolization (as seen in the second attribute above). This need to support control even during differing usage scenarios is the reason that some research has explored using the smartwatch during eyes-free, walking, and encumbered conditions [137, 155, 176, 199, 200].

5. *Attentive to the environment: Wearable devices should provide situational awareness to the wearer.* As the use of wearables goes beyond a traditional desktop setting as is used throughout a person's daily life [125, 148], the smartwatch and its sensors have the potential to recognize these varied usage scenarios and even the wearer's bodily information, incorporating gathered information when and where necessary. For instance, a smartwatch can detect when a user is exercising and adjust its display to show relevant fitness metrics [86].

6. *Communicative to others: These devices should function as communication mediums when desired by the wearer.* More so than other wearables on the market, one of the main uses of the smartwatch is as a communication device [148, 192], providing quick access to notifications and networking capability to allow messaging and other communication. Traditionally, interaction with the smartwatch, such as for communication, has been a challenge due to the small screen size [134]. For this reason, many research works have explored adding varied on- and around-device input capabilities [25, 66, 79, 199].

   However, many of these works often do not take into account commodity smartwatch capabilities and common interaction modalities, such as general touch interactions and the now powerful natural language processing which underlies speech input. Furthermore, many gestural techniques explored can be tiring and slow [4], while some touch techniques often invoke a large number of pre-touch efforts (i.e., lifting and moving your finger to conduct repeated interactions) which should aim to be kept to a minimum [176]. Leveraging simple commodity interactions can provide efficient and effective communication capabilities, allowing intuitive interactions for users to communicate with one another, or perhaps even with a system when exploring data.

2.3 General Smartwatch Usage

Smartwatches are one of the most wide-spread of any wearable currently on the market [168]. It is estimated that in the next four years, until 2028, the smartwatch market will continue to grow about 5.4% from current sales, to reach roughly 230 million units sold worldwide [49]. As smartwatches already offer maturity, convenience, and widespread adoption, this thesis focuses on pushing the boundaries and capabilities of the smartwatch over that of other wearables which offer less potential for interaction with collected health data. In the following paragraphs, we discuss reasons for smartwatch use, general usage details, and long-term usage trends.

Reasons for Adoption and Use People adopt smartwatches for a variety of reasons, including performance expectancy, which is the belief that the smartwatch will enhance their performance [32]. Perceived usefulness also plays a crucial role, as people expect the smartwatch to help them achieve their goals [126]. Additionally, the compatibility of smartwatches with other devices and services encourages adoption [32]. Lastly, aesthetics are an important factor, as the design and appearance of smartwatches contribute to their attractiveness, especially as they are worn on the body [32].

Subsequently, once adopted, research in recent years has abundantly explored the reasons for use of both smartwatches and fitness trackers [34, 69, 88, 111, 125, 148, 165, 192, 193, 203]. These works, spanning six years from early adoption to current trends, found that smartwatches are generally used for time checking, notifications, activity tracking, communication, information management (i.e., weather, stocks, etc.) payment transactions, and navigation.

Increased smartwatch adoption has largely been driven by the smartwatch's perceived capacity for health data collection and usefulness in promoting health and well-being [52, 174]. Not only can the smartwatch track general activity metrics, something for which proximity to the body is a key factor [130], but they also have the potential to aid in detecting COVID-19 symptoms and provide constant health monitoring, making them valuable for individuals of all ages, including those with dementia, depression, and high-stress [113, 206]. Now, smartwatches are expected to soon offer further sophisticated health monitoring features from those mentioned above [74, 89, 93, 99, 186], including broader well-being, stress management, breathing detection, and medication/diet tracking. This holistic health tracking primes the smartwatch for a lead role with respect to not only personal health data collection but also immediate access to the data, for greater usefulness and performance expectancy, by allowing for increased and in-situ insights.

Usage Metrics Interactions with smartwatches are typically brief, with most lasting less than 5 seconds [148, 192], with average of only 1.9 seconds [148]. Activity-related interactions tend to be slightly longer, ranging from 5 to 10 seconds and averaging up to 18 seconds [103, 148, 192]. However, applications such as maps and messaging can greatly extend interaction times, reaching up to 45 seconds [148], suggesting a natural maximum length for individual smartwatch interactions. Visuri et al. [192] categorize these interactions into two types: peeking or glancing at the screen, which accounts for about 65% of interactions, and more engaged interactions, which make up the remaining 35%. They also note that the majority of interactions, 82%, are initiated intentionally by the user, with notifications prompting the remaining 18%. This active and intentional interaction with the smartwatch bodes well for its further integration into daily life, as it demonstrates peoples' willingness to engage with the smartwatch for various purposes beyond just passive notifications. Such engagement underscores the versatility and utility of smartwatches, highlighting their role not only as convenient tools for quick glances but also as intentional parts of a larger device ecosystem.

Smartwatch interactions, whether quick glances or more engaged actions, occur in various locations and during many daily activities, such as walking to and from destinations, commuting, working out, or performing tasks at work [148, 192]. The broad range of usage scenarios and the convenience of a wrist-worn device provide significant advantages over even the smartphone, as smartwatches offer quick and easy access to information. This is particularly beneficial during activities where using a smartphone might be cumbersome or impractical (e.g., during a hike or walking to the transit station). Notably, during activity tracking, Chun et al. [34] have found that interactions with the smartwatch in fact surpass those with a smartphone, emphasizing the smartwatch's ease of use and accessibility during in-situ activities. With this greater integration into daily life, it is important that we provide people as much access to the information that they desire, anytime and anywhere.

Long-Term Usage Trends Interestingly, usage typically declines over time for both activity trackers [38, 41, 70] and smartwatches [131, 142, 173]. This decline in use is often attributed to a mismatch between user expectations/goals and the actual offerings of the smartwatch [38, 41, 70, 142]. Users may start with specific goals in mind, such as improving overall health and fitness, but then find that the smartwatch does not adequately support these goals through captured metrics, accuracy of metrics, potential for exploration, and expected functionality [70, 142, 144]. This limited functionality can lead to overall dissatisfaction, causing users to gradually abandon the smartwatch [131]. Externally, people may even stop using their smartwatch due to a lack of motivation and/or life factors that that can interrupt/de-prioritize data tracking [7].

Research has identified several factors, however, that can motivate continued use of smartwatches. One key factor is perceived usefulness through utilitarian aspects (i.e., tracking quality), where people find value in the tracked data that is collected [33, 78, 142, 173]. However, hedonic aspects (i.e., fun and pleasurable aspects) are also important as they positively contribute to enjoyment and self-expression which in turn increases satisfaction and then continuation of use [31, 78, 173]. Afforded customization and personalization, through utilitarian and hedonic aspects further improve perceived value which in turn factors into continuation of use [117]. Finally, the convenience characteristics of smartwatches, such as their wrist-worn accessibility, ease of use, and quick access to needed information all play a significant role in maintaining continued use [142].

2.4 The Smartwatch and Personal Data Exploration

User Needs and Exploration Goals Research has shown that people have diverse informatics needs and motivations specific to health data exploration on the smartwatch. Common reasons for using the smartwatch as a personal health tool include high-level understanding of activities [58], comparative exploration [3, 161], accountability and participatory interaction [133, 194], motivation [3], as well as goal- and performance-based use [3, 6, 22]. Importantly, the reasons for tracking health data can evolve and change over time [59], further requiring diversity in the tools provided.

Research has also noted that data exploration is desired not only during an activity, but also immediately before and after [3, 148]; this reflection, using a smartwatch, temporally near an activity can be increasingly intertwined with the activity itself, and thus beneficial to influence real-time and actionable decisions [3, 69, 107]. In turn, this can further help to set and achieve meaningful personal health goals. For instance, individuals might want to adjust their activity levels based on their daily step count or improve on their sleep quality and length by analyzing sleep data immediately after waking up.

Data Exploration on Smartwatches: Challenges and Opportunities One of the current challenges in achieving the above mentioned desired exploration, is that current smartwatches standardize visualizations and displayed metrics, and thus a person's underlying needs and goals [180]. Simply put, people with similar smartwatches and operating systems are presented with identical visualizations and metrics (albeit with their own numeric values). Given the diverse and highly unique nature of personal health data exploration, it is essential to develop tools that allow for unique and personalized exploration and understanding of the tracked data.

Of the metrics that are displayed, Spiel et al. [180] highlight that while smartwatches and other fitness trackers are designed to capture quantitative metrics, they often overlook the qualitative experiences, understanding, and reflection associated with these metrics. This disconnect forces users to translate qualitative goals, such as losing weight, into simple quantitative metrics, like increasing daily step counts [138], without reflecting deeper.

From the user's perspective, it remains difficult to find mobile health applications that suit their needs and health goals [161], with people citing a lack of information richness and overall usefulness regarding their smartwatch health applications [145]. Neshati et al. [134] note that smartwatch users are seeking answers to their personal health data queries that are simply not available to them; these types of desired queries are exemplified as *"how am I doing so far?"* or *"how am I doing compared to my friend, Jane?"* [108], and are not currently supported within a time frame that supports common smartwatch use. These missing features, underscored by our limited knowledge of concretely reported queries, hamper full engagement and a broader range of benefits [70].

While a growing body of research revolves around smartwatches and wearable activity trackers, only about 10% of such research is centered around the collected data and the ability to convey appropriate meaning and function [171]; other research focuses on topics such as privacy, acceptance, adoption, and abandonment, behaviour change, and hardware. Of the works that do focus on data exploration, they often highlight the technical complexities of visually rich personal health data represented on a small smartwatch screen which creates an exploration environment with limited usability, customization, and interaction [34].

Much of the limited exploratory capability on the smartwatch stems from the small screen of a smartwatch which presents several challenges, including screen occlusion and fat finger issues [172] during interaction. The limited screen size constraints not only inhibit interaction but also visualization options, making it difficult to display comprehensive data trends and comparisons over time [3, 134]. While displaying data alone is a challenge, further providing filtering options, toggles, etc. can be cumbersome to navigate through potentially involving extensive screens, or lists of options. These challenges importantly present opportunities for innovative interfaces and visualizations tailored specifically for smartwatches. As data collection becomes more advanced, and the smartwatch further matures, there is a growing need for creative solutions that not only condense complex information into digestible, easy-to-understand formats, but also enhance exploration potential, enabling users to gain unique and personal insights with minimal interaction.

Interaction Techniques for Smartwatch Data Exploration While many research works have broadly explored how to expand interaction capabilities on the smartwatch that mitigate the small screen challenges the [1, 48, 95, 137, 141, 155, 162, 181, 204] (only a select few of the many research works), very few focus these interactions for use in data exploration. To note, some work has been explored which utilizes the smartwatch as an input tool within a multi-device exploration environment (i.e., in conjunction with a large wall display) [80, 201]. However, these works do not involve exploration directly on the smartwatch screen.

To our knowledge, Neshati et al. [137] is the only work that directly focuses on creating an input technique for interacting with charts on the smartwatch screen. Their approach successfully utilized a portion of the bezel around the screen to reduce fat-finger and screen occlusion issues, allowing users to accurately target individual data points within a displayed chart. However, this input technique has its limitations, as it primarily enables users to view individual values for specific data points of a chart already on screen, lacking the uniqueness and functionality needed for expanded interaction and even deeper exploration.

Visualization Techniques for Smartwatch Data Exploration Practitioners, designers, and researchers all acknowledge the need to adequately represent self-tracking data on the smartwatch, especially given the typical short interaction times and their small screen sizes. Currently, various visualization techniques are used on commodity smartwatches to reflect complex health data, including bar, donut, radial bar, and line charts [13, 136], as well as text, icons, and pictograms [84]. Each has its pros and cons. Bar charts are widely used and easy to understand but can be space-inefficient. Donut charts are common but not always effective at presenting proportions. Radial bar charts are not efficient for comparing more than a few data points. Line charts provide clear representations of time-series data but can become cluttered if too many data points are displayed.

Regardless of the visualization method chosen, however, ensuring glanceability is paramount for effective data representation on smartwatches. To be truly effective on a smartwatch, visualizations must be glanceable—they should convey essential information within five seconds or less [14, 68]. This quick comprehension is key given the brief nature of smartwatch interactions.

To this effect, case studies of smartwatch visualizations have been conducted, to first better understand people's perceptions and ability to perform common data analytic tasks through visualization on the smartwatch. These include to understand glanceability [13] and visual parameters that make up the visualizations [85, 120]. Blascheck et al. [13] pinpoint this problem by investigating perceptions and efficiencies of bar, donut, and radial bar charts for a common two-point data comparison task. They find that 24 data points represented as a bar or donut chart allowed for quick and efficient comparison. Furthermore, Islam et al. [85] studied a variety of sleep data visualization designs to elicit the most preferred and suggest design guidelines for differing form factors.

Other visualizations techniques have been explored to push the boundaries of the current techniques used. Amini et al. [3] proposed, through both user study and designer sessions, a variety of visualization ideas each which target a data exploration need that people had on the smartwatch. Visualizations proposed included simple text and numbers, standard charts, and icons and symbols. For more involved data analysis on the smartwatch, and recognizing that many data points could be involved, Neshati et al. [135, 136] and Tufte [189] proposed compression of line graphs, often called sparklines. These charts were shown to not only retain salient features necessary, but also remain to allow for the performing of common analytic tasks (i.e., trend detection, max/min detection, and value comparison).

Summary While these works are crucial for supporting glanceable exploration of collected data on smartwatches, and positing interaction capabilities for specific data analysis, our work in this thesis recognizes an opportunity for a complementary approach. Although the average interaction time with a smartwatch is 1.9 seconds and often under 5 seconds, usage time increases significantly during physical activity, averaging 18 seconds, and reaches up to 45 seconds for other interactions

[148]. We believe that, given the diverse exploration goals of smartwatch users, personal health data needs, and perceived advantages of exploration directly on the smartwatch while in-situ, we can and should enhance the exploratory capabilities of smartwatches. By extending, and complementing current solutions to allow for more personalized exploration directly on the smartwatch, we can remain within the typical bounds of smartwatch interaction. This approach can provide users with deeper insights anytime and anywhere, enhancing the utility and impact of health data collection and exploration directly on the smartwatch.

Chapter 3

Databiting: Lightweight, Transient, and Insight Rich Exploration of Personal Data

---

We first conceptualize a form of data exploration that serves as a lightweight and transient bridge between short- and long-form exploration; something we recognize and strive for throughout the following chapters. Conceptualizing, rather than providing a strict definition, acknowledges the fluid boundaries of various data exploration uses and user needs. This chapter resulted in a publication in Computer Graphics and Applications as a Visualization Viewpoint [158]. The article was a collaborative effort among myself, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. The collaboration took shape through countless meetings focused on concept creation and development, as well as editorial work. Any mention of 'we/our' in this chapter refers to my co-authors.

As device hardware and software advance, enabling broader access to personal data, new opportunities for data exploration arise: data exploration has the potential to intertwine with our lived experiences and day-to-day activities. However, conducting data exploration in many scenarios of use poses unique challenges. It is crucial that exploration does not hinder, but rather assists, a wide range of scenarios and contexts in which we find ourselves seeking insight [106]. In this chapter, we advocate for the visualization and personal informatics research communities to focus on the development of lightweight and transient exploration techniques that remain to enable insight rich access to personal data.

Current systems offer one of two approaches to personal data exploration. Glanceable and micro visualizations have been widely adopted in mobile applications [3, 13, 15, 26, 68, 85, 135, 136], at times combined into dashboards (Figure 3.1 left). They provide concise and focused representations of information in a limited space and context for users to easily grasp information at a glance. However, despite their popularity and necessity, they offer only specific insights and allow limited interaction, leaving users without the ability to cater to their personal and situational needs. In contrast, heavyweight applications (e.g., Tableau, Excel, custom script writing) have been designed to enable longer-form and more comprehensive data exploration (Figure 3.1 right). These applications often require considerable time and knowledge to use. These barriers make them at times inaccessible or inconvenient (e.g., during physical activity, while walking a pet, or while cooking). Between these approaches, a significant gap in the field of personal informatics and visualization arises: limited information richness hinders users' ability to better comprehend and leverage personal data through exploration that can be efficiently undertaken during broader contexts.

Within this chapter, we conceptualize and discuss the notion of *databiting*, a term we coined to indicate lightweight, transient, and insight rich exploration. We further delineate five research considerations—contextual factors, interaction modalities, the complementary relationship between databiting and other forms of exploration, personalization, and evaluation challenges—focused towards enabling and understanding databiting. Importantly, these research areas can work in concert to provide lightweight and transient access to richer personal insight anytime and anywhere. By embracing ideas and approaches outlined in this article, we can empower individuals to effortlessly gain insights from their data as needed, transforming the way they explore and interact with their personal data. Together, this recognizes the importance of our interactions with personal data and emphasizes the significance of seamlessly integrating rich personal insights into our daily lives.

As such, the key contributions of this chapter are two-fold:

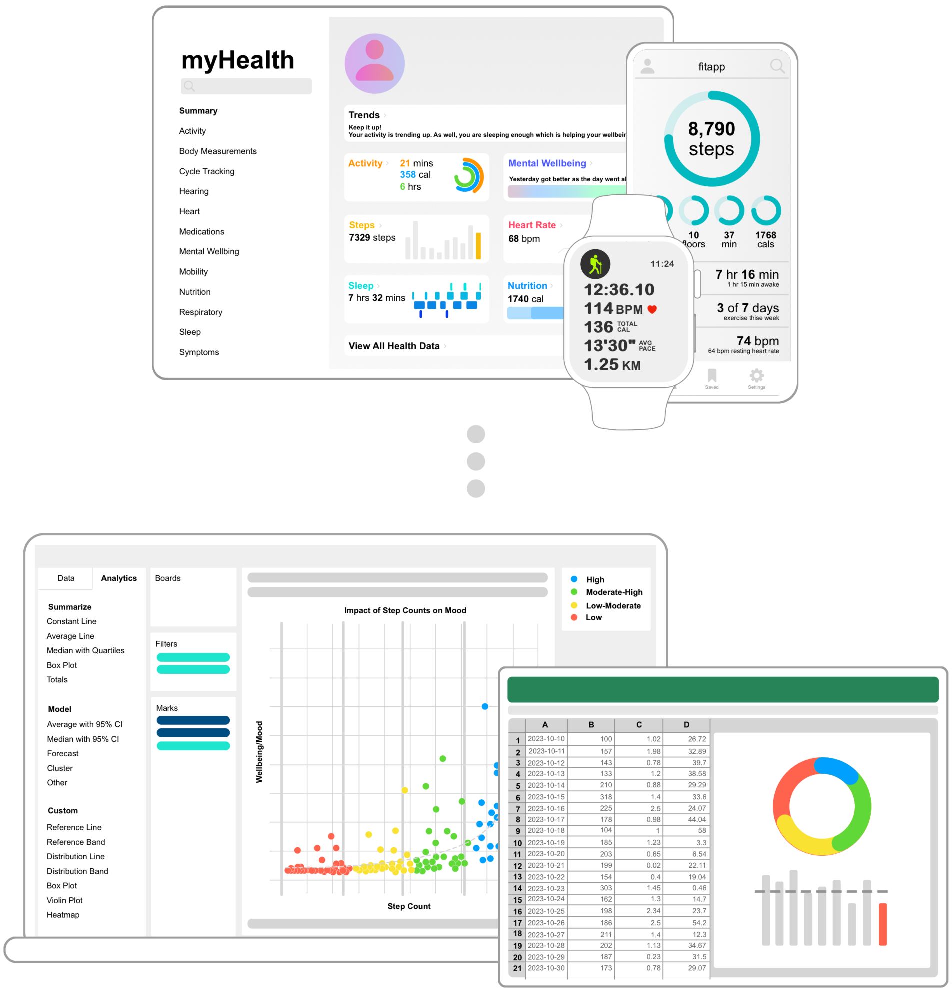


Figure 3.1: Representations of current mobile data exploration applications are highlighted, grouped by general information richness: applications which afford glanceable and micro-visualizations (at times combined into dashboards) (top) and applications which provide potential for heavyweight data analysis (bottom).

C1: An introduction to the term *databiting*, conceptualized as lightweight and transient data exploration, that bridges the gap between glanceable and longer-form data exploration. We provide a conceptualization of *databiting*, highlight its unique characteristics, and discuss potential benefits.

C2: A delineation of open research challenges and goals for the research community to pursue in enabling *databiting*. Specifically, we discuss contextual factors, interaction modalities, the relationship between complementary forms of data exploration, personalization and customization, as well as evaluation challenges.

3.1 The Databiting Concept

We conceptualize *databiting* as the act of interacting with data to gain increasingly rich insight through lightweight and transient exploration. The result is a *databite*, concise personal insight that extends upon what can be derived from glanceable or micro visualizations. Databiting as both a new concept and a topic for research is fluid in nature: Boundaries defining insight and data exploration methods allowing for such insight are not rigidly defined or fixed.

To illustrate this concept, we draw upon analogy. Databiting can be seen as equivalent to eating a small and easily consumable snack. The size of a snack and the number of bites required may vary from person to person and from context to context. Yet, what remains constant is the lightweight and transient nature of snacking compared to consuming a meal (often a reasonably large amount of food). In the context of data exploration, databiting equates to the consumption of bite-sized information that provides rich insights or sustenance in the moment. This builds upon simply viewing a mobile data visualization and does not require more in-depth and long-form data exploration, which can be done later when necessary or more appropriate.

Importantly, databiting is not meant to replace either long-form exploration of data or shorter-form viewing of glanceable visualizations; rather, it is complementary to them. By bridging the gap between these two forms of exploration, databiting offers a new, complementary, form of exploration that pushes the boundaries of what is currently attainable. This integration of exploration methods can foster a more comprehensive, valuable, and unique (i.e., richer) understanding of personal data, yet remain accessible in a lightweight and transient manner. By offering a range of exploration options, across devices and throughout a range of usage scenarios, we expect individuals can derive greater benefits from their data-driven insights anytime and anywhere.

3.1.1 Lightweight, Transient, and Insight Rich Exploration

Time and effort, which should remain small, are key factors in the context of databiting. Glanceable and micro visualizations excel at providing quick and easy data engagement, however, remain limited in their level of insight conveyed.

We envision that there remains a large opportunity to enable access to more information rich insights while maintaining a lightweight and transient approach, see Figure 3.2. Consider a scenario in which a runner is stopped at a traffic light, waiting to cross the street. The primary task is their run and the focus on the surrounding environment. Secondary to this, they look at their heart rate zone data. Seeking a databite they simply tap on a stacked bar chart which highlights their current heart rate zone. This action reveals increasingly detailed insight into the time spent in each heart rate zone, enabling the runner to concentrate on entering or maintaining a specific zone as they proceed with their run. Notably, this is not functionality that is currently available.

As can be seen, databiting can offer richer insight without requiring substantial effort, engagement, or time. This allows data exploration to occur as a secondary task, alongside a primary ongoing activity (e.g., while out for a walk) or during a recurring daily activity (e.g., riding a bus home). Through the prioritization of simplicity rather than detail and complex insight, the small *size* of databites ensures that they can be easily and appropriately consumed.

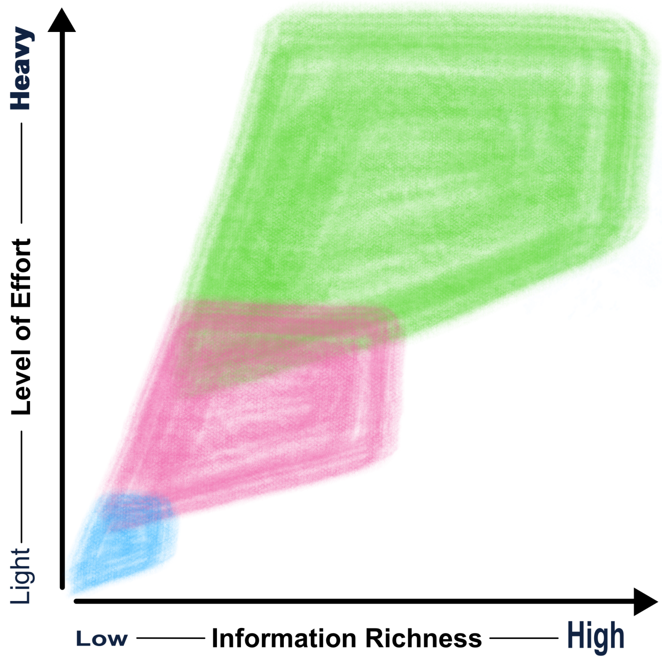


Figure 3.2: In blue (bottom left), we highlight glanceable and micro visualizations for data exploration. In green (top right), we highlight heavyweight data exploration and analysis. In pink (middle), *databiting*, as a concept, promotes the need for increasing information richness while exploration remains lightweight and transient. We encourage the reader to envision how visual data exploration can consume this area of the graph. We present these rectangles as *sketched* illustrations to signify the fuzzy boundaries of these forms of exploration and potential variance within them.

3.1.2 Device Agnostic Exploration

Due to their highly portable nature, the immediate availability of data being collected, and the smaller form factors limiting heavier exploration, databiting is particularly well-suited for smartwatches and other wearable devices. However, the emphasis of databiting is not on the specific device used or insights gained but is in the manner in which data is accessed. As such, databiting can be seen as device agnostic and can be done on any device that grants access to relevant data, ranging from wearables, smartphones, and tablets, to laptops. For instance, before switching to a different task on a laptop, a researcher first quickly checks the current screen-time information. With simple mouse interactions, they learn about a negative trend that results from increasing the duration of one continuous work block and thus decide to take regular breaks, starting with an immediate break. Similarly, a tablet user exploring their financial data while in transit could use the stylus to circle a transaction and draw a line to another to view how they are related. Taking the interaction a step further, the user could employ pre-defined gestures to accomplish further exploration (e.g., displaying a transaction timeline or finding similar transactions).

3.1.3 Data Agnostic Exploration

The versatility of databiting extends beyond that of personal data, making it applicable to a wide range of domains and contexts involving data exploration. While the initial conceptualization revolves around personal data for individual insights, the fundamental principles of lightweight, transient, and insight rich exploration can be seamlessly applied to various data categories. For instance, in the business context, professionals can benefit from quick insights (e.g., during a meeting). Databiting can enable easy access to necessary insights to foster discussions and inform decision-making without the need for extensive data preparation or analysis. In the context of scientific research, researchers running a study can use databiting to garner quick insights about the amount and quality of collected data. This can help them decide to continue or stop and refine, if needed.

3.2 Expected Benefits of Databiting

We discuss the envisioned potential benefits of databiting that are not fully realized with the current capabilities of mobile data exploration. Further study is needed to identify and demonstrate any tangible benefits that may exist.

3.2.1 Introductory and Intermediary Access

As highlighted earlier, databiting has the potential to bridge the gap between brief information access (short-form exploration) and comprehensive knowledge gain (long-form exploration). Offering lightweight, transient, and insight rich access to personal data allows for engagement with smaller bite-sized snippets of timely information. This approach provides an accessible entry point for individuals new to data exploration to begin to explore and understand their data.

Furthermore, a bite-sized approach to exploring data can serve as an intermediary step, providing individuals with a gateway for more in-depth data exploration at a later time. By offering exploration in easily consumed pieces, databiting can spark further interest, familiarity, and excitement. When presented with quick and accessible personal insights, individuals may become more curious and motivated to further explore their data. Over time, this increased engagement may foster a greater sense of familiarity and confidence with data; ultimately facilitating further exploration and a deeper understanding.

3.2.2 Increased In-Situ Insight

Due to the lightweight, transient, and insight rich nature, we anticipate databiting will be beneficial for obtaining data-driven insights during in-situ exploration. In-situ instances of exploration refer to moments in which data analysis and reflection occur closely related to an ongoing activity, enabling immediate and direct impact [108]. Databiting can extend in-situ data access, currently offered through glanceable and micro visualizations, by providing further information richness. This leads to potentially increased actionable insights that are directly relevant to the ongoing activity. Whether it is using a smartphone to explore caloric intake for the day while cooking, asking a smart speaker about your home energy consumption for the day as you arrive from work, or checking productivity on your laptop while studying, databiting can empower individuals to gain valuable, personalized insights that inform immediate decision-making.

3.2.3 Perceived Usefulness

A current challenge concerning devices and applications that allow for personal data access and exploration is a lack of perceived usefulness [131]. Current offerings often fail to meet expectations, resulting in the abandonment of devices, applications, and even the collection of data altogether [38]. While addressing perceived usefulness is a multifaceted challenge, databiting can serve as a promising start. Increased information richness and personalized insights enabled by databiting have the potential to enhance the perceived usefulness of the devices and applications used for databiting, and the value derived from collected data. In turn, this could lead to greater overall outcomes as the collection and exploration of personal data are not outright abandoned.

3.3 Research Considerations

In this section, we discuss how we can begin to enable databiting through research, to achieve increased access to personal data anytime and anywhere. Specifically, we discuss research considerations that aim to overcome existing challenges and benefit from promising opportunities.

3.3.1 Contextual and Attentional Factors

Contextual factors can become increasingly integrated with databiting, shaping the recognition of potential exploration, the integration of relevant data, and even output provided to answer a question asked during a databiting instance. Unlike conventional data exploration, in which context might be analyzed as an additional factor later on, databiting can be seen to easily integrate context. Imagine a traveler, using a mobile app to explore a new city. Basic on-screen insights might reveal location in relation to nearby attractions. However, databiting can leverage contextual data such as time of day, personal preferences, and current location. This integration of contextual data can enhance the overall experience and provide the potential for increasingly relevant in-situ insights. Focusing research on understanding and incorporating contextual factors within data exploration can significantly impact exploration that is possible. For example, we can study techniques to incorporate contextual attributes into data visualizations, develop context-aware recommendation algorithms, and look to gain a better understanding of how context can be incorporated into queries and answers desired for databiting.

Beyond integrating context into the data itself, the situational demands and underlying contexts of in-situ activities can also influence data exploration capabilities. Factors such as physical exertion, cognitive load, attentional capacity, and movement can affect a person's ability to engage with databiting effectively. Research into these dimensions can provide valuable and actionable insights for optimizing databiting to better fit the context of use. For example, leveraging multi-source data streams [110]—such as sensor data (e.g., heart rate, skin conductance, accelerometer) combined with self-reported measures of perceived workload and task difficulty—can help build a comprehensive understanding of an person's exploratory capability in real time. Using this contextual information, we could model expected performance in terms of efficiency and effectiveness, akin to approaches that have been used to measure arm fatigue during mid-air gestures [77].

3.3.2 Interaction Modalities

Enabling databiting requires considering appropriate interaction modalities that are not only efficient but also cater to the unique constraints of in-situ scenarios. This is not straightforward, especially because we aim to increase the level of data accessibility across devices with potentially limited interaction spaces (e.g., smartwatches, smartphones, and other wearables).

To facilitate lightweight, transient, and insight rich exploration of personal data, a multimodal approach through natural language and the device's primary interaction method can be used (see Figure 1.1). The primary input modality,often touch, allows for direct manipulation, discrete selections, and when natural language keywords cannot be remembered [98]. Importantly, natural language (e.g., speech) benefits from enabling fast and flexible expression of complex queries [8]. Recent research in personal health applications on smartphones has demonstrated the benefits of multi-modal touch and speech interactions for gaining insights into personal health data [98]. Notably, individual and combined interactions were often used for differing, yet equally important, components of the data exploration process. Touch and speech combined showed promise for refinement of an initial query or displayed graph, much like databiting may extend upon glanceable or micro visualizations. However, a better understanding of individuals' personal data query requirements needs to be further addressed to fully recognize the interactive needs of databiting.

Furthermore, exploring alternative output modalities can complement currently available data visualizations. Devices such as headphones, earbuds, and home assistants offer opportunities to incorporate natural language responses when databiting. This can appropriately provide access to data when people's visual systems are overloaded [19]. There is limited work on the requirements for natural language responses to personal data queries. Further research can focus on formulating efficient responses, determining the level of conveyed insights, and ensuring the memorability of responses. Integrating these findings with existing data visualization approaches can optimize access to personal data.

3.3.3 Databiting and Broader Exploration

The nature of databiting invites opportunities to consider a relationship with further long-form data exploration. Understanding how people can transition from one to the other and how previous in-depth exploration can inform future databiting will contribute to a cohesive and personalized data exploration experience. However, it is crucial to recognize that the seamless transition between databiting and broader exploration experiences will likely require extensive device and app interoperability. Future research is needed to address these technical hurdles.

Databiting can afford a stepping stone for people to delve into in-depth data exploration when more appropriate. For example, as seen in Figure 1.1, Sam may recognize that their pace was improving for a while, until the most recent hike. To better understand why today's hike had such a decrease in pace, they can save the databite such that they are reminded to further explore for an external cause (e.g., decreased sleep) at a later time (e.g., while at a desk) and on an appropriate device (e.g., on a tablet). It becomes important to understand how individuals can save and manage databites, and any new questions that arise, for future exploration.

Conversely, the results of prior in-depth data exploration have the potential to influence and enrich future databiting. For example, an individual may have taken the time to pore through their credit card statements, itemizing transactions into categories and noting a budget limit for each. When in the store, this past exploration may influence databiting which is directed towards understanding if a purchase can and should be made within a predefined category.

These examples further highlight an underlying challenge concerning interoperability. Currently, fragmentation and non-standardized access to collected data hinders cohesive personal data exploration. For instance, data saved on one device may not seamlessly integrate with a user's preferred in-depth exploration tool on another device (e.g., to analyze collected data from an Apple Watch in Tableau, exporting and importing of data is required). Furthermore, data collected on multiple devices may be stored using different protocols and formats. Without a concerted focus on interoperability, the potential for databiting to seamlessly complement broader exploration methods, and vice versa, may be hampered, limiting overall effectiveness.

3.3.4 Personalization and Customization

Personalization: Studying and supporting personalization is a key aspect when looking to optimize databiting and enhancing the exploration process. By tailoring insights, recommendations, and visualizations to individuals, and their contexts of use, personalization ensures that relevant information is efficiently presented. This reduces the required time and effort, enabling people to quickly access valuable insights and make informed decisions.

Through personalized databiting, technology mediation (i.e., artificial intelligence, machine learning, etc.) can alleviate individuals from manual data analysis. For example, a personalized databite may provide instant workout recommendations or in-situ alerts to metrics trending in a specific direction. A potential reduction in temporal workload allows people to focus on the actionable outcomes of databiting rather than the exploratory process. Furthermore, serendipitous discovery is a benefit of visual data exploration. Personalization can be used to mediate tailored serendipitous insight, even with lightweight and transient exploration.

When developing personalized systems that provide increased insights with minimal interaction, both technical and experiential research questions arise. How can user behavior and preferences be accurately captured and utilized for personalization in databiting? It is crucial to ensure that notifications and information provided align with internal models and desires, as undesired interactions can negatively impact the user experience. This further leads to the question of how personalization impacts user engagement and satisfaction with databiting. Understanding and building around a person's individual preferences is fundamental to delivering effective personalized insights.

Furthermore, identifying the most relevant data and patterns for providing meaningful and timely personalized insights is essential. This further aligns with understanding the contexts of use. For example, a system must determine which data points are most valuable to the user at any given time and adapt to changing contexts and preferences. Algorithms that support personalization must consider a vast amount of information, including user behavior and past interactions, preferences, historical data, contextual factors, and more. Specifically, we need to understand which types of data and patterns are most relevant for providing meaningful and timely insights. By addressing these questions, our future goal will be to develop personalization strategies and algorithms that enhance user experience and make databiting a more intuitive and efficient process through personalization.

Customization: For those who prefer a higher degree of control, offering customization options is a critical aspect of effective databiting. Customization allows users to take the time to preemptively tailor their data exploration applications to their preferences, which can include the creation of custom charts, setting specific data filters, and adjusting the look and feel of their data exploration interface. By enabling users to customize these aspects, explicit databiting (i.e., querying of data) is not necessarily required. Yet, a challenge that remains in this space is how to effectively provide people the ability to create their own exploration experiences.

To address this, future research should explore several key questions: How can customization tools be designed to be intuitive and accessible for all users, regardless of their technical proficiency? What methods can be employed to ensure that customized settings are easily adjustable as user needs and preferences evolve? Additionally, how can we support users in discovering and utilizing customization options to their fullest potential?

One promising solution lies in developing multi-modal systems. By integrating multiple modes of interaction, users can more effectively tailor their databiting experiences to fit their unique needs and contexts. This approach not only enhances the usability of customization features but also ensures that the system remains adaptable and responsive to the diverse ways in which users interact with their data.

Spectrum of Personalization and Customization: Finally, to fully realize the potential of personalization and customization, balancing automated personalization with user control is also essential to maintain system intuitiveness and flexibility. It is unlikely that personalization and customization is a binary choice. Rather, individuals will likely fall along a spectrum between the two. This balance may vary between people, and even for an individual person given varying goals and contexts of use. Research in this space should work to model user preferences that can align a person along this spectrum. By then adapting a user interface to allow for the desired level of personalization and customization, we can afford a greater user experience and efficiency for databiting.

3.3.5 Evaluation Challenges

To realize the full potential of databiting, two areas of evaluation must be considered. First, it is essential, yet challenging, to understand an individual's needs and goals when interacting with personal data while in-situ. Modest methods such as sketching or surveys lack real-world data, while heavyweight approaches through the creation of working mobile applications are costly and time-consuming. Balancing methods through data engagement interviews and Wizard-of-Oz studies can help capture needs in various daily contexts, although in-lab study methods may suffer from recall bias [75].

With the above in mind, accommodating in-situ studies is crucial for a comprehensive understanding of databiting. Current methods such as Experience Sampling [39] and Diary Studies [17], while valuable, have limitations in not always capturing the full range of potential study data. Further developing appropriate research methods within situated environments is required. By capturing real-world usage and contextual factors, we can gain a more nuanced understanding of how individuals engage with databites in their everyday lives, further advancing the field and maximizing potential.

Second, to assess the value of databiting once it has taken place, measuring effectiveness of the exploration conducted and insight enabled by databiting is important. Capturing data as closely tied to the databiting experience can be increasingly valid and important to analyze. This further would allow us to understand the impact of databiting compared to glanceable and micro visualizations, and longer-form data exploration. Longitudinal studies can capture behavior change over time, but studying and capturing the immediate influence of databiting is not straightforward. Given the lightweight and transient nature of databiting, as well as the in-situ instances of interaction, it is essential to develop methods, tools, and scales that assess the immediate effects of databiting without further distracting users from their current activities or being compounded as a part of the databiting interaction. Ensuring that evaluation techniques are minimally intrusive will help gather accurate data on the effectiveness of databiting in real-world scenarios.

Several research questions arise from these challenges: How can immediate impacts of databiting be accurately measured in real-time? What minimally intrusive evaluation techniques can be developed for in-situ databiting interactions? One goal with respect to these research questions would be to understand these impacts without the need to analyze any captured personal health data. Given a variety of privacy concerns, and the increased difficulty in studying raw personal health data, creating methods that do not require such data would be important. Finally, alongside understanding the immediate impact of a databiting interaction, how can we assess the cognitive and emotional impact of databiting on users? Addressing these questions will help in developing robust evaluation strategies that capture the immediate and even perhaps the long-term impacts of databiting, ensuring that these systems are efficient and appropriately designed.

3.4 Summary

In this chapter, we conceptualize *databiting* as the process of extracting increasingly rich insight through lightweight and transient data exploration. Given increasingly powerful devices, coupled with the vast collection of personal data, and in-situ use, this form of exploration has the potential to complement the way people interact with and understand their personal data throughout their daily lives. By bridging the gap between glanceable and micro visualizations (short-form exploration) and heavyweight (long-form) exploration, databiting can provide complementary insight, enabling people to better access their personal data.

We have introduced the concept of databiting, provided examples and expected benefits, and delineated research considerations that remain to be undertaken. We envision that this key, yet under-explored, concept of personal informatics may soon become a reality. This thesis and its later chapters aim to push the boundaries of this beginning, making databiting on the smartwatch a complementary part of personal data exploration. We hope this chapter alone inspires research communities towards the creation of applications and tools that enable databiting. We anticipate that both exciting challenges and opportunities will arise, which in turn will shape the future of databiting and access to personal data anytime and anywhere, for the better.

Chapter 4

Eliciting In-Situ Personal Health Data Queries on the Smartwatch

---

With databiting in mind, we must first gain knowledge of what personal health data queries are even of interest to people throughout their daily lives. Surprisingly, there is little concrete knowledge about the specific queries people desire. Only once queries are known, can we begin to understand and build for them. In this chapter, we now begin to focus our efforts on the smartwatch as a mature tool that has the potential to provide and benefit from our envisioned databiting capability. This chapter resulted in a publication in the journal on Interactive, Mobile, Wearable, and Ubiquitous Technologies [157]. The article was a collaborative effort among myself, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. The collaboration took shape through combined efforts in ideation, study design and editorial work. Furthermore, Charles-Olivier Dufresne-Camaro supported efforts in the data analysis/coding process highlighted throughout this chapter. Any mention of 'we/our' in this chapter refers to my co-authors and colleague.

4.1 Introduction

Understanding user intentions for personal health data exploration and interaction *in-situ* on smartwatches remains largely unexplored. A notable study captured a dataset of desired queries from smartwatch users [154], providing valuable insights, but relied on a survey, which is prone to recall bias [75]. Current research on smartwatch personal health data spans two main domains: first, identifying and addressing the needs of smartwatch users regarding their personal health data [3, 134]; and second, developing data visualization [3, 13, 26, 68, 135, 136] and interaction [137] techniques to overcome the limitations posed by the small screen size of smartwatches. Although these studies offer significant insights, they often lack a broader empirical understanding of the specific data exploration desires of smartwatch users.

To understand, and begin to support databiting on the smartwatch, this chapter first elicits *what* exploratory queries smartwatch users desire throughout their daily lives. Additionally, we reflect on *why* the queries provided were desired. Finally, we look to better understand any implications of being in-situ (i.e., *when/where* data exploration is desired). More specifically, we focus on the following two research questions:

> **RQ1:** What personal health data queries are of interest to people for exploration on their smartwatches throughout their daily lives?
>
> **RQ2:** How does being in-situ influence the desired personal health data queries of smartwatch users?

To answer these research questions, we conducted a week-long study with Apple Watch users (N = 18), and concluded with a final interview. Throughout the week, we elicited a total of 205 desired queries from participants that would allow them to better explore and have access to their collected personal health data on their smartwatch. Using a custom built data collection application, participants reported responses throughout their daily lives, which included a natural language query and current activity information. We learned that participants not only looked to utilize their smartwatch for data exploration temporally surrounding a tracked physical activity [3], or in-situ, but also found it beneficial to use the smartwatch throughout daily activities for immediate and often discrete exploration. Furthermore, with respect to why participants looked to explore data, we highlight a new Preemptive and Proactive insight category, and find expanded Current Status as well as Contextual insight categories when compared to previous works [3, 29, 30, 108].

The key contributions of this chapter are twofold:

C1: An empirical study to capture health data queries desired for exploration on a smartwatch throughout daily life. From this study we provide a dataset of the 205 natural language queries captured through our study[4].

C2: A thorough analysis of the queries in our captured dataset, to provide a better understanding of where, when, and why queries are desired. Combining qualitative insights with quantitative metrics, we uncover patterns and trends that can inform the design and implementation of effective *databiting* on smartwatches.

4.2 Relevant Related Work

4.2.1 In-Situ Data Collection Methods

Typical in-lab data collection methods, such as interviews, surveys, and focus groups are often subject to recall bias [75]. Thus, ecologically valid methods of data collection have been utilized in research across domains. These methods include, Diary study [17, 23, 37, 51, 73, 90, 177] and Experience Sampling Method (ESM) (or the equivalent Ecological Momentary Assessment) [39, 45, 105, 190]. Both methods often utilize a pre-built survey or questionnaire for participants to respond with, but a notable difference between Diary and ESM studies is the nature in which data is meant to be captured. ESM studies utilize notifications sent to participants, where data capture is intended to be done in immediate succession to the alert. On the other hand, Diary studies allow for the collection of in-situ self-report data whenever participants have a response to report. Notifications are then used as reminder rather than a trigger to submit responses. To further reduce data capture

[4] `https://smartwatch-personal-health-data-queries.github.io/`

burdens, researchers have begun to utilize smartphones and smartwatches for successful data collection [76, 99, 202]. The always-available and body-worn nature of a smartwatch allows for easy access to the data collection tool. Furthermore, notifications on smartwatches have a high level of awareness from the participant [24, 148], have shown to allow for higher response rates, and are perceived as being less distracting during daily life [83, 150].

4.3 User Study

Our in-situ data collection spanned the course of a week, in which we employed a diary study method. As part of the study, we installed a custom data collection application on participant's smartwatches, through which they recorded a desired personal health data query using natural language (either spoken or written), as well as additional activity information. Participants provided responses throughout their daily lives, specifically when they felt it would be beneficial to access such information concerning their personal health data directly on their smartwatch. Ethics for this study was obtained from and approved by our institutional review board.

4.3.1 Participants

We recruited 18 participants (P1–P18; ten females and eight males) from Reddit. We advertised our study across subreddits relating to personal health as well as a number of general city subreddits across Canada. Our inclusion criteria were those who (1) were aged 18 years or older; (2) own an Apple Watch Series 3 or newer with watchOS 7 or higher installed, and have a paired iPhone; (3) have access to Zoom and a stable internet connection; (4) are native English speakers; (5) have no motor, visual, or speech impairments; (6) currently collect at least one of the following personal health data: sleep, nutrition, physical activity, steps, calories, women's health, and/or mindfulness data on their Apple Watch; and (7) have been regularly collecting personal health data on their Apple Watch for at least three months.

In appreciation for their time and effort, participants were offered up to $40 CAD. The amount a participant received was not tied to the number of responses, but rather the number of participation days. We provided the equivalent of $9.50 CAD for attending the introductory session, and another equivalent of $9.50 CAD for attending the final interview. During the week long data collection, we added an additional equivalent of $3 CAD for each day for which a participant provided at least one response. We provided compensation after the final interview, or upon withdrawal, in the form of an electronic Amazon.ca gift card.

4.3.2 Data Collection Method

To understand what personal health data queries lay users have, researchers have previously used focus groups, lab based experiments, and questionnaires [3, 154]. While these methods provide a positive general understanding, we aim to mitigate the potential for recall bias through concretely reported, in-the-wild, responses. Within our work, desired responses could arise at any time within a participant's daily life. Therefore, we utilized a Diary study method combined with ESM's random-interval sampling. We note, both fixed interval-based and event-triggered collection methods would have restricted responses to specific and predictable times or to certain activities respectively. A Diary study method combined with ESM's random-interval sampling allows for a wide range of boundary pushing and ecologically valid queries to be captured, during a range of times and activities, without restriction, as they are deemed beneficial to the participant.

4.3.3 Data Collection Application

We created a data collection application[5] for the Apple Watch using Swift, and deployed it via Apple's TestFlight beta program to each participant's individual Apple Watch. We chose only the Apple Watch due to the immensely simple logistics in installing our application remotely. This also ensured consistency in data collection procedures. Our application utilized Google's Firebase Realtime Database to collect and store the responses submitted. As our goal is to capture

[5] `https://github.com/reyb/Personal-Health-Query-Recorder`

queries for personal health data exploration on smartwatches, the application was designed purely for data collection; it did not answer the desired queries from our participants.

Data Collection Questions

Our application consists of up to four questions for participants to respond to. All questions were required to be answered when submitting a response. The questions and input methods were designed to support fast and easy reporting of responses while mitigating interaction difficulties on a small-screen device. This included the use of natural language reporting for open-ended questions (Q1 and Q2), leveraging the flexibility and ease in reporting ability [99], and single screen options for Q3 and Q3-1. Our application captures the following information, illustrated in a flow diagram shown in Figure 4.1 and described below:

Q1 (open-ended): What question or command do you have of your health data? This allowed us to capture the personal health data query from the participant. Q1 elicited a query regarding what participants were interested in quickly exploring or accessing on their smartwatch. Participants could either speak or type a query using the Apple Watch's built in text-entry methods. Upon recording the query, it could be reviewed on-screen in real-time and repeated to correct errors, if needed.

Q2 (open-ended): What are you currently doing? This helped us gain general contextual and daily activity information of the participant's daily life at the time of recording a desired query. For simplicity on the part of the participant, we allowed for natural language input in the same manner as the first question.

Q3 (dichotomous): Is your response related to your current activity? As a personal health query may or may not relate to the current activity being performed, this question allowed us to better understand the association between a participant's current activity and their personal health data query given in Q1. Either a "Yes" or "No" answer could be selected.

Q3-1 (multiple choice): Where are you in your activity? This would appear to the participant only if they selected "Yes" in the previous question. From this question we aimed to gain further knowledge, understanding the in-situ moment surrounding an activity, that a need for exploring personal health data arises. ”Before,” ”During,” or ”After” could be selected. In order to better understand when a desired query may be temporally related to an activity, we ensured participants understood that our definition of relation could also include just before starting and after completion of an activity. For example, a participant could be going to the gym looking to soon start their workout (Before), actively engaged in their workout (During), or heading back to the change room having just finished (After). By further exploring this time distinction, we can introduce a greater level of granularity and aim to understand when to provide exploratory capability or further insight to smartwatch users.

Figure 4.1: A flow diagram of the questions asked within our data collection application.

Reminders

To elicit many responses, our application employed two forms of reminders. First, we used push notifications. We customized these notifications to each participant, based on their self-declared wake and sleep times. Furthermore, the notifications were systematically random in that a notification would be sent at a participant's declared wake time, and then concurrently sent throughout the day between one to two hour intervals (chosen at random) after the prior notification. Notifications in this manner were repeated until the participant's declared sleep time. This method was chosen to ping participants at different times throughout each day, ideally attempting to remind them of the study at different in-situ moments within their daily lives.

Second, our application also leveraged a watch-face widget, which could be used on a range of available watch faces. All participants were asked to update their watch face to incorporate this widget. We offered a small circular widget which was continuously displayed on a participant's home watch-face and additionally provided a counter of the number of responses a participant had submitted throughout the study. Moreover, this widget allowed quick and easy access to our application, and thus recording a response, by simply tapping on the widget.

4.3.4 Study Procedure

Our study included three stages: an introduction and tutorial session, a seven-day in-the-wild data collection period, and a final interview. The procedure and study materials were iterated upon during two pilots with people who were recruited in the same means as our participants, thus meeting our study's inclusion criteria. Participants provided consent at the start of the study.

Introduction and Tutorial Session.

To start the study, the participant joined a Zoom meeting, ~45 minutes long, where we introduced and acknowledged their interest and participation in our study. Participants were encouraged and asked to interject with any comments and/or questions during the meeting. The researcher shared presentation slides (please refer to the supplemental material) via the screen sharing functionality. The goal of the project was expressed to the participant along with other important remarks before they completed a demographic survey. Then, the researcher guided the participant through the installation and setup process of the application on the participant's own Apple Watch to be used throughout the data collection stage of the study. This included setting up notifications and the watch-face widget for reminders. The researcher then gave participants a walk-through of the application through an on-screen emulator running the application on the researcher's computer and shared via screen sharing. Upon completion of the walk-through, the researcher gave participants unrelated-to-the-study practice trials to ensure the application worked smoothly and any technical issues related to data collection were appropriately handled.

Finally, the researcher gave an explanation on the potential queries and purpose of the study. The researcher asked participants to provide queries that they deem as beneficial, without worrying about current technological limitations. Queries would ideally allow participants to better explore or access their personal health data directly on their smartwatch. The researcher also asked participants to only provide a query when it both arose within their daily lives and was deemed to be a data exploration task they would like to perform on their smartwatch. The researcher explained to participants that feedback or answers to their queries would not currently be given, however they should envision receiving this directly on the smartwatch. No specific examples were provided to the participants, so as not to bias their potential queries. However, high-level categories of health data exploration (e.g., history of data, goals/performance data) were discussed to invigorate ideation.

Data Collection.

Participants used the application to submit responses over the course of the next seven days, and were instructed to wear their watch as they normally would. A response could be submitted at any time throughout the day. We did not require a minimum number of responses throughout the study, to not elicit forced responses from participants. Due to the potential for participants to be active when a query arose, we instructed them to only provide a response through the application when it was safe for them to do so. Notifications sent each day acted solely as reminder of the study.

Final Interview.

After the final day of data collection, a Zoom meeting was held where a researcher conducted a semi-structured interview with each participant. The meeting was audio recorded and later transcribed. The goals of the interview were: (1) to gain a better understanding of the smartwatch's role within each participant's health journey; (2) to explore additional information regarding in-situ smartwatch use for personal health data exploration; and (3) to discuss preferences in terms of interaction and visualization when exploring personal health data on a smartwatch. To aid in recollection, a report of each participant's queries were shown to them via Zoom's screen sharing functionality. Finally, the researcher answered any remaining questions from a participant, thanked them, and provided the compensation. Each interview took between 30 and 45 minutes.

4.4 Analysis and Results

4.4.1 Analysis

A total of 229 responses were logged through our application. First, we manually inspected the data, curating a dataset that only included valid responses. Through this process, we discarded 24 responses which fell under three categories: (1) the query had no specific element of collected personal health data (e.g., *"Am I over the food poisoning from yesterday?"*); (2) the query was related to smartwatch functionality rather than personal health data exploration (e.g., *"Is there a better way to track active minutes?", "How much battery does tracking a walk use?"*, *"Is there a way to account for temperature while working out?"*); (3) the query was action based and did not allow for exploration (e.g., *"Set my bedtime for 12 AM and wake me up by 8 AM"*, *"Start outdoor run"*, *"Record 96 ounces of water for* *the day"*). After this process, 205 valid responses remained for analysis.

We qualitatively analyzed the valid responses, containing a query, activity of the participant, and relation to this activity through an open coding process. To do this, we followed the same approach as used by Srinivasan et al. [184]. To specify, two researchers first explored the reported queries and activities for broad themes, subsequently creating a coding schema. Then, the same researchers individually coded random subsets of the data after which they came together to compare results for agreement. They refined the schema and codes, and individually coded a new random subset of data, until an 85% agreement was reached. Once the researchers came to agreement, the data was independently coded in full using the mutually-agreed upon codes, again working together to reach full consensus as needed.

4.4.2 Results

Table 7.1 summarizes the demographic, smartwatch usage, and health data collection information, as well as response counts of our study participants. Participants were aged from 18 to 56 (M = 29.8) and held a range of occupations. At the time of conducting the study, participants had collected personal health data for an average of 39.3 months (SD = 32.7 months) and had used a smartwatch for an average of 31.3 months (SD = 26.0 months). Collected personal health data by participants mainly focused in two categories, physical activity (18/18 participants) and sleep (7/18 participants) data. Of our 205 valid responses there was an average of 11.4 responses per participant (SD = 5.5; Min = 3; Max = 22).

Table 4.1: Summary of demographic information, health data collection, smartwatch usage experience, and number of responses reported from our study participants.

Alias	**Age**	**Gender**	**Occupation**	**Health Data Collection**	**Smartwatch Usage**	**Collected Data**	**Study Responses**
P1	19	F	Customer Service	1y 2m	1y 2m	Sleep, Physical Activity, Women's Health	8
P2	31	F	Care Coordinator/Dementia Counselor	10y 0m	2y 6m	Sleep, Physical Activity, Women's Health	10
P3	38	M	Communications Advisor	5y 6m	5y 6m	Sleep, Physical Activity	16
P4	35	F	Teacher	3y 0m	3y 0m	Physical Activity, Women's Health	9
P5	33	M	Information Security Specialist	5y 1m	1y 0m	Sleep, Physical Activity	9
P6	24	F	Student	1y 0m	0y 8m	Nutrition, Physical Activity	3
P7	42	M	Claim Evaluator	3y 0m	3y 0m	Physical Activity	15
P8	25	F	Scientific Evaluator	0y 3m	0y 3m	Physical Activity	6
P9	36	F	Post Doctoral Fellow	0y 8m	0y 8m	Physical Activity, Women's Health, Mindfulness	20
P10	30	M	Student	7y 0m	7y 0m	Physical Activity	12
P11	23	F	Student	2y 3m	2y 3m	Physical Activity, Mindfulness	16
P12	56	F	Retired Lawyer	1y 9m	1y 9m	Physical Activity	7
P13	22	F	Educational Assistant	6y 4m	6y 4m	Physical Activity, Mindfulness	18
P14	21	M	Student	2y 1m	2y 1m	Physical Activity	22
P15	18	F	Customer Service	0y 5m	0y 5m	Nutrition, Physical Activity, Mindfulness	11
P16	25	M	Software Engineer	6y 7m	6y 7m	Sleep, Nutrition, Physical Activity, Mindfulness	5
P17	40	M	City Planner	1y 11m	1y 11m	Sleep, Physical Activity, Mindfulness	15
P18	18	M	Student	0y 11m	0y 11m	Sleep, Physical Activity	3

We coded our valid responses, leading to the following five dimensions: (1) personal health data insight category, (2) current daily activity, (3) whether the query was related to the activity, (4) the time in activity if related, and (5) the query type. Taken together, these can enable a better understanding of the characteristics surrounding desired smartwatch data queries, combined with the types of insight that are of interest during certain daily activities. Below we detail each dimension and their codes.

Expanded Personal Health Data Insight Categories

Through the queries desired by participants, we coded the overarching insight category for which each query aligned. Table 4.2 shows the codes, descriptions, and selected queries. We expanded upon categories from previous works [3, 29, 30, 108] to assign our collected queries into the following codes: Current Status or Value (89, 42.9%), Historical or Trend (67, 32.7%), Combination or Comparison (59, 28.8%), Goals or Performance (57, 27.8%), Preemptive and Proactive (47, 22.9%), and Contextual (24, 11.7%). These codes are not mutually exclusive, and thus a query can have multiple codes (and as such the percentages above are individually calculated from the 205 total queries). For example, *"How much dancing do I need to do to burn 800 calories?" (P18)* can be seen as fitting into both the Preemptive and Proactive and the Goal and Performance categories. Of the six codes, participants reported queries in a minimum of 3 codes and maximum of 6 (M = 5.0, SD = 1.0).

Table 4.2: Query insight categories. Please note, categories are not mutually exclusive. ** denotes a new insight category found in our work; * denotes an expanded insight category as compared to [3, 29, 30, 108]. Q is a question; C is a command.

Response Category (# of Responses, % of Total, # of Participants)	Description	Example Responses [all directly quoted]
Current Status or Value * (89, 42.9%, 17)	Current, single value, metric that is collected and/or aggregated on the smartwatch to be given to the user.	(Q) Am I over or under my calorie goal at the moment? (P15)* (Q) How many calories did I burn that workout? (P4) (Q) What was my peak heart rate during my workout? (P15)* (C) Give me a report for my readiness for activity. (P16)
Historical or Trend (67, 32.7%, 18)	Previously collected metrics, prior to the current day's or activity's. Can often be used to explore changes over time.	(Q) How long on average does it take me to fall asleep? (P3) (Q) How many steps have I taken this week? (P9) (C) I would like to check a trend in my sleep in the past seven days. (P1) (C) Show me my body weight trends for this month. (P6)
Combination or Comparison (59, 28.8%, 17)	Combine and/or compare two or more different measured values. These can be done over time, between metrics, between activities, or between oneself and others.	(C) Compare my running stats with the same time last year. (P17) (C) Show me a graph of my runs both time and distance in 2021. (P11) (Q) Does my walking pace change when I walk with someone else? (P12) (Q) How many calories were burned in today's work out compared to yesterday? (P6)
Goals or Performance (57, 27.8%, 14)	Goals, such as for steps, calorie intake, calories burned, distance travelled, etc. Performance stems from completing a goal, as well as quality metrics such as fast/slow or best/worst.	(Q) How fast did I finish my 1st kilometer of my hike today? (P6) (Q) What was my best kilometer during my run? (P4) (Q) Which activity had the highest calories burned per minute? (P18) (C) Tell me when I reach a nine minute walking pace. (P12)
Preemptive and Proactive ** (47, 22.9%, 15)	Advice or information that will allow one to be preemptive or proactive in making decisions and/or to prepare for an event in the future.	(Q) How long do I need to run three times per week to achieve the November challenge? (P17) (C) Give me a suggested work out based on my readiness score. (P16) (Q) Is there a day of the week I am more likely to beat (friend) in our fitness challenge? (P12) (Q) How much dancing do I need to do to burn 800 calories? (P18)
Contextual * (24, 11.7%, 9)	Impact and/or affect of an external or collected metric on another.	(Q) How is the air quality affecting my walk? (P16) (Q) Is my cycle affecting my sleep? (P2) (Q) Is my running pace slower in the days following a strength training workout? (P13)* (Q) What is the impact of my sleep in my running? (P17)*

Current Status or Value insight was often about more than just the simple metrics captured. Amini et al. [3] found that participants explored step count, distance, calories, pace, speed, and heart rate during an activity. While this remains to be seen in our study, we found a broader definition of current metrics desired; these included heart rate zones, total values from activities throughout the day, peak values or fluctuations of metrics throughout the activity, and aggregated values such as perceived exertion (see Table 4.2 for specific queries). These examples, which can be seen as increasingly unique-to-user, should be considered within smartwatch health applications to expand the usefulness and benefit to a broader range of smartwatch users.

Contextual exploration was utilized by participants to find cause and effect between a range of data. Choe et al. [30] discussed participants' interest to include and explore the effect that external data such as time of day, location, or weather can have on their own collected personal health data. While our captured queries from participants garnered similar contextual information, we also note that participants were inclined to look for cause and effect using their own collected data as context. For example, *"Is my cycle affecting my sleep?" (P2)*, *"Does my walking pace change when I walk with someone else?" (P12)*, *"What is the impact of my sleep in my running?" (P17)*, and *"Does weightlifting focusing on different muscle groups affect my heart rate?" (P14)*. As participants used the smartwatch to capture a range of health and activity data either automatically or through a discrete input, they desired to explore context surrounding these captured instances on the smartwatch.

Preemptive and Proactive queries, a new form of insight brought forward from our study compared to previous works [3, 29, 30, 108], make up ~20% of our collected data from 15 of 18 participants (see Table 4.2 for specific queries). Smartwatches influence decision-making through on-screen metrics related to activities [3, 108]; for example, a runner glancing at their smartwatch can adjust their pace if it's below their desired value. Our participants, however, were looking for a wider range of influential exploration from their smartwatch, such as to help choose a workout for the day, plan an activity based on goals, or to pick up on elements that they alone may not be able to predict. This form of insight was seen as a means for preparing oneself for a future event rather than simply reflecting on current or past metrics.

Participants were looking to utilize these Preemptive and Proactive insights from their smartwatch to influence immediate and in-situ decision making, as well as for some daily and longer term planning. This longer term planning was often seen as tied to the Goals or Performance category, as 23/47 Preemptive and Proactive queries were also to gain insight on how to achieve immediate or future goals. Finally, while about a half of the queries in this category were temporally related to an activity (i.e., immediate influence), we further note that Preemptive and Proactive queries were desired on the smartwatch throughout the day (i.e., longer-term planning).

Smartwatch Personal Health Data Queries Desired throughout Daily Activities

Analyzing the daily activities of participants led to seven codes. These codes categorize the current activity being performed in a participant's daily life, during which their desire for exploration on the smartwatch arose. These codes include: Physical Activity, Self Care, Work, Leisure, Sleep, Transportation, and Other; the codes, their descriptions, and counts can be seen in Table 4.3. Codes were mutually exclusive, and thus each response was given a single code. Of the seven codes, participants reported responses in a minimum of 2 codes and maximum of 6 (M = 4.1, SD = 1.4).

Table 4.3: Summary of the daily activities participants were undertaking at the time of a response, and the relation of the reported queries to these activities: **not related**, **before**, **during**, **after**.

Daily Activity	Description & Examples	# of Responses (% of Total), # of Participants	Query Related to Daily Activity
Physical Activity	Body movement that requires more energy than resting. (e.g., dancing, walking, running, weight training, yoga, sports, etc.)	70 (34.1%), 16	
Self Care	Activities that pertain to normal day-to-day human function. (e.g., cooking, eating, chores, morning ready routine, etc.)	44 (21.5%), 12	
Work	Fulfilling duties either for job or school. (e.g., job based tasks, studying, attending meetings, etc.)	29 (14.1%), 12	
Leisure	Activities performed for relaxation and fun. (e.g., reading, watching TV, lounging, etc.)	25 (12.2%), 14	
Sleep	The absence of wakefulness. (e.g., napping or nightly rest)	17 (8.3%), 9	
Transportation	Moving from point A to point B. (e.g., driving, taking the bust/metro, taxi, etc.)	10 (4.9%), 5	
Other	Activities that do not fit within the prior categories. (e.g., within the study interview, due to mHealth app notifications)	10 (4.9%), 5	

We asked participants whether their personal health data query was related to their current activity as reported above (Q3), and thus in-situ. This was either a Yes (107, 52.2%) or No (98, 47.8%) answer; examples from our captured data include *"Leaving the gym" - "Show me my heart rate chart from today's gym session"* as being related while *"At work" - "What was my fastest kilometer in my run?"* as not. We see an almost equal distribution overall, however, when combined with the activity we see a distinction. The most drastic example of this distinction is in relation to Physical Activity. Here, 97% of queries reported were related to the Physical Activity being done. In contrast, all other daily activities, aside from Sleep, provided time within daily life for increased insight and reflection that was unrelated to the participant's current activity. Within our captured responses, 16 of 18 participants found it beneficial to report a query both related and unrelated to the current activity they were doing.

When queries were related to the daily activity being done by the participant, we additionally asked whether they were just Before (24, 22.4%), During (38, 35.6%), or After (45, 42.0%) the activity. The results collected follow closely with our demographic survey which asked when participants aim to explore their data on their smartwatch, Before (4/18 participants), During (10/18 participants), and After (15/18 participants).

Participants noted during the interviews that exploration Before activities could only do so much to affect the activity collecting the health data once started. In relation to our insight categories, broader Preemptive and Proactive as well as general Goal insight was most often queried Before an activity. This was seen as a means to help prepare for the activity at hand, rather than exploring current or past data.

During an activity, specifically for Physical Activity, expanded insight beyond simple metrics pulled participants away from being in the moment and focused on the activity at hand, thus was not as desired. P5 mentioned *"It wasn't much [exploration] like during the workout 'cause in the workout I found that is more like just concentrating on whatever I was doing and I didn't really have any questions to ask."* This is reflected in the Current Status or Value insight category being the most queried During an activity. In fact, of the queries that took place surrounding Physical Activity, 26.5% were During the activity, or in-situ, of which 68% of these were simply to understand a Current Status or Value.

Finally, exploration after an activity allowed for immediate reflection to take place which could help influence future activities. Interestingly, participants most often looked to perform data exploration surrounding Physical Activity after the completion of and regarding the activity itself (32/66 queries). This reflection After an activity was mainly of the Current Status or Value, Historical or Trend, and Combination or Comparison insight categories.

From the interviews held with participants, we note that these captured results may not provide the entire picture. First, while our results provide a general understanding of what queries and when these queries are desired, we note that across combined insight categories, daily activity, and relation to the activity, our results show queries reported in 128 unique combinations of these. This highlights the deeply unique and personal aspect to personal health data exploration needs. Second, two participants suggested that while they may utilize different in-situ moments within an activity to explore their data, the activity itself was not always the determining factor when they aim to explore their data on the smartwatch. Notably, current overarching goals set by a participant and how long they had used a smartwatch for tracking data could affect the type and time of their exploration. As P7 discussed, *"Because I've been using it [the smartwatch] for like a couple years, uhm, I think I'm pretty good at like knowing what kind of workout will make me hit my calorie burn goal or get the steps I need or those kinds of things."*

4.4.3 Natural Language for Personal Health Data Exploration by Lay Individuals

Participants had first-hand experience using natural language throughout our in-situ study. This included being in different environments and surrounded by others. We note that all participants utilized speech to record their queries, with two participants typing a subset. This aspect allowed us to discuss during the interview participants' thoughts towards the use of natural language.

Positive Reaction Towards the Use of Speech Based Smartwatch Interfaces

Participants were overall excited for the potential to use speech to explore their personal health data. Sixteen (16) participants suggested they would use a speech-based interface for data exploration on smartwatches if it was available. P13 stated *"I'd prefer it because it's quicker for me. You know, like if I'm in the middle of doing something, it's easy just for me to say it and then move on."* with P16 adding *"the responses get more tailored, like if you ask it and it like answers perfectly, it just makes it even better to use."* P12 discussed how natural language could even benefit the in-situ nature of smartwatch use during walking, *"It works really really well while walking, ... if you have your earbuds in or like if you lift the watch to your mouth you can get a decent signal without much interruption and you don't have to talk super loudly."* These discussions suggested that speech could be a great tool for smartwatch interaction, which has an already limited interaction space and becomes even more limited while mobile [176].

Reduced Privacy Concerns for Speech-Based Personal Data Exploration

Prior research with a focus on natural language and speech as an input modality discusses limitations regarding privacy and social acceptability [55, 98, 116]. This includes disrupting others and the awkwardness of talking to a device. Through our discussion with participants, we found that half (9) had limited to no concerns regarding the use of natural language for personal health data exploration. Natural language was mainly seen by participants as beneficial and acceptable for two reasons. First, natural language is becoming more commonplace, through improved interfaces and the broader use of Bluetooth headphones/earbuds. Second, the data exploration that was often desired was not seen as overly private, personal, or specific to a user; thus, participants were open to speaking the questions and commands around others if needed. P15 summarized this by saying, *"This kind of stuff [the responses provided] is more objective data, which I guess to me, is not as private as like thoughts and opinions, that's more subjective. That I'd be concerned about someone thinking negatively versus like if they see that oh she didn't get out of bed today."* We, however, note that for half of our participants who were concerned with using natural language for personal health data exploration, their queries were seen as being increasingly personal to them, and thus were uncomfortable in public settings. Yet, while this still hinders the potential use of natural language interfaces for some, we remain to see an overall difference in attitude compared to the prior mentioned works.

Additionally, as personal health data exploration has the potential to become increasingly personal, applications will likely become more aware of the unique exploratory needs of users. Previous research has suggested that privacy concerns do not determine whether people use mHealth applications or not [16]. In relation to this, all our participants (18) suggested that, due to the increase in perceived benefit, as long as the application came from a trusted source and had the necessary measures in place [205], they had no problems providing these types of exploratory questions or commands. This follows a privacy pragmatist approach which suggests a person may have strong feelings about privacy, yet they are willing to allow access to their information for their own benefit [87], and is often seen within younger populations [102].

4.5 Discussion and Future Work

4.5.1 In-Situ and Non-In-Situ Preparation-for-Action

Reflecting on personal data using a smartwatch can occur increasingly close to the action, for which the reflection is related, to benefit on-the-fly decisions [3, 69, 107]. In fact, when reflection and action are related, Ploderer et al. [149] suggest there exists reflection-in-action (i.e., real time and Current Status or Value insight) and reflection-on-action (i.e., aggregation of data such as Historical or Trend, or Contextual insight) enabling both maintenance and discovery, respectively [108]. Rather than reflection-on-action only taking place in-situ and immediately after an activity, our results also showcase a need for discovery and reflection while away from an action for which the query would be related. We further postulate that smartwatch users are looking for additional discovery through *preparation-for-action*. This form of exploration was deemed as beneficial by our participants most often for Preemptive and Proactive as well as goal and performance insight. This preparation-for-action not only happens in-situ, immediately prior to an action, but also away from the action for which the reflection was related. Many research works focus on the in-situ exploratory capabilities of the smartwatch [3, 86, 99, 163], however expanding on the smartwatch's capabilities during non-in-situ and prior to in-situ usage scenarios could be critical for further adoption, continued use, and a range of unrealized benefits.

4.5.2 Query Insight Category Dependent on In-Situ Activity

While coding our data, we recognized that identical queries could lead to different insight categories. For these, the insight category of the query was highly dependent on the daily activity being performed and its relation to the query (i.e., whether the query was being reported in-situ or not). As an example from our dataset, *"What is my average walking pace?" (P3)* can imply and elicit different meaning depending on when it is asked. For instance, if the query was asked in-situ while during the middle of a walk, the answer could likely be seen as the average walking pace of only the current walk (Current Status). However, if the query was asked while sitting down at work the answer may require the calculation of the average walking pace across all walks recorded (Historical or Trend). Thus, utilizing contextual, *when* and *where*, information available from the smartwatch's sensors as well as user-initiated activities can at times become a key component in understanding a lay person's personal health data query and information needs on a smartwatch. This can then be crucial in regards to formulating appropriate responses to allow for lightweight and transient data exploration that is beneficial to smartwatch users.

4.5.3 Limitations

We recognize that our strict inclusion criteria resulted in the exclusion of individuals with impairments. Thus, our dataset is not fully representative of all who utilize a smartwatch for personal health data collection, exploration, and health monitoring. While our work largely provides a general understanding of personal health data queries on smartwatches, we suggest these aforementioned user groups should be studied in their own regard.

Due to the early nature of our work, we chose to not provide feedback to our participants (i.e., the answers and data representations in response to queries). This allowed us to capture a broad range of desired smartwatch data queries without influencing further potential responses. However, this design does limit us in understanding any forms of continual and serendipitous exploration on the smartwatch, such as asking a single followup question based on feedback given. During typical data exploration, one explicit query is often not enough to fully represent what is interesting to a person. Feedback given in response to a query can often lead to subsequent and unanticipated queries [129]. Moreover, this does not allow for discovery of new and interesting information beyond what is asked. While the smartwatch may not allow for lengthy data exploration, the limit to which the smartwatch can enable follow-ups, the relationship between the smartwatch and other data exploratory tools (i.e., smartphone, tablet, desktop), and the impact that given feedback has on smartwatch personal health data exploration should be further studied.

Finally, as with any elicitation study there are additional limitations to note. First, we are likely not able to capture all usage scenarios participants may experience, and for which a desire for smartwatch data exploration could arise. Examples of such times include a person preparing for a marathon or someone on vacation. Second, participants often do not realize a device's full potential limiting the range of responses submitted. For example, blood glucose monitoring has a potential future within smartwatches, yet was not a component in any queries reported by our participants. We aimed to mitigate both of these limitations through recruiting eighteen participants, each running the study for one full week (including weekdays and a weekend), and expressed for participants to not worry about current technological limitations instead focusing on queries that they desired to be possible. As such, we believe that our results remain to provide a broad understanding and new insight into queries desired, which in future can be translated to a range of usage scenarios and newly captured data.

4.6 Summary

This chapter offers an empirical understanding of *what* personal health data queries smartwatch users wish to explore throughout their daily lives. Through an in-situ diary study with 18 participants over a week, we captured queries that have the potential to facilitate exploration on a smartwatch in various daily contexts. Through the results of this data collection and a final interview with participants, we offer the elicited queries through a public dataset (i.e., *what*), report on query insight categories (i.e., *why*), and query relation to daily activities (i.e., *when* and *where*). Participants reported a desire to utilize the smartwatch for momentary and immediate personal health data exploration, not only during in-situ moments but also across a range of daily activities. We suggest several key implications for the design of smartwatch mHealth applications; including supporting preemptive and proactive exploration, expanding upon current status and contextual exploration, allowing for exploration away from an in-situ tracked activity, and the offering of natural language interaction.

Chapter 5

Towards Natural Language Interaction for Personal Health Data Queries on Smartwatches

---

With the knowledge of what queries are desired, we can now begin to focus on how to enable interaction for such queries. This chapter, and the next, centers on the following broad research question: How can we expand the expressivity of data exploration on the smartwatch, through both multi-modal input and output modalities, within the confines of the smartwatch's capabilities and data exploration needs? This chapter specifically highlights necessary components to consider when allowing for the handling of the given queries. Work presented in this chapter resulted in a publication in the conference on Mobile Human-Computer Interaction [156]. The article was a collaborative effort among myself, Charles-Olivier Dufresne-Camaro, and Pourang Irani. Charles-Olivier Dufresne-Camaro conducted the coding process alongside myself and provided editorial support. Any mention of 'we/our' in this chapter refers to my co-authors.

5.1 Introduction

Smartwatches, through advancing input modalities such as touch, speech, gesturing, and buttons/dials, have the potential to enable broader interaction with the collected data. Fundamentally, however, our lack of knowledge surrounding the interactive requirements for personal health data queries hinders progress. More specifically, in this chapter, we focus on the following research question:

> **RQ1:** Through analysis of the personal health data queries collected, what components comprise a personal health data query desired for exploration on the smartwatch and provided by lay users?

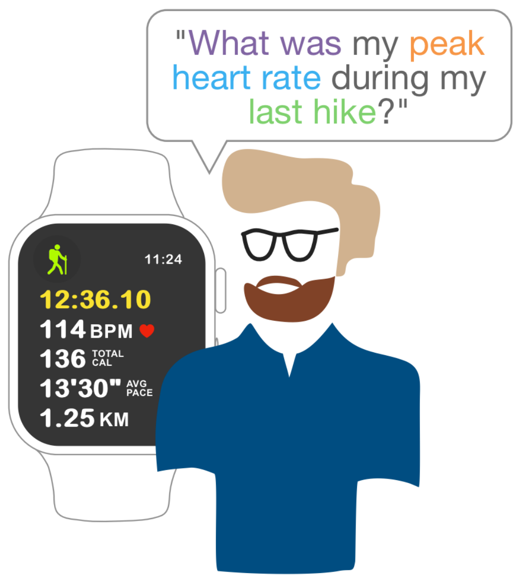


Figure 5.1: Components within a personal health data query explored within this chapter. We highlight various interrogatives (e.g., ”What was”), aggregations (e.g., ”peak”), data attributes (e.g., ”heart rate”), and filters (e.g., ”last hike”) to be considered when processing a query.

In this chapter, we focus on understanding and characterizing components of personal health data queries desired for exploration on the smartwatch. Throughout, our motivation lies in increasing the capability for data exploration directly on the smartwatch. Yet, results conveyed in this chapter can be seen as being beneficial for a range of devices. To achieve our research goal, we extend and compare to work previously done [154], through the analysis of our previously captured dataset. We explore various dimensions, including attributes of data requested, aggregation methods, mechanisms for filtering, and interrogatives used within the queries; see Figure 5.1 for an example of these dimensions combined within a single query. By analyzing across these dimensions, we provide a better understanding of *how* people want to explore and access their personal health data on smartwatches. In turn, the results shared can influence interaction in applications catering to smartwatch data exploration.

Our contribution is as follows:

> **C1:** Identification of key dimensions in personal health data queries, such as interrogatives, data sources, aggregations, and filtering mechanisms. These insights provide a foundation for enhancing the natural language processing capabilities on a range of smartwatches, key to the development of part-of-speech tagging capabilities.

5.2 Relevant Related Work

5.2.1 Natural Language Query Analysis

In this section, we highlight work on natural language query analysis as these often must characterize and dismantle queries for understanding and processing. While we analyze a natural language query dataset in this work, we highlight that we do not propose natural language as the only means of interaction with personal health data on the smartwatch. Instead, we utilize works in this area to simply gain a better understanding of the components involved when querying data and to situate our analysis among related work.

With the increasing capability for devices to perform natural language processing (NLP), even on a smartwatch, many toolkits have become available for use [114, 124]. These toolkits help perform common NLP tasks such as language identification, tokenization, sentiment analysis, named entity recognition, and part-of-speech tagging. However, these toolkits do not explicitly focus on parsing health related information, nor do they offer solution if we do not know what to look for when performing tokenization and part-of-speech tagging. As such, we must first gain a general understanding the components used within potential personal health related queries before using these tools.

Natural language interfaces (NLIs) have become increasingly popular for general interaction (e.g., Siri, Alexa, and Google Assistant have become more pervasive in daily life) [21, 122] and for visual data exploration [98, 183]. While mainly focused towards data experts, research has collected and explored natural language queries across multiple explicit and implicit dimensions. These include data attributes, chart types, data encodings, aggregations, design references, question words, and verb tenses [65, 154, 169, 184, 187]. While not all dimensions are necessary for personal health data querying on the smartwatch (i.e., some of the prior works focus on specific applications such as for visualization creation), these provide insight into required components which we can then code for and subsequently analyze to handle and process a data-driven query.

5.3 Dataset and Analysis

5.3.1 Dataset

We use the previously collected dataset from Chapter 4 in our analysis. Within the dataset, our analysis focuses on the queries themselves while also incorporating other elements of the dataset for granularity (i.e., the relation of the query to a current activity).

5.3.2 Coding Procedure

Analyzing the dataset, we coded components of the queries. To ensure coding accuracy, we followed a procedure used within related work [184] and similar to Chapter 4. Specifically, our procedure was as follows: Two researchers first explored the queries independently, creating a coding schema that would outline the potential dimensions that could be assessed. After discussion, dimensions were chosen for which to analyze and code. As a team, a code book was created. Then, the same researchers individually coded a random subset (10% of all queries available) of the data. The assigned codes were then compared for agreement. Until 85% agreement was reached, the code book was refined and a new subset of data was individually coded. Once agreement was obtained, the entire dataset was independently coded using the finalized and agreed upon code book. Finally, any remaining code disagreements were discussed and resolved until a full consensus was reached and a single code was assigned for each dimension explored.

The dimensions explored within the coding schema, and codes used within, followed closely with prior work exploring natural language query interfaces [154, 184]. Hence, we analyze the data type requested, data attributes, filtering mechanisms, and the interrogatives used.

Throughout, explicit and implicit/semantic codes are utilized to describe aspects of the data. As the dataset was captured through lay-users, we allow synonyms when using the explicit code rather than deferring these to another code. For example, *peak* can be seen as an explicit aggregation for *max*. In contrast, *how many* is an implicit/semantic aggregation of *count*. These will be highlighted further in their respective subsections below. Lastly, when reporting codes, we provide round brackets containing the count of queries the code captures and the percentage of the dataset this represents.

5.4 Results

Exploration through Commands; Information Immediacy through Questions

Guided by the definitions and codes created by Srinivasan et al. [183] for natural language data exploration, we found that participants' queries were framed either as a Question (173, 84.4%) or Command (32,[6] 15.6%).

The majority of Commands, provided by over a half of participants (11/18), were of the Historical or Trend as well as Combination or Comparison insight cat-

[6]We categorized two instances of queries from P18, "Summary of my sleep cycles," as an (implicit) Command.

egories (21/30). Examples include *“Show me a graph comparing my caloric intake over the last week” (P7)*, *“Show me a graph of my runs both time and distance in 2021” (P11)*, *“Give me a report for my readiness for activity” (P16)*, and *“Compare cycle data from today to the same day in my last cycle” (P2)*. Through these Commands, we can see that the desired outcome of the participant is not explicitly clear (i.e., while we can try to provide an optimal visualization, there is not a discrete answer that can be given). Additionally, these Commands often suggested that participants had the intention to further explore or view a larger range of collected health data, often through on-screen visual representation directly on the smartwatch.

Conversely, a Question was often much more direct and closed-ended, with the intended result of the insight seemingly known to the participant. All participants provided queries in the form of a Question. Examples include *“Have I stood up this hour?” (P17)*, *“How long on average does it take me to fall asleep?” (P3)*, *“How many hours did I sit yesterday?” (P9)*, *“How many steps did I get during that 2 kilometer walk?” (P4)*, and *“What was my calories burned in the last 30 minutes?” (P3)*.

These Questions benefit the type and length of interaction that is typically undertaken by smartwatch users [148, 192], as they can allow for direct feedback. P14 discussed during the interview, *“I think the question sort of implies immediacy [...] and I think it’s the immediacy that the watch would be nice if it covered.”* This finding provides us with valuable information surrounding the intent and perceived use of the smartwatch for personal health data exploration, especially while in-situ. Often times, throughout one’s day, discrete and immediate insight is valuable. This insight, while potentially leading to further exploration, does not require it, and thus the use of a smartphone or desktop application is not immediately needed.

5.4.1 Attributes of Requested Data

The data requested can be organized into categories, differentiated through activity and the data that is collected. Figure 5.2 (left) highlights the eight codes used to quantify data types of interest. Not surprisingly, as the smartwatch is generally

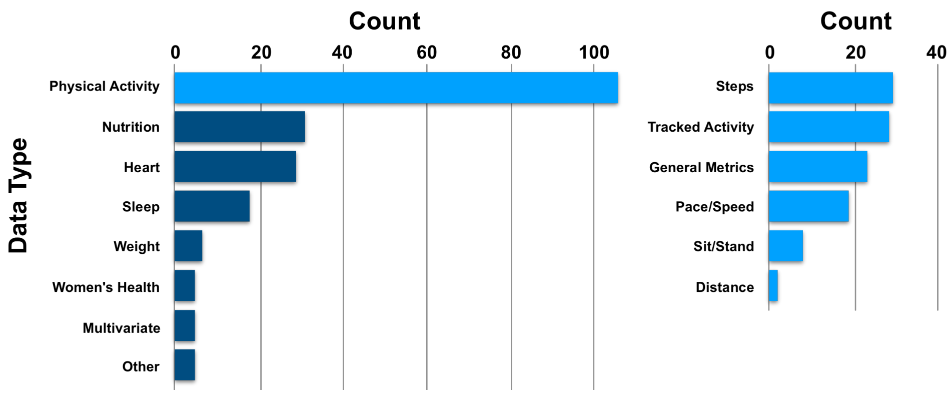


Figure 5.2: Left: Counts of the data types the queries were focused towards. Right: Sub-codes within the Physical Activity code.

used as a fitness and physical activity tracker, physical activity data constituted the majority of queries desired (105, 51%). This includes data such as steps, general tracked activities (e.g., "Show me a history of all my dance workouts." or "How many times have I worked out this week?"), and general metrics (e.g., "Did I close all my rings today?" or "How many active minutes am I at?"); see Figure 5.2 (right) for a complete breakdown.

Furthermore, heart data (e.g., heart rate, ECG) (29, 14%) and sleep data (e.g., sleep time, wake time, duration, sleep stages) (18, 9%) were of interest. For granularity, we chose to code heart data separate from physical activity and sleep as this data is captured across and independent of both of the former mentioned codes. Not surprisingly, data that is currently tracked automatically on the smartwatch (i.e., physical activity, heart, and sleep data) were the most queried. In contrast, other data types such as weight, women's health, and nutrition (e.g., number of calories or meals eaten, water intake) which require discrete input or are not currently supported, were less queried. As more data becomes automatically sensed and calculated for tracking, these can be expected to also have relatively higher interest for data exploration.

Attribute references can be seen as words within a query that correspond to a data attribute or specific data point within the collected available data. These references were either (i) explicit, where the reference in the query was specific to a data point being captured (e.g., “What is my current *heart rate*?” and “How many *steps* did I get during that 2 km walk?”) or (ii) implicit, where the reference to data within the query was too broad, could hold different meaning for different people, or required collection of multiple data points (e.g., “Compare my *running stats* from the same time last year” and “Is my *work out better* at my home gym or commercial gym?”).

From the 205 total queries in the dataset, we find a large majority, 80% (164) of queries, utilize explicit references to data attributes. Only 20% (41) of the queries showcase implicit referencing. Implicit queries were less immediately data-driven and often contained broad interest into a topic and the need to aggregate data from multiple sources (e.g. “What’s the best exercise for me today?”). Queries with implicit referencing of attributes may also present challenges with regard to providing appropriate response. No clear difference in the use of explicit or implicit attribute referencing was seen for queries registered at different times within activity or when away from activity.

5.4.2 Aggregations

Aggregation references include words that would enable the conducting of a mathematical transform on the data. This is common when performing data analysis (e.g., obtaining the sum, count, average, etc.). Our exploration of aggregations was first coded into the type of aggregation requested; see Figure 5.3. We found five forms of aggregation and a sixth non-aggregation form: (i) Count (59, 29%) (e.g., “*How many* runs have I completed thus far in 2021?”), (ii) Average (29, 14%) (e.g., “What is my *average* step count per day”), (iii) Min/Max (10, 5%) (e.g., “What was my *fastest* kilometer in my run?”). Other synonyms include: slowest, highest, lowest, peak, best, and worst. (iv) Total (4, 2%) (e.g., “*How many* miles have I *accumulated* through walking, running, and biking over the course of this year”?), (v) Variance (3, 2%) (e.g., “How much has my pace *fluctuated* during my walk”), and (vi) N/A and Current value (97, 48%), where no aggregation is necessary and a value is simply being requested (e.g., “*What is* my resting heart

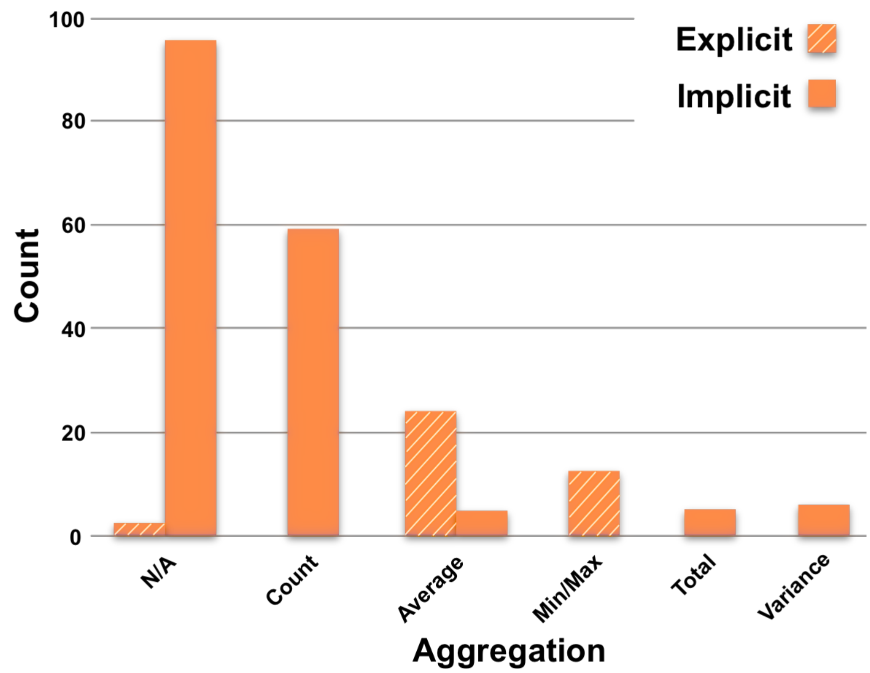


Figure 5.3: Counts of the aggregations (mathematical transforms of the requested data) found within the queries. The aggregations are further noted by explicit and implicit references.

rate?").

We further explored aggregation references through either (i) explicit aggregation, where direct reference to an aggregation transform was used, (37, 18%) (e.g., "What's my *average* walking pace per kilometer" → Average) and (ii) implicit aggregation, when phrasing was used, (168, 82%) (e.g, "How long does it take after a walk to get back to resting heart rate?" → Average and "*How many* calories did I burn in the last 4 hours?" → Count). Notably, when we explore the aggregations through this lens, the vast majority of aggregations are performed utilizing implicit requests. When looking at majority, only for average and min/max did people utilize explicit referencing more than implicit referencing.

5.4.3 Interrogatives

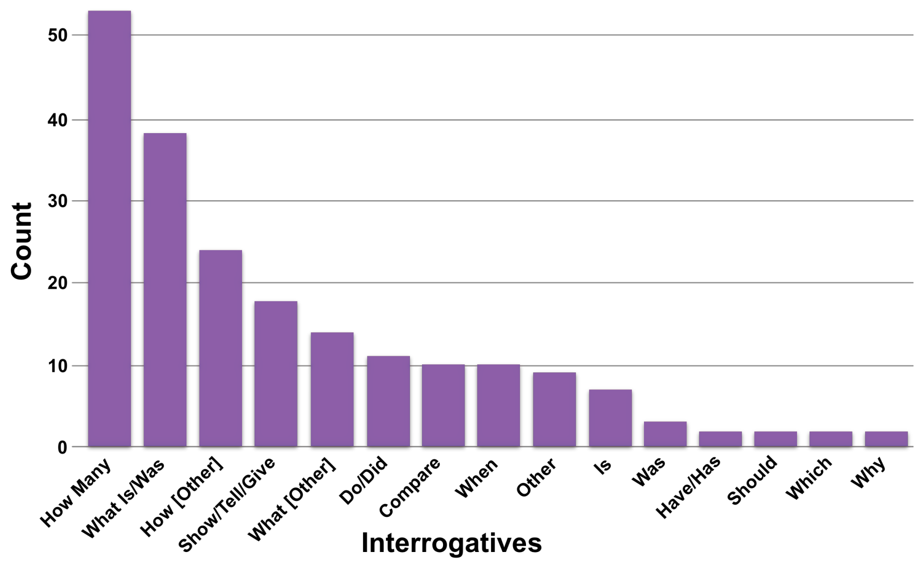


Figure 5.4: Counts of the interrogatives (elements used to express questions/commands and intent) found within the queries.

Interrogatives, or question words, can provide insight into the aggregation desired, indicate questions versus commands, underscore query intent, and hint at appropriate forms of output (e.g., *show me* compared to *tell me*). Figure 5.4 shows the interrogatives coded. Of interest, over 60% of queries contain either "how" (77, 38%) or "what" (52, 25%) question interrogatives. These can be further broken down into the interrogatives "how many" (53/77), often implying a count aggregation, and "what is/was" (38/52), often implying a single value calculation.

5.4.4 Filtering Mechanisms

Filtering of data is a common exploratory task (e.g., "Compare walking pace *September and October*" filters the data to the months of September and October while excluding other data). Four codes were used to differentiate filtering mechanisms used within the queries: (i) N/A (42, 20%), where no filtering was needed as the entire data related to the query would be used, (ii) Current (34, 17%) where the

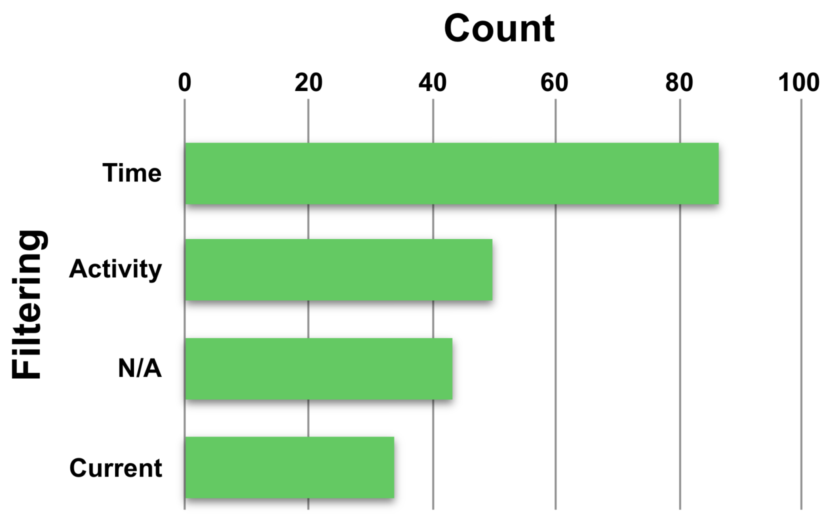


Figure 5.5: Counts of the filtering mechanisms found within the elicited queries (elements use to filter a subset of all data collected).

current or most recent value would be filtered removing data collected in the past, (iii) Time dependent (85, 41%), where a notion of time was used to filter data (e.g., "What was my highest heart rate in the *last hour*?"), and (iv) Activity dependent (49, 24%), where an instance of an activity is used to filter data rather than an explicit notion of time (e.g., "What was my best kilometer *during my run*?", "Show me my heart rate chart from *today's gym session*.", or "Was the *1st km of my hike* faster than the *last kilometer* today?").

Notably, activity dependent filtering is a subset of time dependent filtering, however is referenced with respect to an activity or activities rather than the specific time period. The smartwatch is inherently a device that captures instances of activity (e.g., tracking nightly sleep, individual runs or walks, when food/water is consumed, when weighing yourself using a connected scale). Queries containing activity-dependent filtering allow filtering to be accomplished naturally, by referencing when the smartwatch was used to track data, without the need for a user to remember or further explore specific times an event or activity occurred. Interestingly, the number of queries containing activity dependent filtering shows increase after an activity. This even slightly eclipses time dependent filtering immediately after an activity.

Table Table 5.1 provides a complete breakdown of the filtering, aggregations, and data attributes; this table provides these relation to each other along with the in-situ association as highlighted in Chapter 4.

Table 5.1: Counts of the codes for the attributes, aggregations, and filters are shown, delineated by the in-situ relation for which the query was elicited. We note, that 5 queries contained both time and activity dependent filtering, as such the total count for the filtering columns results in 210.

		Attributes		Aggregation		Filtering			
		Explicit	Implicit	Explicit	Implicit	N/A	Current	Time	Activity
Yes (107)	**Before (24)**	20	4	2	22	13	1	6	4
	During (38)	32	6	7	31	9	8	14	7
	After (45)	36	9	13	32	2	6	16	20
No (98)		76	22	15	83	18	19	44	13
Totals		164	41	37	168	42	34	85	49

5.4.5 Components of a Personal Health Data Query

Looking at the results together, four components can be seen to make up a personal health data query on the smartwatch. These include the interrogative or question word(s), the data subject or attribute(s), the aggregation term, and the filtering mechanism. Importantly, however, not all are needed when querying data. At a minimum, all queries in the dataset contain an interrogative and data subject. This coincides with the results found by Rawassizadeh et al. [154]. Thus, some queries do not contain aggregation or filtering terms, often implying exploration of a current value or of all data captured (e.g., "What is my resting heart rate?"). This is important for interaction of data on the smartwatch, as no matter the interaction modality considered, we must obtain this information at a minimum. Furthermore, we restate that the interrogative can be used as the aggregation term. As such, this is not always explicitly required (e.g., "How many (→ count) steps did I take in the past seven hours?" versus "What was my peak heart rate during my workout?").

5.4.6 Response Expectations

The query provided can in fact also hint at the expected response type. As such, to better understand the potential forms of output required, we further coded the appropriate and expected response type for each of the queries in the dataset.

Throughout, we found four possible response types: 1) **Value** (115, 56.1%), a single value response (e.g., What was my average heart rate in the last hour? - would elicit a single average heart rate to be reported filtered by the condition provided), 2) **Open** (41, 20.0%), a response requiring a list of data/information of any size (e.g., In what workouts does my heart rate reach zone 5? - would elicit a list of workouts for which the provided condition is true), 3) **Binary** (32, 15.6%), a yes/no response (e.g., Is my heart rate higher than normal? - would elicit at its lowest level a simple yes/no response), and 4) **Range** (17, 8.3%), comparative two-value responses (e.g., What is my average heart rate on weekdays compared to weekends? - would elicit the output of two unique average heart rates filtered by the provided conditions).

5.5 Discussion

5.5.1 Lay-Person Exploration of Personal Health Data

Throughout, we target a broader audience who are interested in exploring and gaining insight from their personal health data. Our analysis is one of the few to focus on lay-person queries. The majority of work in this area has focused on individuals who are familiar with data exploration practices and tools (e.g., data analysts) [94, 183, 184], with to our knowledge only one work analyzing lay-person queries for smartwatch data exploration [154].

Within our findings, the influence of lay-person exploration was most evident through implicit aggregation and activity dependent filtering. The majority of queries within the dataset analyzed contained implicit aggregation, a vast difference from the only quarter of queries captured with data experts [184]. While the work done by Rawassizadeh et al. [154] highlights both explicit and implicit notions of time and location, this dimension is not quantified for comparison nor discussed for aggregation. Furthermore, prior work has put focus on time dependent filtering as this is a primary dimension of personal health data [98]; Rawassizadeh et al. also suggest time as the only filtering mechanism [154]. However, activity dependent filtering found in our analysis can be seen as a means for lay-users to more easily recall events performed rather than specific times they were performed. Both these implicit means of exploring data are understandably easier for people, especially on the smartwatch where visual exploration is limited due to the small-screen size and focus may be on a primary in-situ task at hand [14].

5.6 Summary

This chapter provides an understanding of the different components that comprise the queries desired for personal health data exploration on the smartwatch. We identified several dimensions related to these queries, including the requested data types, attributes, aggregation methods, filtering mechanisms, and interrogatives used. Our findings emphasize the need for more comprehensive and lay-user access to a range of differing data aggregation and filtering options. By considering these insights, practitioners and application designers can develop better applications that meet users' specific needs and preferences for interaction with personal health data. The implications of this chapter provide valuable guidelines for future works aimed at enhancing the utilization of smartwatches as effective personal health tracking and exploration devices, including our own work within Chapter 7.

Chapter 6

User Preferences of Voice Assistant Answers to Personal Health Data Queries

---

Continuing our focus on enabling interaction for such queries, this chapter shifts attention to the complement of input: output. Specifically, as there is plenty of smartwatch visualization research conducted and guidelines proposed, we explore how to respond to personal health data queries through spoken, voice assistant, output. Spoken output can provide an efficient and complementary means to respond to personal health data queries, especially while on-the-go when the visual system can be overloaded [19]. Work presented in this chapter resulted in a publication in the conference on Conversational User Interfaces [159]. The article was a collaborative effort among myself, Yumiko Sakamoto, Jaisie Sin, and Pourang Irani. Yumiko Sakamoto collaborated on study design, data analysis, and editorial work. Jaisie Sin collaborated on study design and editorial work. Any mention of 'we/our' in this chapter refers to my co-authors.

6.1 Introduction

The rapid integration of voice assistants (VAs) within an array of devices is transforming the way we interact with technology. VAs now offer seamless interaction across standalone smart-speakers, smartphones, and wearables such as the smartwatch. The current capabilities of VAs allow for the handling of general commands (e.g., controlling entertainment devices) and answering of basic questions (e.g., the current weather).

Now, pushing the boundaries of what a VA is capable of, we refer to the storied example provided in the introduction of this thesis. As Sam asks, "How does today's pace compare to my last six hikes?" - the VA could respond with "Your pace today is slower than your average pace by 2 minutes and 28 seconds per kilometer." Currently, however, this level of question and answering through VA interaction is not possible, yet is understandably very plausible. Furthermore, research has shown that it can even be beneficial to use natural language (i.e., auditory or spoken) interactions when the visual system is overloaded [20]; this occurs often while a person is in-situ and actively engaged in a workout or other activity.

Looking at past work in this area, research has focused on the use of VAs for addressing general health knowledge questions [2, 127] and for their use in healthcare [50]. Furthermore, research has explored VA answer structures, specifically minimal, keyword, and full sentence, however only for common tasks and questions [71]. Ultimately, however, we do not have an understanding of how VA answers should be structured, specifically for personal health data questions. Uncertainty about how to provide answers is one such element that limits the potential of VAs to adequately respond to users' growing personal information needs, and thus the greater potential for databiting while on-the-go. This missing knowledge can even factor into user satisfaction and adoption of these systems all together [175, 188].

In this chapter, we focus on the following research questions:

> **RQ1:** How should voice assistant answers be structured to support spoken responses to personal health data queries?
>
> **RQ2:** How are voice assistants perceived as a tool to query personal

health data compared to general voice assistant interactions?

To address this knowledge gap we implemented a browser-based pseudo-VA which allowed for study participants to ask and receive answers to a variety of personal health data questions. Across two online studies, 82 participants interacted with our VA, asking questions and receiving answers to a combined total of 30 unique question-answer pairs. Within our studies, we explored three answer structures (i.e., Minimal, Keyword, and Full Sentence), each paired with questions from four personal health question response types (i.e., Open, Range, Binary, Value) and six known personal health insight categories (i.e., Contextual, Preemptive and Proactive, Goal and Performance, Combination and Comparison, Historical and Trend, and Current Status).

We found that in contrast to general VA question and answers, where short answers were preferred, participants preferred the Full Sentence answer structure given the response types and insight categories studied. Moreover, Full Sentence answers allowed for clarity in the answer when data was ambiguous, yet remained efficient despite the longer mean response time compared with other answer structures.

Our contributions are two-fold:

C1: Two studies, utilizing a custom-built browser-based pseudo-VA, allowing participants to explore personal health questions and answers across a total of 30 unique personal health question-answer pairs. We offer the implementation of our browser-based pseudo voice assistant that is used within both studies.

C2: Comprehensive user preferences of different answer structures for voice assistants in response to personal health data queries. We gather empirical insights into perceived answer quality, behaviour, comprehensibility, efficiency, and preference for personal health data questions. The findings provide elementary guidelines that complement known guidelines for current voice assistant question and answering capabilities.

6.2 Relevant Related Work

6.2.1 Interaction with Voice Assistants

Voice assistants (VAs) are quickly being adopted for use, now with over half of Americans using a VA in their homes [143]. To provide reason for such adoption, current research on natural language interaction (e.g., speech) has highlighted its utility in facilitating micro and hands-free interaction [5], especially when the visual system is overloaded (e.g., while tracking a walk) [19] and within various simultaneous activities [151]. These reasons for using a VA fit perfectly within the vision of databiting and the necessity for lightweight and transient exploration that does not hinder an ongoing activity, but rather assists and provides complement.

As we interact with a VA using natural language, research understandably explores the idea of personification and human-likeness in VA answers and dialogue [109], as well as social interactions [152]. However, in stark contrast to this, others have highlighted a need for technical systems [53], rather than human. Notably, increased human-likeness tends to increase trust and privacy concerns [27, 119], concerns that are at the forefront for personal health data [191, 207]. The contrast in human-like versus technical can greatly influence the experience of the VA interaction. As such, it is not only important to understand answer structures, but also necessary to continue to uncover preferences and perceptions as new VA interactions, such as for personal health data queries, become possible.

6.2.2 Voice Assistants in Health Contexts

Conversational user interfaces (CUIs) encompass a variety of interactive systems designed to facilitate natural language and conversational interactions. CUIs are an emerging means for people to gain general health-related information [2, 18, 127, 198], to self-report health and fitness data [121, 153], and to fill out health-related forms when health literacy is low [101]. CUIs in health contexts can take various forms, such as text-based chatbots, virtual assistants, and voice-activated platforms [50, 104]. Each of these forms offer unique capabilities for supporting health-related interactions and at times offer solutions to potential hurdles such as mispronunciation and recognition of medical terms and credibility [12, 146].

VAs, as a subset of CUIs, are becoming increasingly adopted as they are an embedded technology in many of our smart devices. VAs in health contexts currently allow users to query general health topics and symptoms [18], often providing links to online sources. As personal health data monitoring and exploration becomes more commonplace [35], along with a rise in Artificial Intelligence and Large Language Models, the intersection of VA interaction and personal health data querying will quickly become a reality. Despite this potential, current VA systems do not fully leverage their capabilities for question-answering tasks. With little actively working technology in this area, we largely do not know how a VA should answer personal health data queries, or how a VA is viewed for such tasks.

6.2.3 Voice Assistant Answer Structures

Investigations into VA communication styles have often focused from user to device. However, the feedback and response styles of the VAs not only play into a major design principle [139], but also a critical role in users' perceptions and adoption of these devices [175]. Yet, the effect of machine-to-human expression has been under investigated. For example, while factors like interruption [56] and conversational repair [46] have been explored, to our knowledge, only one work has explored answered structures for general VA use [71].

By exploring marketed VA answer practices as well as conducting a user study to compare various answer structures, Haas et al. [71] provide a comprehensive analysis of the experiences and user preferences regarding different VA answers. They first found that commercial VAs opt to convey more humanistic and full sentence answers for many common questions and commands. Only for home automation, where the outcome of the VA interaction is also noticeable within ones environment, were shorter keyword answers used. Then, in their own user study, Haas et al. found that minimal answers (those that provide the answer and no other supplemental or contextual information) were preferred for most command-based interactions while keyword answers (those that provide the answer and a brief confirmation of the keyword in the request) were preferred for most query-based interactions. This highlights a more utilitarian use of a VA than is currently recognized, where brief and basic answers suffice. Yet, many of the general use questions and commands explored are not *personal*. More specifically, there is a need for research to explore and understand people's expectations regarding answer structures when querying personal health data through VAs.

6.3 User Studies Methodology

Within two online user studies we used the same apparatus, procedure and data collection described below for both, and derived from previous work [71]. We follow the same procedure as previous work so that we can make direct comparisons and reflections between our findings, focused on personal health data question and answers, and previously published findings, which focus on currently afforded question and answers as well as task-based commands.

6.3.1 Apparatus

We built a browser-based pseudo-VA using Javascript and the WebSpeech API[7]. The WebSpeech API enables two important elements: (1) Speech recognition (i.e., recognizing a personal health data question) and (2) Speech synthesis, where speech (i.e., the answer to the question) can be vocalized, for which we used the "Google US English Female" voice. We chose the female voice for its similarity to the default voices current VAs use. The pseudo-VA used within the user studies can be demoed in Google Chrome[8]. Furthermore, the underlying code provides implementation details which can be used in future.

We utilize the term *pseudo* to describe our VA, as the functionality of our VA was limited to only handle questions desired within our user studies. As such, the VA was not fully functional as we would expect a commercial VA to be (e.g., Google Assistant, Siri, Amazon Alexa, etc.). While this limits capability, it provides experimental control. During interaction with our pseudo-VA, the participant's question would be recognized and then processed by checking for keywords

[7] `https://wicg.github.io/speech-api/`
[8] `https://vaphdqa.github.io/vaphdqa/`

(and varying synonyms) specific to each question. Only when our pseudo-VA recognized all required keywords, was the appropriate answer vocalized. If recognition could not be made (e.g., required keywords were missing) a response would encourage the participant to follow the prompt given and to try again.

When interacting with our pseudo-VA, a button on-screen was used to trigger the start of recognition, rather than utilizing keyword detection as many commercial VAs do. This design choice was done for privacy reasons. As such, participants had control of when recognition would begin, with recognition ending once the WebSpeech API recognized a natural stop in the question spoken. Furthermore, the text in the button would change to inform the participant when the VA was listening.

6.3.2 Procedure

Each study comprised three distinct stages: an introduction, the main trials, and demographic surveys. The procedure, VA recognition, and data collection were refined across two separate pilot phases for each study. In the first pilot phase, a single participant took part in the study while sharing their screen on Zoom and providing think-aloud feedback throughout. The second pilot phase involved three participants, none of whom took part in the first phase, who mirrored the procedures that our main study participants would follow. This process allowed us to refine the methodology and address any potential issues.

Within each of the studies, we include two attention check questions. One question was built into the pseudo-VA, mimicking a typical study trial. Specifically, all participants had to ask about their calorie intake for the day, to which the pseudo-VA responded with instructions that needed to be followed on the next screen. The second attention check question was simply slotted within a demographic survey, asking participants to choose a specific response.

Throughout the studies, the pseudo-VA was embedded into a larger Qualtrics survey and opened in Google Chrome. Once open, participants proceeded through the survey, as described below. The study procedure was approved by our institution's ethics review board. All participants provided informed consent prior to starting a study.

Study Introduction

Before starting the main trials, participants were guided through a study introduction process, designed to optimize interaction and comfort with the browser-based pseudo-VA and the study in general. The study introduction comprised of the following elements: (1) Voice input and output verification. Participants could test their input and output by continually interacting with the pseudo-VA, using preset and generic questions. (2) Study task. We provided a brief overview of the study's context and tasks. (3) Practice trials. Participants experienced three trials, during which they were prompted to check their calendar for the following day. Note, as with the main trials, we did not have access to a participant's personal data; throughout, generic data was used when providing responses. (4) Personal health data preference. To further enhance interest and focus within our study, we asked participants to choose between heart rate and step count data as their preferred question and answer topic for the remainder of the study. These topics for personal health data exploration were chosen for their popularity [3, 156], and can be seen as topics that are of interest throughout many daily contexts [157].

Main Trials

The main trials in our study were structured into blocks, with each block focusing on questions requiring a specific response type (i.e., Open, Range, Binary, and Value) or pertaining to a specific category of personal health data (i.e., Contextual, Preemptive and Proactive, Goal and Performance, Combination and Comparison, Historical and Trend, and Current Status) in each of Study 1 (see Table 6.1) and Study 2 (see Table 6.2) respectively.

To ensure balance and minimize potential order effect, both the question blocks and answer structures within a single question block were ordered using a Latin Square design. Participants were randomly assigned to a question block ordering and for each question block randomly assigned an answer structure ordering. Throughout, we ensured as best we could that an equal number of each order for both the questions and answers was shown across all participants in the study. This study design allowed for a systematic exploration of how different answer structures performed across various dimensions of personal health data queries.

Trials were grouped in threes, corresponding to each question block. Participants asked the same question for each trial in a question block while receiving each of the three differently structured answers. We used a repeated measures design, in which each participant had to ask each of the four or six different questions for each of the three answer structures. This results in a total of 12 or 18 question/answer interactions (trials) for each participant, for Study 1 and Study 2 respectively.

During each trial, participants were prompted to ask a personal health data question using their own words. After providing the question, participants heard an answer given by the VA. Following a successful interaction with the VA for each trial, participants were automatically directed to complete the User Experience Questionnaire Plus (UEQ+) [100, 167]. The UEQ+ survey provides insights into participants' subjective experiences and satisfaction levels with the voice assistant answer structures.

Following the work which created the UEQ+ survey for voice assistants [100], and previous work [71], we paired two semantic scales to compose a single experience quality factor. We chose two semantic differentials with the highest found loading from three scales: 1) Behaviour, consisting of the scales *artificial - natural* and *unlikable - likeable*, 2) Comprehensibility consisting of the scales *complicated - simple* and *unambiguous - ambiguous*, and 3) Efficiency consisting *slow - fast* and *inefficient - efficient*. Finally, we incorporated a fourth scale: 4) Quality, which consisted of the scales with the third and fourth highest loading, yet still recommended by the creators [100]. The Quality scale consisted of the scales *unintelligent - intelligent* and *inappropriate - suitable*. We opted for these scales as the scales with the highest loading for Quality were *not helpful - helpful* and *useless - useful*. Through preliminary discussion, we felt these scales required the use of actual personal health data for a participant to fully evaluate these semantics.

After completing a question block, participants were asked to rank-order their preference for the three answer structures heard during that block. This ranking task aimed to elicit participants' subjective preferences with a forced response akin to selecting a single option if given the choice on their own device.

It is important to again note that all answers provided during the trials were generic, and no personal health data from participants was used in generating VA answers. This approach ensured consistency and privacy in the study design while still allowing for a thorough exploration of participants' preferences with respect to VA answer structures for personal health data queries.

Surveys and Open Feedback

Our study incorporated surveys and open feedback to gather insights into participants' demographics, personality traits, and preferences regarding voice assistant usage. We gathered information about age, gender, background, VA usage, and personal health data collection practices. Additionally, participants completed the Ten-Item Personality Inventory (TIPI) [67] to assess personality traits and the Attitudes For Technology Interaction (ATI) [64] scale to assess their attitudes and comfort with technology. We also explored attitudes towards voice assistants for both general and personal health data use, separating these surveys to mitigate carry-over responses. Finally, participants were given the opportunity to provide open feedback to express their thoughts, suggestions, and/or concerns.

6.3.3 Question Answer Structures

The questions prompted for participants to ask within each study were derived from a public dataset of personal health data queries captured in-the-wild from experienced smartwatch users [157]. The questions within our studies were carefully selected to represent the elicitation of different response types in Study 1 and various desired categories of personal health insights in Study 2. Within each study section below, we further highlight the questions used and how they were derived.

For both studies, the questions were answered using three answer structures: Minimal, Keyword, and Full Sentence. These answer structures have been utilized in previous work [71], and mimic the *Full* and *Brief* options offered by Google's Assistant. Minimal answers solely contain the information required to answer the question. Keyword answers provide the answer and confirmation of the question asked. Full sentence answers provide full sentence responses emulating human-like sentence structure. Notably, Minimal and Full Sentence answer structures follow humanistic response behaviours, while Keyword does not. As we allowed for both step count and heart rate as question topics to promote interest and engagement within the study, we aimed to ensure as much consistency as possible between the questions and answers for either data source. This included ensuring as much commonality between answers as possible while also ensuring answers were of similar lengths; see Table 6.1 and Table 6.2 for all questions and answers used (for both step count and heart rate topics). No matter the question topic chosen by participants, each participant saw the same number of question and answer trials during their respective study (i.e., 12 in Study 1 and 18 in Study 2).

6.3.4 Participant Recruitment

Recruitment of participants was done through Prolific[9]. Prior to participation, potential participants completed an eligibility survey to ensure they met our inclusion criteria: participants were required to (1) Have their first language be English. Therefore, answer structures could be properly evaluated by native speakers; (2) Have used a VA before. As such, participants held experiences either good or bad with the use of VAs and their answers; (3) Currently collect and/or explore personal health data. By having experience with personal health data, participants could have defined expectations and preferences.

When partaking in the main study, we asked participants to place themselves in a room with as little distraction as possible. We further required the use of a desktop/laptop computer with Google Chrome installed and for participants to have working microphone and speaker/headphones. No matter their responses, all participants who took part were paid 0.25 GBP for completing the eligibility and 6.5 GBP for completing the study.

[9] `https://www.prolific.com/`

6.3.5 Data and Analysis

The data from each study is analyzed across four key areas: (1) UEQ+ scores for each question-answer pair, (2) rank order preferences, (3) attitudes towards VAs, and (4) open feedback.

Important differences are highlighted for each study below with means and the confidence interval boundaries listed in text and graphically presented within figures. Throughout, we opt to utilize confidence intervals rather than relying on *p*-values. Graphically presenting confidence intervals allows us to systematically assess any effects at play while also gauging *practical significance* [47, 54]. Reporting on confidence intervals rather than *p*-values has become increasingly popular in HCI literature [11]. Confidence intervals offer greater understanding for a broader audience and do not suffer from the illusion of truth sometimes provided by a *p*-value [54]. As no statistical testing is being performed, we explicitly do not utilize *significance* terminology. Rather we find and report on differences through confidence intervals which do not overlap [47, 54]. Before calculating the means and confidence intervals all outliers for each pairwise analysis were removed as any data point outside three standard deviations.

As the creators of the UEQ+ scale do not test for inter-item reliability, we calculated the Cronbach's α for each pairwise comparison using .70 as a cutoff [140]. When the α did not reach the cutoff for any UEQ+ scale, we report on the combined UEQ+ scale, while further exploring its semantic differentials separately.

6.4 Study 1 - Response Types

6.4.1 Questions

For Study 1, the prompted questions were categorized by the type of response the question would elicit. The possible response types, as seen in Chapter 5, are: 1) **Value**, a single value response, 2) **Open**, a response requiring a list of data/information of any size, 3) **Binary**, a yes/no response, and 4) **Range**, comparative two-value responses.

Table 6.1 shows all questions and answers used within the study. Answers took an average of 2.8, 4.5, and 6.3 seconds to convey to the participant for each of Minimal, Keyword, and Full Sentence answer structures respectively.

Table 6.1: Questions and answers used in Study 1. The forward slash denotes the separation between the choice of heart rate or step count topics, one of which was chosen by the participant for use throughout the study.

Response Type	**Expected Question**	**Minimal**	**Keyword**	**Full Sentence**
Value	What was my average heart rate in the last hour? / What was my step count in the last hour?	71 beats per minute / 1,375 steps	Average heart rate last hour, 71 beats per minute / Step count last hour, 1,375 steps	In the last hour, your average heart rate was 71 beats per minute / In the last hour, your step count was 1,375 steps
Open	In what workouts does my heart rate reach zone five? / In what workouts does my step count reach 2,000 steps?	Indoor running, outdoor cycling, and rowing / Indoor running, outdoor running, and hiking	Workouts reaching zone five, indoor running, outdoor cycling, and rowing / Workouts reaching 2,000 steps, indoor running, outdoor running, and hiking	In the past, indoor running, outdoor cycling, and rowing workouts have brought your heart rate into zone five / In the past, you have reached 2,000 steps during indoor running, outdoor running, and hiking workouts
Binary	Is my heart rate higher than normal? / Is my step count higher than normal?	Yes / Yes	Yes, heart rate higher than normal / Yes, step count higher than normal	Your current heart rate is higher than your normal heart rate / Your current step count is higher than your normal step count
Range	What is my average heart rate on weekdays compared to weekends? / What is my average step count on weekdays compared to weekends?	85 beats per minute compared to 76 beats per minute / 9,820 steps compared to 10,680 steps	Weekdays, 85 beats per minute. Weekends, 76 beats per minute / Weekdays, 9,820 steps. Weekends, 10,680 steps	Your average heart rate during the week is 85 beats per minute. While on the weekends, your average heart rate is 76 beats per minute / Your average step count during the week is 9,820 steps. While on the weekends, your average step count is 10,680 steps
	Mean and Standard Deviation of Response Times (seconds):	M=2.6, SD=1.3 / M=3.0, SD=1.6	M=4.2, SD=1.1 / M=4.8, SD=1.4	M=5.7, SD=2.0 / M=6.7, SD=2.4

6.4.2 Participants

Thirty-four participants took part, with one participant failing the attention checks. Of the 33 participants whose data we used for analysis their ages ranged from 18 to 63 years old (*M* = 35.0, *SD* = 11.7; 24 Females, 9 Males). Furthermore, 19 participants self-identified as White, nine as Black/African, three as Hispanic, and two as Asian. Twenty-two (22) participants indicated they use Google's Assistant, 17 use Apple's Siri, 11 use Amazon's Alexa , six use Samsung's Bixby, and two use Microsoft's Cortana. As marketed VAs offer similar responses [71], we were not concerned with bias from using a specific VA. Fifteen (15) participants stated using a voice assistant more than once per day, two once per day, eight a few times a week, two once a week, and six less than once per week. On average, participants had been collecting personal health data for 56.5 months (*SD* = 32.4 months). On average, Study 1 took 24.9 minutes to complete (*SD* = 11.2 minutes). Eleven (11) participants chose heart rate while 22 participants chose step count as their data type to explore.

6.4.3 Results

Quality, Behaviour, Comprehensibility, Efficiency

Mean participant ratings with 95% confidence intervals for answer quality, behavior, comprehensibility, and efficiency are shown in Figure 6.1.

Quality For answer Quality (a composite of *helpfulness* and *usefulness*), differences were observed within Range, Binary, and Value response types, but not for Open. Within the Range response type, participants rated the Full Sentence answer structure as having the highest Quality (*M* = 6.36, *CI* [6.11, 6.61]) compared to both Keyword (*M* = 5.77, *CI* [5.48, 6.07]) and Minimal (*M* = 4.94, *CI* [4.56, 5.32]). Furthermore, Keyword answers were rated higher than Minimal. For the Binary response type, the Full Sentence answer structure received higher ratings (*M* = 6.11, *CI* [5.84, 6.38]) compared to both Keyword (*M* = 4.95, *CI* [4.56, 5.35]) and Minimal (*M* = 4.38, *CI* [3.79, 4.97]). For the Value response type, participants rated the Keyword answer structure lower (*M* = 5.74, *CI* [5.43, 6.06]) when compared to the Full Sentence answer structure (*M* = 6.33, *CI* [6.09, 6.57]). Only when the response type was Open, the answer structure did not influence the perceived Quality.

We compared the quality ratings of Keyword answer structures across the four response types. In the Binary response type, Keyword answers were rated lower (*M* = 4.95, *CI* [4.56, 5.35]) when compared to other response types. Similarly, for the Minimal answer structure, Binary responses received the lowest rating (*M* = 4.38, *CI* [3.79, 4.97]) when compared to Open and Value response types. Furthermore, Full Sentence answers consistently received higher ratings across all response types, suggesting a potentially higher perceived quality.

Behaviour In examining the answer Behaviour (a mean composite of *naturalness* and *likability*), a similar pattern for Full Sentence was evident. Regardless the response type, Full Sentence always yielded the highest ranking; differences between Full Sentence and the other answer structures within each response type were found for all except Value. Moreover, the means for Full Sentence did not vary across the four response types. This was consistent with both other two answer structure types (i.e., the Behaviour ratings were not influenced by Response Type). Thus, to further explore, we created mean scores for each answer structure type. As anticipated, the Full Sentence structure yielded the highest mean (*M* = 6.19, *CI* [5.99, 6.41]) while Minimum Sentence (*M* = 5.28, *CI* [4.97, 5.59]) and Keyword (*M* = 5.57, *CI* [5.32, 5.80]) did not vary.

Comprehensibility In terms of answer Comprehensibility (a mean composite of *simplicity* and *ambiguity*), only one answer structure effect was found across response types. For the Range response type, Full Sentence and Keyword answer structures were seen as equally comprehensible while Minimal was seen as the least comprehensible (*M* = 5.06, *CI* [4.56, 5.56]). For other response types, the answer structure did not affect the level of comprehensibility. Noticeably, using a Minimal answer structure for both Range (*M* = 5.06, *CI* [4.56, 5.56]) and Binary (*M* = 5.23, *CI* [4.67, 5.79]) response types was seen as less comprehensible than if

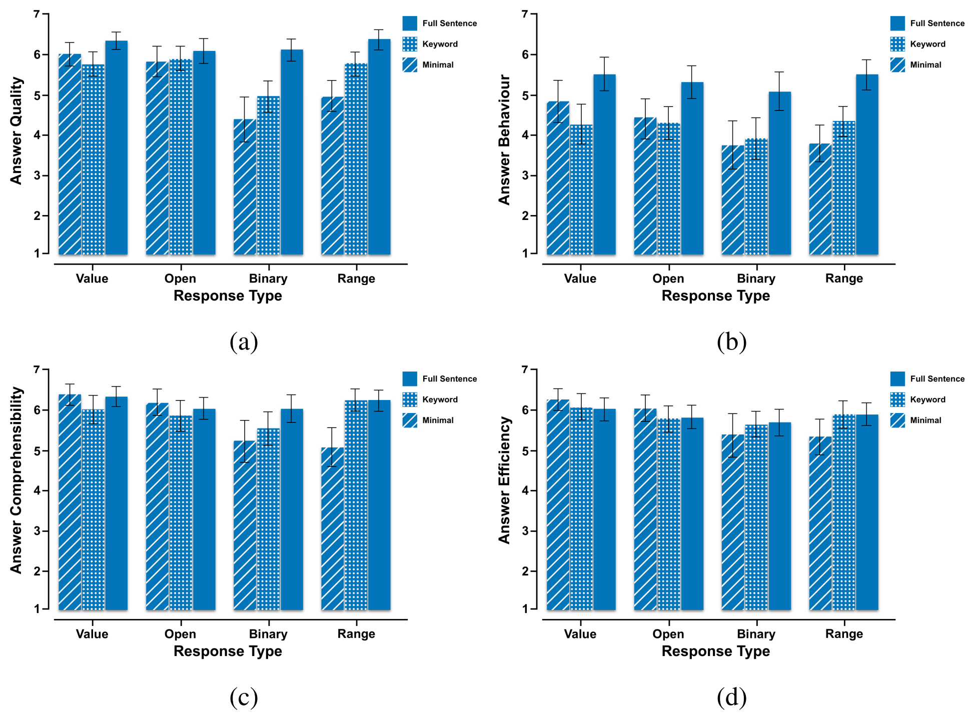
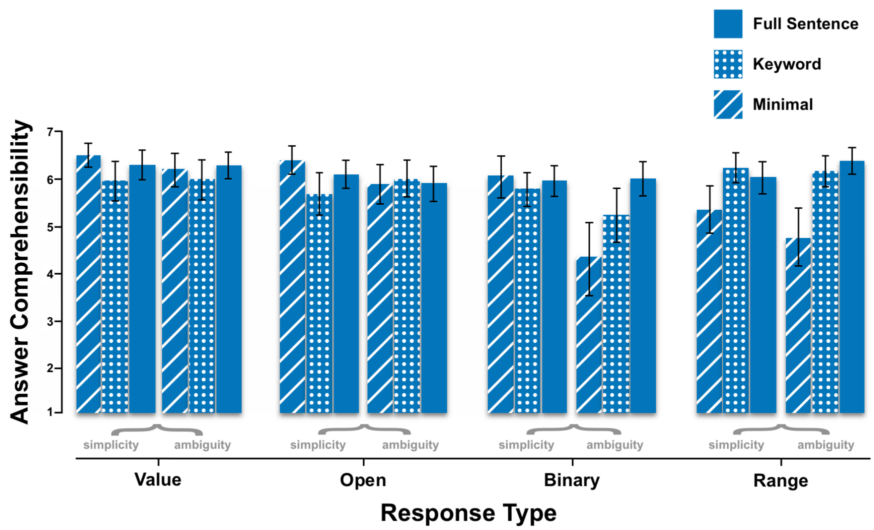
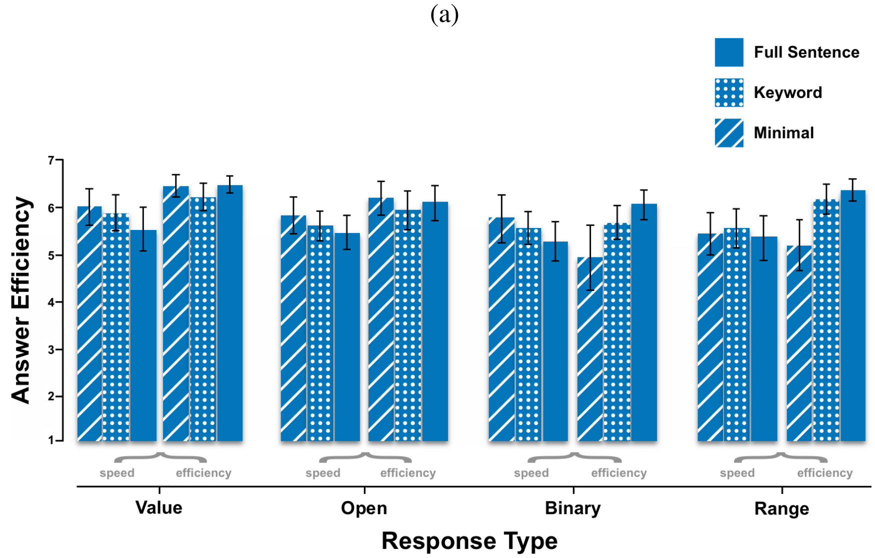


Figure 6.1: Study 1 mean UEQ+ ratings with 95% confidence intervals: Quality (a), Behaviour (b), Comprehensibility (c), and Efficiency (d). Ratings are compared by the response types (Value, Open, Binary, Range) and answer structures explored in the study.



(a)



(b)

Figure 6.2: Study 1 mean UEQ+ ratings with 95% confidence intervals for a) *Comprehensibility* - separated by simplicity and ambiguity; and b) *Efficiency* - separated by speed and efficiency. Ratings are compared across the response types (Value, Open, Binary, Range) and answer structures explored in the study.

used for Open and Value.

The level of Cronbach's α for Comprehensibility was 0.66. As such we separated the semantic differentials used and further explored *complicated* - *simple* and *ambiguous* - *unambiguous* separately; see Figure 6.2a. Differences across the semantic differentials are present for the Minimal answer structure in the Binary response type. More specifically, the Minimal answer was rated as being highly simple but is different when compared to ambiguity, suggesting the response is simple yet ambiguous.

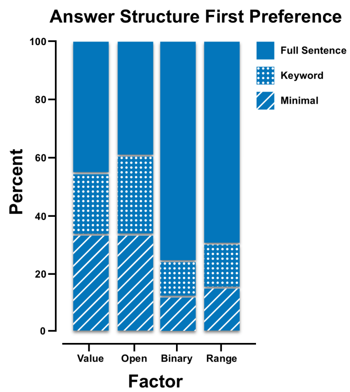
Efficiency For answer Efficiency (a mean composite of *efficient* and *fast*), no differences were found.

However, the level of Cronbach's α for Efficiency was 0.65. As such we separated the semantic differentials used and further explored *slow* - *fast* and *inefficient* - *efficient* separately; see Figure 6.2b. Full Sentence answers were rated as more efficient than they were fast for the Range, Binary, and Value response types. This suggests that while an answer does not have the fastest mean response times, as is the case with Full Sentence, they are still viewed as efficient by participants.

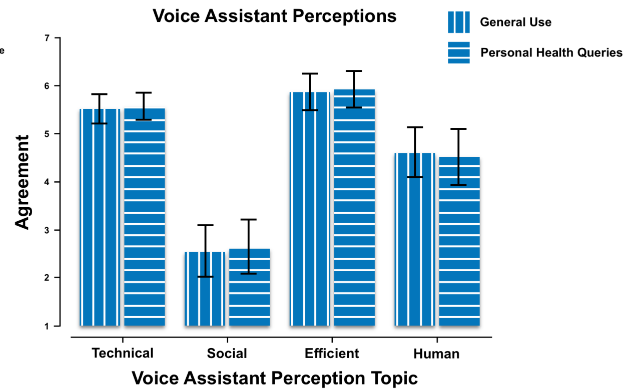
Interestingly, we realized that the sentence structure did not affect the perceived speed. VA's answers using Full Sentence, Keyword, and Minimal structures were perceived equally fast regardless of the actual response time (see Table 6.1 for mean response times).

Preference and Attitudes Towards Voice Assistants

Participants generally favored the Full Sentence answer structure; see Figure 6.3a. If Full Sentence answers were not preferred, then Minimal was often the preferred answer structure. This preference pattern suggests that participants prioritize responses that exhibit human structuring of answers. Notably, while participants demonstrated clear preferences for the Full Sentence answer structure throughout, only Range and Binary saw the majority of participants chose Full Sentence as their preferred answer structure. This is likely due to a need for additional context and clarity within these response types. In contrast, Value and Open



(a)



(b)

Figure 6.3: Participant's preference of answer structure for each response type (Value, Open, Binary, Range) (a). Participant's perceptions of voice assistants (b).

response types allow for implicit interpretation and internal verification given an answer (i.e., if asking what workouts a person took 2000 steps within, cycling is an obvious wrong answer). This internal verification leads participants to perceive a slightly lesser need for Full Sentence responses; instead, favoring greater flexibility and brevity in the answer structure.

Participants perceived voice assistants for personal health data exploration and general use similarly, indicating a consistent perception across these question domains; see Figure 6.3b. Interestingly, participants rated highly that a VA both was viewed as a technical system and should prioritize efficiency. Attributes related to human likeness and social interaction were not as strongly desired, with a VA social companion being even less preferred than human likeness. This suggests that participants value technical capabilities and efficiency in VAs, and may not necessarily expect or desire human-like or social qualities (e.g., as a trainer or coach).

Open Feedback

Across 18 unique comments, the open feedback provided by participants revealed three main topics. Firstly, two participants expressed a newfound interest in utilizing voice assistants for personal health data exploration. Second, regarding the study procedures, comments were generally positive, with seven participants expressing enjoyment and satisfaction. However, one participant suggested a need for a slower speed during interactions with the voice assistant. Additionally, two participants suggested the use of a different voice, specifically male, during interactions; no other interactive comments, such as recognition issues were mentioned. Finally, in terms of preferences, participants interestingly expressed diverse and opposing opinions. Two participants suggested favoring concise and clear responses, while another further mentioned they found human-like responses to be unsettling. Conversely, three participants preferred longer answers, particularly to confirm that the voice assistant understood their questions and attributing human-like qualities to these responses.

6.5 Study 2 - Insight Categories

In our second study, we focus on insight categories of personal health data questions rather than broader response types. Notably, a question asked within a personal health data insight category can result in most response types, depending how the question is asked. Given the very few differences in UEQ+ scores between answer structures in both Value and Open responses in Study 1, we extend Study 1 by choosing to study Value responses in Study 2. This decision was driven by the versatility of Value responses, which are applicable across all insight categories, whereas Open responses are not. This study then offers a more nuanced understanding of user preferences in voice assistant answers for personal health data questions.

6.5.1 Questions

The prompted questions were chosen to represent known personal health insight categories, both from previous work [3, 29**?** , 30] and that of our own findings in Chapter 4. These categories include: 1) **Current Status** (CSV), derived from a current measured value, 2) **Historical and Trend** (HT), provides insight into past data, 3) **Combination and Comparison** (CC), derived from combining/comparing data sources, time periods, and/or activities, 4) **Goal and Performance** (GP), derived from user goals and performance metrics, 5) **Preemptive and Proactive** (PP), provides insight into a future action, and 6) **Contextual** (CT), provides context to gain insight. When choosing questions from within our previously captured dataset, we first opted for questions that were categorized into a single insight category. However, we note that some insight categories overlap with the Historical and Trend category.

Answers took an average of 1.8, 3.0, and 4.0 seconds for Minimal, Keyword, and Full Sentence answer structures respectively. As per our study goal, all answers invoked a Value response type. All questions and answers used within this study can be seen in Table 6.2.

Table 6.2: Questions and answers used in Study 1. The forward slash denotes the separation between the choice of heart rate or step count topics, one of which was chosen by the participant for use throughout the study.

Insight Category	**Expected Question**	**Minimal**	**Keyword**	**Full Sentence**
Current Status (CSV)	What is my current heart rate? / What is my current step count?	68 beats per minute / 7,350 steps	Current heart rate, 68 beats per minute / Current step count, 7,350 steps	Your heart rate is currently at 68 beats per minute / Your step count is currently at 7,350 steps
Historical and Trend (HT)	What was my average daily heart rate last week? / What was my average daily step count last week?	78 beats per minute / 10,270 steps	Average daily heart rate, 78 beats per minute / Average daily step count, 10,270 steps	Your daily average heart rate last week was 78 beats per minute / Your daily average step count last week was 10,270 steps
Combination and Comparison (CC)	Is my heart rate different from my average? / Is my step count different from my average?	Higher than average / Higher than average	Current heart rate, higher than average / Current step count, higher than average	Your current heart rate is higher than your average heart rate / Your current step count is higher than your average step count
Goal and Performance (GP)	Which day of the week is my heart rate the highest? / Which day of the week is my step count the highest?	Saturdays / Saturdays	Highest heart rate, Saturdays / Most steps taken, Saturdays	Your heart rate is the highest on Saturdays / Your step count is the highest on Saturdays
Preemptive and Proactive (PP)	How long should I control my breathing to get to my resting heart rate? / How far should I walk to get to 10,000 steps?	Two minutes / 2.5 kilometres	To reach your resting heart rate, two minutes / To reach 10,000 steps, 2.5 kilometres	To reach your resting heart rate, you should control your breathing for two minutes / To reach 10,000 steps, you should walk a distance of 2.5 kilometres
Contextual (CT)	Is my heart rate lower in the morning, afternoon, or evening? / Is my step count lower in the morning, afternoon, or evening?	Evening / Afternoon	Lowest heart rate, evening / Lowest step count, afternoon	Your heart rate is the lowest in the evening / Your step count is the lowest in the afternoon
	Mean and Standard Deviation of Response Times (seconds):	M=1.6, SD=0.4 / M=1.9, SD=0.7	M=2.7, SD=0.4 / M=3.2, SD=0.7	M=3.6, SD=0.8 / M=4.3, SD=1.0

6.5.2 Participants

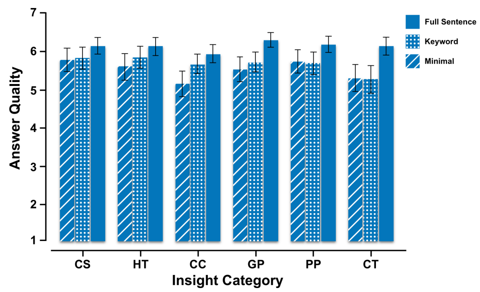
Of the 52 participants who took part, three participants were removed for not passing our attention checks. Of the 49 participants whose data we used for analysis their ages ranged from 18 to 67 years old (*M* = 34.5, *SD* = 11.2; 30 Females, 19 Males). Furthermore, 31 participants self-identified as White, 13 as Black/African, one as Hispanic, one as Asian, two as Multiracial, and one as Middle Eastern. Twenty-eight (28) participants indicated they use Google's Assistant, 23 use Apple's Siri, 26 use Amazon's Alexa, five use Samsung's Bixby, and four use Microsoft's Cortana. Twenty (20) participants stated using a voice assistant more than once per day, five once per day, 16 a few times a week, three once a week, and five less than once per week. Participants had been collecting personal health data for an average of 51.5 months (*SD* = 34.2 months). On average, Study 2 took 24.9 minutes to complete (*SD* = 9.4 minutes). Ten (10) participants chose heart rate while 39 chose step count as their data type to explore.

6.5.3 Results

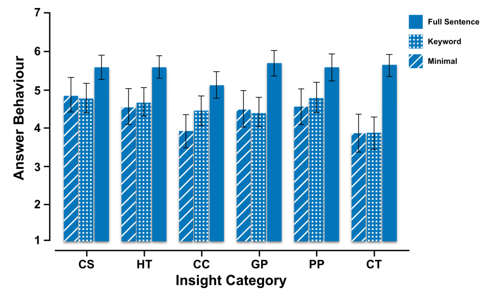
Quality, Behaviour, Comprehensibility, Efficiency

Mean participant ratings with 95% confidence intervals for answer quality, behaviour, comprehensibility, and efficiency are shown in Figure 6.4.

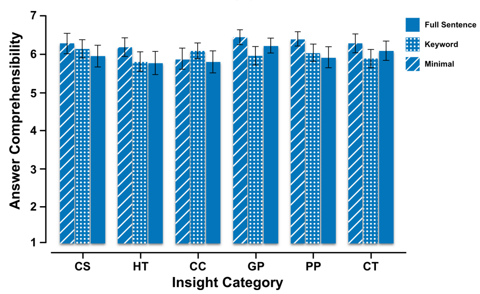
Quality Within insight categories, no differences in perceived quality are seen for Minimal and Keyword answer structures. However, Full Sentence shows a higher mean Quality for Contextual (*M* = 6.15, *CI* [5.91, 6.38]) and Goal Performance (*M* = 6.31, *CI* [6.11, 6.5]) insight categories compared to other answer structures. As well, Full Sentence (*M* = 5.94, *CI* [5.7, 6.18]) shows a higher quality than Minimal (*M* = 5.15, *CI* [4.81, 5.5]) in the Combination and Comparison insight category. Notably, across insight categories, there are no differences for each of the individual answer structures, suggesting that the insight category does not change perceived Quality of an answer structure.



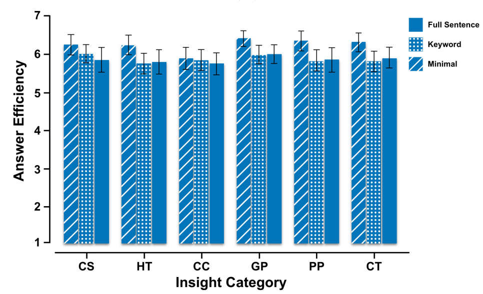
(a)



(b)



(c)



(d)

Figure 6.4: Study 2 mean UEQ+ ratings with 95% confidence intervals: Quality (a), Behaviour (b), Comprehensibility (c), and Efficiency (d). Ratings are compared by the insight categories (Current Status and Value (CSV), Historical and Trend (HT), Combination and Comparison (CC), Goal and Performance (GP), Pre-emptive and Proactive (PP), and Contextual (CT)) and answer structures explored in the study.

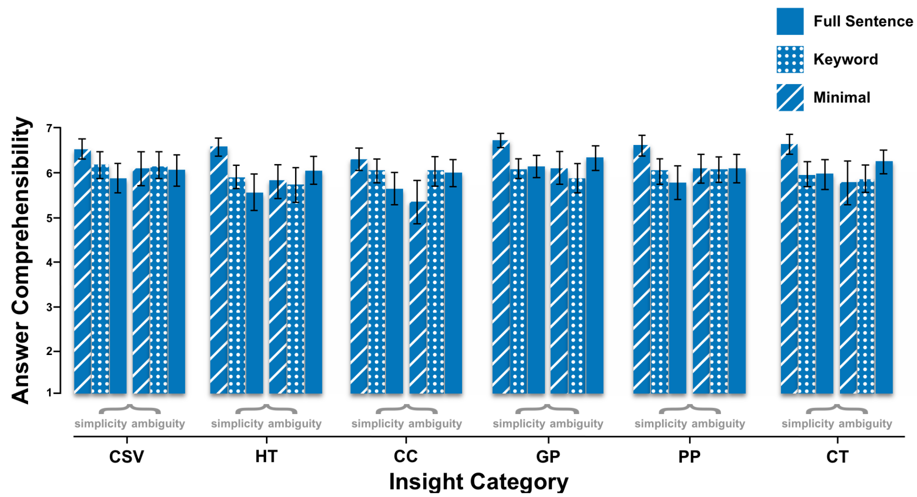
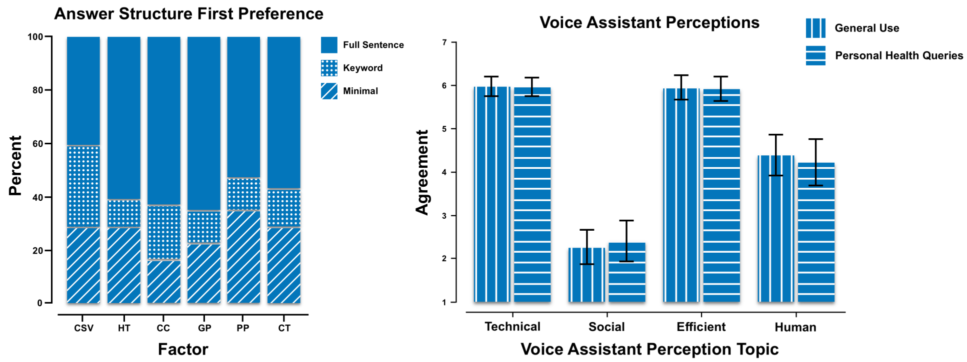


Figure 6.5: Study 2 mean UEQ+ ratings with 95% confidence intervals for Comprehensibility - separated by the simplicity and ambiguity. Ratings are compared across the insight categories (Current Status and Value (CSV), Historical and Trend (HT), Combination and Comparison (CC), Goal and Performance (GP), Preemptive and Proactive (PP), and Contextual (CT)) and answer structures explored in the study.

Behaviour In parallel to Study 1, we once again noticed a general trend wherein Full Sentence resulted in the highest scores for Behaviour (*naturalness* and *likability*). However, Minimal Sentence and Keyword, which generally scored lower than Full Sentence, did not vary from one another.

Comprehensibility The level of Cronbach's α for Comprehensibility was 0.49. Thus, we investigated the semantics (*complicated - simple* and *ambiguous - unambiguous*) independently; see Figure 6.5. Noticeably, as in Study 1, it is the Minimal answer structure which provides differences comparing across the two semantics in the Contextual, Goal and Performance, Combination and Comparison, as well as Historical and Trend insight categories. Each time the answer is rated as more simple while being ambiguous.



(a) (b)

Figure 6.6: Participant's preference of answer structure for each insight category (Current Status and Value (CSV), Historical and Trend (HT), Combination and Comparison (CC), Goal and Performance (GP), Preemptive and Proactive (PP), and Contextual (CT)) (a). Participant's perceptions of voice assistants (b).

Efficiency For Efficiency (a mean composite of "*efficient*" and "*fast*") no differences were found. The answer structure did not influence the levels of perceived efficiency. No other effects were found.

Preference and Attitudes Towards Voice Assistants

As with Study 1, we again see the Full Sentence answer structure as being the preferred majority for the remaining insight categories. Only for the Current Status insight category is this less pronounced. As complexity in the question increases, from that of a Current Status question, Full Sentence seems to be preferred mainly for its Quality and Behaviour. Finally, perceptions of VAs across both studies were comparable.

Open Feedback

Across 19 unique comments the same three topics arose from Study 1. First, three participants expressed an interest in using a VA for personal health data exploration. However, two participants shared that they prefer to perform visual data analysis. These comments are important; our aim is not to replace visual data analysis. Instead, our work aims to diversify approaches to explore personal health data. Second, nine study procedure comments highlighted that the study went well. One participant mentioned the VA could slow down its answers while only three participants expressed the VA was *sensitive* which caused some interpretation issues throughout (sensitivity was expressed as being due to background noise, a learned accent, and an illness influencing speech). Finally, only one participant commented that they preferred concise and clear answers so as minimize the time taken for the interaction.

6.6 Discussion

6.6.1 Implications for the Design of VA Interactions for Personal Health Data Queries

Comparisons with General VA Interactions

Our results indicated that users preferred Full Sentence answers for their personal health data queries. This runs counter to prior work, which has suggested that Minimal and Keyword responses are ranked positively, and sometimes preferred, for common VA tasks (i.e., knowledge queries, home automation, reminders, calendar queries) [71]. We believe this discrepancy arises due to the brevity of Minimal and Keyword responses which fall short in conveying the level of comprehension required for many personal health data queries. For example, if a person asks ”What is on my calendar tomorrow?”, the VA could respond using a Minimal answer structure, stating ”Lunch with Tyler, 1PM. Games with Danica, 7PM.” The content in the answer itself provides a connotation of calendar events and does not produce an answer that is ambiguous.

In contrast, if a person asks ”What is my average daily step count in the last week?”, providing a Minimal answer, such as ”10,320 steps”, does little to convey that the question was properly understood. Many other possibilities exist for a similar answer (e.g., average daily step count in the last month or current step count.) Such ambiguity has been previously noted as a barrier in using personal health data for clinical purposes [196], and now appears to be a common barrier for end-users exploring their own personal health data.

These findings highlight the importance of tailoring answer structures to specific questions, both general and health-related. Furthermore, contextual information plays a key role in VA personal health data exploration, where users comprehend the information provided holistically, rather than focusing solely on single numerical or categorical values. As such, design considerations for VA personal health data interactions should prioritize confirmation and inclusion of key aspects of the data.

Efficient Full Sentence Answers

Despite being longer with respect to response time, Full Sentences were perceived as equally efficient as Minimal and Keyword answers. We contemplate several reasons for this observation. First, contextual information provided within a Full Sentence may contribute to a more comprehensive understanding of the answer, thereby reducing the need for follow-up questions or clarification. This can ultimately enhance efficiency, even if individual responses take longer. Second, the context in which the answer is given may influence its perceived efficiency. In distracting environments, concise responses may be preferred, while in quieter settings, more detailed answers may be deemed appropriate. Thus, the threshold of an efficient answer may vary depending on the situational context, allowing for flexibility in response length without compromising perceived efficiency.

Not only do our findings indicate that Full Sentence is perceived to be efficient, but they suggest that there may be room to augment Full Sentence answer content without sacrificing perceived efficiency. This is due to the fact that we observed that participants *feel* Full Sentence answers were equally as fast as Keyword and Minimal answers. Looking ahead, leveraging the capabilities of Full Sentence answers could allow for serendipitous information, akin to visual data exploration. For example, if a person asks "What is my average daily step count in the last week?", the VA could provide a Full Sentence answer stating, "Your average daily step count in the last week is 10,320 steps. This is higher than the previous week, keep it up!"

Moreover, future investigations could aim to enhance the depth of information conveyed, the number of data points included, and/or the influence of certain answers to enrich user interactions with VAs. For example, if a person asks "Is my heart rate lower in the morning, afternoon, or evening", the VA could respond with "Your heart rate is the lowest in the evening, and is roughly 15 beats per minute lower than other times of day." Each of these areas of exploration should explore the context in which VA interactions take place (i.e., at home or while on a walk). Future work in these areas can build from the results found in our work while then aiming to provide more comprehensive guidelines for personal health data questions and answers, allowing for VA interactions that suit people's needs and preferences and could provide greater influence.

6.6.2 Human Emulation and Unwavering Perceptions

Our results highlighted unwavering and confident perceptions within both of our studies for the use of VAs for personal health data question and answering. This can be seen in the highly similar responses captured in our VA perception survey questions (see Figure 6.3b and Figure 6.6b, and follow the results of previous work exploring general VA use [71]. As seen from this reported data, striking a balance between technicality and efficiency, while providing answers that emulate full sentences (and therefore human likeness) is key. Importantly, however, we must be cognizant that VA interactions should remain to invoke as little social interaction as possible (e.g., the VA should not emulate a fitness coach). Notably, VA responses which emulate human behaviour, in part conveyed through Minimal and preferred Full Sentence answer structures, have been shown to raise people's expectations of the VA [115].

While expectation and capability may be a concern in the earlier life cycles of VAs, rapid improvements to VA performance will likely mitigate these concerns over time. Of more interest, is that human-like responses can lead to incorrect and inappropriate use of a VA [160]. Furthermore, trust and privacy become concerns when the VA is seen as increasingly human-like [27, 119], a concern that is amplified with respect to personal data [191, 207] over that of general knowledge (e.g., the weather or population of the USA). Therefore, we encourage designers of VA technologies to pay close attention to the balance required to accommodate these preferences.

6.6.3 Comparisons With Commercial Personal Health Data Question and Answering

To better situate this chapter's findings we once more explore current capabilities with respect to VA personal health data question and answering. As such, we again asked Siri a range of personal health data questions. As functionality is still limited (i.e., Siri can not answer many of the questions within our dataset and this our study), we explored questions, and slight variations of questions promoted by Apple[10]. Notably, the use of Siri for the exploration of personal health data is coupled with a display, rather than through a standalone device (e.g., in a standalone smart speaker and as seen in our study). As such, we provide this information for discussion purposes only.

[10] `https://tinyurl.com/siriphdqa`

Table 6.3: Recorded personal health data question and answers using Siri on an iPhone 14 Pro running iOS 17.3.

Question	**Vocal Response**	**Supplementary Information Displayed**
What is my current step count?	[Step Count] steps	Steps Today
What's my heart rate?	As of February 2, 2024 11:48 AM, it was 77 BPM	Heart Rate [Month] [Day] [Year] [Time]
How far did I walk yesterday?	[Distance] km	Walking + Running Distance Yesterday
How far have I walked this week?	[Distance] km	[Distance] km Daily Average Walking + Running Distance [Month] [Day] - [Month] [Day] [Year]
What is my move ring at?	You've burned [Current Calorie Burn] out of your [Calorie Burn Goal] calorie goal	Move Ring Today [Time]

Our study findings reveal preferred differences compared to how Siri answers questions. While Siri predominantly uses Minimal vocal responses and Keyword information displayed on screen, our studies highlight a preference for Full Sentence answers. Notably, Siri rarely employs Full Sentences, except for activity ring data, where multiple data points are conveyed, and when there are contextual deviations from the question (see Table 6.3, "What's my heart rate?" - heart rate data was not current and Siri responded with the last sample recorded). This divergence from our study findings in VA practice highlights varied strategies in voice assistant design, with Siri prioritizing brevity and visual support. However, the presence or use of a screen may not always be optimal (as in the example shared in the Introduction where the focus should remain on the hiking environment). Such insight and discussion sheds light on the diverse approaches that can be adopted by voice assistants in managing personal health data and more importantly underscores the importance of understanding user preferences, expectations, and contexts.

6.6.4 UEQ+ Semantic Differentials

In our study, despite using an adapted version of the the UEQ+ survey to measure VA user experience [100], we observed conflicting semantic differentials for Efficiency and Comprehensibility. The lack of correlation found suggests that the semantic differentials may not be effectively capturing the same intended user experience factor. These findings highlight the need for future work to refine the UEQ+ scale for VA interactions, aligning semantic differentials more closely with user experiences. Specifically, better assessing combinations of the semantic differentials while also exploring the potential for new overall user experience factors could enhance the scale's utility and effectiveness.

While it can be argued that two items do not need to directly relate to provide a worthwhile assessment, as we do in this work, the divergent performance of these factors raises some concerns. For example, a response may be both simple and ambiguous, resulting in a lower overall comprehensibility. However, it is better to isolate factors that measure the same element of user experience for a more comprehensive and nuanced understanding. By doing so, practitioners can pinpoint specific areas for improvement and create more targeted solutions. This approach not only enhances the accuracy of evaluations but also leads to more effective and preferred interactions, ultimately improving overall user satisfaction and interaction quality.

6.6.5 Limitations

Our study has three main limitations. First, its online nature restricts the generalizability of findings to real-world VA interactions, which can be potentially influenced by contextual factors and day-to-day use. Such examples include asking a personal health data question during a walk or conversely while sitting at home relaxing. Thus, future research should extend our findings through in lab and real-world settings for enhanced validity and reliability. Second, our participant pool consisted of individuals familiar with VAs and personal health data (i.e., potential sampling bias) While we are confident that focusing initially on this demographic provides insights into general results, we acknowledge it may not fully represent the broader population for whom VAs could be used. To address this, future research could include participants with varying levels of VA familiarity and increasingly diverse demographic backgrounds (e.g., older adults). Third, while Study 2 focused on insight categories, only Value-based responses were utilized. While this aligned with the study's goal, it does leave other combinations of insight category and response type to still be evaluated which in turn could provide increasingly fine-tuned guidelines for VA answers. Furthermore, while our study offered heart rate and step count as data types of choice throughout each study, to engage participants within the study, we recognize that questions pertaining to specific data types could result in different desired answers structures. Future work could perform a comparative study across the many data types captured within one's personal health data. Our studies offer initial insights into answer structures for VA interactions involving personal health data questions. By exploring diverse response types and insight categories, notably applicable to any data type, we lay the groundwork for designing and developing VA interactions involving personal health data.

6.7 Summary

Through the use of a custom-built browser-based pseudo-voice assistant (VA), this chapter investigates differing answer structures in response to personal health data queries. Combined with knowledge gained from the previous chapters, and known guidelines for smartwatch visual output design, we now have the tools necessary to begin building a smartwatch application which can appropriately handle a range of personal health data queries desired by people.

Specifically in this chapter, two user studies involving a total of 82 participants were conducted, during which participants interacted with our VA, posing questions and ranking their experiences and preferences of three distinct answer structures: Minimal, Keyword, and Full Sentence. We provide empirical findings that reveal a notable preference for full sentence answers, which consistently demonstrated higher quality, behavior, comprehensibility, and efficiency across various response types (Open, Range, Binary, and Value) and personal health insight categories (Contextual, Preemptive and Proactive, Goal and Performance, Combination and Comparison, Historical and Trend, and Current Status). These results come at a contrast to previous work which explore answer structures for general VA use. Our results suggest that full sentence answers offer less ambiguity, and despite their longer response time, full sentence answers were perceived as equally efficient. Along with other findings, such as a desire for VAs to be efficient and technical rather than social entities (e.g., as a fitness coach), we provide design implications in line with these results that offer insight into future VA systems handling personal health data queries.

Chapter 7

DATAWATCH: Enabling In-Situ Databiting on the Smartwatch Leveraging Touch and Speech

---

In this chapter, we present our most focused approach to databiting on the smartwatch. Building on the knowledge gained from previous chapters, we now explore the use of databiting for in-situ exploration of tracked workout data through a custom-built smartwatch application, DataWatch. DataWatch specifically targets tracked workout data, which holds the greatest potential for in-situ exploration, as seen in Chapter 4, and was the most frequently queried type of data, as noted in Chapter 5. Work presented in this chapter has not been submitted to a venue, however, we do anticipate submitting this work in future. Building of the DataWatch application, study design, and analysis are entirely my own work. However, discussions surrounding the work put forward in Chapter 3 and Chapter 4 influenced aspects of the DataWatch application and study methodology.

7.1 Introduction

We present DataWatch, a novel and custom-built smartwatch workout tracking application designed to facilitate databiting—lightweight and transient data exploration—specifically for tracked fitness data while in-situ. We designed and developed DataWatch (Figure 7.1) to enable users to interact with their fitness data directly on their smartwatch, supporting three types of exploration: 1) *Single Value* – displaying aggregated textual values, 2) *Browse* – showing charted values from past workouts, and 3) *Compare* – comparing past workout data with current metrics. Users can explore data during three workout phases: 1) *Before*, 2) *During*, and 3) *After* a workout. Additionally, DataWatch supports two filtering methods: 1) *Temporal* and 2) *Activity* filtering.

Smartwatches, due to their limited screen space, cannot or do not accommodate touch-based controls and options alongside their glanceable visualizations. For instance, the current Apple Watch applications restrict users to viewing data by predefined time segments, such as the current day or week. A lack of screen space makes it impractical to navigate to separate pages to adjust widgets, options, and query types. For example, time-based interactions, such as entering specific dates, times, and ranges, are tedious on smartwatches, leading most current health applications to limit exploration to static glanceable visualizations.

Yet, in certain everyday situations, such as when people are in-situ, having the flexibility for lightweight and transient data exploration directly on the smartwatch can greatly benefit individuals. As such, and to address the challenges of touch noted above, DataWatch draws inspiration from previous research advocating the benefits of multimodal interaction [98]. Specifically, we incorporate speech as a primary input modality to overcome the limitations of touch-based controls previously mentioned. Speech-based interaction takes little screen space, is flexible, and allows users to express their intent easily [8, 42], making it an ideal solution for smartwatches.

To better understand how such databiting capability is perceived and utilized, we conducted an exploratory, observational, study with 12 individuals who track workouts regularly and are interested in their personal fitness data. Participants in our study downloaded DataWatch to their own Apple Watch, and used DataWatch to track their workouts over the period of one week, while also databiting when desired.

Our contributions are three-fold:

Figure 7.1: DataWatch supports in-situ exploration of past tracked workout data through multi-modal interaction. People can track and control their workouts while viewing their currently collected metrics ①, as is typical with current smartwatch health applications. Different to these applications, people can then at any time long-press elements on screen ② (elements which can be long-pressed, and thus queried, are denoted with a blue dot) to initiate a spoken query regarding the interacted element ③. Upon providing a query, DataWatch processes the query to show the resulting output to the person ④. In this Figure, a databiting instance involving *comparative* exploration and *temporal filtering* is queried *during* a person's workout. Please note, screens which appear beside and below the smartwatch are screens which can be swiped/scrolled to, from the current displayed screen.

C1: The design and implementation of DataWatch which represents a novel approach to smartwatch data exploration. To our knowledge, DataWatch is the first app to leverage touch and speech to enable exploration of past tracked workout data directly on the smartwatch. It facilitates databiting—lightweight and transient data exploration—allowing users to interact with their personal fitness data while in-situ.

C2: An empirical exploratory study was conducted with 12 Apple Watch users using DataWatch over the course of one week. Through quantitative and qualitative analysis, we provide insights into how people explore their past collected workout data, their motivations for such exploration, the perceived actionable insights and usefulness, and their general reactions to DataWatch.

C3: Reflecting on our observations, and participants' feedback, we discuss implications of DataWatch's design and provide opportunities for further development of databiting. Through this discussion, our aim is to continue to better support personal data exploration on the smartwatch in the future.

7.2 DataWatch

To enable flexible and personalized in-situ fitness data exploration on the smartwatch, we designed and developed DataWatch. DataWatch is an Apple Watch workout application that leverages speech as well as combined touch and speech interactions for lightweight and transient exploration of previously collected fitness data. In this section, we describe the design rationales for DataWatch, highlight the user interface and interactions, and discuss technical aspects of DataWatch's implementation.

7.2.1 Design Rationales

DR1: Utilize Intuitive Layouts, Interactions, and Visualizations to Support Lay Persons DataWatch targets everyday smartwatch users who track workouts and want to further explore their fitness data. As such, we recognize that users may not have expertise in data visualization or analytics. Therefore, DataWatch prioritizes intuitive design, interactions, and visualizations.

DataWatch is designed to visually and interactively mirror Apple's default Workouts application. This familiarity can greatly reduce any effort and time needed when learning to use a new application. Thus, people can focus their attention on the novel exploration capabilities, and not the interface itself. As such, DataWatch employs only one additional interaction above and beyond that of Apple's Workouts application; a long-press (i.e., touch and hold), which can be done on a variety of on-screen elements denoted by a blue dot (see Figure 7.1 ②), to initiate a query. This then brings forward a keyboard/microphone for which users can use to provide a query (see Figure 7.1 ③).

DataWatch utilizes line charts, which are common and intuitive for lay users [136], and text-based representations of data. Furthermore, text is used to highlight key insights, when queried, within a line chart. This can further simplify comprehension of data and provide immediate insight from the requested data. For example, Figure 7.1 ④ highlights the use of both a line chart and textual representations of data as seen in DataWatch.

DR2: Leverage the Complementary Nature of Speech and Touch Interactions on the Smartwatch The limited screen real estate of a smartwatch limits both the visual components that can be displayed as well as the input mechanisms afforded. To address the limited screen real estate and input limitations, we leverage natural and familiar modalities: speech as well as touch and speech. Touch interactions provide familiar and direct manipulation of on-screen elements [98], such as the metrics displayed during a workout (see Figure 7.2 ②), while speech offers a more expressive way to convey intent and a higher freedom of expression [8, 42, 182].

Our goal was to utilize these complementary modalities for lightweight and transient exploration, rather than provide complete parity between touch and speech interactions. This complementing of input modalities is best exemplified in Figure 7.2, where the touched metric is used as input for the processing of the query subsequently provided. Utilizing a combined touch and speech approach recognizes that touch alone would not suffice for databiting on a smartwatch, while also keeping databiting as close to the data of interest and without needing additional screen real estate for widgets, options, and exploratory controls. We further detail how we utilize both touch and speech in Sections 7.3.2 and 7.3.3 below.

Figure 7.2: DataWatch supports in-situ exploration *before* starting a workout. On the homepage, scrolling through the many workouts available for tracking, people can long-press on a workout type of interest to initiate a query ①. In this Figure, a databiting instance using an *activity filter* and *browse* exploration capabilities is demonstrated ②. Please note, the list of workouts available extends beyond what is shown below the smartwatch on the left, and is reduced here for space preservation.

DR3: Support Comprehensive In-Situ Exploration Reflecting on personal data using a smartwatch can occur increasingly close to the action, for which the reflection is related, to benefit on-the-fly decisions [3, 69, 107]. The term in-situ exploration is often thought to only encompass exploration of data that is currently being tracked while performing the activity tracking that data. However, we take a broader stance, considering in-situ exploration to include any interaction with data that occurs in the context of and environment where the activity takes place. As such, DataWatch supports exploration before, during, and after a tracked workout, which allows for a comprehensive in-situ exploratory experience.

Before starting a workout, people can benefit from insights that help them prepare and set goals (i.e., preparation-for-action). This preparation phase is key for setting realistic and achievable targets, ensuring users are well-informed. Figure 7.2 ① highlights this before activity querying of data.

During a workout, real-time data exploration can provide actionable insights, insights which can potentially be influential to the activity being performed (i.e., reflection-in-action [149]). Displayed metrics on screen update continually throughout the workout to show currently captured and calculated values associated with the workout. These can provide actionable information if people know how the current value displayed relates to a previously defined goal or historical data. However, if they do not, databiting can provide actionable insights previously unavailable to the person. Figure 7.1 highlights an example of during workout querying of data.

After completing a workout, reflection on the collected data (i.e., reflection-on-action [149]) allows users to understand their performance, identify areas for improvement, and track progress over time. This reflective phase is essential for longer-term motivation and continuous improvement. As such, a summary screen shows the aggregate metrics captured within the workout just completed and Figure 7.3 highlights after workout querying of data.

DR4: Enable Flexible Temporal Filtering to Support Visually-Constrained Exploration In addition to traditional temporal filtering (e.g., "The last six months", "January 1 to February 1", "Last month", "Since June 1st", etc.), supporting activity-based filtering allows people to explore and temporally filter their data implicitly. Temporal filtering, while useful, can be limiting when users cannot remember the exact dates of their past activities and when visuals on-screen do not highlight these dates for further filtering. By incorporating activity-based filtering, users can explore their data based on the last number of activities performed. This enhances

Figure 7.3: DataWatch supports in-situ exploration *after* completing a workout. On the summary screen, shown post workout, people can long-press on a metric of interest to initiate a query ①. In this Figure, a databiting instance using a *temporal filter* and *single value* exploration capabilities is demonstrated ②. Please note, content displayed around the smartwatch on the left, can be scrolled to and from.

accessibility and ease of use, ensuring that valuable data is not overlooked simply because the user cannot pinpoint when it was collected. Figures 7.2 and 7.3 highlight activity-based filtering, while Figure 7.1 highlights a temporal filter.

7.2.2 User Interface and Interaction Design

Base Functionality At its core, DataWatch is a workout tracking application. As such, DataWatch allows participants to choose an open-goal workout type to track (e.g., indoor walk, outdoor cycle, hike, swim, elliptical, yoga, etc.), see Figure 7.2 (left). Upon starting a workout, live metrics are captured and shown on screen across a variety of workout-specific metrics pages which can be viewed by scrolling up/down, see Figure 7.1 ①. Controls, including pause/resume, end, and a microphone for querying are available by swiping left from any of the metrics pages, while music playback controls are available by swiping to the right. Upon completion of a workout, the workout is saved, and a summary screen provides aggregated collected metrics of the workout, see Figure 7.3 (left).

Databiting DataWatch currently supports databiting of five captured metrics: 1) *Duration*, 2) *Distance*, 3) *Pace*, 4) *Calories Burned*, and 5) *Heart Rate*. These metrics are not only common across most all workout types and in many mHealth applications [97], but are also some of the most desired for exploration by smartwatch users according to our findings in Chapter 5.

Of this data, people can conduct databiting using three types of exploration: 1) *Single Value* - provides an aggregated value textually represented on-screen; see Figure 7.3 ②. People can access single value exploration through keywords such as “what is”, “what was”, “how many”, etc., 2) *Browse* - provides a line chart of all queried data as well as any supplemental aggregate information through text and a corresponding dashed line within the chart; see Figure 7.2 ②. People can access Browse exploration through keywords such as “show”, “explore”, “highlight”, etc., and 3) *Comparison* - provides similar exploration to that of Browse, however, also includes the currently captured metric; see Figure 7.1 ④. As a comparison chart is displayed, the currently captured metric is highlighted as a larger line mark and updates live within the chart as the captured metric changes. People can access Comparison exploration through keywords such as “compare”, “combine”, etc.

Query Input The use of each of touch and speech input modalities can be accessed throughout varying parts of DataWatch. Speech alone can be initiated from the microphone button located in the controls screen during a workout, see Figure 7.1 ①. To utilize touch and speech, a person can long-press any on any interactive element denoted by a blue dot (i.e., an activity before a workout or a metric during and/or after a workout, see Figures 7.1 ②, 7.2 ①, and 7.3 ①). Importantly, DataWatch recognizes the element touched and incorporates this into the spoken query; this can be seen in the example in Figure 7.1 ② through ④, where heart

Figure 7.4: DataWatch's input screens are shown demonstrating live query input. By default the keyboard is brought forward (left), which can be switched to dictation capture (right) by selecting the microphone icon at the bottom right of the screen ①. Input can either be cancelled or sent for processing through the controls at the top of either input screen ②.

rate is not mentioned within the query yet is understood as the data desired for exploration due to the initial touch interaction.

Furthermore, DataWatch recognizes the specific workout type being interacted with, currently tracked, or having just been tracked, when databiting occurs. This applies before starting a workout when a query is initiated by long-pressing on a workout type, and during or after a specifically chosen workout. This workout recognition within a query is best exemplified in Figure 7.1 ④, where DataWatch shows data from only walking workouts within the time period provided. This level of recognition allows people to focus on data pertinent to a particular workout type while ensuring that the databiting process is in-situally relevant.

Upon initiating a query, a keyboard is overlaid on the currently visible screen. By default the keyboard is shown, allowing for a natural language query to be typed out, while the microphone button can also be tapped to initiate speech input; see

Figure 7.4 ①. When using either the keyboard or microphone, the query entered by the participant is seen live as they provide input. When finished, a user can click “Done” which sends the query for processing and resolution, or, if desired, the person can cancel the input to go back to the screen from which the query interaction was initiated; see Figure 7.4 ②.

Feedback When DataWatch can process the query and is able to resolve the data needed, the exploration result is overlaid on top of the currently displayed activity, metric, control, or summary screen. To end the exploration, the person taps on “Close” (see Figures 7.1 ④, 7.2 ②, 7.3 ②). Conversely, when DataWatch is unable to process or resolve the query, a contextual message is overlaid. After reading the message the person can tap “OK” (see Figure 7.5) to go back to the screen from which they initiated the input.

Figure 7.5: When a provided query is either not valid or resolving the data was not successful, a contextual message is displayed. Three potential messages are shown here, while others exist depending on the issue faced.

7.2.3 Interacting with DataWatch

We illustrate DataWatch’s interface and interactions through a storied usage scenario: Sam is starting their day with an indoor walk on their gym’s track while using DataWatch to track their activity. We note, this scenario does not encompass all queries possible.

Before Starting the Walk Before starting his walk, Sam uses DataWatch to get motivated and prepare-for-action (as illustrated in Figure 7.2). He opens DataWatch and rather than tapping on Indoor Walk to get started, he first taps and holds on the desired workout. He then states "Show me my longest walk in the last 10 walks." DataWatch quickly handles his command and displays that his longest walk from his last 10 was 24 minutes. Feeling good today, Sam is motivated by this data to surpass his previous record and wants to aim for 30 minutes.

During the Walk A few minutes into the walk, Sam is curious about his heart rate's performance during a specific period (as illustrated in Figure 7.1). First, he notices his current heart rate is around 99 BPM. Curious about how this compares to the latter part of last year when he was more regularly at the gym, Sam taps and holds the heart rate metric and simply says, "Compare to my average from July 1st to Dec 31st." The resulting chart shows his current heart rate is roughly alongside his average but on a slight downward trend. Realizing he isn't pushing himself as much as he used to, and could, Sam decides to increase his effort to reach the higher values he saw earlier in the explored period. Throughout his walk, he keeps an eye on his BPM, ensuring he pushes himself to his target value. This reflection-in-action helps Sam adjust his effort to meet his new-found goal.

After Ending the Walk After completing his walk, Sam becomes curious about his recent performance while scrolling through his summarized metrics. Sam sees his distance metric and decides to tap and hold to ask a question (as illustrated in Figure 7.3). Unsure of exactly when he has been to the gym in the last week or so, Sam asks DataWatch, "How far have I walked in the last five walks?" The response indicates he has walked 7.2 km in total. This was lower than Sam expected; he had hoped it would be higher. Taking this reflection-on-action as motivation, he sets a new goal for his next walk: to walk 2.8 km, which would bring his total to 10 km over the last six walks. This quick post-workout reflection helps Sam set future targets and stay motivated to further improve his fitness.

7.2.4 Implementation

DataWatch leverages the Apple Watch and watchOS platform, specifically designed for Series 4 watches onwards running watchOS 9.0 or later. Thus, DataWatch can run on 10 generations of Apple Watches dating back to 2018; this is in fact only limited by Apple's own watchOS restrictions which inhibit older watches from installing watchOS 9.0, required for some of the framework access utilized within DataWatch. DataWatch incorporates two parts: a primary watch application, which handles workout tracking, user queries, and interaction, and a companion iPhone application, which handles the query processing and resolving in the background. To ensure compatibility with accompanying iPhones, iOS 17.0 or above is required. Both applications were built entirely in the Swift programming language.

Apple's HealthKit[11] framework is used to read and write workout data within DataWatch while maintaining user privacy and control. Prioritizing user privacy, DataWatch first seeks permission to read and write workout data. Furthermore, people have the ability to later rescind or allow permission of this data if desired.

Since all personal tracked data is stored on the user's iPhone, the current version of DataWatch requires the iPhone to be unlocked to resolve a query. This is an Apple-imposed security measure to protect personal health data. Only the most recent week's data is stored on the Apple Watch, which is insufficient for most queries. Therefore, unlocking the iPhone grants access to the necessary data. Note that the companion iPhone app operates in the background and does not need to be opened for DataWatch to function.

To enable natural language interaction, a custom Natural Language Processor (NLP) and Resolver (NLR) were built. When a user interacts with DataWatch, their query, and additional contextual information about the workout and interaction, are securely transferred to the iPhone via Watch Connectivity[12]. Once transferred to the iPhone, the NLP parses the query using part-of-speech tagging, identifying parameters such as the desired data source, query intent (through interrogatives used), data filters (temporal and activity), and aggregations. Chrono, a natural language

[11] `https://developer.apple.com/documentation/healthkit`

[12] `https://developer.apple.com/documentation/watchconnectivity`

time-period parsing library originally built in JavaScript[13] and modified for Swift[14] was further custom-modified for use within DataWatch to handle temporal filtering of data. Following the NLP's completion, resulting in a set of found tags, an NLR analyzes the extracted tags alongside the current watch screen information and user interaction details to then fully resolve the interaction, lookup the necessary data, and perform any needed calculations. Finally, the results are shared back to the watch, again through Watch Connectivity, for appropriate display.

7.3 Study

7.3.1 Recruitment

We recruited participants from Reddit and through advertised posters on our University's two campuses. We advertised our study across subreddits relating to personal health, apple watch use, as well as a number of general city subreddits across Canada. Our inclusion criteria were those who (1) were aged 18 years or older; (2) own an Apple Watch Series 4 or newer with watchOS 9.0 or higher installed, and have a paired iPhone; (3) have access to Zoom and a stable internet connection; (4) are native English speakers; (5) currently live in Canada; (6) use the Apple Workout application for at least one of the following workout tracking (indoor walk, outdoor walk, indoor run, outdoor run, elliptical, hike, indoor cycle, stair stepper, rower, yoga, functional strength training, core training, Pilates, dance, tai chi, cool down); and (7) have been regularly collecting personal health data on their Apple Watch for at least three months.

In appreciation for their time and effort, participants were offered up to $20 CAD. The amount a participant received was not tied to the number of responses, but rather their overall participation. We provided the equivalent of $5 CAD for attending the introductory session, and another equivalent of $5 CAD for attending the final interview. During the week long data collection, we added an additional equivalent of $10 CAD if the participant used DataWatch to track a workout at

[13] `https://github.com/wanasit/chrono`

[14] `https://github.com/quire-io/SwiftyChrono`

least once. Notably, this was not tied to databiting instances, rather only general app use. We provided compensation after the final interview, or upon withdrawal, in the form of an electronic Amazon.ca gift card.

7.3.2 Study Procedure

Our study included three stages, similar to that of the study conducted in Chapter 4: an introduction and tutorial session, a seven-day in-the-wild data collection period, and a final interview. The procedure and study materials were iterated upon during two pilot participants with people who were recruited in the same means as our participants, thus meeting our study's inclusion criteria. All participants provided consent at the start of the study, in accordance to our approved ethics protocol reviewed by our institution's ethics review board.

Introduction and Tutorial Session.

To start the study, the participant joined a Zoom meeting, ~30 minutes long, where we introduced and acknowledged their interest and participation in our study. Participants were encouraged and asked to interject with any comments and/or questions during the meeting. The researcher shared presentation slides (please refer to the supplemental material) via the screen sharing functionality. The goal of the project was expressed to the participant along with other important remarks before they completed a demographic survey. Then, the researcher guided the participant through the installation and setup process of the application on the participant's own Apple Watch to be used throughout the data collection stage of the study. The researcher then gave participants a walk-through of the application through a live demo from the researcher's Apple Watch which was mirrored to the researcher's iPhone and the iPhone's screen shared on Zoom. Upon completion of the walk-through, the researcher reiterated the general functionality of DataWatch through slides, also highlighting queries that were not possible. Finally, participants tried a query on their Apple Watch to ensure everything was working and any technical issues related to interaction, data access, etc. were handled.

Data Collection.

Participants used DataWatch over the course of the next seven days to track their workouts as usual, while being encouraged to use the databiting capabilities as desired. We did not require a minimum number of databiting instances throughout the study, so as to not elicit forced uses and queries from participants. Due to the potential for participants to be active (e.g., during a run) when a desired databiting instance arose, we instructed them to only provide a response through the application when it was safe for them to do so.

Final Interview.

After the final day of data collection, a Zoom meeting was held where a researcher conducted a semi-structured interview with each participant. The meeting was audio recorded and later transcribed. The goals of the interview were: (1) to garner general reactions and experiences of using DataWatch, (2) to explore perceptions towards the use of speech as an input modality for personal health data queries, (3) to discuss insights gained as a result of databiting, and (4) better understand motivations behind databiting in-situ. To aid in recollection, a report of each participant's queries were shown to them via Zoom's screen sharing functionality. Finally, the researcher answered any remaining questions from a participant, thanked them, and provided the compensation. Each interview took ~30 minutes.

7.4 Results

Table 7.1 summarizes the demographic, smartwatch usage, and health data collection, workout tracking details, as well as databiting counts of our study participants. Of our total participants, one was removed from analysis as they did not use DataWatch to track a single workout throughout the week of data collection, nor use it for databiting even prior to a workout. Of the remaining 12 participants, participants were aged from 20 to 44 (*M* = 27.9) and held a range of occupations. At the time of conducting the study, participants had collected personal health data for an average of 46.2 months (*SD* = 41.6 months) and had used a smartwatch for an average of 36.0 months (*SD* = 33.0 months) while wearing the smartwatch for an average of 6.6 days per week (*SD* = 0.8 days). Participants largely tracked walks, runs, hikes, cycles, and functional strength training workouts, while including other workouts such as swimming, soccer, elliptical, Pilates, dance, yoga, core, and high intensity interval training.

In the following subsections, we report on quantitative results from the use of DataWatch, as well as emergent themes found from our qualitative analysis of the interview scripts.

Table 7.1: Summary of demographic information, health data collection, smartwatch usage experience, and number of tracked workouts and responses reported from our study participants.

Alias	Age	Gender	Occupation	Health Data Collection	Smartwatch Usage (Total / Days per Week)	Number of Workouts Tracked	Databiting Interactions
P1	22	M	Student	0y 8m	0y 8m / 5	2	0
P2	21	M	Student	7y 4m	3y 9m / 7	3	13
P3	38	M	Consultant	6y 4m	6y 4m / 7	2	13
P4	44	M	Professor	13y 0m	10y 0m / 7	7	4
P5	23	M	IT Administrator	0y 9m	0y 8m / 5	9	11
P6	38	F	Student	2y 0m	1y 0m / 7	9	28
P7	39	M	Senior Solution Architect	4y 1m	4y 1m / 7	3	4
P8	22	M	Student	0y 4m	0y 4m / 7	11	8
P9	24	F	Communications & Program Coordinator	4y 0m	4y 0m / 7	7	8
P10	20	M	Student	1y 9m	1y 9m / 7	2	17
P11	23	F	Research Assistant	2y 6m	1y 0m / 6	2	7
P12	21	F	Sales Associate	3y 5m	2y 5m / 7	1	5

7.4.1 Interaction Usages

Within our study, 58 total workouts were tracked using DataWatch (*Avg* = 4.8, *Min* = 1, *Max* = 11, *SD* = 3.4), including walks, runs, cycles, gym sessions, core training, and rowing. In and around these workouts, a total of 120 databiting interaction attempts were made (*Avg* = 9.8, *Min* = 0, *Max* = 28, *SD* = 7.1). Figures 7.6, 7.7, and 7.8 highlight all databiting interactions for each participant. Notably, this distribution of databiting interactions follows very closely with the results found in Chapter 4, see Table 4.3 the first row corresponding to Physical Activity.

Among the 120 databiting interactions, no critical errors occurred (i.e., a message or result was always provided). However, 18 (15.0%) databiting interactions failed due to uncertainty of the term *length/long* and *much* (i.e., "How *long* was my last workout" - DataWatch requested that a specific data type be included as *long, length, and much* have connotations of both distance and time metrics), 12 (10.0%) failed due to an unresolved error related to queries that included an edge case. This edge case filtered out the most recent workout through activity-based filtering, such as (*P6* - "What's my highest heart rate during the *last exercise*"), 14 (10.8%) were simply unsupported queries (i.e., comparisons before a workout where no current metrics were available to compare, or using multiple aggregations in a query). As a result, 76 databiting interaction attempts were successfully executed.

Input, Processing, and Exploration Time Input took an average of 12.7 seconds (*Min* = 5.2, *Max* = 44.4). This includes the time from when the keyboard was brought forward on the watch, to when the participant clicked *done* to send the query for processing. Then, our processor and resolver worked quickly, processing and resolving the queries in an average of 0.3 seconds (*Min* = 0.01, *Max* = 8.5). Finally, participants viewed the resulting visualization for an average of 7.4 seconds (*Min* = 1.4, *Max* = 30.4). Combining times for input, processing and resolving, as well as viewing results, we found a total average exploration time of 21.7 seconds (*Min* = 7.4, *Max* = 58.8).

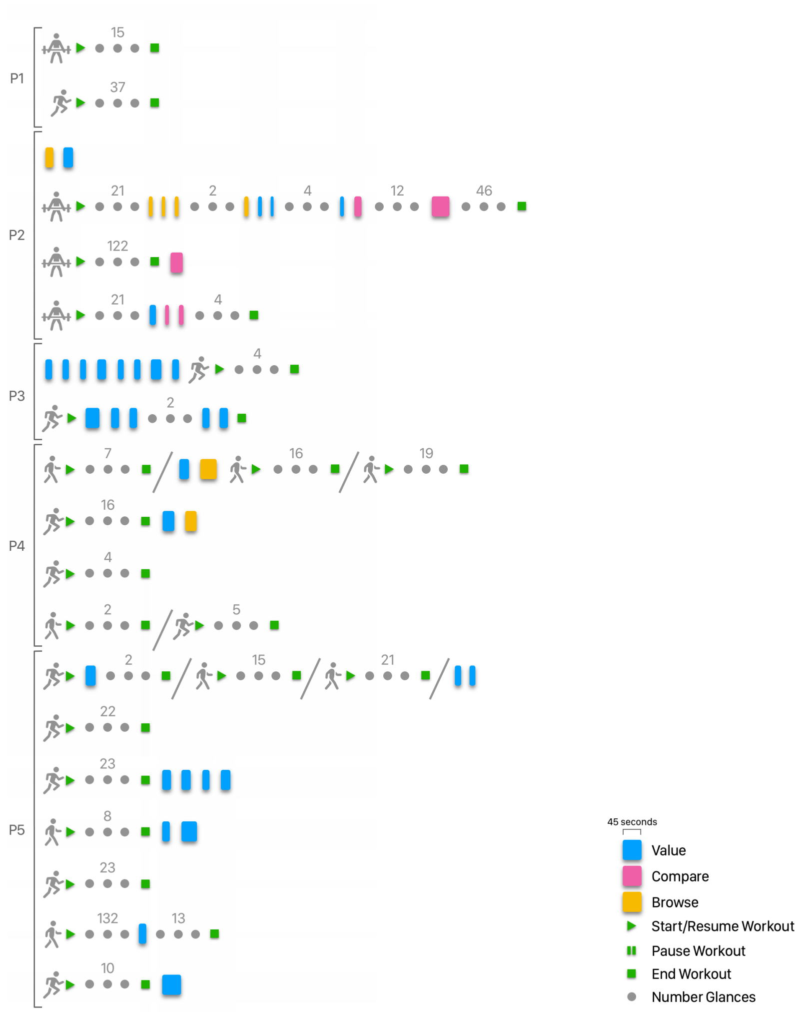


Figure 7.6: Instances of databiting (represented by colored blocks) are highlighted within workout timelines and organized by participants (P1-P5) and days of use (each row). Note that the spacing between icons and the physical length of the workouts (beginning to end) do not correspond to any specific results or times. However, the size of the colored databiting blocks reflects the total exploration time for each instance.

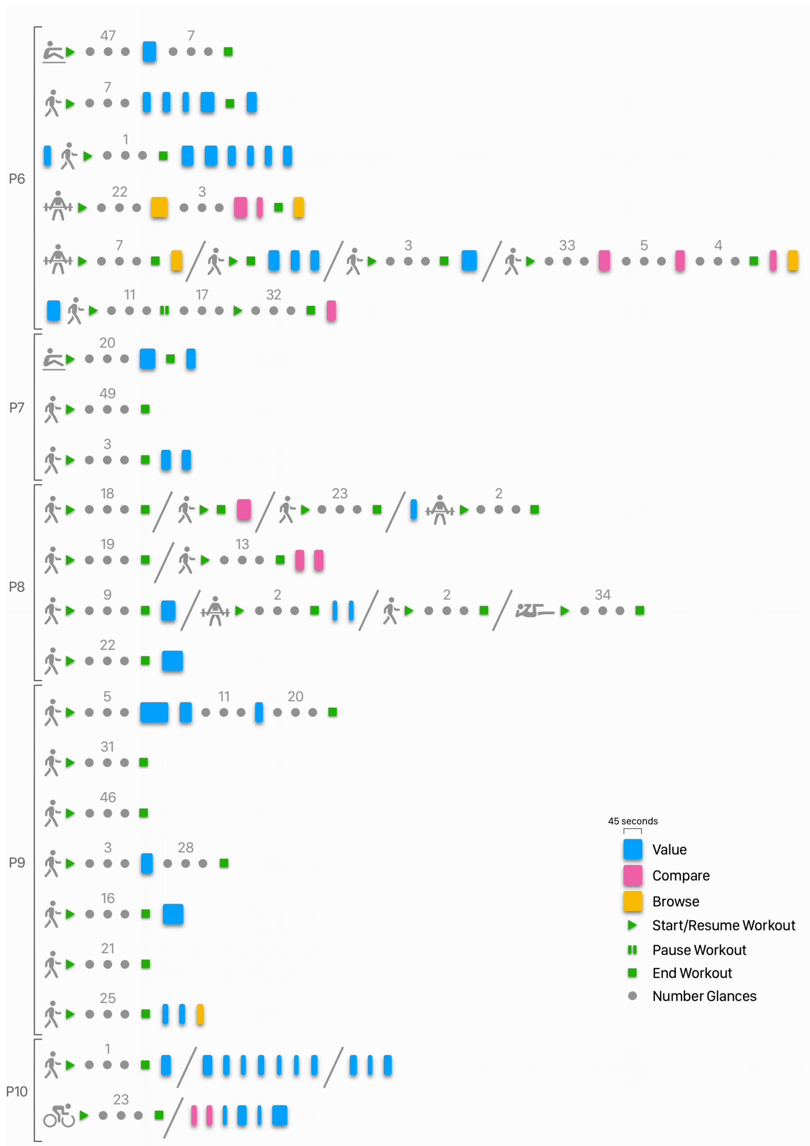


Figure 7.7: Instances of databiting (represented by colored blocks) are highlighted within workout timelines and organized by participants (P6-P10) and days of use (each row). Note that the spacing between icons and the physical length of the workouts (beginning to end) do not correspond to any specific results or times. However, the size of the colored databiting blocks reflects the total exploration time for each instance.

Figure 7.8: Instances of databiting (represented by colored blocks) are highlighted within workout timelines and organized by participants (P11-P12) and days of use (each row). Note that the spacing between icons and the physical length of the workouts (beginning to end) do not correspond to any specific results or times. However, the size of the colored databiting blocks reflects the total exploration time for each instance.

Exploration Instances To trigger input on the watch, participants almost exclusively used the activities and metrics, directly interacting with the data they were interested in exploring. However, one databiting interaction was initiated using the microphone button on the controls screen during a workout (*P5* - "Tell me my average pace from the last five runs"). This occurred early on in the participant's workout, likely when a pace was not yet displayed on the metrics screen given the limited captured metrics during the started workout (see Figure 7.6, P5's first workout on the first day).

Participants largely preferred the Value exploration type, with 89 (74.2%) databiting interactions requesting this type of visualization. Value was followed by Compare which was used for 16 (13.3%) databiting interactions, while Browse was used for 15 (12.5%). While a clear preference of exploration type was seen, participants conducted databiting instances almost equally Before (44, 36.7%), During (34, 28.3%), and After (42, 35.0%). While the majority of databiting interactions were done in-situ, some participants (P2, P5, P10, P12) also used DataWatch to explore data away from a workout (25 instances of the 120 — highlighted above in the Before counts as the list of workouts screen was interacted with, 20.8%). No clear distinction between the exploration type and time of exploration was found, however Browse exploration was slightly more often conducted before and after a workout.

Finally, we saw almost half of databiting interactions use activity filtering (53, 44.2%). Temporal filtering was used in 46 (38.3%) of databiting interactions, while the remaining required all data collected to be used as no filter was provided. There again seemed to be no pattern evident for the use of a specific filter for varying exploration types or during varying exploration times.

7.4.2 Motivation for Exploration

Two major themes arose which suggested reasons for databiting in-situ on the smartwatch: verification of performance (P3-7, P9, P10) and curiosity (P3-4, P6-8). Participants used databiting to quickly check their workout metrics before and after an exercise to first understand their performance and to then later ascertain their progress towards a goal. For example, P8 mentioned *"I mean I'd see that [time walked] and I'd be like, oh, that's interesting, 9 minutes. Usually I feel like it's 10. Let me just check how does this compare with my other times?...it is faster."* While during a workout, databiting interactions pertaining to heart rate were often used as a gauge for effort and performance. Heart rate was seen as a metric that participants could control and *"feel"* (P2, P6-8) in-situ, more so, for example, than calories burned or distance. To this effect P6 stated *"I felt that for this exercise...that my heart rate is very high in this exercise and I wanted to verify that."* P8 then went as far to say *"I only monitor [on the smartwatch] only when I feel that my heart rate is really going up."*

Curiosity, piqued by the emergence of the databiting functionality, also motivated exploration on DataWatch for a handful of participants (P3-4, P6-8), even for data types previously unexplored. P8 highlighted *"Like my workout at the gym, I don't really focus too much on my heart rate, especially when I'm not doing cardio. But I figured you know, since I have it there...I was surprised to see that it did get pretty high."* P7 added their experience with the distance metric, *"On average, walk for two kilometres. That was a surprise. I never measure distance, I just walk until I feel that I don't want to anymore."*

7.4.3 Speech Interaction for Personal Health Data Queries

All participants, except two (P9, P11) agreed that speech interaction on the smartwatch was an appropriate method for databiting interaction. It was seen as *"comfortable"* (P3), *"easy"* (P4), and a '*'very feasible option and especially given the small screen of the the watch"* (P7).

The most salient discomfort of speech interaction, noted by a range of participants (P1-6, P8, P10-12), was the physical act of raising the smartwatch to one's mouth, which can feel awkward in the absence of Bluetooth headphones. P2 summarized this by saying *"So I did kind of feel a little bit self self-conscious of like kind of talking my watch because I think it kind of looks kind of funny cause people are like, why are you talking to watch? I would say I was cognizant...I definitely was like in the back of my mind like, hey, I'm like talking to my watch right now. Just kind of looks funny and people might see me talking, but it didn't really matter that much."* P6 further added that *"I feel that I need to explain what I'm doing."* However, we postulate, the increasing prevalence of voice assistant usage through headphones mitigates this issue in future. As voice interaction continues to integrate seamlessly into daily routines, it not only enhances the practicality of smartwatches for health data exploration but also aligns with evolving user behavior patterns.

A secondary theme that arose from two participants (P5 and P6) considered workouts as a group behaviour (e.g., walking with a partner) and how speech interactions with the watch can hinder the social context in which databiting may be desired. For instance, P5 said *"Obviously, like if I'm right next to someone and they're walking with me to go somewhere I might not want to do that just to, you know, be more engaged in that conversation."* This emphasizes the importance of developing touch-only and more versatile multi-modal solutions for effective databiting on the smartwatch.

Otherwise, our discussions with participants highlighted minimal concerns about asking data-oriented personal health questions in public settings (P2-4, P6-10). Three reasons for this arose throughout our interviews. First, questions available to ask in DataWatch were not deemed to be overly intimate. To this effect P2 shared *"I* *wasn't really worried and I feel like personally the data wasn't specific or personal enough to really warrant me caring."* While this was the case for the data available to be explored in DataWatch, we recognize that other queries (e.g., queries regarding blood sugar could be viewed as more private than those about heart rate) may provide differing perceptions. Second, participants (P2, P7, P10) felt indifferent to the presence of others, stating it is not other's business nor would they likely care beyond hearing them. P7 mentioned *"It's nobody's business and if they're curious to listen you're welcome to. But I'm sure they forget about it."* Finally, it was noted that speech interaction was acceptable as long as the responses were not audibly broadcasted. P8 added *"I'm not out loud saying my whole health data, so I'm the only one that sees it in the end."* While the question was perhaps not intimate, we must recognize that the data in response to the databiting interaction remains as an intimate element.

7.4.4 Benefits of Databiting

In Chapter 3, we discuss potential benefits of databiting, including actionable insight and motivation for further exploration. Here, we discuss each of these to better understand the benefits of databiting through DataWatch.

Actionable Insight

Given the increased insight available through databiting, one envisioned benefit is that of actionable insight. While metrics on screen can provide this, it is limited, especially if a specific goal or value is not predetermined. By databiting, current metrics can be associated with past data, providing a more complete view of trends and comparisons. Throughout, over half of our participants (P3, P5-9, P12) discussed instances where databiting in DataWatch provided actionable insight beyond what was currently achievable in the commodity workout tracking applications that do not support databiting.

Some participants found that databiting encouraged them to push or reserve themselves. For example, P3 highlighted how databiting caused them to push their heart rate *"What's the average heart rate like?...it was lower than I expected, so*

I thought, oh, I can push myself more. I'm like I should be able to get that up to 130." This was also the case for P6 and P8, as well as P9 who conversely realized their peak heart rate was higher than desired and aimed to keep it within a lower range; P9 mentioned *"I could show like the highest and I think it was like 182 or something, which I was never aware of that when I was working out. I was being more mindful and then I was focused more on my heart rate."* P8 also highlighted how databiting influenced their pace and pushed them to try and go faster.

Rather than pushing or reserving a value of a certain metric, participants also found actionable insight by using resulting answer to then encourage them to maintain values. P5 mentioned *"Am I exerting myself to the degree I wanted to? I understand that I have a training plan set out, but in the moment you kind of just forget like how how much am I actually doing. While I was thinking through that in my head, I could just ask that...I felt like I got the answer very quickly and I felt that it was very helpful [to keep me on track]."*

For some, however, actionable insight was workout, and even current tracking goal, dependent (P1, P2, and P10). P1, a participant that in fact did not conduct a single databiting interaction, discussed specific queries that would be beneficial at different instances in their training, but were not currently relevant. Specifically P1 discussed *"I think seeing my maximum would be pretty helpful, yeah, because that way I can kind of get a sense of like what I've done previously...what I'm kind of capable of, and like, am I able to do more today in this particular day. I think for me it can definitely like motivate me...I'm always like oh let's maybe push it another couple like points."*

P2 discussed how databiting interactions during gym sessions were not as actionable for them, however they imagined databiting with respect to pace during a run to be influential (P2 was dealing with an injury preventing them from running during the time of the study), *"It's whenever I have something like this [databiting], it's not for the sake of interest, it's more sort of the sake of can it help me right? It's that the data that I was provided when I was doing, like a weightlifting workout, wasn't really relevant to me. I didn't really care what my heart was. I didn't really care how many calories I was burning cause that's not why I'm working out right. Something like a run, for example, would be nice to compare my* *pace. If I'm trying to keep pace, that makes sense because I can actually adjust my workout as I'm doing it. It's very workout dependent, any workout that you're going to be consistently checking your stats that makes a lot more sense to be to explore your data further.*

Further Exploration

Another benefit discussed previously is that of using databiting as a stepping stone to further explore one's data in a more traditional manner. While the majority of participants did not feel the need to further explore their data, interestingly, we found through discussion that four participants (P5, P8, P9, and P12) in fact did conduct additional data exploration due to a databiting interaction.

Each participant had varying reasons: P5 found that databiting reinvigorated their motivation to explore their data (they had previously exported data to Notion and done their own data analysis). P8 used databiting and initial curiosity to explore their pace for the first time, which in turn sparked a new-found interest in exploring and monitoring pace. They mentioned *"I actually got interested in looking at my pace, like I'm going to use that as a metric now. Because it was so easily accessible and easily seen on the app I was like, oh, wait like, why don't I check my paces for the last few months on my phone in more detail."* P12 was similar, in that databiting allowed for initial exploration, which resulted in curiosity that initiated further exploration on the smartphone. Finally, P9 further explored their data as a means to verify a performance value they uncovered through databiting while on a walk. They had found that during a databiting interaction, their heart rate was significantly lower than expected, and used their phone to further look into their collected workouts to better understand why: *"Yeah, I did have to do that [pull out their phone] because I was like this is so weird. Why is it so low? Because like whenever I do my workout my average is usually at least above 100 something, so I was like really confused. But then I pulled it up [on the phone] and I was okay that makes sense because we did like a really really slow paced walk [with a friend that likes to casually walk]."*

7.4.5 General Reactions to DataWatch

In general, all participants expressed excitement and positivity towards the databiting feature included in DataWatch, expressing their belief that it should be functionality included by default. For instance, P8 highlighted *"I think this is something that smartwatch should absolutely have. I mean, a big reason I wanted to get a watch was to not have to pull up my phone out of my pocket, especially when I'm being active or when I'm like at the gym or whatever."* While P9 further contrasted their experience with DataWatch to that of Apple's default Workout application; they mentioned *"I did appreciate the additional data that it was able to give me so that was really cool because I never really thought about it. But you actually don't get any of that with Apple's [Workout app]. So I thought it was really, really cool."*

DataWatch notably enabled quick and easy access to information, our desired goal; as P5 highlights *"After a workout, you're exhausted. You don't really have that kind of mental frame of mind to go through the past, but if you just have a quick thought and you're like, OK, I want to ask that I thought it was super helpful in that. It's more convenient, it's faster and it's less effort...I don't really want to be always going through my fitness data on my phone."* P8 also mentioned, echoed further by P4, that the interactions were well designed and suited the watch, saying *"Especially when I'm being active or when I'm like at the gym or whatever, I'd rather be able to do it [exploration] my watch, just as interactions are a lot shorter, I don't get distracted or I don't have to, like, push through my pocket or my backpack to get it [the phone]."*

However, a few drawbacks still remained and were noted by participants. First, an overarching topic that came up in most interviews was the hindrance of having to unlock the iPhone in order to conduct databiting on the watch. P3 mentioned *"It ended up a bit more frustrating just because, as you said, you know the the hardware and software limitations...oh, like your phone is not unlocked, your phone is locked, unlock it and then ask."* While this was a unfortunate aspect that could not be avoided, we do believe this is a factor that can be simply mitigated in future operating system updates. Second, it was discussed a handful of times that certain workouts simply offered too much friction for databiting during the workout itself.

P5 mentioned *"I have the thought of the question I have in mind during the run, it is just the execution of that process...there's some friction there. And so during the middle of that run, it just felt like okay maybe I'll come back to this later at some point when I have the time."* Supported by our recorded uses of DataWatch, no databiting was conducted during a run that was not immediately after starting a run or immediately before ending the run. This highlights that databiting, as currently implemented, may need to be even more lightweight for increasingly physical and cognitively demanding workouts.

Lastly, while no formal question was asked with respect to whether participants would want to continue using DataWatch, over that of the Apple Workout application, three participants (P1, P4, P10) asked us, unsolicited, if they could keep DataWatch and continue to use the app on their smartwatch outside of the study. This unsolicited interest suggests a preference for DataWatch, highlighting its perceived benefits and user satisfaction. It indicates that the databiting features and usability of DataWatch may offer a more appealing or effective experience than existing alternatives, promoting the idea of databiting and its inclusion in applications on the smartwatch.

7.5 Discussion

7.5.1 Feasibility of Databiting on the Smartwatch

Previous research has established natural interaction times for various smartwatch application categories [148, 192]. Specifically, fitness applications have been shown to have interaction times on average of 18 seconds. Furthermore, smartwatch interaction has been seen to have a maximum acceptable boundary extending to 45 seconds for tasks involving navigation and communication. In our study databiting interactions using DataWatch were on average 21.7 seconds (Min = 7.4, Max = 58.8). This finding suggests that databiting interactions not only remain well within the acceptable limit of 45 seconds, they in fact align rather closely with the typical interaction times observed for fitness applications. This adherence to established interaction boundaries, alongside the strong general reactions to DataWatch, underscores our applications effectiveness in providing a seamless and efficient experience for in-situ exploration of fitness data on smartwatches.

In addition to the quantitative and temporal aspects, all participants agreed that databiting functionality on the smartwatch was both feasible and highly desired. This consensus underscores several key points: First, databiting offers a level of exploration that accommodates a wide range of users. In our study, participants varied in their exploration experience—from those who used the smartwatch primarily for general activity tracking to others who exported their data into other programs for in-depth analysis. Second, databiting was not perceived as a hindrance to the activities being tracked. This collective endorsement for databiting on the smartwatch highlights the potential for databiting to enhance the smartwatch experience, with respect to fitness tracking and data exploration, making it a versatile tool for all users.

7.5.2 Reflecting on Design Guidelines

DR1: Utilize Intuitive Layouts, Interactions, and Visualizations to Support Lay Persons All participants agreed that DataWatch effectively emulated the Apple Workout application, making it intuitive and easy to adopt without requiring significant upfront effort. However, several areas for improvement were identified. Participants noted the absence of features such as automatic recognition of workout start and end times, as well as more complex metrics like power and cadence, which are now offered by Apple. Incorporating these features in future updates would enhance the intuitive and expected behavior of the application, further supporting its usability for laypersons.

With respect to visualization, we opted for line charts to display resulting information. Participants commented on the benefit of viewing data through the charts presented. However, our choice could be enhanced by incorporating sparklines [135] and bar charts, which may provide clearer and more accessible representations. While we aimed to facilitate data exploration through visual representation, and wanted to incorporate all data points involved with the query, the limited number of queries resulting in this type of exploration underscores the complexity of the task. Most queries requested a small subset of filtered data, however, as we saw 19 queries which still requested all captured data. Simplifying the represented information could promote more efficient and intuitive exploration directly on the smartwatch and in-situ. We may in turn also allow for a greater level of multi-modal input, where data could be algorithmically aggregated (e.g., data captured in the past year could be grouped by months) and a single bar could be tapped to provide details-on-demand. By refining our visualization approach, we can better support users in making sense of larger amounts of data quickly and effectively.

DR2: Leverage the Complementary Nature of Speech and Touch Interactions on the Smartwatch While speech interaction was widely adopted within our study, the occasional use of the phone for text input and remaining concerns of social awkwardness encourage us to think about potential areas for improving the multi-modality of DataWatch. Additionally, participants noted that speech interaction could be cumbersome at times, particularly during workouts such as running. While speech interaction is flexible, touch interaction capabilities could be improved to offer set exploration features throughout. Enhancing touch interactions could involve more detailed metrics/visualizations accessible through tapping on respective metrics. Furthermore, list pickers could allow for preset adjustments to the already on-screen visualizations. In future, implementing an increasingly multi-modal approach offers a best of both worlds, allowing all to feel comfortable exploring their data in-situ on the smartwatch.

DR3: Support Comprehensive In-Situ Exploration Our study further revealed that participants had a relatively equal interest in exploring fitness data not only during workouts but also before and after them. This was compounded by the unintentional capturing of databiting interactions that occurred away from a workout itself. This comprehensive, in-situ, exploration desire aligns with our findings in Chapter 4, where users expressed a need for continuous access to their health data throughout their daily routines. This result further highlights the need to broaden the scope of smartwatch applications beyond in-workout tracking and monitoring of metrics to encompass more holistic health monitoring and data-driven decision-making throughout the day and surrounding one's workouts.

Moreover, participants' interest in exploring data before workouts again suggests a preemptive and proactive approach to fitness data exploration and understanding, where users seek quick and immediate information from the smartwatch to plan and prepare. This finding also highlights the potential for smartwatches to support preemptive health behaviors and goal setting, enhancing user engagement beyond during and after-workout analysis.

DR4: Enable Flexible Temporal Filtering to Support Visually-Constrained Exploration Activity-based filtering was predominantly utilized throughout our study, allowing us to largely achieve our design goal of offering implicit temporal filtering through activity-based filters (i.e., ...last 5 runs). During the study, however, we observed instances where participants expressed interest in more granular filtering options beyond generalized activity instances (e.g., different running distances or cycling routes). For example, some users indicated a desire to filter data by specific running distances, such as by the last seven 5 km or 10 km routes, or by geographic cycling routes/locations. Furthering the activity-based filtering capability could cater to increasingly diverse user needs and provide more nuanced insights into fitness data trends.

7.5.3 Limitations

One limitation of the current version of DataWatch is its reliance on the user's iPhone being unlocked to access personal health data. This requirement adds an extra step to the interaction (not captured in our times reported above) and disrupts the in-situ smartwatch experience. While we believe that this limitation was not enough to impact our study findings, it was a frustration noted by many participants. In future, we simply expect to not be limited by this. Companies in control can simply offer increased storage directly on their smartwatches, provide alternative wearable authentication methods, and/or improve synchronization between devices when data is requested.

Another notable limitation is the absence of the Speech framework supported directly on the Apple Watch[15]. Because of this, DataWatch could not in fact provide auditory feedback in response to databiting queries, as studied in the previous chapter. This limitation reduces DataWatch's versatility in-situ, as auditory feedback could benefit user interaction while performing other tasks [20]. Given that the majority of databiting interactions involved single Value explorations, auditory feedback could have been structured as per our previous Chapter's findings, and been easily incorporated. Future iterations, once Speech is supported, could include this element. Including auditory responses could broaden the application's appeal, offering intuitive and increasingly multimodal interaction while in-situ.

7.6 Summary

DataWatch is a novel, custom-built, smartwatch fitness application that enables databiting—lightweight and transient data exploration—while in-situ of a tracked workout. In this chapter, we highlighted design guidelines as well as the user interface and interaction design of DataWatch, which was built from the results of previous chapters.

To better understand and support the argument for databiting as a viable exploration paradigm on the smartwatch, we then conducted an exploratory, observational study with 10 participants. Participants used DataWatch to track their workouts over the period of one-week, databiting when desired. Our findings, including many instances of databiting, underscore that databiting interactions can occur within a reasonable amount of time directly on the smartwatch and while in-situ. Moreover, unanimous participant agreement emphasizes the utility and desirability of databiting as a feature that should be integrated into smartwatch fitness applications.

[15] `https://developer.apple.com/documentation/speech`

Chapter 8

Discussion

8.1 Revisiting Lightweight and Transient Interaction for Insight Rich Exploration

Throughout this thesis, we intentionally refrain from quantifying lightweight and transient. As discussed in Chapter 3, the concept of databiting can vary significantly between individuals and contexts. This inherent variability creates a *fuzzy* boundary, making precise quantification challenging. However, one constant remains: within any given person or context, databiting tends to be more lightweight and transient compared to comprehensive data exploration. Conversely, it requires more effort and time than visually assessing a glanceable visualization. Yet, databiting must remain feasible in-situ and while on-the-go. Analogously, in terms of eating, snacking is undoubtedly more lightweight and transient than consuming a full meal yet requires slightly more effort than simply viewing and assessing food in front of you. Our goal with databiting was to offer data exploration equivalent to snacking

Without quantifying lightweight and transient, a question remains: Did we achieve our goal in enabling databiting—lightweight and transient data exploration for increasingly rich insight? We believe there are results from within our final study which support the success of databiting. Figure 8.1 highlights aspects of transient and lightweight exploration, which we provide for discussion purposes only and not as a concrete means for measure.

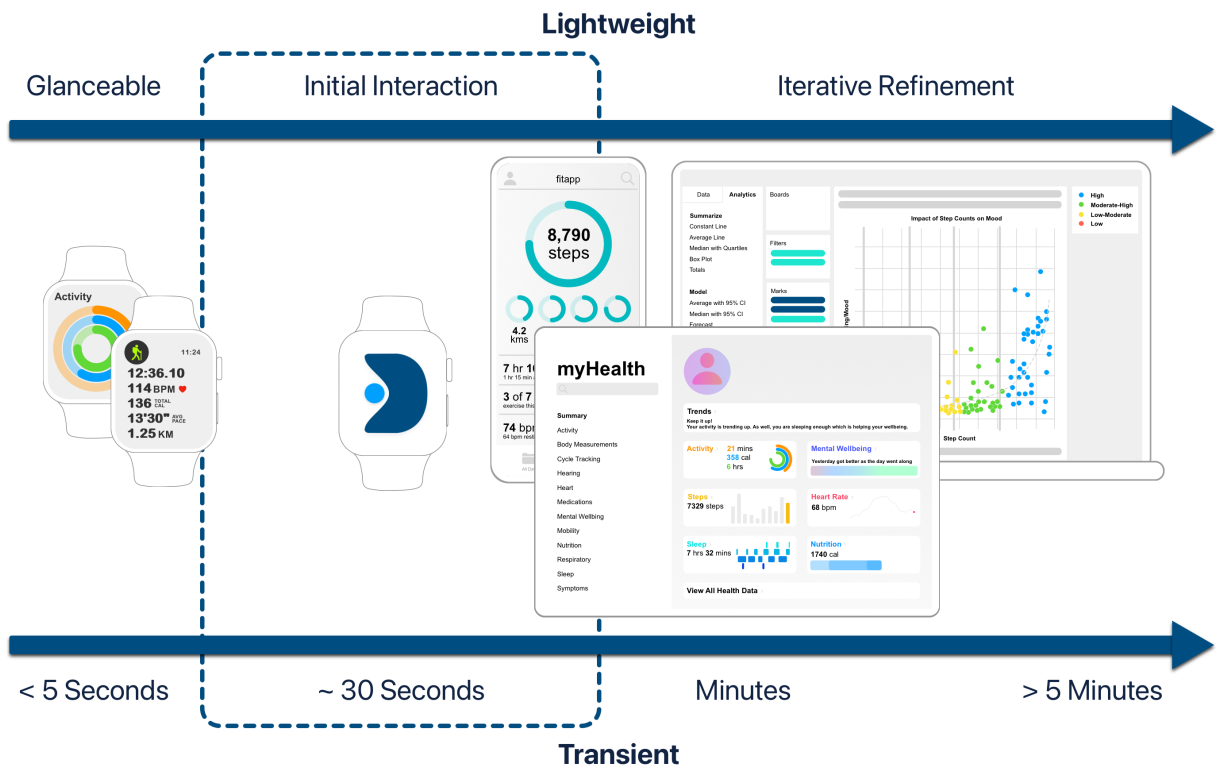


Figure 8.1: For discussion purposes we introduce the above framing of lightweight and transient data exploration. We provide this here, rather than in Chapter 3, due to the fuzzy and hard-to-quantify nature of these boundaries. As such, with this figure we aim to provoke thoughtful discussion rather than establish definitive measures early on. While we recognize that the smartphone, tablet, and laptop can provide glanceable and initial interactions we place them on the right of the figure as this is where many interactions are meant to take place.

First, transience is evident in the short duration of databiting interactions observed in our final study. Participants were able to quickly integrate their data, through input of a query, and even interpret the result without prolonged engagement. In fact, the entire databiting interaction took less than 30 seconds on average. While less transient than a glanceable visualization, importantly, this remains well within the typical interaction time on a smartwatch [148, 192] while also remaining largely under the typical interaction time seen on the smartphone, tablet, and desktop which often involves interaction that are minutes at a time [132, 132, 179].

Lightweight interactions are more difficult to quantify. As such, we recognize a lightweight interaction as one that simply does not require multiple and iterative inputs from the person. Furthermore, this interaction should not hinder the primary task at hand. From our interviews with participants, and the endorsement of databiting by all, we argue that the lightweight capability was largely achieved. However, while databiting was in fact conducted during workouts, further supporting its lightweight nature, we also recognize that the current implementation of databiting was not seen as lightweight enough to be conducted during certain workouts that require more attention and are faster paced (i.e., running). This finding suggests that further work needs to be done to reduce the effort required to achieve a databite under such circumstances. Such possibilities were discussed by our participants in our final study and include simplifying the databiting interactions through pre-meditated design of visualizations or selection of saved databiting queries that are readily available.

Our goal, given that the interaction needed to be transient and lightweight, was to ensure that people could still achieve insight rich generation of knowledge beyond what was currently possible. While some insight was actionable, participants still expressed finding new insights, verifying unknown values, and aligning their performance with their goals which was not a possibility prior given the displayed metrics on the smartwatch. This allowed users to make informed decisions on-the-go without the need for more time-consuming analysis. As a result, DataWatch successfully bridged the gap between glanceable information and detailed exploration, enhancing the overall utility and user experience of the data collected by the smartwatch.

The next logical question is how far can databiting go? In other words, how much insight can be provided within a transient time frame and without much effort. Regarding this question, we believe we are nearing the leading edge of databiting interaction. Beyond the transient and lightweight boundaries discussed in previous paragraphs, our studies in Chapters 5 and 7 reveal that value-based databiting was most commonly desired and queried. Value-based responses are intriguing as they do not offer the ability to further explore information or gather serendipitous insight afforded by a visualization. This may have been influenced by an understanding of the smartwatch's limitations by our participants, where they recognize other devices as more efficient for such exploration. Together, considering the temporal databiting findings, the lightweight maximum that was seen as too much during certain workouts, and the value responses, this indicates that we have likely found a reasonable maximum for databiting, providing quick, easy, actionable insights while maintaining ease of use and efficiency on the smartwatch.

8.2 Towards Speech Interaction for Personal Health Data Exploration

Through DataWatch's speech interaction capability in Chapter 7, and that of our elicitation app in Chapter 4, spoken queries were integrated into everyday scenarios of our participants. This included being at the gym, walking alongside others, and even in public settings such as a sports match. Notably, and for the large majority, we found a significant shift towards a new, more positive, view of using speech as in input modality for providing personal health data queries. This is largely impactful, as our studies were conducted in-the-wild using working prototypes. Previous work that has found opposite, more negative, viewpoints has largely been conducted in isolated, in-lab, or survey environments [55, 98, 116].

This shift was largely driven by a few common ideologies: First, bluetooth headphones and interactions with voice assistants are becoming common place, making the act of interacting less socially awkward than previously noted. Second, the immediate benefit and ease of interacting through speech, especially while in-situ, was recognized by participants and outweighed drawbacks that may otherwise arise. Third, data exploration of common metrics, as long as the answer is not made aware to the public, was not seen as personal enough to cause concern. For example, simply asking if one's heart rate has been above their average is not a query that gives away valuable and personal information.

However, it is important to recognize that this positive attitude towards speech interaction is not universal. Half of our participants in our elicitation study in Chapter 4 still had reserved privacy concerns, while one participant in our DataWatch study in Chapter 7 did not use speech at all. Future research should continue to explore ways to address these concerns, perhaps through increased multi-modal interactive capabilities.

8.3 Semi-Longitudinal Study Methodologies

The choice of a *one-week* in-the-wild study methodology, used twice within this thesis, was primarily motivated by the need to balance *participant engagement* with the feasibility of *accurate recall*. Pilot testing revealed that in one week participants could comfortably integrate the study application into their daily lives, while periods longer than one week led to noticeable memory recall issues. Therefore, a one-week study provided an optimal balance, allowing for natural use and reliable data collection. Below we discuss benefits and drawbacks that were either explicitly considered or a resulting lesson of the methodology used.

Benefits A variety of benefits, discussed below, arose from use of a one-week in-the-wild study methodology, which were underscored by results presented in previous chapters.

First, conducting a one-week study offers significant advantages in terms of feasibility and practicality compared to increasingly longitudinal studies. The shorter, one-week, duration is easier to manage and demands fewer resources, which benefits both us as researchers and the participants who are giving their time. With a one-week duration, the need for intermediate check-in sessions and prolonged support is minimized, simplifying the study process. Logistical challenges, such as scheduling and maintaining participant compliance, are also reduced, making it easier to coordinate and conduct the studies. Additionally, the shorter study period lowers overall costs, making it a more economical choice. Yet, considering these benefits, a one-week study remains to ensure that the necessary data is collected without overhead unnecessary to the shorter-term goals of the projects.

Second, the practicality of the one-week study duration allows for quick and frequent iteration of the study design. As demonstrated in the pilots conducted in Chapters 4 and 7, the full and complete study methodology was able to be employed, providing us with comprehensive insights regarding not only the introductory tutorials, but also the technical limitations of the applications, and the structure and flow of our interview. Together, having a complete understanding of the study methodology, and even *true* data (i.e., data captured in the exact means our study participants would report data) from which to analyze, we were able to iterate components of our study design in a timely, yet complete, manner.

Third, a one-week study duration remained to allow for participant recall of interactions and experiences. Over a longer duration, recall issues can set in, leading to less reliable quantitative reporting. In fact, during the interviews at the end of the week, we observed that some participants occasionally needed a brief moment to recall contextual information related to an interaction, query, or use of the applications that happened earlier in the week. While no complete recall issues took place, we recognize the one week duration as bound to consider for recall from participants using an application throughout their daily lives. By limiting the study to one week, and importantly providing details of responses and interactions within the interview (an element that was iterated upon in the study conducted in Chapter 4 in part to the practical time frame used), participants were better able to accurately recall details about their usage patterns, behaviors, and experiences with the study applications. This overall recall accuracy results in higher quality data, as participants' reports are more likely to be precise and reflective of their true interactions. Consequently, the qualitative insights gained from our one-week studies can be considered reliable, allowing us to be confident in the results reported.

Finally, we believe that a one-week period strikes an optimal balance for participant engagement, as it is long enough for users to become familiar with the novel application they are using, beyond that of the tutorials provided, while also providing a chance to fully incorporate it into their daily lives. In fact P9 from our final study supported this aspect by discussing *"Yeah, I think my first couple of ones were just like, okay, yeah, let's just play around and see. But then, afterwards I guess, the numbers I was looking for were a little more meaningful."* We believe that the relatively short time frame also helps to maintain participant interest and motivation, as it avoids the fatigue and disengagement that can occur in longer studies. By keeping the study period concise, participants remain focused, interested, and committed; of our 30 total participants across the studies in Chapters 4 and 7 we only had one participant drop out of the the study in Chapter 4, due to external reasons which they felt impeded their ability to participate.

Drawbacks While there are a variety of benefits, a one-week in-the-wild study methodology can have drawbacks which we discuss below.

First, while a one-week period provides valuable insights into initial and short-term databiting interactions, it may overlook long-term usage patterns and/or the effects of sustained use over extended periods. For instance, changes in usage frequency (i.e., number of databiting interactions during a workout), or shifts in user preferences as interactions find a norm, may not come to be within just one week. This is an acknowledged limitation (which we further discuss below in our thesis limitations discussion). Increasingly longitudinal approaches are needed to comprehensively understand how databiting evolves over time and the benefits that are thus derived. Balancing the immediate understanding of short-term insights with the need for long-term understanding is a key direction for future work which can aim to provide a more holistic and comprehensive view of user behavior as it pertains to databiting.

Second, the novelty effect present when using a new application, or simply being involved in a study itself, presents a drawback that must considered. Initial enthusiasm and curiosity can influence participants' interactions and behaviors during the early stages of using a new app or device. Examples of this within our work may include elicited queries which were provided simply due to participation in the study but were not queries that regularly arise, or databiting interactions in

DataWatch which would not be initiated beyond the novelty period of using the application. Some participants, especially those that may have conducted grouped exploration away from, before, or after a workout, are likely not to show this behaviour over an extended period of use (in Chapter 7—P3, P6, P10). Consequently, data collected during the one-week study periods may not accurately reflect long-term, habitual, usage patterns as previously discussed above.

To mitigate this drawback within our study design, we had participants interact with DataWatch during the tutorial session (an aspect that was iterated upon to support participants in their novelty periods); often, many started a workout in DataWatch to view the metric screens, while we also encouraged participants to ask questions both of us and to the DataWatch application (of which the data was excluded from our analysis). We further provided remuneration that was not tied to a number of responses, encouraged use only when it was deemed beneficial, and reinforced the notion that interactions were not being evaluated against others in the study. Combined, we believe this allowed for immediate impacts of novelty to be mitigated, however, we must remain mindful of this potential bias and consider common strategies to mitigate its impact. Further strategies for mitigation include conducting follow-up studies after the novelty has worn off or adjusting data analysis to account for initial enthusiasm outside a certain bounds.

Finally, capturing data in-the-wild poses unique challenges, especially when we consider our desired lightweight and transient nature of databiting. As such, a handful of elements were discussed throughout the creation of our final study designs. Natural use was deemed as paramount; we felt that interactions during the studies should act as close to the desired interactions as possible. Consequently, intrusive methods such as frequent surveys (even if short) after every databiting interaction was to be avoided to maintain the authenticity of user interactions and perceptions of use from participants. Furthermore, the number of questions asked within our elicitation study were in fact reduced to better coincide with in-situ use. Achieving comprehensive data collection is of course crucial to understand our research questions. As we are losing information that can be gained immediately and in-situ, a blend of quantitative and quantitative insights, including through the use of semi-structured interviews, ensures a comprehensive understanding of user experiences. To further supplement the data collection, we could in future include notifications for supplemental data capture throughout the day, when may be more appropriate, reflecting back on use since the last reporting. These mixed methods approaches can further enrich data capture in-the-wild while not disrupting the intended interactions.

8.4 Thesis Limitations and Future Work

Chapter-specific limitations and future work were highlighted within each previous chapter where applicable. Furthermore, Chapter 3, Section 3.3, highlights in detail future research directions in pursuit of enabling databiting. In this thesis, we largely focused on the interaction modalities research challenge, necessary to first enable and validate such databiting exploration on the smartwatch. As previously discussed, other future research directions include 1) incorporating contextual factors, 2) the relationship between databiting, glanceable visualizations, and longer-form exploration, 3) personalization and customization, and 4) evaluation challenges. Here, we provide more general thesis-level limitations and future work as a result of these limitations.

8.4.1 Who is Databiting?

Throughout this thesis, an absent challenge is that of *who* is interested in databiting. More specifically, we do not focus on how varying populations may require different databiting interactions to explore data, and even the queries that are of importance. As this is early work aimed at first understanding and then building this form of lightweight and transient data exploration in-situ on the smartwatch, our focus is on the larger general population that has been collecting data for a minimum of three months. This criterion ensures that participants have some knowledge of their needs and data, providing a more informed basis for our research.

However, this approach notably excludes individuals. As example, those who have only recently purchased a smartwatch may have very different health queries given their initial exploration of the smartwatch's capabilities. Moreover, those being introduced to the collection of their personal health data may have varying queries and databiting needs. Additionally, our work does not account for specific groups such as athletes, who may have unique performance-related queries, or older adults and other individuals with chronic health conditions who might prioritize different types of data. By not addressing these varied user groups, our work may not fully capture the diverse range of queries and needs for databiting across different demographic groups.

Future research should aim to explore these differences to create more inclusive and tailored databiting solutions that cater to the specific needs of a broader population. This includes conducting studies with diverse demographic groups, considering factors such as age, fitness level, health status, health goals, experience collecting and exploring data, socioeconomic background, and cultural context. By doing so, we can ensure that databiting interactions evolve to meet the varied and specific needs of everyone, making data exploration on smartwatches more accessible and impactful.

8.4.2 Study With Apple Device Users

Throughout this thesis, we focused our study on Apple Watch users. This decision was driven by two main factors: First, the time-consuming nature of developing for individual watch platforms led us to choose a single platform. Unlike smartphones, where frameworks like React Native can be used across multiple platforms (e.g., Android and iOS), smartwatches do not offer similar cross-platform development tools. Second, we chose the Apple Watch because it has the largest user base in North America, making participant recruitment easier. Although this decision excludes users of other smartwatch brands (e.g., Google, Samsung, Garmin), we believe that the common sensors, interaction modalities, and typical usage patterns across all smartwatches minimize the potential diversity of queries that might have been elicited and any differences in the functionality of the final application. While perhaps a query regarding a derived metric (i.e., a combined score value encompassing a range of raw captured metrics such as a Readiness Score or Energy Level) would arise, it is likely to remain to fall within an overarching insight category. Therefore, while our focus on the Apple Watch is a limitation, it is unlikely to significantly impact the generalizability of our findings, where as *who* (as discussed above) is likely to provide greater influence.

8.4.3 Dataset of Desired Queries

Throughout, our objective was to understand and enable databiting on the smartwatch. One key aspect of this understanding was to first gather *what* queries were of desire. Much of the work put forward in this thesis then relies on the dataset captured in Chapter 4. Specifically, the dataset captured contains 205 queries from 18 participants (on average 11.4 queries per participant). The closest work to this by Rawassizadeh et al. [154], provides a dataset of 716 queries from 131 participants (on average 5.5 queries per participant). However, study methodology should be carefully considered. While it is feasible to collect queries quickly and efficiently through online survey, producing a large dataset in the process, collecting data with real-world validity is much more cumbersome, yet valuable.

A natural question that arises from any elicitation study is whether we have captured all possibilities, either specific queries, broader insight categories, and/or mechanisms used within a query. We cannot say with certainty that we captured all possibilities, and it is entirely the case that given additional participants and device capabilities that a specific query would arise that has not been previously mentioned. An example of such a non-reported query could relate to outlier detection (e.g., "Were there any instances of irregular heartbeats or atrial fibrillation in the last day?"). However, we do believe that our dataset provides a comprehensive and importantly real-world report of desired queries. Specifically, when coding insight categories of the queries, not only did we find examples of insight categories previously noted in research, we in fact expanded upon categories while also introducing a new category. Furthermore, when comparing our results to the dataset captured by Rawassizadeh et al. [154], we found identical components that make up a query. Again, however, we were able to expand upon filtering mechanisms and showcase divergence of interrogatives given our real-world context of the provided queries.

As our goal was not to train a model for understanding and resolving of queries, a large dataset is not needed to accomplish our research goals. By coding queries into overarching structures and insight categories, we were able to derive meaningful patterns and insights from the collected data. Furthermore, the real-world context of our data collection adds to the validity and applicability of our results, providing a strong foundation for not only our successful development of DataWatch, but also for future research to build upon which has not been previously seen. However, more data would undoubtedly enhance the robustness and generalizability of our findings. Future studies should aim to expand the dataset further, capturing a wider variety of queries and participant demographics to validate and extend the insights derived from our current dataset. This would help to ensure that the databiting interactions we have identified are representative of a broader user base and can be more effectively utilized in diverse real-world contexts while also beginning to allow for training of models which can support such personal health data exploration.

8.4.4 Part-of-Speech Tagging Versus Large Language Models

Deploying large language models (LLMs) on smartwatches for personal health data exploration raises significant challenges that are not easily mitigated. These include ethical concerns surrounding data ownership and control, model training requiring substantial data, transparency of model features, and compliance with regulatory frameworks [72]. Moreover, LLMs pose risks such as biased responses, potential over-reliance or unsafe use, and exploitation of user trust to gain private information [195], all of which could negatively impact personal health outcomes. Additionally, on-board computation, though improving year over year, remains a limitation for older smartwatch models still in circulation. For these reasons, and to prioritize creating an immediate and accessible impact, we focused on implementing a part-of-speech tagging approach.

Despite these challenges, LLMs are rapidly emerging as a highly capable and promising tool. Thus, they should not be ignored. Unlike part-of-speech tagging, LLMs offer advanced query processing and contextual understanding. For instance, a complex query such as, "Compare my heart rate and pace at each kilometer of my run yesterday and my last 5K", exceeds the capabilities of our current approach. Furthermore, queries which require understanding, such as "How has my sleep affected my running pace?" can not be currently handled. Recent research has demonstrated the potential of LLMs in data informatics tasks [96, 178]. By leveraging their ability to synthesize and interpret complex natural language queries, LLMs could unlock more dynamic and flexible interactions with personal health data. Looking ahead, future work in personal data exploration and interface customization will undoubtedly benefit from integrating LLMs while building from and incorporating our findings throughout.

8.4.5 Enabling Preemptive and Proactive Insight

Preemptive and Proactive exploration was a new insight category found within our collected dataset of desired queries. We believe that this was observed due to our study setup where we emphasized to our participants that they should not worry about the capabilities of existing technologies; whereas previous data exploration work has been conducted with a working prototype, such as DataWatch was later on, despite its limited functionality [30, 63, 98].

Currently, mHealth applications and smartwatch operating systems provide a primitive form of preemptive and proactive insight. Notifications, suggestions, or motivational reminders are often utilized to proactively encourage people to stand up, breathe, or move after sedentary periods. However, these are limited in their expressiveness and ability, and do not allow for data exploration to be included. Our elicited queries were much more involved. The queries within this category were aimed towards people utilizing their personal health data to prepare for future events through system recommendations, see Chapter 4 Section 4. This included, querying when to workout, how long to workout for, and specific workouts to do.

We actively chose not to support such queries within DataWatch at this time. The looming limitation is that of ethical concerns in providing misguided information which could in turn have adverse impacts. For example, if a person asks for a distance they should run, we must confidently be able to justify the response. A value of 5 kms versus 10 kms is a difference that could significantly impact a person's physical well-being, potentially leading to injury or overexertion if the recommendation is not appropriately tailored to the individual.

In the pursuit of enabling this form of insight in the future, many questions arise; to showcase these, we highlight two queries from our dataset: *"How much dancing do I need to do to burn 800 calories?" and "Give me a suggested workout based on my readiness score?"*. First, how should the answer be calculated? Calculating the appropriate answer can be a challenge, especially when little prior data is available or standards are unknown. Second, when providing an answer, how can we convey uncertainty and variance to the user? These queries often do not have a discrete answer available, with more factors and external data needing to be considered. Finally, is this form of insight ethically possible? With the ability to recommend, and ultimately have a person act upon an answer, this carries with it the importance of not misleading a person which could have ramifications. These areas of future study are important in enabling this form of beneficial and desired insight on a smartwatch, and even for a broader set of devices.

8.4.6 User Adoption and Behavior Change

Our work is limited when considering the scalability and long-term use of DataWatch and the underlying databiting interactions that it is built upon. This thesis primarily focuses on initial user intentions and preferences, which may not accurately reflect long-term user behavior. While this initial study allows us to explore and argue the feasibility of such interactions, it does not support the notion of long-term use. Given an extended period of time, people may exhibit different engagement patterns, as we already saw within our week long studies. This differing use can be influenced by evolving personal goals, changing contexts of use, and the perceived benefit that is gained from such interactions. However, as we see from the lived personal informatics model proposed by Epstein et al. [59], exploration practices revolve and evolve as time goes on. This suggests that initial databiting interactions may be targeted again over time even after a period of non-use.

Furthermore, sustained interactions with a novel system such as DataWatch are inherently challenging to achieve. While initial adoption may be driven by novelty and curiosity, maintaining consistent usage requires the technology to integrate seamlessly into users' daily routines and provide continuous value. While we argue that integration in-situ is wholly feasible, sustained value is something we did not explore. Factors such as usability, perceived usefulness, and ongoing user engagement strategies play critical roles in fostering long-term interest of databiting practices to the benefit of the people conducting such exploration.

Overall, while databiting and the following DataWatch application shows promise and feasibility for lightweight and transient access to personal data, further research is needed to address the limitations related to long-term user adoption and behavior change as a result of interactions and insights gained. Investigating how users interact with the technology over extended periods, understanding the factors that influence sustained engagement, the influence of databiting on further exploration and vice versa, and quantifying long-term behaviour change as a result of databiting practices will all provide further valuable insights for improving the the scalability and long-term usability of databiting.

Chapter 9

Conclusion

The research, and its outcomes, in this thesis provide an understanding and implementation of *databiting*, including the conceptualization of this lightweight and transient approach to data exploration (Chapter 3), elicitation of queries from smartwatch users in their daily lives (i.e., *what*, *when*, *where*, and *why* queries are desired—Chapter 4), and an analysis of input and output requirements (i.e., *how* we can support databiting on the smartwatch—Chapters 5 and 6). This foundation of knowledge then supported the development of DataWatch (Chapter 7), a smartwatch workout tracking application that facilitates in-situ databiting exploration of past workout fitness data through lightweight and transient interactions using touch and speech, enhancing peoples' ability to gain insights during various phases of their workouts.

The findings from this research provide overarching insight that carries importance for the future of smartwatch and broader on-the-go mobile health applications. Specifically, the potential for smartwatches to offer deeper, more interactive engagement with personal health data is not only desired but wholly achievable. People seek more from their smartwatches than passive data collection; they want insightful and in-situ engagement that informs and motivates their fitness activities. DataWatch exemplifies how smartwatch health applications can evolve to begin to meet these expectations by facilitating in-situ exploration of past workout data through lightweight and transient touch and speech interactions—known as databiting. This approach to data exploration was demonstrated to be feasible on the smartwatch while in-situ and offers benefits including introducing increased actionable insights and motivation to further engage with collected data.

9.1 Summary Contributions of the Thesis

In summary, the overarching contributions of this thesis are as follows:

C1: An introduction of the term *databiting*, conceptualized as lightweight and transient data exploration, that bridges the gap between quick, glanceable data visualizations and more extensive, detailed data analysis. Further to providing a conceptualization of *databiting*, we also discuss potential benefits and highlight future, and necessary, research directions. This foundation sets the stage for developing more intuitive and effective ways for people to engage with their personal data, both on smartwatches and other mobile platforms.

C2: An elicited, and now publicly available, dataset of 205 personal health data queries desired for exploration on a smartwatch throughout daily life. This dataset, coupled with a thorough qualitative and quantitative analysis, offers valuable insights into what personal health data query needs people have, including when, where and why these queries arise.

C3: Interaction requirements for lightweight and transient exploration of personal health data on smartwatches. We identify key dimensions of our captured queries, such as interrogatives, data sources, aggregations, and filtering mechanisms, which are crucial for enhancing input methods, particularly natural language processing capabilities. Additionally, we gather insights on user preferences for different output structures from voice assistants, focusing on perceived quality, behavior, comprehensibility, and efficiency. These findings provide essential guidelines for optimizing both input and output interactions, ensuring that data exploration on smartwatches can be supported in being lightweight and transient.

C4: Development and validation of DataWatch, an Apple Watch application that demonstrates the practical implementation of *databiting*. By integrating multimodal interactions, specifically touch and speech, DataWatch enhances in-situ exploration of past workout data, demonstrating the feasibility and user benefits of providing lightweight and transient data exploration capabilities directly on the smartwatch.

Bibliography

[1] Sunggeun Ahn, Jaeyeon Lee, Keunwoo Park, and Geehyuk Lee. Evaluation of edge-based interaction on a square smartwatch. *International Journal of Human-Computer Studies*, 109:68–78, 2018. → pages 24

[2] Emily Couvillon Alagha and Rachel Renee Helbing. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of alexa, google assistant and siri. *BMJ health & care informatics*, 26(1):e100075, 2019. → pages 8, 85, 87

[3] Fereshteh Amini, Khalad Hasan, Andrea Bunt, and Pourang Irani. Data representations for in-situ exploration of health and fitness data. In *Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare*, pages 163–172, 2017. doi: 10.1145/3154862.3154879. → pages xviii, 2, 6, 8, 14, 22, 23, 25, 28, 44, 45, 47, 56, 57, 58, 64, 65, 91, 104, 124

[4] Ahmed Sabbir Arif and Ali Mazalek. A survey of text entry techniques for smartwatches. In *Human-Computer Interaction. Interaction Platforms and Techniques: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17-22, 2016. Proceedings, Part II 18*, pages 255–267. Springer, 2016. → pages 18

[5] Daniel L Ashbrook. *Enabling mobile microinteractions*. Georgia Institute of Technology, 2010. → pages 87

[6] Stavros Asimakopoulos, Grigorios Asimakopoulos, and Frank Spillers. Motivation and user engagement in fitness tracking: Heuristics for mobile

healthcare wearables. In *Informatics*, volume 4, page 5. MDPI, 2017. → pages 22

[7] Christiane Attig and Thomas Franke. Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition. *Computers in Human Behavior*, 102:223–237, 1 2020. ISSN 07475632. doi: 10.1016/j.chb.2019.08.025. → pages 21

[8] Jillian Aurisano, Abhinav Kumar, Alberto Gonzales, Khairi Reda, Jason Leigh, Barbara Di Eugenio, and Andrew Johnson. Show me data": Observational study of a conversational interface in visual data exploration. In *IEEE VIS*, volume 15, page 1, 2015. → pages 7, 37, 120, 123

[9] Karlin Bark, Jason Wheeler, Gayle Lee, Joan Savall, and Mark Cutkosky. A wearable skin stretch device for haptic feedback. In *World Haptics 2009-Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 464–469. IEEE, 2009. → pages 17

[10] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5):1–27, 2013. → pages 14, 15

[11] Lonni Besançon and Pierre Dragicevic. The continued prevalence of dichotomous inferences at chi. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–11, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359719. doi: 10.1145/3290607.3310432. URL `https://doi.org/10.1145/3290607.3310432`. → pages 95

[12] Timothy W Bickmore, Ha Trinh, Stefan Olafsson, Teresa K O'Leary, Reza Asadi, Nathaniel M Rickles, and Ricardo Cruz. Patient and consumer safety

risks when using conversational assistants for medical information: an observational study of siri, alexa, and google assistant. *Journal of medical Internet research*, 20(9):e11510, 2018. → pages 88

[13] Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, and Petra Isenberg. Glanceable visualization: Studies of data comparison performance on smartwatches. *IEEE transactions on visualization and computer graphics*, 25(1):630–640, 2018. doi: 10.1109/TVCG.2018.2865142. → pages 2, 8, 24, 25, 28, 44

[14] Tanja Blascheck, Frank Bentley, Eun Kyoung Choe, Tom Horak, and Petra Isenberg. Characterizing glanceable visualizations: From perception to behavior change. In *Mobile Data Visualization*, pages 151–176. Chapman and Hall/CRC, 2021. → pages 17, 25, 82

[15] Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, Alaul Islam, Tingying He, and Petra Isenberg. Studies of part-to-whole glanceable visualizations on smartwatch faces. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*, pages 187–196, 2023. doi: 10.1109/PacificVis56936.2023.00028. → pages 2, 8, 28

[16] Nadine Bol, Natali Helberger, and Julia CM Weert. Differences in mobile health app use: a source of new digital inequalities? *The Information Society*, 34(3):183–193, 2018. doi: 10.1080/01972243.2018.1438550. → pages 64

[17] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1):579–616, 2003. URL `https://doi.org/10.1146/annurev.psych.54.101601.145030`. → pages 41, 45

[18] Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. An empirical study of older adult's voice assistant use for health information seeking. *ACM Trans. Interact. Intell. Syst.*, 12(2), jul 2022. ISSN 2160-6455. doi:

10.1145/3484507. URL `https://doi.org/10.1145/3484507`. → pages 87, 88

[19] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, page 473–480, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136307. doi: 10.1145/642611.642694. URL `https://doi.org/10.1145/642611.642694`. → pages 17, 37, 84, 87

[20] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 473–480, 2003. doi: 10.1145/642611.642694. → pages 8, 85, 151

[21] Laura Burbach, Patrick Halbach, Nils Plettenberg, Johannes Nakayama, Martina Ziefle, and André Calero Valdez. "hey, siri", "ok, google", "alexa". acceptance-relevant factors of virtual voice-assistants. In *2019 IEEE International Professional Communication Conference (ProComm)*, pages 101–111, 2019. doi: 10.1109/ProComm.2019.00025. → pages 71

[22] Ana Isabel Canhoto and Sabrina Arp. Exploring the factors that support adoption and sustained use of health and fitness wearables. *Journal of Marketing Management*, 33(1-2):32–60, 2017. → pages 22

[23] Scott Carter and Jennifer Mankoff. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 899–908, 2005. doi: 10.1145/1054972.1055098. → pages 45

[24] Marta E Cecchinato, Anna L Cox, and Jon Bird. Always on (line)? user experience of smartwatches and their role within multi-device ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing*

Systems, pages 3557–3568, 2017. doi: 10.1145/3025453.3025538. → pages 46

[25] Xiang'Anthony' Chen, Tovi Grossman, and George Fitzmaurice. Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 615–620, 2014. → pages 18

[26] Yang Chen. Visualizing large time-series data on very small screens. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*, pages 37–41, 2017. doi: 10.2312/eurovisshort.20171130. → pages 2, 8, 28, 44

[27] Eugene Cho, S. Shyam Sundar, Saeed Abdullah, and Nasim Motalebi. Will deleting history make alexa more trustworthy? effects of privacy and content customization on user experience of smart speakers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. doi: 10.1145/3313831.3376551. URL `https://doi.org/10.1145/3313831.3376551`. → pages 87, 113

[28] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1143–1152, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557372. URL `https://doi.org/10.1145/2556288.2557372`. → pages 14

[29] Eun Kyoung Choe, Bongshin Lee, et al. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications*, 35(4):28–37, 2015. doi: 10.1109/MCG.2015.51. → pages xviii, 14, 45, 56, 57, 58, 104

[30] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 173–182, 2017. doi: 10.1145/3154862.3154881. → pages xviii, 14, 45, 56, 57, 58, 104, 164

[31] Jaewon Choi and Seongcheol Kim. Is the smartwatch an it product or a fashion product? a study on factors affecting the intention to use smartwatches. *Computers in Human Behavior*, 63:777–786, 2016. → pages 21

[32] Veerisa Chotiyaputta and Donghee Shin. Explicating consumer adoption of wearable technologies: a case of smartwatches from the asean perspective. *International Journal of Technology and Human Interaction (IJTHI)*, 18(1): 1–21, 2022. → pages 19

[33] Stephanie Hui-Wen Chuah. You inspire me and make my life better: Investigating a multiple sequential mediation model of smartwatch continuance intention. *Telematics and Informatics*, 43:101245, 2019. doi: 10.1016/j.tele.2019.101245. → pages 21

[34] Jaemin Chun, Anind Dey, Kyungtaek Lee, and Seung Jun Kim. A qualitative study of smartwatch usage and its usability. *Human Factors and Ergonomics In Manufacturing*, 28:186–199, 7 2018. ISSN 15206564. doi: 10.1002/hfm.20733. → pages 1, 16, 17, 19, 21, 23

[35] Arlene E Chung, Ashley C Griffin, Dasha Selezneva, and David Gotz. Health and fitness apps for hands-free voice-activated assistants: content analysis. *JMIR mHealth and uHealth*, 6(9):e9705, 2018. → pages 88

[36] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A Kientz, and Sean A Munson. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pages 770–786, 2016. → pages 14

[37] Karen Church and Barry Smyth. Understanding the intent behind mobile information needs. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 247–256, 2009. doi: 10.1145/1502650.1502686. → pages 1, 2, 45

[38] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. No longer wearing: Investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 647–658, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335744. doi: 10.1145/2750858.2807554. URL `https://doi.org/10.1145/2750858.2807554`. → pages 21, 35

[39] Sunny Consolvo and Miriam Walker. Using the experience sampling method to evaluate ubicomp applications. *IEEE pervasive computing*, 2(2):24–31, 2003. doi: 10.1109/MPRV.2003.1203750. URL `https://doi.org/10.1109/MPRV.2003.1203750`. → pages 41, 45

[40] Sunny Consolvo, Beverly Harrison, Ian Smith, Mike Y Chen, Katherine Everitt, Jon Froehlich, and James A Landay. Conducting in situ evaluations for and with ubiquitous computing technologies. *International Journal of Human-Computer Interaction*, 22(1-2):103–118, 2007. → pages 1, 14

[41] Coorevits, Lynn and Coenen, Tanguy. The rise and fall of wearable fitness trackers. In *Academy of Management*, page 24, 2016. → pages 21

[42] Kenneth Cox, Rebecca E Grinter, Stacie L Hibino, Lalita Jategaonkar Jagadeesan, and David Mantilla. A multi-modal natural language interface to an information visualization environment. *International Journal of Speech Technology*, 4(3):297–314, 2001. doi: 10.1023/A:1011368926479. → pages 7, 120, 123

[43] Kate Crawford, Jessa Lingel, and Tero Karppi. Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable

device. *European Journal of Cultural Studies*, 18(4-5):479–496, 2015. → pages 16

[44] Patrick Crowley, Pascal Madeleine, and Nicolas Vuillerme. Effects of mobile phone use during walking: a review. *Critical Reviews™ in Physical and Rehabilitation Medicine*, 28(1-2), 2016. → pages 17

[45] Mihaly Csikszentmihalyi and Reed Larson. Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*, pages 35–54. Springer, 2014. doi: 10.1007/978-94-017-9088-8_3. → pages 45

[46] Andrea Cuadra, Shuran Li, Hansol Lee, Jason Cho, and Wendy Ju. My bad! repairing intelligent voice assistant errors improves interaction. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–24, 2021. → pages 88

[47] Geoff Cumming and Sue Finch. Inference by eye: confidence intervals and how to read pictures of data. *American psychologist*, 60(2):170, 2005. → pages 95

[48] Rajkumar Darbar, Prasanta Kr Sen, and Debasis Samanta. Presstact: Side pressure-based input for smartwatch interaction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2431–2438, 2016. → pages 24

[49] Statista-Empowering People With Data. Number of users of smartwatches worldwide from 2019 to 2028, 2023. URL `https://www.statista.com/forecasts/1314339/worldwide-users-of-smartwatches`. Online; accessed 2024-05-30. → pages 1, 19

[50] Caroline de Cock, Madison Milne-Ives, Michelle Helena van Velthoven, Abrar Alturkistani, Ching Lam, and Edward Meinert. Effectiveness of conversational agents (virtual assistants) in health care: Protocol for a systematic review. *JMIR Res Protoc*, 9(3):e16934, Mar 2020. ISSN 1929-

0748. doi: 10.2196/16934. URL `https://www.researchprotocols.org/2020/3/e16934`. → pages 85, 87

[51] David Dearman, Melanie Kellar, and Khai N Truong. An examination of daily information needs and sharing opportunities. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 679–688, 2008. doi: 10.1145/1460563.1460668. → pages 45

[52] Milad Dehghani. Exploring the motivational factors on continuous usage intention of smartwatches among actual users. *Behaviour & Information Technology*, 37(2):145–158, 2018. doi: 10.1080/0144929X.2018.1424246. → pages 1, 20

[53] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. Mapping perceptions of humanness in intelligent personal assistant interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368254. doi: 10.1145/3338286.3340116. URL `https://doi.org/10.1145/3338286.3340116`. → pages 87

[54] Pierre Dragicevic. *Fair Statistical Communication in HCI*, pages 291–330. Springer International Publishing, Cham, 2016. ISBN 978-3-319-26633-6. doi: 10.1007/978-3-319-26633-6_13. URL `https://doi.org/10.1007/978-3-319-26633-6_13`. → pages 95

[55] Aarthi Easwara Moorthy and Kim-Phuong L Vu. Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, 31(4):307–335, 2015. doi: 10.1080/10447318.2014.986642. → pages 63, 155

[56] Justin Edwards, Christian Janssen, Sandy Gould, and Benjamin R. Cowan. Eliciting spoken interruptions to inform proactive speech agent design. In *Proceedings of the 3rd Conference on Conversational User Interfaces*, CUI '21, New York, NY, USA, 2021. Association for Computing Machinery.

ISBN 9781450389983. doi: 10.1145/3469595.3469618. URL `https://doi.org/10.1145/3469595.3469618`. → pages 88

[57] Chris Elsden, David S Kirk, and Abigail C Durrant. A quantified past: Toward design for remembering with personal informatics. *Human–Computer Interaction*, 31(6):518–557, 2016. → pages 14

[58] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 667–676, 2014. doi: 10.1145/2598510.2598558. → pages 14, 22

[59] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 731–742, 2015. → pages xx, 13, 14, 15, 22, 165

[60] Daniel A Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, et al. Mapping and taking stock of the personal informatics literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–38, 2020. → pages 15

[61] Daniel A Epstein, Tanja Blascheck, Sheelagh Carpendale, Raimund Dachselt, and Jo Vermeulen. Challenges in everyday use of mobile visualizations. In *Mobile Data Visualization*, pages 209–240. Chapman and Hall/CRC, 2021. → pages 7

[62] Andrey Esakia, D. Scott McCrickard, Samantha Harden, and Michael Horning. Fitaware: Promoting group fitness awareness throughglanceable smartwatches. In *Proceedings of the 2018 ACM International Conference on Supporting Group Work*, GROUP '18, page 178–183, New York, NY, USA, 2018. Association for Computing Machinery. ISBN

9781450355629. doi: 10.1145/3148330.3148343. URL `https://doi.org/10.1145/3148330.3148343`. → pages 1

[63] Clayton Feustel, Shyamak Aggarwal, Bongshin Lee, and Lauren Wilcox. People like me: Designing for reflection on aggregate cohort data in personal informatics systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–21, 2018. doi: 10.1145/3264917. → pages 164

[64] Thomas Franke, Christiane Attig, and Daniel Wessel. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ati) scale. *International Journal of Human–Computer Interaction*, 35(6):456–467, 2019. → pages 93

[65] Siwei Fu, Kai Xiong, Xiaodong Ge, Siliang Tang, Wei Chen, and Yingcai Wu. Quda: natural language queries for visual data analytics. *arXiv preprint arXiv:2005.03257*, 2020. → pages 71

[66] Jun Gong, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang'Anthony' Chen, Xiaojun Bi, and Xing-Dong Yang. Wristext: One-handed text entry on smartwatch using wrist gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018. → pages 18

[67] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003. → pages 93

[68] Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A Munson, and Marc Hassenzahl. 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*, pages 144–155, 2016. doi: 10.1145/2971648.2971754. → pages 2, 8, 17, 25, 28, 44

[69] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. Activity tracking in vivo. volume 2018-April. Association for Computing Machin-

ery, 4 2018. ISBN 9781450356206. doi: 10.1145/3173574.3173936. → pages 16, 19, 22, 64, 124

[70] Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. Fostering engagement with personal informatics systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, page 286–300, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340311. doi: 10.1145/2901790.2901803. URL `https://doi.org/10.1145/2901790.2901803`. → pages 1, 2, 21, 23

[71] Gabriel Haas, Michael Rietzler, Matt Jones, and Enrico Rukzio. Keep it short: A comparison of voice assistants' response behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3517684. URL `https://doi.org/10.1145/3491102.3517684`. → pages 8, 85, 88, 89, 92, 93, 98, 111, 113

[72] Philipp Hacker, Andreas Engel, and Marco Mauer. Regulating chatgpt and other large generative ai models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, page 1112–1123, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701924. doi: 10.1145/3593013.3594067. URL `https://doi.org/10.1145/3593013.3594067`. → pages 163

[73] Jonna Häkkilä, Farnaz Vahabpour, Ashley Colley, Jani Väyrynen, and Timo Koskela. Design probes study on user perceptions of a smart glasses concept. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 223–233, 2015. doi: 10.1145/2836041.2836064. → pages 45

[74] Tian Hao, Chongguang Bi, Guoliang Xing, Roxane Chan, and Linlin Tu. Mindfulwatch: A smartwatch-based system for real-time respiration monitoring during meditation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–19, 2017. → pages 20

[75] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, 11(6):838–854, 2016. doi: 10.1177/1745691616650285. → pages 6, 41, 44, 45

[76] Javier Hernandez, Daniel McDuff, Christian Infante, Pattie Maes, Karen Quigley, and Rosalind Picard. Wearable esm: differences in the experience sampling method across wearable devices. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*, pages 195–205, 2016. doi: 10.1145/2935334.2935340. → pages 46

[77] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1063–1072, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557130. URL `https://doi.org/10.1145/2556288.2557130`. → pages 36

[78] Jon-Chao Hong, Pei-Hsin Lin, and Pei-Chi Hsieh. The effect of consumer innovativeness on perceived value and continuance intention to use smartwatch. *Computers in Human Behavior*, 67:264–272, 2017. → pages 21

[79] Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. Splitboard: A simple split soft keyboard for wristwatch-sized touch screens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1233–1236, 2015. → pages 18

[80] Tom Horak, Sriram Karthik Badam, Niklas Elmqvist, and Raimund Dachselt. When david meets goliath: Combining smartwatches with a large vertical display for visual data exploration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018. → pages 24

[81] Ya Huang, Kuanming Yao, Jiyu Li, Dengfeng Li, Huiling Jia, Yiming Liu, Chun Ki Yiu, Wooyoung Park, and Xinge Yu. Recent advances in multimode haptic feedback technologies towards wearable interfaces. *Materials Today Physics*, 22:100602, 2022. → pages 17

[82] Ira E Hyman Jr, S Matthew Boss, Breanne M Wise, Kira E McKenzie, and Jenna M Caggiano. Did you see the unicycling clown? inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology*, 24(5):597–607, 2010. → pages 17

[83] Stephen Intille, Caitlin Haynes, Dharam Maniar, Aditya Ponnada, and Justin Manjourides. μema: Microinteraction-based ecological momentary assessment (ema) using a smartwatch. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1124–1128, 2016. doi: 10.1145/2971648.2971717. → pages 46

[84] Alaul Islam, Anastasia Bezerianos, Bongshin Lee, Tanja Blascheck, and Petra Isenberg. Visualizing information on watch faces: A survey with smartwatch users. In *2020 IEEE Visualization Conference (VIS)*, pages 156–160, 2020. doi: 10.1109/VIS47514.2020.00038. → pages 24

[85] Alaul Islam, Ranjini Aravind, Tanja Blascheck, Anastasia Bezerianos, and Petra Isenberg. Preferences and effectiveness of sleep data visualizations for smartwatches and fitness bands. In *CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2022. → pages 2, 8, 25, 28

[86] Alaul Islam, Tanja Blascheck, and Petra Isenberg. Context Specific Visualizations on Smartwatches. Posters of the European Conference on Visualization (EuroVis), June 2022. URL `https://inria.hal.science/hal-03694026`. Poster. → pages 8, 18, 65

[87] Carlos Jensen, Colin Potts, and Christian Jensen. Privacy practices of internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, 2005. doi: 10.1016/j.ijhcs.2005.04.019. → pages 64

[88] Hayeon Jeong, Heepyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. Smartwatch wearing behavior analysis. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1:1–31, 9 2017. doi: 10.1145/3131892. → pages 19

[89] Ravinder Jerath, Mohammad Syam, and Shajia Ahmed. The future of stress management: integration of smartwatches and hrv technology. *Sensors*, 23 (17):7314, 2023. → pages 20

[90] Tero Jokela, Jarno Ojala, and Thomas Olsson. A diary study on combining multiple information devices in everyday activities and tasks. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3903–3912, 2015. doi: 10.1145/2702123.2702211. → pages 45

[91] Simon L Jones. Exploring correlational information in aggregated quantified self data dashboards. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 1075–1080, 2015. → pages 15

[92] Simon L Jones and Ryan Kelly. Dealing with information overload in multifaceted personal informatics systems. *Human–Computer Interaction*, 33 (1):1–48, 2018. → pages 15

[93] Haruka Kamachi, Tahera Hossain, Fuyuka Tokuyama, Anna Yokokubo, and Guillaume Lopez. Prediction of eating activity using smartwatch. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, pages 304–309, 2021. → pages 20

[94] Jan-Frederik Kassel and Michael Rohs. Valletto: A multimodal interface for ubiquitous visual analytics. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2018. doi: 10.1145/3170427.3188445. → pages 82

[95] Frederic Kerber, Tobias Kiefer, Markus Löchtefeld, and Antonio Krüger. Investigating current techniques for opposite-hand smartwatch interaction. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–12, 2017. → pages 24

[96] Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. Diarymate: Understanding user perceptions and experience in human-ai collaboration for personal journaling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642693. URL `https://doi.org/10.1145/3613904.3642693`. → pages 164

[97] Yoojung Kim, Bongshin Lee, and Eun Kyoung Choe. Investigating data accessibility of personal health apps. *Journal of the American Medical Informatics Association*, 26(5):412–419, 03 2019. ISSN 1527-974X. doi: 10.1093/jamia/ocz003. URL `https://doi.org/10.1093/jamia/ocz003`. → pages 127

[98] Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. Data@hand: Fostering visual exploration of personal data on smartphones; leveraging speech and touch interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445421. URL `https://doi.org/10.1145/3411764.3445421`. → pages 14, 37, 63, 71, 82, 120, 123, 155, 164

[99] Young-Ho Kim, Diana Chou, Bongshin Lee, Margaret Danilovich, Amanda Lazar, David E Conroy, Hernisa Kacorri, and Eun Kyoung Choe. Mymove: Facilitating older adults to collect in-situ activity labels on a smartwatch with speech. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2022. → pages 20, 46, 48, 65

[100] Andreas M. Klein, Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. Construction of ueq+ scales for voice quality: measuring user experience quality of voice interaction. In *Proceedings of Mensch Und Computer 2020*, MuC '20, page 1–5, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375405. doi: 10.1145/3404983.3410003. URL `https://doi.org/10.1145/3404983.3410003`. → pages 92, 116

[101] Rafal Kocielnik, Raina Langevin, James S. George, Shota Akenaga, Amelia Wang, Darwin P. Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T. Hsieh, Kabir Yadav, Herbert Duber, Gary Hsieh, and Andrea L. Hartzler. Can i talk to you about your social needs? understanding preference for conversational user interface in health. In *Proceedings of the 3rd Conference on Conversational User Interfaces*, CUI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450389983. doi: 10.1145/3469595.3469599. URL `https://doi.org/10.1145/3469595.3469599`. → pages 87

[102] Michelle Kwasny, Kelly Caine, Wendy A Rogers, and Arthur D Fisk. Privacy and technology: folk definitions and perspectives. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pages 3291–3296. 2008. doi: 10.1145/1358628.1358846. → pages 64

[103] Ricardo Langner, Lonni Besançon, Christopher Collins, Tim Dwyer, Petra Isenberg, Tobias Isenberg, Bongshin Lee, Charles Perin, and Christian Tominski. An introduction to mobile data visualization. In *Mobile Data Visualization*, pages 1–32. Chapman and Hall/CRC, 2021. → pages 20

[104] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y S Lau, and Enrico Coiera. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258, 07 2018. ISSN 1527-974X. doi: 10.1093/jamia/ocy072. URL `https://doi.org/10.1093/jamia/ocy072`. → pages 87

[105] Reed Larson and Mihaly Csikszentmihalyi. The experience sampling method. In *Flow and the foundations of positive psychology*, pages 21–34. Springer, 2014. doi: 10.1007/978-94-017-9088-8_2. → pages 45

[106] Bongshin Lee, Eun Kyoung Choe, Petra Isenberg, Kim Marriott, and John Stasko. Reaching broader audiences with data visualization. *IEEE Computer Graphics and Applications*, 40(2):82–90, 2020. doi: 10.1109/MCG.2020.2968244. URL `https://doi.org/10.1109/MCG.2020.2968244`. → pages 27

[107] Ian Li, Anind Dey, and Jodi Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 557–566, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605589299. doi: 10.1145/1753326.1753409. URL `https://doi.org/10.1145/1753326.1753409`. → pages 13, 14, 15, 22, 64, 124

[108] Ian Li, Anind K. Dey, and Jodi Forlizzi. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, page 405–414, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450306300. doi: 10.1145/2030112.2030166. URL `https://doi.org/10.1145/2030112.2030166`. → pages xviii, 14, 23, 34, 45, 56, 57, 58, 64

[109] Gesa Alena Linnemann and Regina Jucks. 'can i trust the spoken dialogue system because it uses the same words as i do?'—influence of lexically aligned spoken dialogue systems on trustworthiness and user satisfaction. *Interacting with Computers*, 30(3):173–186, 2018. → pages 87

[110] Li Liu, Yuxin Peng, Shu Wang, Ming Liu, and Zigang Huang. Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors. *Information Sciences*, 340-341:41–57, 2016. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2016.

01.020. URL https://www.sciencedirect.com/science/article/pii/S0020025516000311. → pages 36

[111] Xing Liu, Tianyu Chen, Feng Qian, Zhixiu Guo, Felix Xiaozhu Lin, Xiaofeng Wang, and Kai Chen. Characterizing smartwatch usage in the wild. pages 385–398. Association for Computing Machinery, Inc, 6 2017. ISBN 9781450349284. doi: 10.1145/3081333.3081351. → pages 19

[112] Ramon Llamas. Global shipments of wearable devices saw modest growth in 2023 with improved demand expected in 2024 and beyond, according to idc, Mar 2024. URL https://www.idc.com/getdoc.jsp?containerId=prUS51975524. → pages 16

[113] Kok Yin Long, Kamalanathan Shanmugam, and Muhammad Ehsan Rana. An evaluation of smartwatch contribution in improving human health. In *2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pages 1–4. IEEE, 2023. → pages 20

[114] Edward Loper and Steven Bird. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*, 2002. → pages 70

[115] Ewa Luger and Abigail Sellen. "like having a really bad pa": The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 5286–5297, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450333627. doi: 10.1145/2858036.2858288. URL https://doi.org/10.1145/2858036.2858288. → pages 113

[116] Yuhan Luo, Young-Ho Kim, Bongshin Lee, Naeemul Hassan, and Eun Kyoung Choe. Foodscrap: Promoting rich data capture and reflective food journaling through speech input. In *Designing Interactive Systems Conference 2021*, pages 606–618, 2021. doi: 10.1145/3461778.3462074. → pages 63, 155

[117] Yumei Luo, Lei Yang, Qiongwei Ye, and Qichen Liao. Effects of customization and personalization affordances on perceived value and continuance intention of smartwatch use. *Technological Forecasting and Social Change*, 194:122752, 2023. → pages 21

[118] Deborah Lupton. *The quantified self*. John Wiley & Sons, 2016. → pages 13

[119] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. Social boundaries for personal agents in the interpersonal space of the home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. doi: 10.1145/3313831.3376311. URL `https://doi.org/10.1145/3313831.3376311`. → pages 87, 113

[120] Kent Lyons. Visual parameters impacting reaction times on smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 190–194, 2016. → pages 25

[121] Raju Maharjan, Darius Adam Rohani, Per Bækgaard, Jakob Bardram, and Kevin Doherty. Can we talk? design implications for the questionnaire-driven self-report of health and wellbeing via conversational agent. In *Proceedings of the 3rd Conference on Conversational User Interfaces*, CUI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450389983. doi: 10.1145/3469595.3469600. URL `https://doi.org/10.1145/3469595.3469600`. → pages 87

[122] Suresh Malodia, Nazrul Islam, Puneet Kaur, and Amandeep Dhir. Why do people use artificial intelligence (ai)-enabled voice assistants. *IEEE Transactions on Engineering Management*, pages 1–15, 2021. doi: 10.1109/TEM.2021.3117884. → pages 71

[123] Steve Mann. Wearable computing as means for personal empowerment. In

Proc. 3rd Int. Conf. on Wearable Computing (ICWC), pages 51–59, 1998. → pages 16

[124] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014. → pages 70

[125] Donald McMillan, Barry Brown, Airi Lampinen, Moira McGregor, Eve Hoggan, and Stefania Pizza. Situating wearables: Smartwatch use in context. volume 2017-May, pages 3582–3594. Association for Computing Machinery, 5 2017. ISBN 9781450346559. doi: 10.1145/3025453.3025993. → pages 18, 19

[126] Saumya Misra, Rachana Adtani, Yuvraj Singh, Simran Singh, and Devanshu Thakkar. Exploring the factors affecting behavioral intention to adopt wearable devices. *Clinical Epidemiology and Global Health*, 24:101428, 2023. → pages 19

[127] Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67, 2019. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2019.03.054. URL `https://www.sciencedirect.com/science/article/pii/S0957417419302283`. → pages 8, 85, 87

[128] Jimmy Moore, Pascal Goffin, Jason Wiese, and Miriah Meyer. Exploring the personal informatics analysis gap:"there's a lot of bacon". *IEEE Transactions on Visualization and Computer Graphics*, 28(1):96–106, 2021. → pages xx, 15

[129] Jimmy Moore, Pascal Goffin, Jason Wiese, and Miriah Meyer. An interview method for engaging personal data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–28, 2021. doi: 10.1145/3494964. → pages 66

[130] Bobak Mortazavi, Ebrahim Nemati, Kristina VanderWall, Hector G Flores-Rodriguez, Jun Yu Jacinta Cai, Jessica Lucier, Arash Naeim, and Majid Sarrafzadeh. Can smartwatches replace smartphones for posture tracking? *Sensors*, 15(10):26783–26800, 2015. → pages 20

[131] Suhaib Mujahid, Giancarlo Sierra, Rabe Abdalkareem, Emad Shihab, and Weiyi Shang. An empirical study of android wear user complaints. *Empirical Software Engineering*, 23:3476–3502, 2018. doi: 10.1007/s10664-018-9615-8. URL `https://doi.org/10.1007/s10664-018-9615-8`. → pages 1, 2, 21, 35

[132] Hendrik Muller, Jennifer Gove, and John Webb. Understanding tablet use: A multi-method exploration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 1–10, 2012. → pages 154

[133] Sean A Munson, Hasan Cavusoglu, Larry Frisch, and Sidney Fels. Sociotechnical challenges and progress in using social media for health. *Journal of medical Internet research*, 15(10):e226, 2013. → pages 22

[134] Ali Neshati, Yumiko Sakamoto, and Pourang Irani. Challenges in displaying health data on small smartwatch screens. In *ITCH*, pages 325–332, 2019. doi: 10.3233/978-1-61499-951-5-325. → pages 18, 23, 44

[135] Ali Neshati, Yumiko Sakamoto, Launa Leboe-Mcgowan, Jason Leboe-McGowan, Marcos Serrano, and Pourang Irani. G-sparks: Glanceable sparklines on smartwatches. In *45th Conference on Graphics Interface (GI 2019)*, pages 1–9, 2019. doi: 10.20380/GI2019.23. → pages 2, 8, 25, 28, 44, 148

[136] Ali Neshati, Fouad Alallah, Bradley Rey, Yumiko Sakamoto, Marcos Serrano, and Pourang Irani. Sf-lg: Space-filling line graphs for visualizing interrelated time-series data on smartwatches. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, Mobile-HCI '21, New York, NY, USA, 2021. Association for Computing Ma-

chinery. ISBN 9781450383288. doi: 10.1145/3447526.3472040. URL https://doi.org/10.1145/3447526.3472040. → pages 2, 8, 24, 25, 28, 44, 123

[137] Ali Neshati, Bradley Rey, Ahmed Shariff Mohommed Faleel, Sandra Bardot, Celine Latulipe, and Pourang Irani. Bezelglide: Interacting with graphs on smartwatches with minimal screen occlusion. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445201. URL https://doi.org/10.1145/3411764.3445201. → pages 18, 24, 44

[138] Jasmin Niess and Paweł W. Woźniak. Supporting meaningful personal fitness: The tracker goal evolution model. volume 2018-April. Association for Computing Machinery, 4 2018. ISBN 9781450356206. doi: 10.1145/3173574.3173745. → pages 23

[139] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013. → pages 88

[140] Jum Nunnally. Psychometric theory. *(No Title)*, 1994. → pages 95

[141] Masa Ogata and Michita Imai. Skinwatch: skin gesture interaction for smart watch. In *Proceedings of the 6th Augmented Human International Conference*, pages 21–24, 2015. → pages 24

[142] Obi Ogbanufe and Natalie Gerhart. Watch it! factors driving continued feature use of the smartwatch. *International Journal of Human–Computer Interaction*, 34(11):999–1014, 2018. doi: 10.1080/10447318.2017.1404779. → pages 21, 22

[143] Kenneth Olmstead. Nearly half of americans use digital voice assistants, mostly on their smartphones. *Pew Research Center*, 12, 2017. → pages 87

[144] Debajyoti Pal, Suree Funilkul, and Vajirasak Vanijja. The future of smartwatches: assessing the end-users' continuous usage using an extended

expectation-confirmation model. *Universal Access in the Information Society*, 19:261–281, 2020. → pages 21

[145] Debajyoti Pal, Anuchart Tassanaviboon, Chonlameth Arpnikanondt, and Borworn Papasratorn. Quality of experience of smart-wearables: From fitness-bands to smartwatches. *IEEE Consumer Electronics Magazine*, 9: 49–53, 1 2020. ISSN 21622256. doi: 10.1109/MCE.2019.2941462. → pages 2, 23

[146] Adam Palanica, Anirudh Thommandram, Andrew Lee, Michael Li, and Yan Fossat. Do you understand the words that are comin outta my mouth? voice assistant comprehension of medication names. *NPJ digital medicine*, 2(1): 55, 2019. → pages 88

[147] Sabrina Paneels and Jonathan C Roberts. Review of designs for haptic data visualization. *IEEE Transactions on Haptics*, 3(2):119–137, 2009. doi: 10.1109/TOH.2009.44. → pages 8, 17

[148] Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. Smartwatch in vivo. pages 5456–5469. Association for Computing Machinery, 5 2016. ISBN 9781450333627. doi: 10.1145/2858036.2858522. → pages 1, 17, 18, 19, 20, 22, 26, 46, 73, 147, 154

[149] Bernd Ploderer, Wolfgang Reitberger, Harri Oinas-Kukkonen, and Julia van Gemert-Pijnen. Social interaction and reflection for behaviour change, 2014. → pages 64, 125

[150] Aditya Ponnada, Caitlin Haynes, Dharam Maniar, Justin Manjourides, and Stephen Intille. Microinteraction ecological momentary assessment response rates: Effect of microinteractions or the smartwatch? *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 1 (3):1–16, 2017. doi: 10.1145/3130957. → pages 46

[151] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. Voice interfaces in everyday life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12,

New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi: 10.1145/3173574.3174214. URL https://doi.org/10.1145/3173574.3174214. → pages 17, 87

[152] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. "alexa is my new bff": Social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, page 2853–2859, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346566. doi: 10.1145/3027063.3053246. URL https://doi.org/10.1145/3027063.3053246. → pages 87

[153] Pallavi Rao Gadahad and Anirudha Joshi. Wearable activity trackers in managing routine health and fitness of indian older adults: Exploring barriers to usage. In *Nordic Human-Computer Interaction Conference*, NordiCHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450396998. doi: 10.1145/3546155.3546645. URL https://doi.org/10.1145/3546155.3546645. → pages 87

[154] Reza Rawassizadeh, Chelsea Dobbins, Manouchehr Nourizadeh, Zahra Ghamchili, and Michael Pazzani. A natural language query interface for searching personal information on smartwatches. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 679–684, 2017. doi: 10.1109/PERCOMW.2017.7917645. → pages 6, 44, 47, 70, 71, 72, 81, 82, 162

[155] Bradley Rey, Kening Zhu, Simon Tangi Perrault, Sandra Bardot, Ali Neshati, and Pourang Irani. Understanding and adapting bezel-to-bezel interactions for circular smartwatches in mobile and encumbered scenarios. *Proc. ACM Hum.-Comput. Interact.*, 6(MHCI), sep 2022. doi: 10.1145/3546736. URL https://doi.org/10.1145/3546736. → pages 2, 18, 24

[156] Bradley Rey, Charles-Olivier Dufresne-Camaro, and Pourang Irani. Towards efficient interaction for personal health data queries on smartwatches.

In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction*, MobileHCI '23 Companion, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399241. doi: 10.1145/3565066.3608700. URL `https://doi.org/10.1145/3565066.3608700`. → pages 68, 91

[157] Bradley Rey, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. Investigating in-situ personal health data queries on smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(4), jan 2023. doi: 10.1145/3569481. URL `https://doi.org/10.1145/3569481`. → pages 43, 91, 93

[158] Bradley Rey, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. Databiting: Lightweight, transient, and insight rich exploration of personal data. *IEEE Computer Graphics and Applications*, 44(2):65–72, 2024. doi: 10.1109/MCG.2024.3353888. → pages 27

[159] Bradley Rey, Yumiko Sakamoto, Jaisie Sin, and Pourang Irani. Understanding user preferences of voice assistant answer structures for personal health data queries. In *ACM Conversational User Interfaces 2024*, CUI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705113. doi: 10.1145/3640794.3665552. URL `https://doi.org/10.1145/3640794.3665552`. → pages 84

[160] Lionel Robert. The growing problem of humanizing robots. *Robert, LP (2017). The Growing Problem of Humanizing Robots, International Robotics & Automation Journal*, 3(1), 2017. doi: 10.15406/iratj.2017.03.00043. → pages 113

[161] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. Personal tracking as lived informatics. pages 1163–1172. Association for Computing Machinery, 2014. ISBN 9781450324731. doi: 10.1145/2556288.2557039. → pages 22, 23

[162] Léa Saviot, Frederik Brudy, and Steven Houben. Wristband. io: expanding

input and output spaces of a smartwatch. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2025–2033, 2017. → pages 24

[163] Alexander Schiewe, Andrey Krekhov, Frederic Kerber, Florian Daiber, and Jens Krüger. A study on real-time visualizations during sports activities on smartwatches. In *19th International Conference on Mobile and Ubiquitous Multimedia*, MUM '20, page 18–31, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450388702. doi: 10.1145/3428361.3428409. URL `https://doi.org/10.1145/3428361.3428409`. → pages 65

[164] Bastian Schildbach and Enrico Rukzio. Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 93–102, 2010. → pages 17

[165] Steven Schirra and Frank R Bentley. " it's kind of like an extra screen for my phone" understanding everyday uses of consumer smart watches. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2151–2156, 2015. → pages 19

[166] Anna Schneider and Rene Arnold. Wearables from head to toe: Are they friend or foe? an empirical landscaping of health and fitness wearables and apps in six countries to identify emerging policy challenges. *An Empirical Landscaping of Health and Fitness Wearables and Apps in Six Countries to Identify Emerging Policy Challenges (July 31, 2022)*, 2022. → pages 16

[167] Martin Schrepp and Jörg Thomaschewski. Design and validation of a framework for the creation of user experience questionnaires. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5 (7):88–95, 12/2019 2019. ISSN 1989-1660. doi: 10.9781/ijimai.2019.06.006. URL `https://www.ijimai.org/journal/sites/default/`

files/files/2019/06/ijimai20195_7_9_pdf_19082.pdf. → pages 92

[168] Suranga Seneviratne, Yining Hu, Tham Nguyen, Guohao Lan, Sara Khalifa, Kanchana Thilakarathna, Mahbub Hassan, and Aruna Seneviratne. A survey of wearable devices and challenges. *IEEE Communications Surveys & Tutorials*, 19(4):2573–2620, 2017. → pages 19

[169] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, page 365–377, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341899. doi: 10.1145/2984511.2984588. URL https://doi.org/10.1145/2984511.2984588. → pages 71

[170] Ren-Jay Shei, Ian G Holder, Alicia S Oumsang, Brittni A Paris, and Hunter L Paris. Wearable activity trackers–advanced technology or advanced marketing? *European Journal of Applied Physiology*, 122(9):1975–1990, 2022. → pages 16

[171] Grace Shin, Mohammad Hossein Jarrahi, Yu Fei, Amir Karami, Nicci Gafinowitz, Ahjung Byun, and Xiaopeng Lu. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review, 5 2019. ISSN 15320464. → pages 23

[172] Katie A Siek, Yvonne Rogers, and Kay H Connelly. Fat finger worries: how older and younger users physically interact with pdas. In *Human-Computer Interaction-INTERACT 2005: IFIP TC13 International Conference, Rome, Italy, September 12-16, 2005. Proceedings 10*, pages 267–280. Springer, 2005. → pages 23

[173] Carolin Siepmann and Pascal Kowalczuk. Understanding continued smartwatch usage: the role of emotional as well as health and fitness factors. *Elec-*

tronic Markets, 31(4):795–809, 2021. doi: 10.1007/s12525-021-00458-3. → pages 21

[174] Adriana Silva and Bráulio Alturas. The benefits of the use of smartwatches in promoting regular physical activity. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, 2021. doi: 10.23919/CISTI52073.2021.9476473. → pages 20

[175] Jaisie Sin, Dongqing Chen, Jalena G Threatt, Anna Gorham, and Cosmin Munteanu. Does alexa live up to the hype? contrasting expectations from mass media narratives and older adults' hands-on experiences of voice interfaces. In *Proceedings of the 4th Conference on Conversational User Interfaces*, pages 1–9, 2022. → pages 85, 88

[176] Gaganpreet Singh, William Delamare, and Pourang Irani. D-swime: A design space for smartwatch interaction techniques supporting mobility and encumbrance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi: 10.1145/3173574.3174208. URL `https://doi.org/10.1145/3173574.3174208`. → pages 18, 63

[177] Timothy Sohn, Kevin A Li, William G Griswold, and James D Hollan. A diary study of mobile information needs. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 433–442, 2008. doi: 10.1145/1357054.1357125. → pages 45

[178] Inhwa Song, SoHyun Park, Sachin R. Pendse, Jessica Lee Schleider, Munmun De Choudhury, and Young-Ho Kim. Exploreself: Fostering user-driven exploration and reflection on personal challenges with adaptive guidance by large language models, 2024. URL `https://arxiv.org/abs/2409.09662`. → pages 164

[179] Yang Song, Hao Ma, Hongning Wang, and Kuansan Wang. Exploring and exploiting user search behavior on mobile and tablet devices to improve

search relevance. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1201–1212, 2013. → pages 154

[180] Katta Spiel, Fares Kayali, Louise Horvath, Michael Penkler, Sabine Harrer, Miguel Sicart, and Jessica Hammer. Fitter, happier, more productive? the normative ontology of fitness trackers. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*, pages 1–10, 2018. → pages 22

[181] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. Watchsense: On-and above-skin input sensing through a wearable depth sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3891–3902, 2017. → pages 24

[182] Arjun Srinivasan and John Stasko. Natural language interfaces for data analysis with visualization: Considering what has and could be asked. In *Proceedings of the Eurographics/IEEE VGTC conference on visualization: Short papers*, pages 55–59, 2017. doi: 10.2312/eurovisshort.20171133. → pages 7, 123

[183] Arjun Srinivasan, Bongshin Lee, Nathalie Henry Riche, Steven M Drucker, and Ken Hinckley. Inchorus: Designing consistent multimodal interactions for data visualization on tablet devices. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020. doi: 10.1145/3313831.3376782. → pages 71, 72, 82

[184] Arjun Srinivasan, Nikhila Nyapathy, Bongshin Lee, Steven M Drucker, and John Stasko. Collecting and characterizing natural language utterances for specifying data visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2021. doi: 10.1145/3411764.3445400. → pages 7, 54, 71, 72, 82

[185] Thad Starner. How wearables worked their way into the mainstream, 2014. → pages 15

[186] Marc Strik, Sylvain Ploux, Daniel Weigel, Joske van der Zande, Anouk Velraeds, Hugo-Pierre Racine, F Daniel Ramirez, Michel Haïssaguerre, and Pierre Bordachar. The use of smartwatch electrocardiogram beyond arrhythmia detection. *Trends in Cardiovascular Medicine*, 2023. → pages 20

[187] Melanie Tory and Vidya Setlur. Do what i mean, not what i say! design considerations for supporting intent and context in analytical conversation. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 93–103, 2019. doi: 10.1109/VAST47406.2019.8986918. → pages 71

[188] Milka Trajkova and Aqueasha Martin-Hammond. "alexa is a toy": exploring older adults' reasons for using, limiting, and abandoning echo. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020. → pages 85

[189] Edward R Tufte and Edward Rolf Tufte. *Beautiful evidence*, volume 1. Graphics Press Cheshire, CT, 2006. → pages 25

[190] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)*, 50 (6):1–40, 2017. doi: 10.1145/3123988. → pages 45

[191] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kévin Huguenin, and Mauro Cherubini. Are those steps worth your privacy? fitness-tracker users' perceptions of privacy and utility. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(4), dec 2022. doi: 10.1145/3494960. URL `https://doi.org/10.1145/3494960`. → pages 87, 114

[192] Aku Visuri, Zhanna Sarsenbayeva, Niels Van Berkel, Jorge Goncalves, Reza Rawassizadeh, Vassilis Kostakos, and Denzil Ferreira. Quantifying sources and types of smartwatch usage sessions. volume 2017-May, pages 3569–3581. Association for Computing Machinery, 5 2017. ISBN 9781450346559. doi: 10.1145/3025453.3025817. → pages 1, 18, 19, 20, 21, 73, 147, 154

[193] Aku Visuri, Niels van Berkel, Jorge Goncalves, Reza Rawassizadeh, Denzil Ferreira, and Vassilis Kostakos. Understanding usage style transformation during long-term smartwatch use. *Personal and Ubiquitous Computing*, 2021. ISSN 16174917. doi: 10.1007/s00779-020-01511-2. → pages 19

[194] Ryan Vooris, Matthew Blaszka, and Susan Purrington. Understanding the wearable fitness tracker revolution. *International Journal of the Sociology of Leisure*, 2:421–437, 2019. → pages 22

[195] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zac Kenton, Sasha Brown, Will Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William Isaac, Sean Legassick, Geoffrey Irving, and Iason Gabriel. Ethical and social risks of harm from language models, 2021. → pages 163

[196] Peter West, Max Van Kleek, Richard Giordano, Mark J. Weal, and Nigel Shadbolt. Common barriers to the use of patient-generated data across clinical settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi: 10.1145/3173574.3174058. URL `https://doi.org/10.1145/3173574.3174058`. → pages 112

[197] Mark Whooley, Bernd Ploderer, and Kathleen Gray. On the integration of self-tracking data amongst quantified self members. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014*, pages 151–160. British Computer Society BCS, 2014. → pages 14

[198] Jayme L Wilder, Devin Nadar, Nitin Gujral, Benjamin Ortiz, Robert Stevens, Faye Holder-Niles, John Lee, and Jonathan M Gaffin. Pediatrician attitudes toward digital voice assistant technology use in clinical practice. *Applied clinical informatics*, 10(02):286–294, 2019. → pages 87

[199] Pui Chung Wong, Kening Zhu, and Hongbo Fu. Fingert9: Leveraging

thumb-to-finger interaction for same-side-hand text entry on smartwatches. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2018. → pages 18

[200] Pui Chung Wong, Kening Zhu, Xing-Dong Yang, and Hongbo Fu. Exploring eyes-free bezel-initiated swipe on round smartwatches. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020. → pages 18

[201] Yiqi Xiao and Lu Liu. User-defined interactions for visual data exploration with the combination of smartwatch and large display. *IEEE Access*, 2024. → pages 24

[202] Xinghui Yan, Shriti Raj, Bingjian Huang, Sun Young Park, and Mark W Newman. Toward lightweight in-situ self-reporting: An exploratory study of alternative smartwatch interface designs in context. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4): 1–22, 2020. doi: 10.1145/3432212. → pages 46

[203] Zhice Yang, Zeyu Wang, Jiansong Zhang, Chenyu Huang, and Qian Zhang. Wearables can afford: Light-weight indoor positioning with visible light. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 317–330, 2015. → pages 19

[204] Junhan Zhou, Yang Zhang, Gierad Laput, and Chris Harrison. Aurasense: enabling expressive around-smartwatch interactions with electric field sensing. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 81–86, 2016. → pages 24

[205] Leming Zhou, Jie Bao, Valerie Watzlaf, and Bambang Parmanto. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. *JMIR mHealth and uHealth*, 7(4):e11223, 2019. doi: 10.2196/11223. → pages 64

[206] Tingting Zhu, Peter Watkinson, and David A Clifton. Smartwatch data help

detect covid-19. *Nature biomedical engineering*, 4(12):1125–1127, 2020. → pages 20

[207] Michael Zimmer, Priya Kumar, Jessica Vitak, Yuting Liao, and Katie Chamberlain Kritikos. 'there's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. *Information, Communication & Society*, 23(7):1020–1037, 2020. doi: 10.1080/1369118X.2018.1543442. URL `https://doi.org/10.1080/1369118X.2018.1543442`. → pages 87, 114

Appendices

Appendix A: Study Tutorials

Chapter 4 Tutorial Slides

The following slides were used during the tutorial portion of the study conducted in Chapter 4. A live demo of the application was included during the tutorial.

Study Information, Setup, and Tutorial

Study Overview

- This study will have you:
  - Meet via Zoom for introduction, application setup, and a short demographic survey (~45 minutes)
  - Use a smartwatch application to simply report responses throughout your day (1 week / 7 days)
  - Meet via Zoom to do a one-on-one interview (~45 minutes)

- We will do everything in our power to provide a good experience for you during the study and hope for your involved participation throughout which will compensate up to $30 USD for your time by way of an Amazon gift card

205

Overview

Online Consent - In Progress

Study Overview

Demographic Survey

Download Smartwatch Application and Tutorial

Practice, Final Thoughts, and Participate!

Online Consent

Please go to the link below to read and, if comfortable, sign the online consent form.

If any questions arise, do not hesitate to ask.

QR Code removed for anonymity purposes

Overview

Online Consent - Done

Study Overview - In Progress

Demographic Survey

Download Application and Tutorial

Practice, Final Thoughts, and Participate!

Study Goals

Our high level goal is to allow you to better explore and understand your personal health data on your smartwatch.




Study Goals

- To explore what questions and/or commands people have of their personal health data on their smartwatch

- To explore when and how people wish to interact with their personal health data on their smartwatch

Important Remarks

- We are NOT collecting ANY of your personal health data or other metrics/data behind the scenes, such as your location; the application can NOT access this information

- We are NOT testing for specific responses or evaluating your performance in the application and throughout in the questionnaire and interview

Important Remarks

- We are only COLLECTING what you provide within the application itself:
  1. The question or command you have of your personal health data
  2. Your current activity
  3. If the question relates to your activity
     a. If so, the time within that activity
- We are INTERESTED in any questions or commands you have of your personal health data
- We are looking for HONEST responses throughout
- Your SAFETY when using the application is paramount, so please only respond if and when it is safe to do so

Overview

Online Consent - Done

Study Overview - Done

Demographic Survey - In Progress

Download Application and Tutorial - Done

Practice, Final Thoughts, and Participate!

Demographic Survey

Please go to the link below to fill out the demographic survey.

If any questions arise, do not hesitate to ask.

I will let you know your participant ID to include in the survey.

QR Code removed for anonymity purposes

Overview

Online Consent - Done ☑

Study Overview - Done ☑

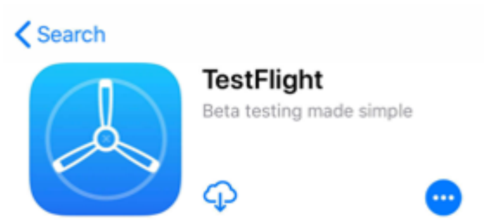
Demographic Survey ☑

Download Application and Tutorial - In Progress

Practice, Final Thoughts, and Participate!

Download the Application

1. Install TestFlight on your **iPhone**
   - This is Apple's Beta software testing application and allows us to share our application with just you and not everyone in the world



Download the Application

2. Read through and accept Apple's terms and conditions if comfortable
   - We will not be accepting any crash report information, as such we will **not** collect any background information from your device

Download the Application

3. Simply open the invitation email sent to your email address

Download the Application

4. This will open the TestFlight application. Now, simply click **Install.**
   You are good to go!

Application First Steps

- Let's set up the following so that the application is ready to go for the study:
  1. Watch face component
     - This will provide a home-screen reminder of the study for the next week
  2. Notification times
     - Notifications will occur randomly throughout the provided start and end times to help remind you of the study in progress

Application Usage

- Please watch as I demonstrate the application. Please let me know if you have any questions!

Application Usage

- When you open the application, you will see the home screen
  - This is the screen where you can record your questions about your data
  - Tap the “Tap to Record” button to speak or write your question

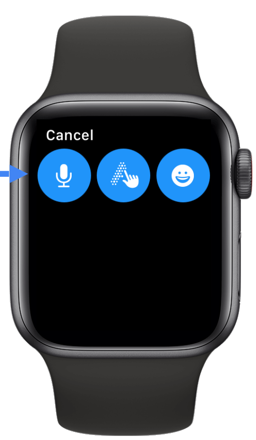
Press this button to record or write your question

If any questions or concerns arise, please contact Bradley at reyb@myumanitoba.ca

Application Usage

- Here, you can select speech or writing to record your question

Select either speech or writing



You can tap cancel at any time

Tap ‘done’ when finished recording. This brings you to the next page

Your question appears here

Application Usage

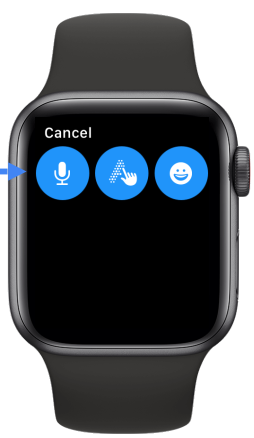
- Activity
  - This will ask you to record or write your current activity

Press this button to record or write your answer

Application Usage

- Here, you can select speech or writing to record your question

Select either speech or writing



You can tap cancel at any time

Tap 'done' when finished recording. This brings you to the next page

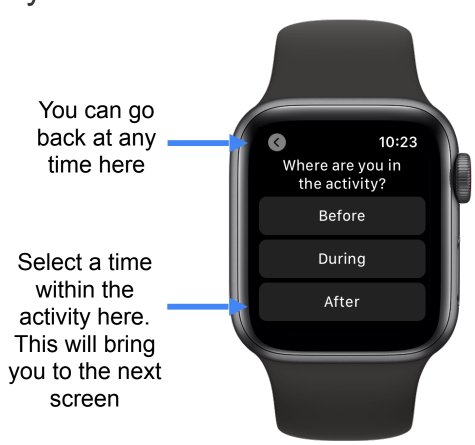
Your question appears here

Application Usage

- Question Related to your Activity
    - This question will ask if your question is related to your current activity (Yes) or not (No)

Application Usage

- Time in the Activity
    - If your question is related to your activity, we will ask where in the activity you currently are



Application Usage

- Final Screen
  - This screen allows you to review your response before submitting

You can go back at any time here

Submit your response here. Tapping this brings you back to the first home screen

Scroll (slide your finger on the screen or use the crown) to see more if the screen can not show everything

Overview

Online Consent - Done

Study Overview - Done

Demographic Survey - Done

Download Application and Tutorial - Done

Practice, Final Thoughts, and Participate! - In Progress

Practice Responses

1. First practice response
    - Open the study application on your watch
    - Record ‘This is my first response for the study’, select ‘done’ when done
    - Record ‘Application Training’, select ‘done’ when done
    - Select ‘Yes’
    - Select ‘During’
    - If happy, select ‘Submit’
    - Congratulations you just quickly and easily submitted your first response!

Practice Responses

2. Second practice response
    - Open the study application on your watch
    - Record ‘This is my second practice trial’, select ‘done’ when done
    - Record ‘An activity not related to training’, select ‘done’ when done
    - Select ‘No’
    - If happy, select ‘Submit’
    - Congratulations you just quickly and easily submitted your second response!

What We Are Interested In

- We are interested in any questions or commands related to:
  - Personal health data that you capture/track on your smartwatch and wish to explore further directly on your smartwatch
  - We want your responses to ideally allow you to better explore and access your personal health data

What We Are Interested In

- Possible categories of questions or commands you may be interested in include:
  - Current status of data
  - History of data
  - Goals, performance, and/or record based
  - Missed data or discrepancies
  - Contextual
  - Combinations and/or comparisons of data

What We Are Interested In

- Don't worry about current technology or applications
- Express whatever questions or commands you have, however feels natural to you, that you want answers to!
  - Duplicates are okay, if they arise multiple times within the week
  - We just ask that you only provide responses when they are relevant to you
  - We do want the response to have an element of personal health data to it
  - The responses should be something you want to be able to access on your smartwatch
- The questions/responses you provide will help create better technologies and applications for you in the future!

What We Are Not Interested In

- We aren't interested in questions related to day-to-day smartwatch function
- These would be typical questions you might ask of Siri:
  - What is the weather today?
  - What is the time?
  - Has my friend messaged me back?
  - Do I have any emails?

Final Thoughts

- You can **use the application any time** you have a question or command of your data
  - You do not need to wait for a notification or specific time
- **Enjoy this opportunity** to be able to ask all the questions and say commands of your data on your smartwatch you've ever wanted, or think would be helpful, for a better future experience for you!
- Please **wear your watch as you normally would** throughout the study
- Only submit responses **when it is safe** to do so

Thanks!

Thanks for your time in participating in our study, again if there are any questions or concerns throughout, please do not hesitate to contact us!

Chapter 7 Tutorial Slides

The following slides were used during the tutorial portion of the study conducted in Chapter 7. A live demo of the application was included during the tutorial.

Study Information, App Setup, and Tutorial

Thank you!

Consent Form

Please use the provided qr code shared onscreen to complete the consent form

Today's Intro Session Overview

1. General Study Overview (~5 minutes)

2. Demographic Survey (~10 minutes)

3. Smartwatch App Installation (~5 minutes)

4. Smartwatch App Tutorial (~10 minutes)

1. General Study Overview

- This study consists of three major components
  1. Introductory study session (currently taking place)
     - Learn about our study goals and how you can use speech and touch to explore your personal health data on the smartwatch
  2. One-week use of our smartwatch application
     - Explore your personal health data on your smartwatch throughout daily life and recorded workouts
  3. Final interview session
     - Interview about your experience using the smartwatch application

1. General Study Overview

- Our high level goal is to allow you to better explore your personal health data directly on your smartwatch

- We are interested in how using speech and touch can allow for exploration

1. Important Remarks

- We are NOT collecting and recording ANY of your personal health data, location, or audio recordings
- We are NOT testing for specific responses or evaluating your performance throughout the study (i.e., you are NOT being tested and there are no right or wrong answers)

1. Important Remarks

- We are INTERESTED in how speech and touch can be used to access your own personal health data on the smartwatch
- We are INTERESTED in the questions you ask and when you ask these questions

Any questions so far?

Today’s Intro Session Overview

1. General Study Overview (~5 minutes) ✅
2. Demographic Survey (~10 minutes)
3. Smartwatch App Installation (~5 minutes)
4. Smartwatch App Tutorial (~10 minutes)

2. Demographic Survey

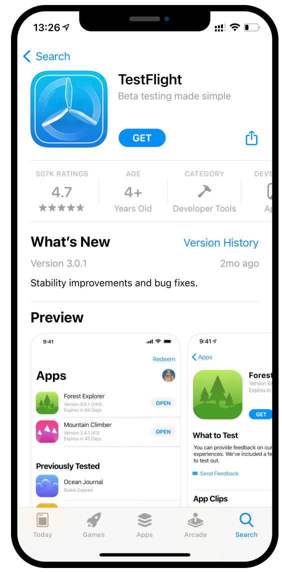
Please use the provided qr code shared onscreen to complete the survey

Today's Intro Session Overview

1. General Study Overview (~5 minutes) ✅
2. Demographic Survey (~10 minutes) ✅
3. Smartwatch App Installation (~5 minutes)
4. Smartwatch App Tutorial (~10 minutes)

3. Smartwatch Application Installation

Install TestFlight



- Install TestFlight **on your iPhone** through the App Store
  - This is Apple's Beta software testing application and allows us to share our application with just you and not everyone in the world
- After installing TestFlight, open TestFlight and accept the terms and conditions
  - We will not be recording crash report information, as such are NOT collecting background data from your device

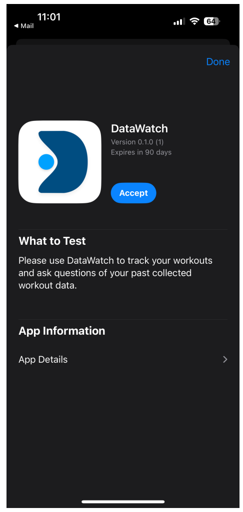
3. Smartwatch Application Installation

Install Our Smartwatch Application

- Scan the QR code below with your iPhone
  - This will open TestFlight and allow you to install our app on your Apple Watch

3. Smartwatch Application Installation

Install Our Smartwatch Application



Today's Intro Session Overview

1. General Study Overview (~5 minutes) ✅
2. Demographic Survey (~10 minutes) ✅
3. Smartwatch App Installation (~5 minutes) ✅
4. Smartwatch App Tutorial (~10 minutes)

4. Smartwatch Application Tutorial

- Why do a tutorial?
  - To familiarize you with our smartwatch application
  - To help you smoothly interact with and explore your data during the study
  - To answer any questions that you may have regarding the application

4. Smartwatch Application Tutorial

Tutorial Shown On-Screen With Live Application

Exploration Times

- You can explore data at three unique times within a workout
    - Before starting a workout – long press on the workout type that you are interested in exploring data for
    - During a workout – long press on a metric that has a blue dot beside it and that you'd like to explore, or swipe to the controls screen and use the microphone
    - After a workout – long press on a metric that has a blue dot beside it and that you'd like to explore

Exploration Modes

- You can explore data in three unique ways
    - Value – Use "What Is/Was, Highlight, Find" in your question
    - Browse – Use "Show, Explore, Browse, View" in your question
    - Compare – Use "Compare, Combine, Contrast" in your question

Example Time Expressions

"This month" → June 1 – June 23, 2024

"From March 1 to 20" → March 1 – March 20, 2024

"Summer 2020" → June 1 – August 31, 2020

"Since March 1" → March 1 – June 23, 2024

Around last Christmas → December 22 – December 28, 2024

Last/past week/month/year… → …

September 2023 → September 1 – September 30, 2023

…

Example Activity Expressions

"Last time" → Your last tracked workout

"Last workout" → Your last tracked workout

"Last 5 times" → Your last 5 tracked workouts

"Last 9 times" → Your last 9 tracked workouts

"Past 7 walks" → Your last 7 walks

"Past 4 bike rides" → Your last 4 bike rides

…

Calculated Values

- You can ask for the average, min, max, total of data
  - Average is calculated by default if nothing is mentioned. To get an average you can also use average, mean, overall
  - To get a min/max, you can use min/max, shortest/longest, slowest/fastest, etc.
  - To get a total of data added up, you can use total, how many/much/far/long, count

Data

- You can ask questions about the following data
  - Duration
  - Distance
  - Pace
  - Calories Burned
  - Heart Rate

A Couple Things To Note

- DataWatch can not answer questions that make inferences of your data
  - Examples of questions not supported:
    - Is my pace faster than last week?
    - Have I walked farther than last month?
    - Am I on track to meet my goal?
    - Should I run 5km today?

A Couple Things To Note

- DataWatch also does not support exploring multiple time periods
  - Examples of questions not supported:
    - Compare my pace from last month to April 2023
    - What was my total distance walked this January and last January?
    - Show my average heart rate this month to May last year

A Couple Things To Note

- DataWatch does support the ability to view your collected data and/or get a calculated value from your collected data
  - Examples of questions supported:
    - Compare to my pace last year
    - Show me my distance walked in the last month
    - What is my average heart rate in the last five walks?

Some Other Questions to Get You Started

- Here are some additional examples of questions that DataWatch supports:
    - How does this compare to my average in the last five workouts?
    - What was the farthest I walked in the last month?
    - How far have I walked in 2024?
    - Show me my fastest pace April

One Caveat

- **To ask a question on the watch, your iPhone needs to be unlocked**
  - This is a security feature from Apple which does not allow access to your collected health data unless your device is unlocked
  - We recommend simply holding your iPhone in your watch wearing hand, unlock your phone, and then interact on the watch

Any questions?

Final Thoughts

- We hope you enjoy this opportunity to further explore your personal health data on your smartwatch
- Please use our application to record any workouts you do throughout the one-week study period
- Please wear your smartwatch as you normally would throughout the study
- Please only ask questions and/or commands when it is safe to do so

Thank you!

Please reach out with any questions or concerns if needed.

Appendix B: Dataset of Collected Queries

Chapter 4 Collected Data

The following is the publicly available dataset captured from the smartwatch application during our study. The dataset has been made available, as linked in our publication at `https://smartwatch-personal-health-data-queries.github.io`.

activity	query	relatedToActivity	time
Eating lunch	How much dancing do I need to do to burn 800 calories	Yes	During
On break	Summary of my sleep cycles	No	
Dancing	During which song did I burn the most calories	Yes	After
On break	Summary of my sleep cycles	No	
Lying down	Which activity had the highest calories burned per minute	No	
Taking a break from work	How many times was I awake last night and for how long	No	
Finished taking a practice final exam	When during the last two hours with my highest heart rate	Yes	After
Dancing	Compare calories burned this dance session to past sessions	Yes	After

About to go on a walk	How many more minutes of exercise do I need to reach my exercise goal for the day	Yes	Before
Working	Based on my activity during the month of October how should I adjust my three goals for the month of November	No	
About to go on a walk	How many steps have I taken so far today	Yes	Before
Sitting down	Did I close all three of my rings for the day	No	
I'm sitting	Did I close all of my rings today	No	
Sitting	How many steps did I take today	No	
Just finished a workout	How many times have I worked out this week	Yes	After
On a walk	How many steps have I taken so far today	Yes	During
Working	How many times did I work out in the past week and how does this number compare to that the number of times I worked out the week before that	No	

Working	How many times did I close all three rings during the month of October	No	
At work	how many days left to my period	No	
I am doing a little bit of work out	compare the heart rate of my friends and mine	Yes	During
Doing homework	I would like to check a trend is my sleep in the past seven days	No	
Preparing dinner	What is my current calorie intake	Yes	Before
Going to work	Should I go for a run today?	No	
Getting ready for work	Am I over or under my calorie goal at the moment	No	
Going out	How many calories that I login for breakfast	No	
At work	How much uninterrupted sleep did I get last night?	No	

Working out	What is my current bloodpressure	Yes	During
Bus	Did I complete my water goal yesterday?	No	
studying in library	How many calories did I burn in the last 4 hours?	No	
Brushing teeth	What's my average sleeping time this past week	No	
Getting out of bed	Do I need more sleep?	Yes	After
Arriving home	How many steps left do I need to complete my goal	No	
Going to the gym	When was the last time I ran and what was the time	Yes	Before
Cleaning up after workout	What was my peak heart rate during my workout	Yes	After
Brushing teeth	Did I reach my water intake goal?	No	

Walking home	How many steps did I take in the past seven hours	Yes	During
In bed	Was my sleep longer than last time?	Yes	Before
Getting ready to sit down for the evening	What type of workout do I burn the most calories per minute historically	No	
Out on a walk trudging through the snow	What is my average walking pace in the winter versus the summer	Yes	During
Laying on the couch watching a documentary about how efficient coyotes gate is	Has my heart rate during walking decreased since I started tracking walks	No	
Just woke up from a nap at 10:57 and didn't have my standing minute yet	When am I most and least likely to get and miss my standing minutes	No	
Sitting in the car at 6:30 PM with several hundred calories left clues my move ring	What kind of workout do I usually do if I have 200 or more calories to close my moving ring	No	
Just returned home from a short walk	What is my weekly average walking kilometres with the walking workout?	Yes	After
Hockey game intermission	Do I burn more calories when I work out in the morning afternoon or evening?	No	

Sitting in the meeting with Bradley	Is my running pace slower in the days following a strength training workout?	No	
Final study meeting with Bradley	Is my first 2 km of walking pace faster or slower than my last two	No	
Getting ready to go to the gym	Give me a report for my readiness for activity	Yes	Before
Gym	What's my blood sugar level	Yes	After
Getting ready to go to the gym	Why is my Sleep report from last night so bad	No	
Gym	Give me a suggested work out based on my readiness score	Yes	Before
Dinner preparations	If I have tacos tonight how will that affect my sleep and readiness score for tomorrow	No	
Grocery shopping	How will not going to the gym during the holiday weekend affect me	No	
Grocery shopping for Thanksgiving	How has the stress from the last two hours of grocery shopping for Thanksgiving affected my readiness score for today	Yes	During

Outdoor walk	How is the air quality affecting my walk	Yes	During
Gym	How did today's gym session compared to the last one given that I was sick this time	Yes	After
Walking to the car	Is my heart rate slower or faster than normal	Yes	During
Drinking lots of beer	Does drinking beer impact my sleep	Yes	Before
Laying on the sofa sick	Does my heart rate look any different compared to the average	Yes	During
Just finished exercising	How many steps did I take my outdoor walk	Yes	After
Finished walking	How many steps did I do on that short walk	Yes	After
Sitting	When is my period going to start	No	
Standing	What is my resting heart rate	No	

Sitting	How many hours did I sit yesterday	Yes	During
Sitting at my desk	How much of my day have I spent sitting	Yes	During
Going to bed	What was my average resting heart rate today	No	
Working	What has my resting heart rate been since I put my watch on	Yes	During
Finished mopping	How many calories did I burn mopping	Yes	After
Standing up	How many hours was I sitting for today	No	
Sitting	How many steps have I taken this week	No	
Walking	How many steps did I just take	Yes	After
Standing	How many calories did I burn today so far	No	

Getting ready for bed	How many hours did I spend sitting today	No	
Getting ready for bed	What is my average step count per day	No	
Running	How many steps did I take during my run	Yes	After
Getting ready for bed	What was my peak heart rate of the day?	No	
Getting ready to exercise	How many runs have I completed thus far in 2021	Yes	Before
Eating lunch after run	Show me a graph of my runs both time and distance in 2021	Yes	After
Just finished my work out	How many times and for how long did my heart rate rise above 130 today?	Yes	After
Homework	How many miles have I accumulated through walking, running, biking over the course of this year	No	
Making breakfast	Over the last three months how many days have I beat Marianne on step count	No	

Walking	Tell me when I reach a nine minute walking pace	Yes	During
Standing	What is the average time I closed my stand ring during last year	Yes	During
I just got up	Tell me the specific days of the week in the last three months where my weight tended to trend up or down	No	
Making breakfast	Over the last year when I breathe when directed by my watch does it lower my heart rate	No	
Getting ready to go for a walk	Tell me when I reach a nine minute kilometre pace	Yes	Before
Getting dressed	What has Marianne's most active day of the week been over the last three months and what has my most active day of the week been on average	No	
Getting dressed	How many times have I tested for signs of atrial fib relation since I got my watch	No	
Having lunch	What is the average time I start tracking my steps in the day over the last year	No	
Watching the news	How many times have I not closed my rings in the last year	No	

I just got up	Is my heart rate consistently the same when I first put my watch on in the morning	Yes	During
I just finished taking an ECG reading	What have my ECG readings been sense August 1	Yes	After
Walking	Tell me when I reach a nine minute per kilometre pace	Yes	During
Reading a notification about Marianne on our fitness challenge	Is there a day of the week I am more likely to beat Marianne in our fitness challenge	Yes	During
I am leaving on a solo walk	Does my walking pace change when I walk with someone else	Yes	Before
Walking	How much has my pace fluctuated during my walk	Yes	During
Cooking breakfast	What are the average steps I take on a Sunday	No	
Yardwork	What was my highest heart rate in the last hour	Yes	After
Weighing myself	What was my average weight in October	Yes	After

Walking	What was my average resting heart rate in October	No	
Watching TV	How many steps do I have left to make my goal today	No	
Starting my day	Show me a chart of my weight loss over the last two weeks, my steps, and my caloric intake	No	
About to eat	Can I eat squash soup and two pieces of toast and still be below my calorie budget	Yes	Before
Waking up starting my day	How much sleep did I get last night	Yes	After
About to lay down	How long have I been lying down today since first waking up	Yes	Before
Getting my day started	Show me a graph comparing my caloric intake over the last week	No	
Working	How many calories do I have left on my budget for the day	No	
Working	Show me the history of my weight over the last two years	No	

Working	How long does it normally take me to walk 5000 steps	No	
About to go walking	How many steps do I have left to do today	Yes	Before
Working	How many calories do I have left in my bank for today?	No	
Driving	How many steps do I have left to make my goal	No	
Planning my next run while working	When is a good time for me to run after the rain stops	No	
breakfast	what is the impact of rain in my running performance	No	
working	when do I need to exercise to improve my vo2 max	No	
waking up	what is the rowing equivalent for my running	No	
Rowing	What is my stroke rate while rowing	Yes	During

Having breakfast	How can I improve my resting heart rate	No	
Eating	What are the effects of BMI	Yes	During
looking at HRV indicator	what does heart rate variability mean	Yes	During
Working	What is my performance compared to same time last year	No	
Outdoor walk	How long do I need to run three times per week to achieve the November challenge	Yes	During
preparing to sleep	what helps me get better sleep?	Yes	Before
waking up	why am I sleeping bad lately?	Yes	After
Having breakfast	How can I increase my VO2 Max	No	
having lunch	what is the impact of my sleep in my running?	No	

Watching a video	Have I stood up this hour	No	
having breakfast	compare my running stats with the same time last year	No	
Weight lifting	Am I in the right heart rate range for this type of work out	Yes	During
Outdoor walk	How does my walking pace change depending on the length of a walk	Yes	During
Walking outside	How long does it take after a walk to get back to resting heart rate	Yes	After
Weightlifting	Is my work out better at my home gym or commercial gym	Yes	After
Weightlifting	Does weightlifting focussing on different muscle groups affect my heart rate	Yes	After
Stationary bike	How long does it take for my heart to reach optimum heart rate during a cycling work out	Yes	During
Sitting	How much lower is my activity today because it's Friday	Yes	During

Working out	Compare the heart rate of this work out to last work out	Yes	After
Waking up and going to the gym	What workout do I have planned for today	Yes	Before
Laundry	How many kilometers do I need to walk to get 10,000 steps	No	
Making lunch	What was my best kilometer during my run	No	
Driving	What was my fastest kilometer in my run	No	
Driving	When was the last time I did a full 5K	No	
Driving passenger	Show me a graph of my 5 km outdoor runs over time	No	
Just out of the shower	How many steps did I get during my dance workout today	No	
Just out of the shower	Show me a history of all my dance workouts	No	

Going for a walk	How long should I walk to get 10,000 steps today	Yes	Before
Watching TV	When should I be expecting my period	No	
Dance workout	How many more steps do I need to get to 10,000	Yes	Before
Getting ready to go for a walk	Give me a 2 km walk route starting from home	Yes	Before
Back from walk	How many steps did I get during that 2 km walk	Yes	After
End of work out	How many calories did I burn that work out	Yes	After
Just getting off the treadmill	How many steps am I at today	Yes	After
On a zoom meeting	How many calories have I eaten today	No	
Dance workout	How many active minutes am I at	Yes	During

Eating lunch	How many steps have I taken today	No	
Walking	What's my average walking pace per kilometer	Yes	During
Eating supper	How many minutes of aerobic activity would it take me to burn 200 cal?	No	
Working out	What's my daily calories burned this week?	Yes	After
Going for a walk	What's the fastest that I walked 2 km outdoors?	Yes	Before
Walking	Show me a graph of my average walking speed during outdoor walk	Yes	After
Waking up	How many hours of deep sleep did I get last night?	Yes	After
Waking up	Show me a graph of my sleep quality from last night	Yes	After
Eating breakfast	What's my blood oxygen level when I'm sleeping?	No	

At the park	What's my average steps per day?	Yes	During
Eating supper	What are my calories burned on weekends versus weekdays?	No	
Working out	Compare max heart rate from this work out to Monday's work out.	Yes	After
Work out	What are average calories burned for HIIT work out?	Yes	After
Working out	What are my average non-active calories burned per day?	Yes	After
Work out	Alert me when my heart rate gets back to my average resting heartbeat	Yes	After
I am at work	How long on average does it take me to fall asleep?	No	
Working	How much time do I spend sitting on average Monday to Friday?	Yes	During
Working	What's my average calories burned in the morning versus afternoon versus evening?	No	

Playing with kids.	What was my calories burned in the last 30 minutes?	Yes	During
Working at my desk	How many calories have a burned since 8 AM?	No	
Walking	What is my average walking pace?	Yes	After
Walking	Compare walking pace September and October	Yes	After
Sitting at my desk	What is my blood sugar at	No	
Going to the gym	Do I need to eat something before I go to the gym	Yes	Before
Walking home from the gym	Should I stop trying for linear growth based on my last workout	Yes	After
Sitting home	What are the factors most likely impeding my squat growth	No	
Sitting on my desk	Based on my current biometrics when will be the best time for me to work out today	No	

Dead lifting	How many reps was my last set	Yes	During
Walking to the gym	What's the best exercise for me today	Yes	Before
Squat	What was my RpE perceived exertion	Yes	During
Walking home	What is my projected one rep maximum	No	
Sitting at home	About how long will it take me to reach my squat deadlifts and benchpress goals	No	
Watching TV	How many calories did I just eat	No	
Sitting around being tired	Is my cycle affecting my sleep	Yes	During
Trying to sleep	Compare data points from nutrition and heartrate and sleep	No	
Nothing	Compare cycle data from today to the same day in my last cycle	No	

Reviewing the days activities	Compare data points from different types of activity	Yes	After
Reviewing the day of activity	Compare data points across different activities within a cycle of a specific number of days	Yes	After
Thinking about what to make for dinner	What is the recommended calorie intake for dinner today	Yes	Before
Just got home relaxing	Was the 1st km of my hike faster than the last kilometre today	No	
Walking home from the gym	How many calories did I burn playing badminton today	Yes	After
Walking home from the gym	Show me my heart rate chart from today's gym session	Yes	After
Walking	How many steps did I do yesterday versus today	Yes	During
Weighing myself	Show me my weight loss trends in the past month	Yes	During
Thinking about what to make for breakfast	What is my calorie budget for today	Yes	Before

Reflecting on my fitness journey and today's goals	How many calories were burned in today's work out compared to yesterday	No	
Walking home from the gym	On average was my heart rate lower during today's run compared to yesterday	Yes	After
Eating a banana	How many calories is in a medium size banana	Yes	During
Doing homework	How many steps should I take today and is it less than yesterday	No	
Weighing myself	Show me my bodyweight trends for this month	Yes	During
Journalling	What was my average heart rate during weight training today	No	
Journalling	How many calories over budget am I today	No	
Just got home relaxing now	How fast did I finish my 1st km on my hike today	No	